scientific reports

Check for updates

OPEN

Comparative transcriptome analysis of bull X- and Y-spermatozoa

Sofi Imran Ul Umar[1,2], Sushil Prasad[2], Soumen Naskar[1]✉, Pranab Jyoti Das[3], Mridula Sharma[4], Arunava Pattanayak[1], Dhanu Kumar Murasing[1,4], Vijai Pal Bhadana[1] & Sujay Rakshit[1]

Sex-sorted bovine semen is a major breakthrough in the dairy sector. The current flow cytometer-based technology for sex-sorting of bovine semen is proprietary, slow, expensive, and associated with lower conception rates. Therefore, developing alternative technologies is crucial for advancing milk production globally through use of sexed semen. However, lack of a comprehensive omics dataset hinders the ability to comprehend significant differences between the sorted sperm types and develop robust biomarkers for sex-sorting. Here, we reported the RNA-Seq analysis of unsorted, X- and Y-sperm in *Bos indicus* cattle. The differential gene expression analysis revealed significant upregulation of 47 genes in bovine Y-sperm and downregulation of 20 genes in comparison to X-sperm (adjusted *p* value < 0.05). Sixteen percent of the transcripts were unique to X-sperm, whereas 20.7% were unique to Y-sperm. The top 22 differentially expressed genes (DEGs) were validated using qPCR. A significant up- or down-regulation was detected in 21 of the 22 genes when comparing bovine Y-sperm to X-sperm ($p < 0.01$ and $p < 0.05$). The transcriptome dataset, the first in the league of bovine X- and Y-sperm, will aid in biomarker discovery for sex-sorting of bovine semen and improved fertility outcomes.

Keywords Sex-sorted semen, Dairy, RNA-seq, Gene ontology, Pathway enrichment

Abbreviations

DGE	Differential gene expression
X-sperm	X-chromosome bearing spermatozoa
Y-sperm	Y-chromosome bearing spermatozoa
lncRNA	Non-coding RNA
qPCR	Quantitative polymerase chain reaction
cDNA	Complementary DNA
PBS	Phosphate buffer saline
TCEP	Tris(2-carboxyethyl) phosphine
fg	Femtogram
SE	Standard error
HISAT	Hierarchical indexing for spliced alignment of transcripts
FPKM	Fragments per kilobase of transcript per million read pairs
CDH1	Cadherin 1
PRM1	Protamine 1
gDNA	Genomic DNA
PTPRC	Protein tyrosine phosphatase receptor type C
KIT	Tyrosine–protein kinase
PCA	Principal component analysis
GO	Gene ontology
MF	Molecular function
BP	Biological process
CC	Cellular component
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
ICG	Internal control gene

[1]ICAR-Indian Institute of Agricultural Biotechnology, Garhkhatanga, Ranchi 834003, India. [2]Birsa Agricultural University, Kanke, Ranchi 834006, India. [3]ICAR-National Research Centre on Pig, Rani, Guwahati 781131, India. [4]G. B. Pant University of Agriculture and Technology, Pantnagar 263145, India. ✉email: snrana@gmail.com

KEGG	Kyoto encyclopaedia of genes and genomes
COVID-19	Coronavirus disease-19
MAPK	Mitogen-activated protein kinases
cAMP	Cyclic adenosine 3′, 5′-monophosphate
GPCRs	G protein coupled receptor
MRP4	Multidrug resistance protein 4

Sex-sorted semen in bovines is a groundbreaking development in the dairy industry. This technology allows for a greater proportion of milk-producing animals (~90%) than conventional (unsorted) semen (~50%), thereby increasing profits for the dairymen[1]. Additionally, it enhances resource-use efficiency, accelerates genetic gain, and promotes animal welfare by reducing the number of undesired male calf births[2]. However, the current flow cytometer-based technology for sex sorting of bovine semen is proprietary, slow, expensive, and associated with lower conception rates[3]. Therefore, developing alternative technologies are crucial for advancing milk production as a global public good. Ongoing research and developmental efforts aim to refine this technology, making it more accessible and efficient and further solidifying its role in sustainable livestock production[2,3]. There are significant gaps in the omics knowledge bases concerning X- and Y-chromosome-bearing spermatozoa (X- and Y-sperm) of cattle. The lack of a comprehensive omics dataset hampers the ability to draw meaningful comparisons across livestock species and to develop robust biomarkers for sex-sorting of bovine semen with improved conception rates. While some progress has been made using advanced techniques such as RNA sequencing and mass spectrometry, these studies often focus on model organisms, leaving a gap in our knowledge of agriculturally important species. Additionally, the dynamic nature of sperm and their interactions with the female reproductive tract are not well understood, further complicating efforts to develop more efficient sex-sorting methods. A detailed understanding of the molecular biology of X- and Y-sperm will not only address these gaps but also contribute to broader fields of reproductive health and biology[4,5].

Recent studies on transcriptome profiles of unsorted sperm from various species such as *Bos indicus* have debunked the notion that sperm is transcriptionally dormant. Comparative male fertility has been the primary focus of these transcriptomics investigations[6–9]. However, there is a paucity of research on the differential expression of genes in bovine X- and Y-sperm. Suppression subtractive hybridization and cDNA microarray analysis revealed 31 differentially expressed upregulated genes in bovine X- (27 numbers) and Y-sperm (4 numbers), respectively[10]. The advent of advanced RNA sequencing technologies has revolutionized transcriptome studies by minimizing the requirement of RNA quantities. By RNA sequencing study, it was deciphered that 492 genes are encoded by mouse X-chromosome as against only 15 genes by the Y-chromosome[11].

Earlier, we reported the proteomics of unsorted sperm and identified differential expression of plasma membrane-associated proteins between X- and Y-sperm of indicus cattle[12]. Here, we hypothesise differential gene expression in bovine X- and Y-sperm which can serve as the basis for their segregation. Therefore, our objective was to study the comparative transcriptome of unsorted and sexed bovine X- and Y-sperm with gene ontology and pathway enrichment analysis and validate the differential gene expression. This report provides novel insights into the complexities of sex differences in dairy cattle.

Methods

General information

Semen samples from bulls for unsorted, sex-sorted X and sex-sorted Y (n = 3 bulls each for RNA-Seq) were obtained from reputed bull semen stations. The semen samples of pedigreed Sahiwal bulls, a prized cattle breed of India and Pakistan, of 4–6 years of age and high genetic merit were used for the present study. The collected ejaculates qualified the minimum standards required for production of frozen semen as prescribed by the Department of Animal Husbandry & Dairying, Ministry of Fisheries, Animal Husbandry & Dairying, Government of India[13]. The sperm were sorted following the principle of Beltsville Sperm Sexing Technology[14]. A high-speed BD influx cell sorter was used for sperm sorting process. The purity of the sexed semen (both X- and Y-sorted) was approx. 90%. Six biological replicates (bulls) and three technical replicates were used for each of the groups for qPCR validation. A minimum of two number of ejaculates per bull was taken, and the amount of each ejaculate per bull was balanced for downstream applications. All methods were carried out in accordance with the relevant guidelines and regulations and are reported in accordance with ARRIVE guidelines (https://arriveguidelines.org).

Experimental design and sampling

Three biological replicates (bulls) for each group viz., Conventional/unsorted (C1, C2, C3), sex-sorted X-sperm (T1A, T1B, T1C) and sex-sorted Y-sperm (T2A, T2B, T2C) were used for RNA-Seq analysis. C versus T1, C versus T2 and T1 versus T2 were the experimental conditions.

RNA extraction

RNA isolation was performed on pooled ejaculates from various bulls. The samples were first purified and washed with phosphate buffer saline (PBS, pH 7.2). Total RNA was isolated via a protocol described previously by us[15]. In brief, sperm samples (10 million cells for each biological replicate) were homogenized and lysed with a lysis buffer cocktail [0.1 ml of lysis buffer from the RNeasy Plus Mini Kit (Qiagen, USA), 0.9 ml of Qiazol (Qiagen) and 0.1 ml of TCEP (Sigma, USA)]. This was followed by phase separation with 0.2 ml of chloroform. The remaining steps were performed according to RNeasy Plus Mini Kit (Qiagen) protocol. The DNase treatment was included. To rule out contamination by somatic cells, we performed molecular screening of the RNA used for RNA-Seq and qPCR. For this purpose, we followed methods outlined by Selvaraju et al.[8]. Spermatozoal gDNA contamination was checked using a set of intron-spanning primers for protamine 1 (*PRM1*). To confirm that the

RNA was free from other types of contamination, cell-specific intron-spanning primers for the Cadherin1 gene (*CDH1*, for epithelial cells), protein tyrosine phosphatase receptor type C gene (*PTPRC*, for leukocytes) and *KIT* oncogene (*KIT*, for germ cells) were used. The primer sequences are provided in Supplementary Table 1. The RNA was stored at − 80 °C for subsequent applications.

RNA-seq analysis

The NEBNext Ultra II Directional RNA Library Prep Kit (New England Biolabs, USA) was used to prepare high-quality libraries according to the manufacturer's protocols, and paired-end sequencing reads of 150 bp were generated with the Illumina HiSeq X sequencing platform. The raw reads were checked via FastQC[16]. Fastp, an ultrafast FASTQ preprocessor with useful quality control and data-filtering features, were used[17]. A cut-off of 30 was set for the Phred quality score, and only high-quality reads were retained.

Mapping and alignment of reads to the Bos indicus genome

The reads were aligned against the NCBI Reference genome of *Bos indicus* (assembly Bos_indicus_1.0) by using a fast and sensitive alignment program HISAT2[18]. To understand the alignment quality, we checked several parameters, including the percentage of mapped reads. The mapped reads were subsequently considered for transcript assembly and quantification of transcript abundance by StringTie[19].

Transcript classification

The transcript count obtained from each sample was further used for the analysis of the differential expression of transcripts between alternate conditions using DESeq2[20]. The transcripts whose p value and adjusted p value were less than 0.05 were selected for further analysis. The log2-fold change cut-offs of (+ 2) and greater for upregulated transcripts and (− 2) and lesser for downregulated transcripts were used in all three conditions (C vs T1, C vs T2 and T1 vs T2). The transcripts with FPKMs present in only one group were considered unique. The differentially expressed spermatozoa transcripts between the alternate conditions were expressed as heatmaps, volcano plots and PCA plots using online tools.

Gene ontology and pathway analysis

The obtained spermatozoa transcripts were subjected to gene ontology (GO) classification via the Panther classification system (PANTHER 18.0) and the Database for Annotation, Visualization, and Integrated Discovery (DAVID) Bioinformatics Resources[21]. The transcripts were classified into four categories: molecular function (MF), biological process (BP), cellular component (CC), and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway. Highcharts was used to create a donut pie chart representing the BPs, CCs, and MFs of genes that are unique to bovine X-sperm and Y-sperm. Pathway enrichment was performed via enriched KEGG functions and clusterProfiler. Using the open-source Cytoscape (version 3.10.1) platform, the ClueGo (version 2.5.4) and Cluepedia (version 1.5.4) plugins were employed to analyse gene interactions, comprehensive networks of integrated GO categories, and pathway analysis[22]. We found several configurations representing a network of interactions between related genes. *Bos indicus* genome served as the background for each analysis.

Validation of differential gene expression by qPCR

The top 22 differentially expressed genes (DEGs) in bovine Y-sperm (in comparison with X-sperm) were selected and validated by qPCR using SYBR Green chemistry (Supplementary Table 1). Six or more bulls for each of the experimental groups (unsorted, C; X-sorted, T1; and Y-sorted, T2) were used for qPCR analysis. *GAPDH* was used as a reference gene in the qPCR for determining relative expression levels. The choice of *GAPDH* as a reference gene was based on previous work[23] that evaluated the expression stability of 10 commonly used housekeeping genes in bovine spermatozoa and revealed *GAPDH* as the most stable internal control gene (ICG) on the basis of the analysis by Genorm, NormFinder, Delta Ct and the comprehensive ranking using RefFinder. *GAPDH* was also found to be stable across the experimental groups on the basis of Ct values and melt curve analysis in the present study. The primer sequences for selected 22 genes are provided in Supplementary Table 1. Statistical analysis of the gene expression data from the qPCR was performed by Student's t-test using Microsoft Excel.

Results

RNA yield and quality

An average yield of 60 fg of RNA per sperm cell was obtained, and it was confirmed that this RNA was devoid of gDNA and other RNA impurities from somatic cells, leucocytes, and germ cells by molecular screening. DNA from sperm cells served as a positive control (Supplementary Fig. 1).

RNA-seq analysis

The total filtered sequence counts for each biological replicate in the three groups used for the study are shown in Supplementary Fig. 2. The average number of filtered reads obtained from unsorted, sex-sorted X- and Y-sperm were 53.39 ± 3.9, 47.75 ± 2.2, and 26.80 ± 2.6 million, respectively. The processed reads were aligned against the NCBI reference genome of *Bos indicus* (assembly Bos_indicus_1.0) via the ultrafast splice-ware aligner HISAT2[18]. Details of the mapping read percentage per sample are given in Supplementary Table 2. The mapped reads were further considered for transcript assembly and quantification of transcript abundance via stringtie[19]. Transcript counts obtained from each sample were further used for analysis of the differential expression of transcripts between alternate conditions via DESeq2[20]. The numbers of upregulated and downregulated transcripts found in the three alternate conditions are given in Table 1. The list of all the differentially expressed transcripts in Y-sperm vis-à-vis X-sperm is given in Supplementary Table 3, and the list of differentially expressed transcripts

Experiment	Upregulated		Downregulated	
	$p < 0.05$	adj. $p < 0.05$	$p < 0.05$	adj. $p < 0.05$
C versus T1	540	47	371	20
C versus T2	2000	1128	2504	1722
T1 versus T2	737	77	1016	91

Table 1. The number of differential transcripts obtained via DSEQ2 analysis for p and adjusted (adj.) *p* values less than 0.05 for each experimental condition. C—Unsorted sperm, T1- X-sorted sperm and T2-Y-sorted sperm.

in X- and Y-sperm of *Bos indicus* bulls compared to unsorted sperm are given in Supplementary Tables 4 and 5. The volcano plot generated by using a DESeq2 dataset, with default log fold-change thresholds of − 2 and + 2 and an adjusted p value threshold of 0.05, is shown in Fig. 1. The top 50 differentially expressed sperm transcripts were plotted via a heatmap (Fig. 2). Principal component analysis (PCA) of differentially expressed genes was performed on biological replicates of each group to evaluate variance (Fig. 3). A Venn diagram of the unique gene expression data is shown in Fig. 4.

Transcript distribution in X- and Y-sperm

We found that 16% of the transcripts were unique to cattle X-sperm and 20.7% were unique to Y-sperm. The percentage of common transcripts between cattle X- and Y-sperm was 63.3% (Fig. 4c). A total of 175 transcripts were differentially expressed in cattle Y-sperm compared with X-sperm of which nine transcripts were classified as noncoding (IncRNA) (five nos. upregulated and four nos. downregulated), 152 as protein-coding (67 nos. upregulated and 85 nos. downregulated), 10 as pseudogenes (four nos. upregulated and six nos. downregulated), and four as transcribed pseudogenes (one no. upregulated and three nos. downregulated). Among the transcripts unique to Y-sperm, 53 nos. were located on the *Bos indicus* Y-chromosome, while among the transcripts unique to X-sperm, 200 nos. were located on the *Bos indicus* X-chromosome. The complete set of SRA data of sex-sorted and unsorted sperm of indicus cattle is available at NCBI (BioProject PRJNA976949).

Gene ontology (GO) analysis

The GO analysis of the upregulated genes revealed their involvement in various molecular functions (MFs), biological processes (BPs), cellular components (CCs), and KEGG pathways across different comparisons: for C versus T1, the genes were involved in 6 MFs, 3 BPs, 4 CCs, and 5 KEGG pathways (Fig. 5a); for C versus T2, they were involved in 8 MFs, 13 BPs, 11 CCs, and 15 KEGG pathways (Fig. 5b); and for T1 versus T2, they were involved in 7 MFs, 5 BPs, 4 CCs, and 5 KEGG pathways (Fig. 5c). Additionally, GO analysis of genes unique to X-sperm showed their involvement in 14 MFs, 11 BPs, and 6 CCs (Fig. 6a), while genes unique to Y-sperm were involved in 8 MFs, 10 BPs, and 4 CCs (Fig. 6b).

Pathway enrichment analysis

The pathway enrichment of the transcripts upregulated in cattle X-sperm in comparison with the unsorted sperm alluded to involvement in the coronavirus disease-COVID-19 pathway (15 counts) and ribosome pathway (11 counts). The transcripts upregulated in cattle Y-sperm compared with unsorted sperm indicated their involvement in the herpes simplex virus 1 infection pathway (88 counts) and glycerophospholipid metabolism pathway (24 counts). The transcripts upregulated in cattle Y-sperm in comparison with X-sperm indicated their involvement in metabolic pathways (111 counts) and the regulation of the actin cytoskeleton pathway (19 counts).

The pathway enrichment of genes unique to cattle X-sperm revealed their involvement in pathways related to cancer (KEGG: 05200, 54 counts) and the MAPK signalling pathway (KEGG: 04010, 31 counts) (Fig. 7). The transcripts unique to Y-sperm indicated involvement in purine metabolism (KEGG: 00230, 10 counts), the calcium signalling pathway KEGG:04020, 7 counts) and the cAMP signalling pathway (KEGG: 04024, 7 counts) (Fig. 8).

Validation of differential gene expression in cattle X- and Y-Sperm

The 22 differentially expressed genes (DEGs) were validated using qPCR. Ten upregulated and 12 downregulated genes in cattle Y-sperm compared with X-sperm were selected based on log2fold changes from DSEQ2 analysis. The biological significance of the validated genes is presented in Table 2. The primers for the selected genes were designed by Primer BLAST from NCBI (Supplementary Table 1) and validated using SYBR Green chemistry. The log2fold change was calculated by using the $2^{-\Delta\Delta Ct}$ method[46]. The means ± SEs of the fold changes in gene expression are shown in Fig. 9. The results revealed significant up-/down-regulation of the selected genes (21 out of 22) in bovine Y-sperm compared with X-sperm ($p < 0.01$, except for the *NAPRT*, *VWC2* and *MVP* genes, $p < 0.05$).

Discussion

The discernible difference between X chromosome-bearing sperm and Y chromosome-bearing sperm is the quantity of DNA in the sex chromosomes[47]. Since the X chromosome has more DNA than the Y chromosome does in mammals, there may be differences in the amount of resultant RNA. Previous works have demonstrated that transcripts or gene products are shared between X- and Y-sperm through intercellular bridges during

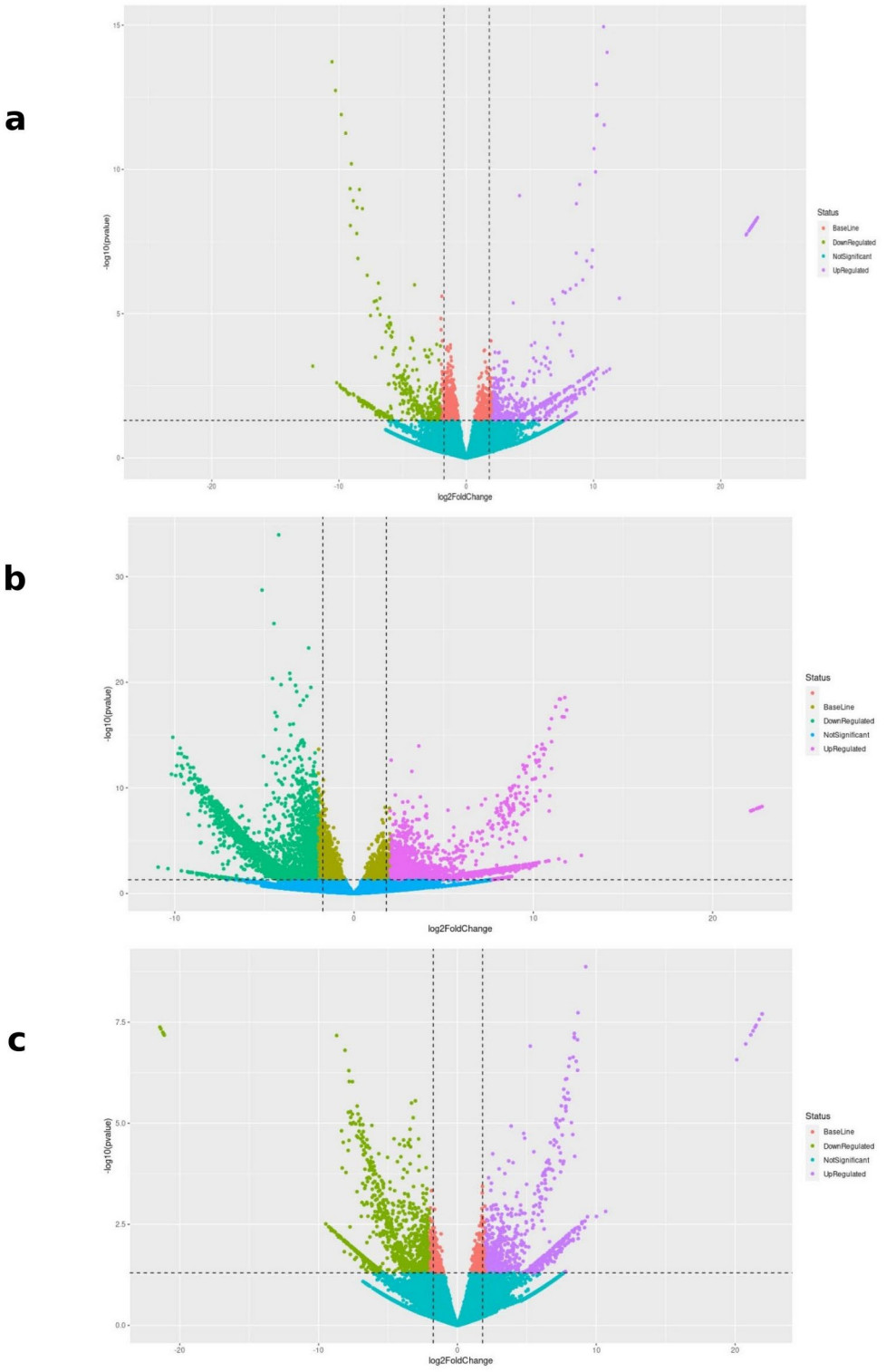

Fig. 1. Visualization of the differential transcripts via a volcano plot. (**a**) Unsorted and X- sorted cattle sperm. (**b**) Unsorted and Y-sperm. (**c**) X- and Y-sperm. The X-axis represents logarithmic fold changes in expression (log2FC), whereas the Y-axis represents the negative decimal logarithm of the P values. The horizontal line refers to the negative logarithmic *P* value cut-off ($P = 2$). The vertical lines mark the fold change cut-offs ($\log2FC > 2$).

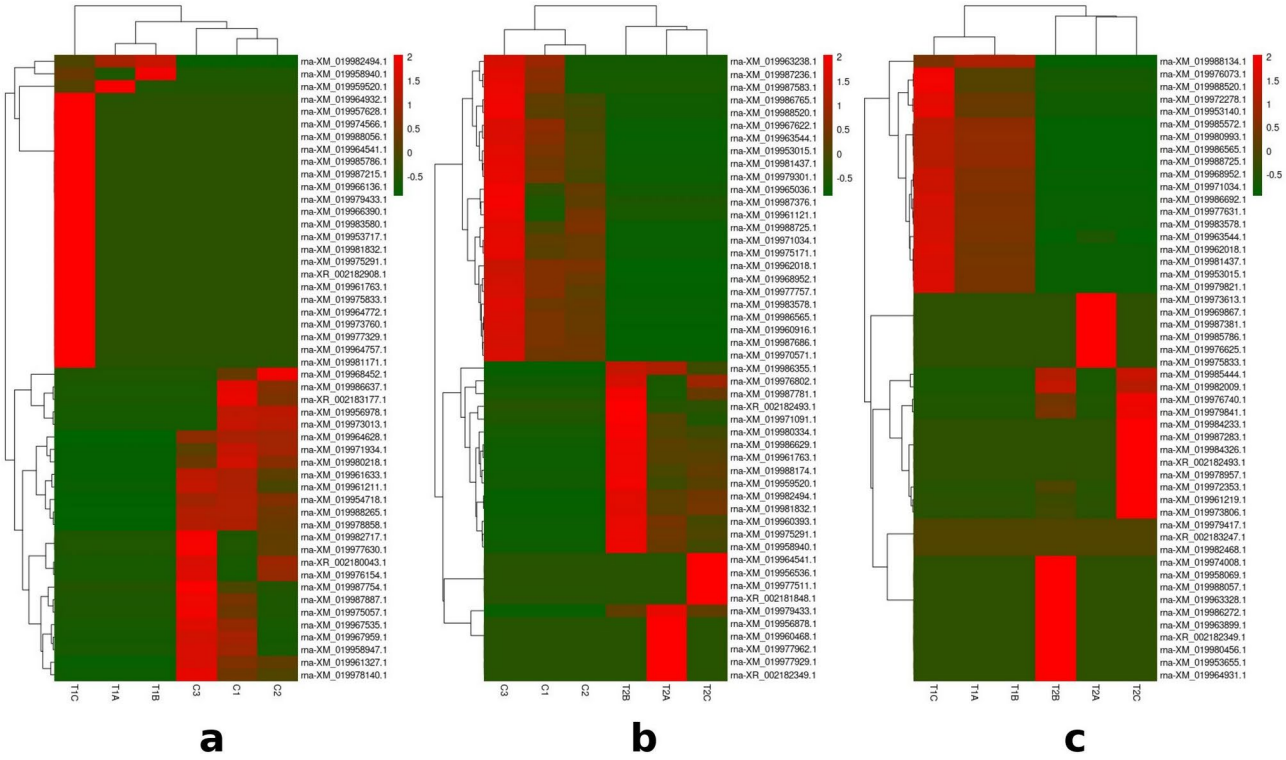

Fig. 2. Heatmap of the top 50 differentially expressed transcripts between (**a**) unsorted (C1, C2, C3) and X-sperm (T1A, T1B, T1C), (**b**) unsorted and Y-sperm (T2A, T2B, T2C), and (**c**) X- and Y-sperm.

spermatogenesis that may or may not be carried forward to mature sperm[48]. However, not all of these products are likely transferred across the intercellular bridge[49]. The current study reports differential and unique gene expression in X- and Y-sperm of bulls, which are original inputs for the differences at the gene/transcript level between X- and Y-sperm. The protein coding transcripts Cathepsin B, Histone H2A, Calmodulin and Glyceraldehyde-3-phosphate dehydrogenase, testis-specific were found unique to bovine X-sperm whereas protein coding transcripts Elongation factor 1-alpha 1 and Chromodomain-helicase-DNA-binding protein 1 were upregulated in bovine Y-sperm. These are in agreements with our proteomic studies[12,50]. Previously, a combination of suppression subtractive hybridization (SSH), cDNA microarray, and sequence-homology analysis identified 27 and four genes upregulated in bovine X- and Y-sperm (*Bos taurus*), respectively[10]. The abundance and diversity of small noncoding RNA (SncRNA) profiles were reported to vary between bull X- and Y-sperm[51].

The GO analysis revealed that the cellular component (60%) was the major portion of the GO categories enriched with unique genes in X-sperm while the unique genes in Y-sperm were enriched in molecular functions (67.3%). The same trend was observed for the genes whose expression was upregulated in Y-sperm compared with X-sperm. The top GO term for the genes unique to Y-sperm was G-protein coupled receptor activity (GO: 0004930, 24.5%), which is associated with MF. The G protein-coupled receptors (GPCRs) constitute the largest family of receptors with seven transmembrane domains and regulate various physiological processes. The known sperm-associated GPCRs are olfactory receptors. They are also involved in fertility and induce bovine sperm acrosome reactions[52]. Our findings suggest that certain olfactory receptors may be exclusive to cattle Y-sperm (Fig. 6B). These receptors may serve as targets for various interactions involving ligands and receptors, which could be utilised for segregation of bovine Y-sperm.

The pathway analysis of genes unique to bovine Y-sperm revealed their involvement in signalling pathways (calcium signalling pathway (KEGG: 04020) and cAMP signalling pathway (KEGG: 04024)), which are important for sperm motility. Purine metabolism (KEGG: 00230, 10 counts), which has been shown to potentially influence bull fertility, also involves genes unique to Y-sperm[53]. The essential role of cAMP signalling pathways is in the activation of sperm motility and in the induction of the vigorous asymmetrical movement (i.e., hyperactivated motility) necessary for the fertilization of sperm. cAMP efflux through Multidrug resistance protein 4 (MRP4) regulates sperm motility in bull spermatozoa[54,55].

Twenty-two DEGs were validated in the present study. The genes upregulated in Y-sperm are reported to be involved in energy metabolism (*PKLR, D2HGDH*), membrane transport (*ATP13A2, TMEM143, TMEM168*), and transcriptional regulation and RNA processing (*HNRNPUL1, RUNX1T1, ELAVL4*) (Table 2). This suggests that the Y-sperm has robust metabolic rate, ion transport or membrane dynamics than X-sperm. Upregulation of the above genes in Y-sperm may explain their higher velocity compared to X-sperm[56]. In contrast, genes downregulated in Y-sperm are associated with structural components (*NUP214, EMILIN2, MINPP1, LRRC27,*

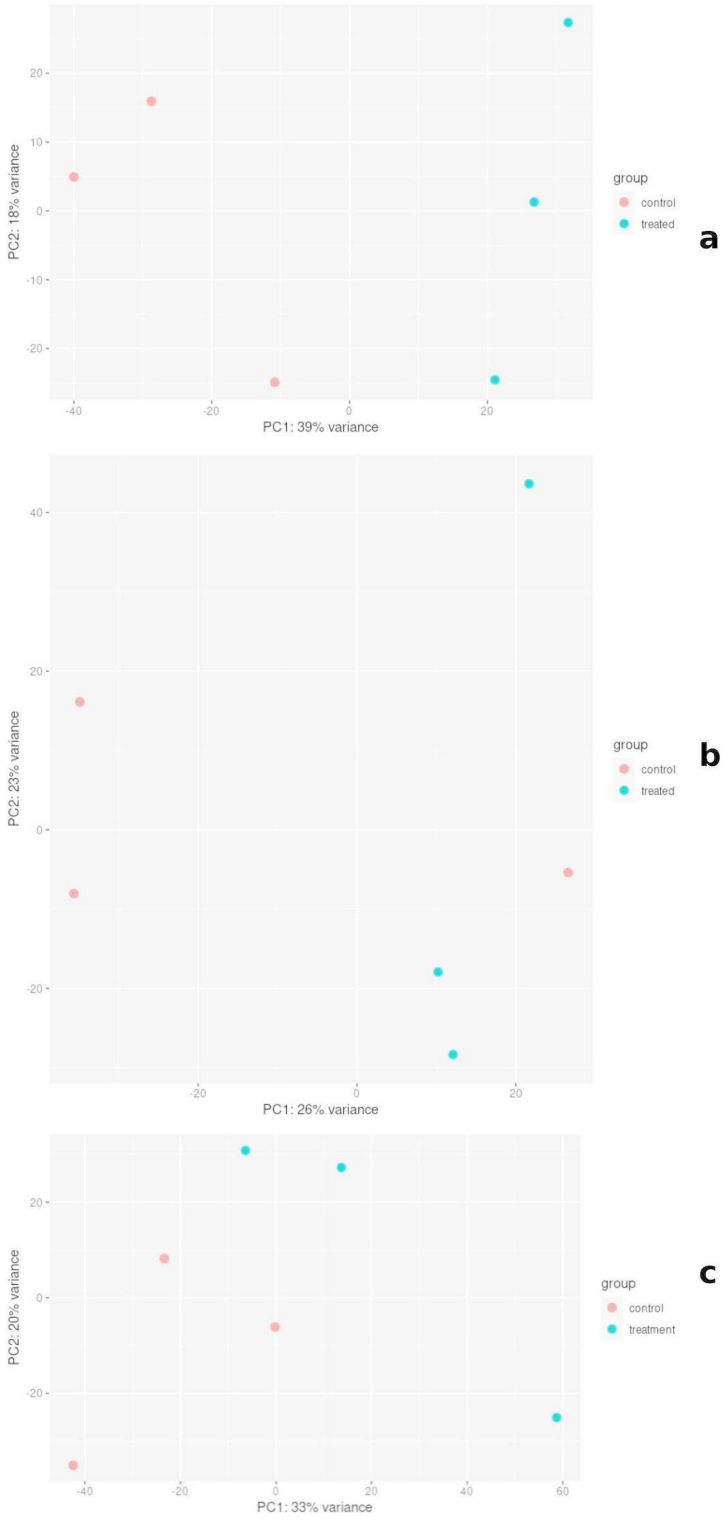

Fig. 3. Principal component analysis (PCA) plots generated from DeSeq2 showing variation within and between groups. (**a**) Unsorted (control) and X-sperm (treatment). (**b**) Unsorted (control) and Y-sperm (treatment). (**c**) X-(control) and Y-sperm (treatment).

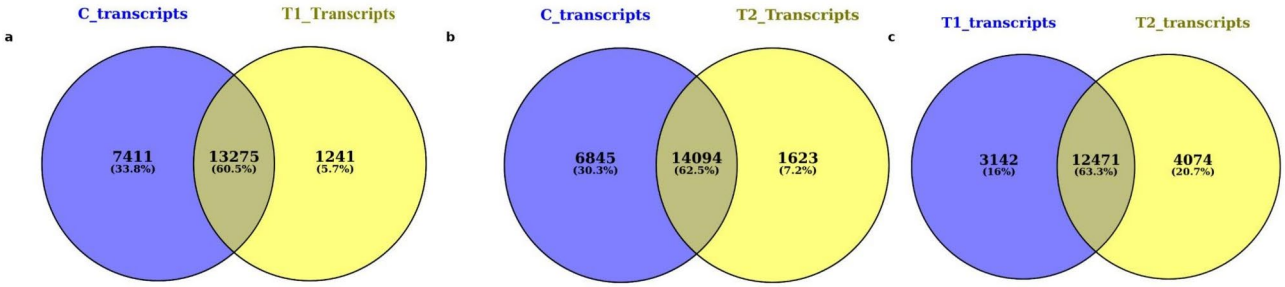

Fig. 4. Venn diagram representing the sperm transcripts in (**a**) Unsorted (C) and X-sperm (T1), (**b**) Unsorted (C) and Y-sperm (T2), and (**c**) X- (T1) and Y-sperm (T2).

DZIP1L) and cellular stress response (*NAPRT, MVP, CA14*). This suggests that Y-sperm may have lower stress resilience as they are structurally less robust, making them more vulnerable to unfavorable conditions[58].

Conclusion

We deciphered the transcriptomes of X- and Y-chromosome-bearing sperm of indicus cattle (*Bos indicus*) and identified and validated the differential transcripts between them. Among the transcripts found unique to Y-sperm, 53 genes were located on the *Bos indicus* Y chromosome, while among the transcripts unique to X-sperm, 200 genes were located on the *Bos indicus* X chromosome. Gene Ontology (GO) analysis revealed that unique genes in cattle X-sperm are mostly associated with cellular components, whereas the unique Y-sperm genes are more involved in molecular functions. Pathway analysis revealed that unique cattle Y-sperm genes are involved in critical signalling pathways, such as calcium and cAMP signalling, which are vital for sperm motility and fertility. Our omics data will help biomarker discovery for sex sorting of bovine semen.

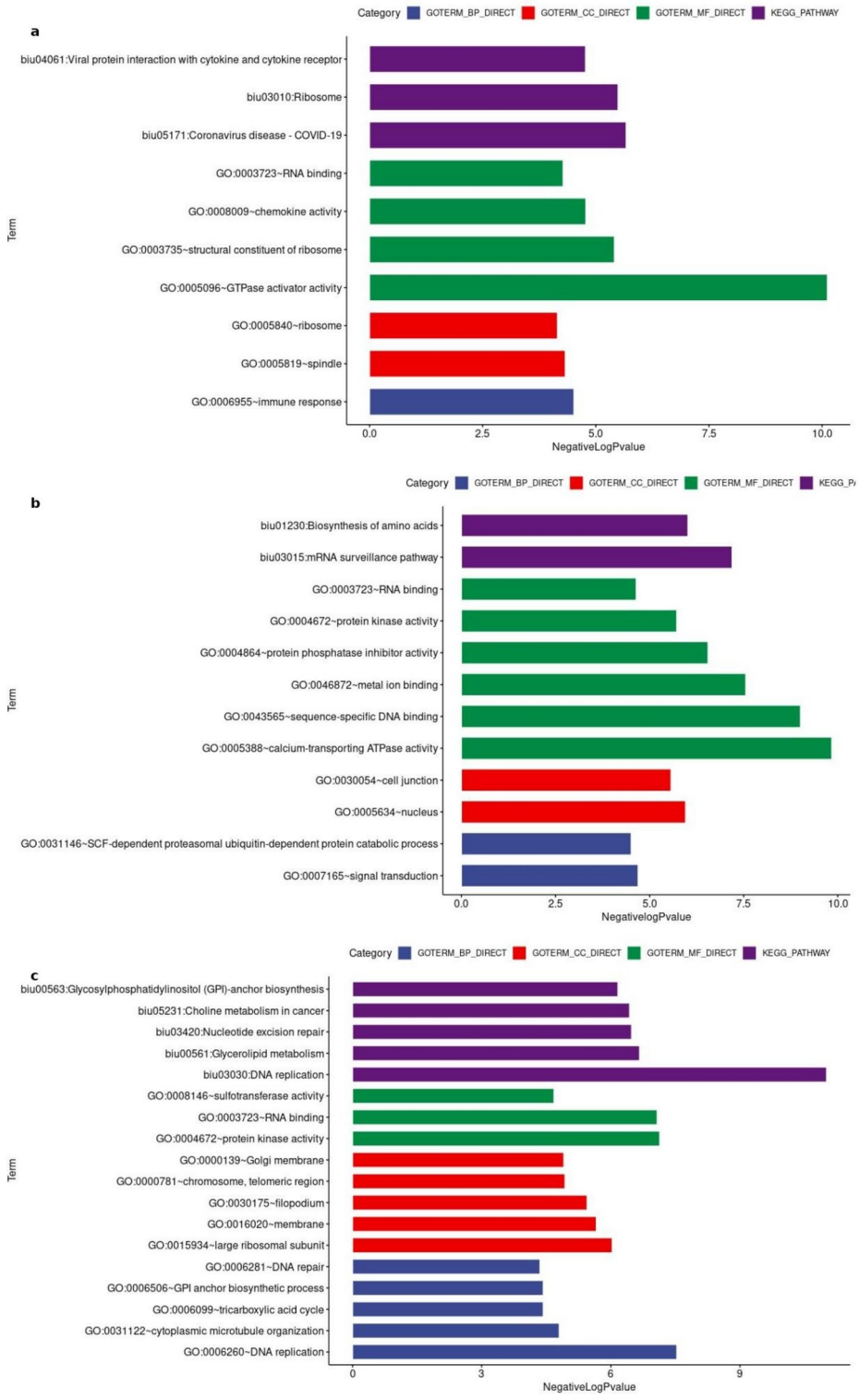

Fig. 5. Top Gene Ontology terms in different categories: (**a**) unsorted and X-sperm, (**b**) unsorted and Y-sperm, (**c**) X- and Y-sperm.

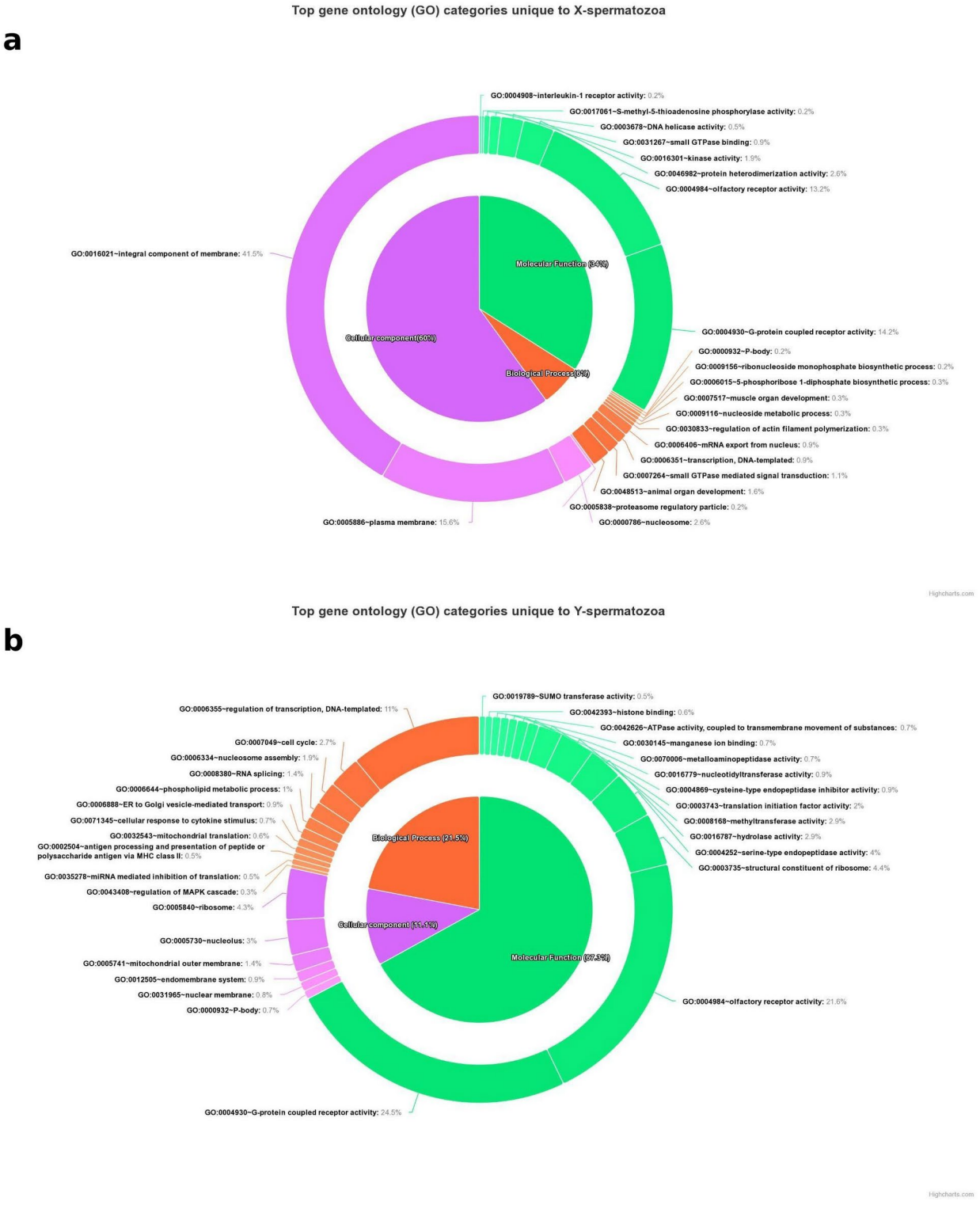

Fig. 6. Top Gene Ontology terms in different categories: (**a**) unique to X-sperm, and (**b**) unique to Y-sperm.

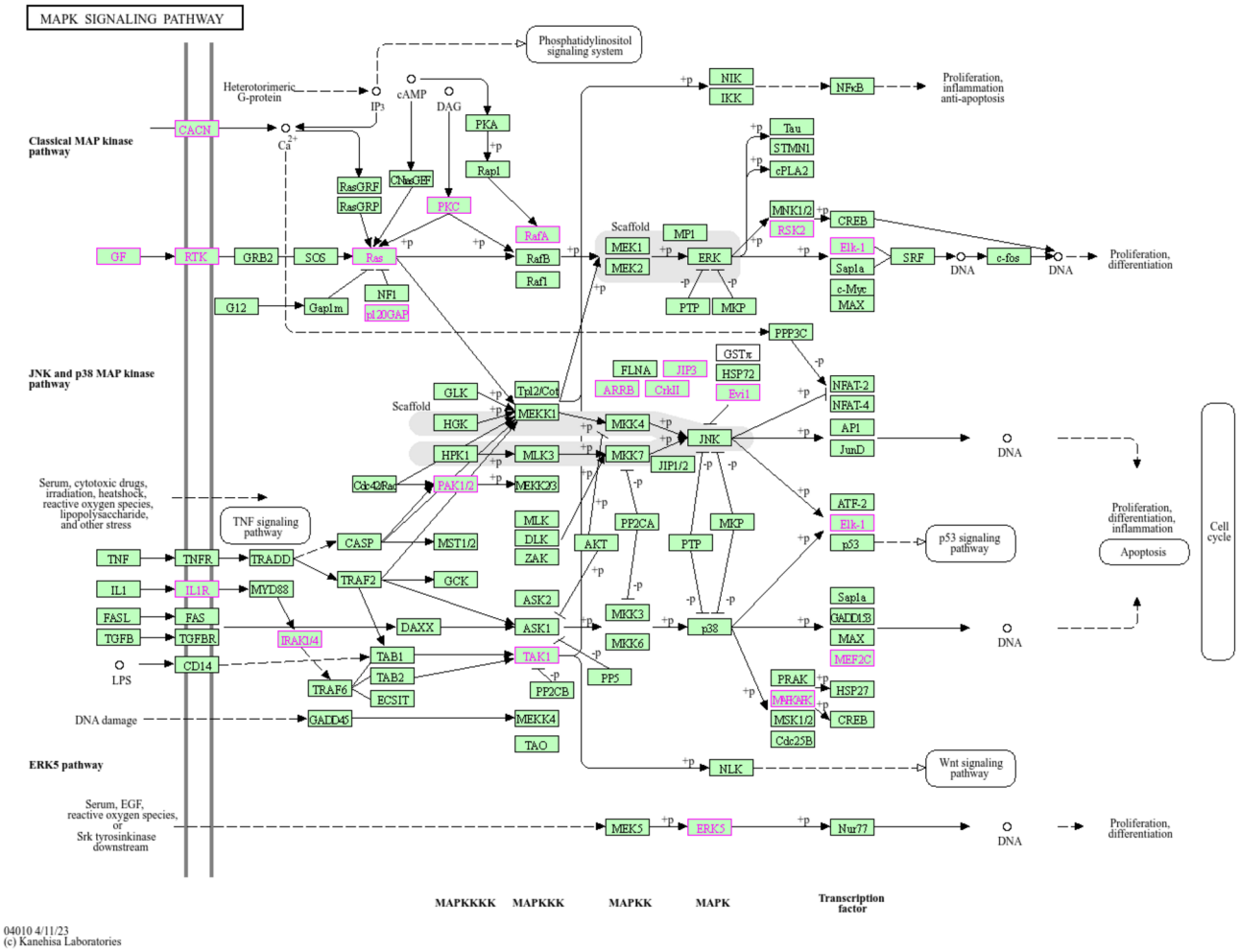

Fig. 7. MAPK signalling pathway (KEGG:04,010, 31 counts; https://www.kegg.jp/pathway/map04010) with transcripts unique to X-sperm (ARAF, ARRB1, CACNG1, CACNG2, CRK, ELK1, FGF10, FGF13, FGF14, FGF2, FGFR1, FGFR3, IKBKG, IL1RAP, IRAK1, KIT, KRAS, MAP3K7, MAPK7, MAPK8IP3, MAPKAPK2, MECOM, MEF2C, PAK1, PDGFRA, PGF, PRKCB, RASA1, RPS6KA2, RPS6KA3, and VEGFC)[57].

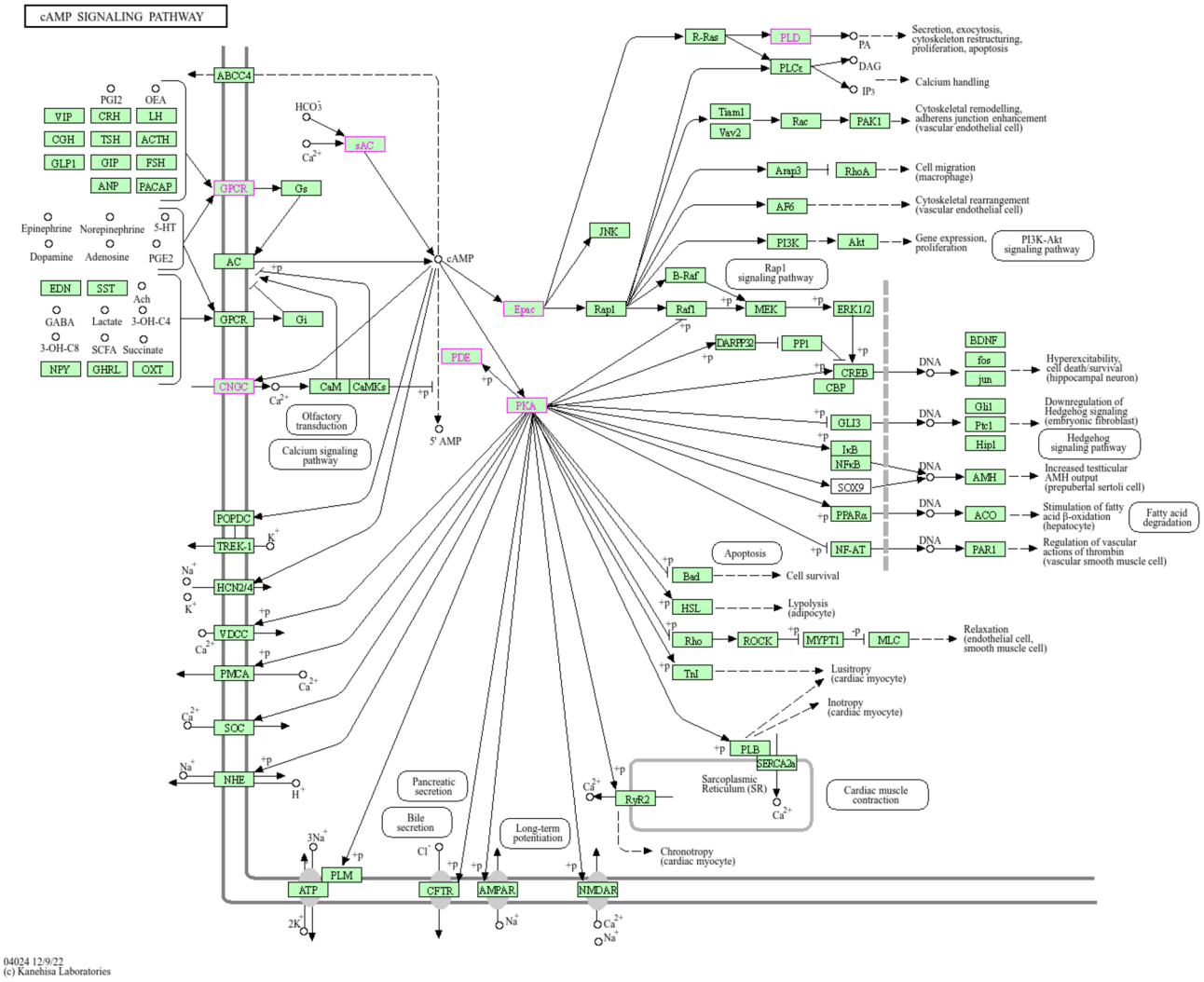

Fig. 8. cAMP signalling pathway (KEGG:04,024, 7 counts; https://www.kegg.jp/pathway/map04024) with transcripts unique to Y-sperm (ADCY10, CNGA1, FSHR, PDE4D, PLD1, PRKACB, and RAPGEF4)[57].

Sl. no	Gene	Gene name	Biological significance
1	D2HGDH	D-2-Hydroxyglutarate Dehydrogenase	D2HGDH is a mitochondrial enzyme that catalyses the oxidation of D-2-HG to α-KG[24].
2	ATP13A2	ATPase Cation Transporting 13A2	Polyamine-transporting ATPase 13A2 is highly expressed gene in spermatocytes and spermatids with possible role in male fertility[25].
3	PKLR	Pyruvate Kinase L/R	This gene catalyses the transphosphorylation of phosphoenolpyruvate into pyruvate and ATP, which is the rate-limiting step of glycolysis. Pyruvate kinase M in germ cells is essential for sperm motility and male fertility[26].
4	TMEM143	Transmembrane Protein 143	Functions as a critical signalling component in mediating NF-kappa-B activation. Limited information is available regarding TMEM143's function in sperm or reproductive processes[27].
5	CHST11	Carbohydrate Sulfotransferase 11	CHST11 is suggested to play an important role in cartilage and bone development in embryos[28].
6	HNRNPUL1	Heterogeneous Nuclear Ribonucleoprotein U-Like 1	This gene encodes a nuclear RNA-binding protein of the heterogeneous nuclear ribonucleoprotein (hnRNP) family. hnRNP family proteins in adult male germ cells are highly expressed in the male reproductive system, suggesting their significant function in male germ cell development[29].
7	RUNX1T1	RUNX1 Translocation Partner 1	This gene encodes a member of the myeloid translocation gene family which interact with DNA-bound transcription factors and recruit a range of corepressors to facilitate transcriptional repression. This gene was found to be expressed in low-fertile buffalo bull sperm suggesting its involvement in fertility[30].
8	TMEM168	Transmembrane Protein 168	Transmembrane proteins in sperm play a critical role in sperm-oocyte interaction and fusion, essential for fertilization, by mediating binding, signalling, and ultimately, the fusion of sperm and egg membranes[31,32]. The role of *TMEM168* in sperm or reproductive biology has not been characterized yet
9	ELAVL4	ELAV Like RNA Binding Protein 4	Enables mRNA 3'-UTR AU-rich region binding activity; poly(A) binding activity; and pre-mRNA intronic pyrimidine-rich binding activity. ELAVL4/HuD ameliorates Alzheimer's disease-related molecular changes[33]
10	SPATA21	Spermatogenesis Associated 21	Spermatogenesis-associated (SPATA) genes are correlated with infertility[34]. However, there is an absence of specific investigations elucidating the role of SPATA21
11	NAPRT	Nicotinate Phosphoribosyltransferase	Nicotinic acid (NA; niacin) is converted by nicotinic acid phosphoribosyl transferase (NAPRT; EC 2.4.2.11) to NA mononucleotide (NaMN), which is then converted to NA adenine dinucleotide (NaAD) and finally to NAD, which is a coenzyme in cellular redox reactions and essential to many cellular metabolism processes, including stress response. Nicotinamide phosphoribosyl transferase levels are found to be increased in immature sperm compared with mature sperm[35,36].
12	CA14	Carbonic Anhydrase 14	Carbonic Anhydrases II and IV in sperm are key enzymes in the regulation of sperm motility and essential for the HCO3 – -mediated beat frequency increase during early sperm activation[37]
13	VWC2	Von Willebrand Factor C Domain Containing 2	VWC2 (Von Willebrand Factor C Domain Containing 2) is a protein coding gene having role in sperm motility by regulating REC8 in chicken[38]
14	NUP214	Nucleoporin 214	The nuclear pore complex is a massive structure that extends across the nuclear envelope, forming a gateway that regulates the flow of macromolecules between the nucleus and the cytoplasm. Nucleoporins are the main components of the nuclear pore complex in eukaryotic cells. These complexes have a crucial role as spermatogenesis leads to a reduced nuclear pore structure and function[39]
15	EMILIN2	Elastin Microfibril Interfacer 2	Expressed in a variety of tissues, including the early conceptus and embryo mesenchyme and is predicted to enable extracellular matrix constituent conferring elasticity[27].
16	MVP	Major Vault Protein	This gene encodes the major component of the vault complex. The encoded protein may play a role in multiple cellular processes by regulating the MAP kinase, JAK/STAT and phosphoinositide 3-kinase/Akt signalling pathways[27].
17	MINPP1	Multiple Inositol-Polyphosphate Phosphatase 1	This gene encodes multiple inositol polyphosphate phosphatase; an enzyme that removes 3-phosphate from inositol phosphate substrates. Potentially related to the decreased efficiency of sperm storage tubules with aging[40].
18	ZNF341	Zinc Finger Protein 341	Enables DNA binding activity and DNA-binding transcription activator activity. The zinc finger gene family represents one of the largest in the mammalian genome, with several of these genes reported to be involved in spermatogenesis[41].
19	DZIP1L	Leucine Rich Repeat Containing 27	Involved in primary cilium formation and have a possible anchoring role in sperm flagellum formation together with Fam92[42].
20	LRRC27	XM_019953140.1	Predicted to be active in cytoskeleton. The Leucine-rich repeat protein (LRRC) family of genes have been associated with male fertility (LRRC46). However, LRRC27 have not been characterized yet[43].
21	CMKLR1	Chemokine Like Receptor 1	The CMKLR1 gene, which encodes the chemokine-like receptor 1 (ChemR23), is implicated in reproduction, particularly within the chemerin signalling system that influences energy metabolism and female reproductive activities, potentially affecting male fertility[44].
22	RBM20	RNA Binding Motif Protein 20	The RBM20 gene encodes an RNA-binding protein that modulates mRNA splicing, especially in cardiac and skeletal muscles, with mutations associated with dilated cardiomyopathy (DCM) and possibly hypertrophic cardiomyopathy (HCM). The characterisation and function of RBM20 gene in sperm are still unknown[27,45].

Table 2. The biological significance of the differentially expressed genes validated through qPCR.

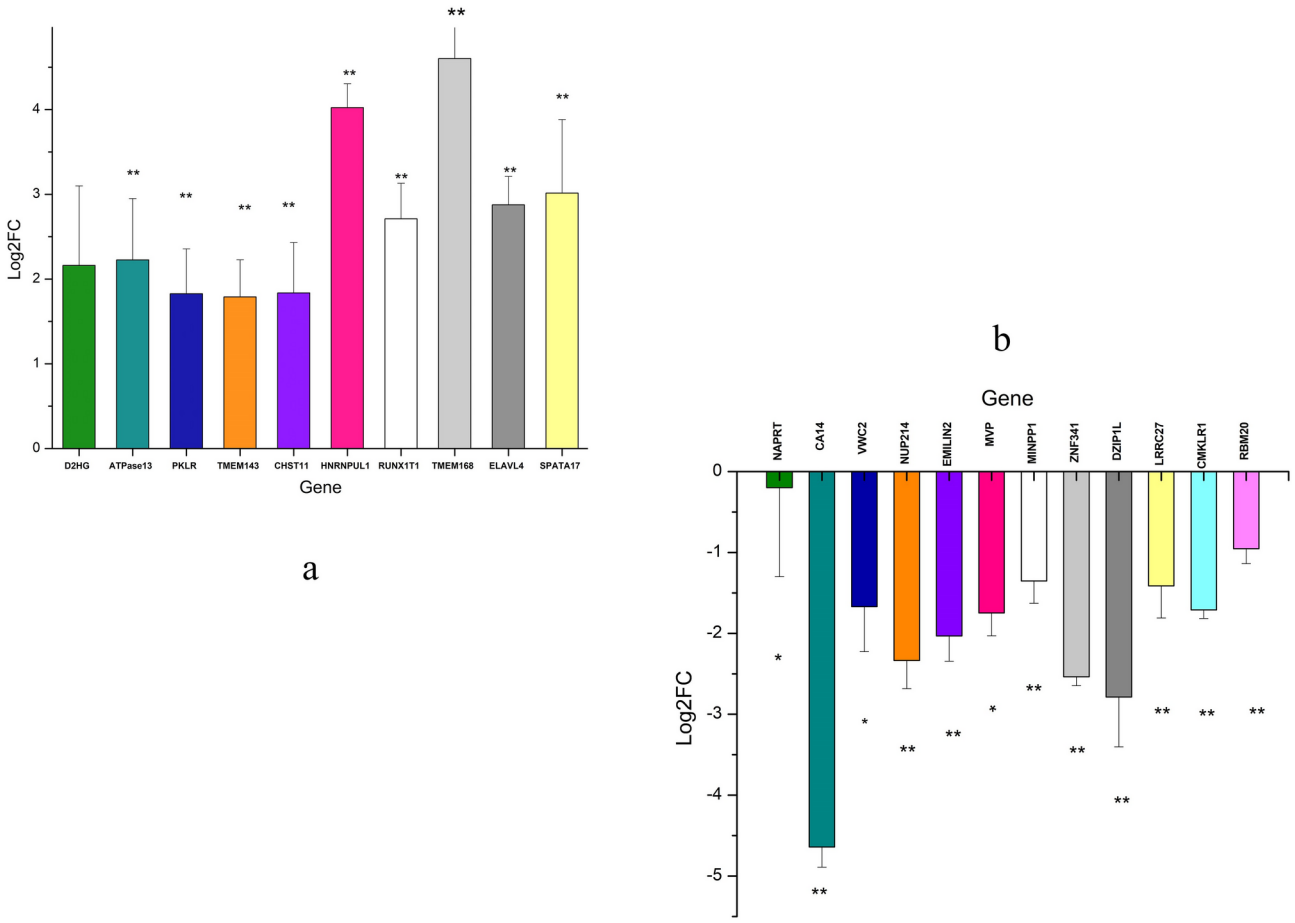

Fig. 9. Log2-fold change in the expression of selected (**a**) upregulated and (**b**) downregulated genes in bull Y-sperm (in comparison to X-sperm). $^{*}p < 0.05$; $^{**}p < 0.01$.

Data availability

Sequence data that support the findings of this study have been deposited in the National Center for Biotechnology Information (NCBI) with the BioProject code PRJNA976949, and other data are provided within the manuscript and supplementary files.

Received: 28 August 2024; Accepted: 21 April 2025
Published online: 26 April 2025

References

1. Cerchiaro, I., Cassandro, M., Dal Zotto, R., Carnier, P. & Gallo, L. A field study on fertility and purity of sex-sorted cattle sperm. *J. Dairy Sci.* **90**(5), 2538–2542. https://doi.org/10.3168/jds.2006-694 (2007).
2. Holden, S. A. & Butler, S. T. Review: Applications and benefits of sexed semen in dairy and beef herds. *Animal* **12**, 97–103. https://doi.org/10.1017/S1751731118000721 (2018).
3. Ruelle, E., Shalloo, L. & Butler, S. T. Economic impact of different strategies to use sex-sorted sperm for reproductive management in seasonal-calving, pasture-based dairy herds. *J. Dairy Sci.* **104**(11), 11747–11758. https://doi.org/10.3168/jds.2021-20150 (2021).
4. Wolfner, M. F., Suarez, S. S. & Dorus, S. Suspension of hostility: Positive interactions between spermatozoa and female reproductive tracts. *Andrology* **11**(5), 943–947 (2023).
5. Miller, D. J. The epic journey of sperm through the female reproductive tract. *Animal* **12**, 110–120 (2018).
6. Prakash, M. A. et al. RNA-Seq analysis reveals functionally relevant coding and non-coding RNAs in crossbred bull spermatozoa. *Anim. Reprod. Sci.* **222**, 106621. https://doi.org/10.1016/j.anireprosci.2020.106621 (2020).
7. Indriastuti, R. et al. Sperm transcriptome analysis accurately reveals male fertility potential in livestock. *Animals* **12**(21), 2955. https://doi.org/10.3390/ani12212955 (2022).
8. Selvaraju, S. et al. Occurrence and functional significance of the transcriptome in bovine (*Bos taurus*) spermatozoa. *Sci. Rep.* **7**, 42392. https://doi.org/10.1038/srep42392 (2017).
9. Raval, N. P., Shah, T. M., George, L. B. & Joshi, C. G. Insight into bovine (*Bos indicus*) spermatozoal whole transcriptome profile. *Theriogenology* **129**, 8–13. https://doi.org/10.1016/j.theriogenology.2019.01.037 (2019).
10. Chen, X. et al. Identification and characterization of genes differentially expressed in X and Y sperm using suppression subtractive hybridization and cDNA microarray. *Mol. Reprod. Dev.* **81**(10), 908–917. https://doi.org/10.1002/mrd.22386 (2014).
11. Umehara, T., Tsujita, N. & Shimada, M. Activation of Toll-like receptor 7/8 encoded by the X chromosome alters sperm motility and provides a novel simple technology for sexing sperm. *PLoS Biol.* **17**(8), 3000398. https://doi.org/10.1371/journal.pbio.3000398 (2019).

12. Laxmivandana, R., Patole, C., Sharma, T. R., Sharma, K. K. & Naskar, S. Differential proteins associated with plasma membrane in X- and/or Y-chromosome bearing spermatozoa in indicus cattle. *Reprod. Domest. Anim.* **56**(6), 928–935. https://doi.org/10.1111/rda.13936 (2021).
13. Department of Animal Husbandry and Dairying, Ministry of Fisheries, Animal Husbandry and Dairying, Government of India (2022). Minimum Standards required for production of Frozen Semen., https://dahd.gov.in/sites/default/files/2024-04/MSP-2022-for-production-of-bovine-frozen-semen.pdf. Accessed 15 Feb, 2025.
14. Johnson, L. A., Welch, G. R. & Rens, W. The Beltsville sperm sexing technology: High speed sperm sorting gives improved sperm output for in vitro fertilization and A.I. *J. Anim. Sci.* **77**(2), 213–220. https://doi.org/10.2527/1999.77suppl_2213x (1999).
15. Umar, S. I. et al. Development and optimization of an efficient RNA isolation protocol from bovine (*Bos indicus*) spermatozoa. *BB Reports.* **40**, 101862. https://doi.org/10.1016/j.bbrep.2024.101862 (2024).
16. Andrews S. FastQC: A quality control tool for high throughput sequence data (2010). http://www.bioinformatics.babraham.ac.uk/projects/fastqc
17. Chen, S., Zhou, Y., Chen, Y. & Gu, J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**(17), 884–890. https://doi.org/10.1093/bioinformatics/bty560 (2018).
18. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**(8), 907–915. https://doi.org/10.1038/s41587-019-0201-4 (2019).
19. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**(3), 290–295. https://doi.org/10.1038/nbt.3122 (2015).
20. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* https://doi.org/10.1186/s13059-014-0550-8 (2014).
21. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**(1), 44–57. https://doi.org/10.1101/gr.1239303 (2009).
22. Shannon, P. et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**(11), 2498–2504. https://doi.org/10.1101/gr.1239303 (2003).
23. Prakash, M. A. et al. Comparative transcriptomic analysis of spermatozoa from high- and low-fertile crossbred bulls: Implications for fertility prediction. *Front. Cell Dev. Biol.* https://doi.org/10.3389/fcell.2021.647717 (2021).
24. Yang, J., Zhu, H., Zhang, T. & Ding, J. Structure, substrate specificity, and catalytic mechanism of human D-2-HGDH and insights into pathogenicity of disease-associated mutations. *Cell Discov.* **7**(1), 3 (2021).
25. Park, Y. J. & Pang, M. G. Mitochondrial functionality in male fertility: From spermatogenesis to fertilization. *Antioxidants.* **10**(1), 98 (2021).
26. Qian, G. Q., Wang, X. C., Zhang, X., Shen, B. & Liu, Q. Pyruvate kinase M in germ cells is essential for sperm motility and male fertility but not spermatogenesis. *Asian J. Androl.* **26**(2), 212–219 (2024).
27. National Center for Biotechnology Information (NCBI). Gene [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/gene. Accessed 21 Mar 2025.
28. Townsend, J., Braz, C. U., Taylor, T. & Khatib, H. Effects of paternal methionine supplementation on sperm DNA methylation and embryo transcriptome in sheep. *Environ. Epigenet.* **9**(1), dvac029 (2023).
29. Wang, X. L., Li, J. M. & Yuan, S. Q. Characterization of the protein expression and localization of hnRNP family members during murine spermatogenesis. *Asian J. Androl.* **25**(3), 314–321 (2023).
30. Paul, N. et al. Transcriptomic profiling of buffalo spermatozoa reveals dysregulation of functionally relevant mRNAs in low-fertile bulls. *Front. Vet. Sci.* **7**, 609518 (2021).
31. Hao, H. et al. The vertebrate-and testis-specific transmembrane protein C11ORF94 plays a critical role in sperm-oocyte membrane binding. *Mol. Biomed.* **3**(1), 27 (2022).
32. Hernández-Falcó, M., Sáez-Espinosa, P., López-Botella, A., Aizpurua, J. & Gómez-Torres, M. J. The role of sperm proteins IZUMO1 and TMEM95 in mammalian fertilization: A systematic review. *Int. J. Mol. Sci.* **23**(7), 3929 (2022).
33. Van der Linden, R. J. et al. RNA-binding protein ELAVL4/HuD ameliorates Alzheimer's disease-related molecular changes in human iPSC-derived neurons. *Prog. Neurobiol.* **217**, 102316 (2022).
34. Sujit, K. M. et al. Increased DNA methylation in the spermatogenesis-associated (SPATA) genes correlates with infertility. *Andrology* **8**(3), 602–609 (2020).
35. Riammer, S. et al. Nicotinamide phosphoribosyltransferase production in human spermatozoa is influenced by maturation stage. *Andrology* **4**(6), 1045–1053 (2016).
36. Agliano, F., Karginov, T. A., Ménoret, A., Provatas, A. & Vella, A. T. Nicotinamide breaks effector CD8 T cell responses by targeting mTOR signaling. *iScience* **25**(3), 103932 (2022).
37. Wandernoth, P. M. et al. Normal fertility requires the expression of carbonic anhydrases II and IV in sperm. *J. Biol. Chem.* **290**(49), 29202–29216 (2015).
38. Du, Z. et al. Transcriptome analysis of the testes of male chickens with high and low sperm motility. *Poult. Sci.* **101**(12), 102183 (2022).
39. Santos, Á. D. et al. Human spermatogenesis leads to a reduced nuclear pore structure and function. *BioRxiv.* **30**, 10 (2024).
40. Yang, L. et al. Transcriptome analysis and identification of age-associated fertility decreased genes in hen uterovaginal junction. *Poult. Sci.* **100**(3), 100892 (2021).
41. Zhang, S. et al. The shorter zinc finger protein ZNF230 gene message is transcribed in fertile male testes and may be related to human spermatogenesis. *Biochem. J.* **359**(3), 721–727 (2001).
42. Lapart, J. A. et al. Dzip1 and Fam92 form a ciliary transition zone complex with cell type specific roles in drosophila. *Elife* **8**, e49307 (2019).
43. Yin, Y. et al. LRRC46 accumulates at the midpiece of sperm flagella and is essential for spermiogenesis and male fertility in mouse. *Int. J. Mol. Sci.* **23**(15), 8525 (2022).
44. Estienne, A. et al. Chemerin impairs in vitro testosterone production, sperm motility, and fertility in chicken: Possible involvement of its receptor CMKLR1. *Cells* **9**(7), 1599 (2020).
45. Zhang, Y. et al. Disruption of the nuclear localization signal in RBM20 is causative in dilated cardiomyopathy. *JCI insight.* **8**(13), e170001 (2023).
46. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCt method. *Methods* **25**(4), 402–408. https://doi.org/10.1006/meth.2001.1262 (2001).
47. Johnson, L. A., Flook, J. P. & Hawk, H. W. Sex preselection in rabbits: Live births from X and Y sperm separated by DNA and cell sorting. *Biol. Reprod.* **41**, 199–203 (1989).
48. Fawcett, D., Ito, S. & Slautterback, D. The occurrence of intercellular bridges in groups of cells exhibiting synchronous differentiation. *J. Cell Biol.* **5**(3), 453–460. https://doi.org/10.1083/jcb.5.3.453 (1959).
49. Ventelä, S., Toppari, J. & Parvinen, M. Intercellular organelle traffic through cytoplasmic bridges in early spermatids of the rat: Mechanisms of haploid gene product sharing. *Mol. Biol. Cell.* **14**, 2768–2780. https://doi.org/10.1091/mbc.e02-10-0647 (2003).
50. Sharma, V., Verma, A. K., Sharma, P., Pandey, D. & Sharma, M. Differential proteomic profile of X- and Y- sorted Sahiwal bull semen. *Res. Vet. Sci.* **144**, 181–189. https://doi.org/10.1016/j.rvsc.2021.11.013 (2022).
51. Zhou, H. et al. Differences in small noncoding RNAs profile between bull X and Y sperm. *PeerJ* https://doi.org/10.7717/peerj.9822 (2020).

52. Yang, D. et al. *G protein-coupled receptors: Structure- and function-based drug discovery signal transduction and targeted therapy* Vol. 6 (Springer Nature, 2021). https://doi.org/10.1038/s41392-020-00435-w.
53. Talluri, T. R. et al. Integrated multi-omics analyses reveals molecules governing sperm metabolism potentially influence bull fertility. *Sci. Rep.* https://doi.org/10.1038/s41598-022-14589-w (2022).
54. Chiarante, N. et al. Cyclic AMP efflux through MRP4 regulates actin dynamics signalling pathway and sperm motility in bovines. *Sci. Rep.* https://doi.org/10.1038/s41598-020-72425-5 (2020).
55. Buffone, M. G., Wertheimer, E. V., Visconti, P. E. & Krapf, D. Central role of soluble adenylyl cyclase and cAMP in sperm physiology. *Biochim. Biophys. Acta Mol. Basis Dis.* **1842**, 2610–2620. https://doi.org/10.1016/j.bbadis.2014.07.013 (2014).
56. Cui, K. H. & Matthews, C. D. X larger than Y. *Nature* **366**(6451), 117–118. https://doi.org/10.1038/366117b0 (1993).
57. Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y. & Ishiguro-Watanabe, M. KEGG: Biological systems database as a model of the real world. *Nucleic Acids Res.* **53**(D1), D672–D677 (2025).
58. You, Y. A. et al. Sex chromosome-dependent differential viability of human spermatozoa during prolonged incubation. *Hum. Reprod.* **32**(6), 1183–1191. https://doi.org/10.1093/humrep/dex080 (2017).

Acknowledgements

The authors thank ICAR-Indian Institute of Agricultural Biotechnology, Ranchi for all-round support; Paschim Banga Go-Sampad Bikash Sanstha (a Govt. of West Bengal organization), Kolkata for experimental samples; Bionivid Technology Private Limited, Bengaluru for the sequencing services; and Dr. Laxmivandana Rongala and Shashi Kumar for their technical support.

Author contributions

SIUU: Investigation, Formal analysis, Software, Validation and Original draft preparation. SP: Resources and Supervision (to SIUU). SN: Conceptualization, methodology, writing, review and editing, funding acquisition, and supervision. PJD: Methodology, Review and editing. MS, AP, VPB: Resources. DKM: Investigation. SR: Review and Resources.

Funding

The work is carried out with funding support received from the Science and Engineering Research Board (Dept. of Science and Technology, Govt. of India) through research projects 2019/000437 and 2023/000377 (to SN), and ICAR-Indian Institute of Agricultural Biotechnology, Ranchi.

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval and consent to participate

The experimental protocol was approved by ICAR-Indian Institute of Agricultural Biotechnology, Ranchi *vide* PF/18/2016/07052019.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-99438-2.

Correspondence and requests for materials should be addressed to S.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.