

REVIEW

Open Access

Long non-coding RNAs (lncRNAs) in spermatogenesis and male infertility



Meghali Joshi and Singh Rajender*

Abstract

Background: Long non-coding RNAs (lncRNAs) have a size of more than 200 bp and are known to regulate a host of crucial cellular processes like proliferation, differentiation and apoptosis by regulating gene expression. While small noncoding RNAs (ncRNAs) such as miRNAs, siRNAs, Piwi-interacting RNAs have been extensively studied in male germ cell development, the role of lncRNAs in spermatogenesis remains largely unknown.

Objective: In this article, we have reviewed the biology and role of lncRNAs in spermatogenesis along with the tools available for data analysis.

Results and conclusions: Till date, three microarray and four RNA-seq studies have been undertaken to identify lncRNAs in mouse testes or germ cells. These studies were done on pre-natal, post-natal, adult testis, and different germ cells to identify lncRNAs regulating spermatogenesis. In case of humans, five RNA-seq studies on different germ cell populations, including two on sperm, were undertaken. We compared three studies on human germ cells to identify common lncRNAs and found 15 lncRNAs (LINC00635, LINC00521, LINC00174, LINC00654, LINC00710, LINC00226, LINC00326, LINC00494, LINC00535, LINC00616, LINC00662, LINC00668, LINC00467, LINC00608, and LINC00658) to show consistent differential expression across these studies. Some of the targets of these lncRNAs included CENPB, FAM98B, GOLGA6 family, RPGR, TPM2, GNB5, KCNQ10T1, TAZ, LIN28A, CDKN2B, CDKN2A, CDKN1A, CDKN1B, CDKN1C, EZH2, SUZ12, VEGFA genes. A lone study on human male infertility identified 9879 differentially expressed lncRNAs with three (lnc32058, lnc09522, and lnc98497) of them showing specific and high expression in immotile sperm in comparison to normal motile sperm. A few lncRNAs (Mrhl, Drm, Spga-lncRNAs, NLC1-C, HongrES2, Tsx, lncRNA-tcam1, Tug1, Tesra, AK015322, Gm2044, and lncRNA033862) have been functionally validated for their roles in spermatogenesis. Apart from rodents and humans, studies on sheep and bull have also identified lncRNAs potentially important for spermatogenesis. A number of these non-coding RNAs are strong candidates for further research on their roles in spermatogenesis.

Keywords: Long non-coding RNA, lncRNA, lncRNA regulation, lncRNA databases, Transcriptome, Spermatogenesis, Male infertility

* Correspondence: rajender_singh@cdri.res.in

Division of Endocrinology, Central Drug Research Institute, Lucknow, UP, India



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

According to the central dogma, DNA is transcribed into mRNA, which is translated into protein, and the latter perform most of the functional, structural, and regulatory roles in various cellular processes. Nevertheless, it was found that protein coding genes in both lower and higher organisms are similar and are not enough to account for the complexity of higher organisms, hence challenging the protein centric view. With the advent of high-throughput technology, it was found that the majority of the genome in humans is transcribed and out numbers the current gene annotations. Moreover, a majority of these transcripts do not get translated into proteins. These transcripts which are not translated into proteins are known as the noncoding RNAs (ncRNAs) [1, 2]. Earlier, ncRNAs were considered as ‘junk’, but we realized the importance of this junk chunk only recently [3]. Non-coding RNAs can be classified on the basis of their size. Non-coding RNAs having size less than 200 bp are called small non-coding RNAs (sncRNAs) and those above 200 bp are called long non-coding RNAs (lncRNAs) [4]. MicroRNA (miRNA), small interfering RNAs (siRNAs) and Piwi-interacting RNA (piRNA) come under the category of sncRNAs (Fig. 1). Several studies have documented the role of small non-coding

RNAs in male germ cell development [5, 6]. Research on lncRNAs has picked up the pace only recently with reports on the role of lncRNAs in regulating important genes in human and rodent embryonic stem cell (ESC) models [7, 8]. LncRNAs perform a wide variety of functions in mammals; for instance, they regulate the genes involved in chromatin modification, inactivate the X-chromosome by direct binding, inhibit antisense mRNA expression, promote antisense mRNA degradation, act as competing endogenous RNA (ceRNAs) to inhibit the function of miRNAs and bind to the DNA binding proteins to inhibit their interaction with the target genes [9].

Long non-coding RNAs have been reported to regulate a variety of cellular processes such as differentiation, proliferation and apoptosis. The differential expression of lncRNAs in human diseases highlights their roles in various biological processes. It has been reported that three lncRNAs: Peril, Evf2 (Dlx6os1) and linc-Brn1b are important for neural development [10, 11]. Several lncRNAs play important role in the development of organs during embryogenesis, such as lncRNA-Fendrr, which is transcribed divergently from the transcription factor Foxf1. The mouse knock-out model of Fendrr was seen to have defects in lungs, heart and gastrointestinal

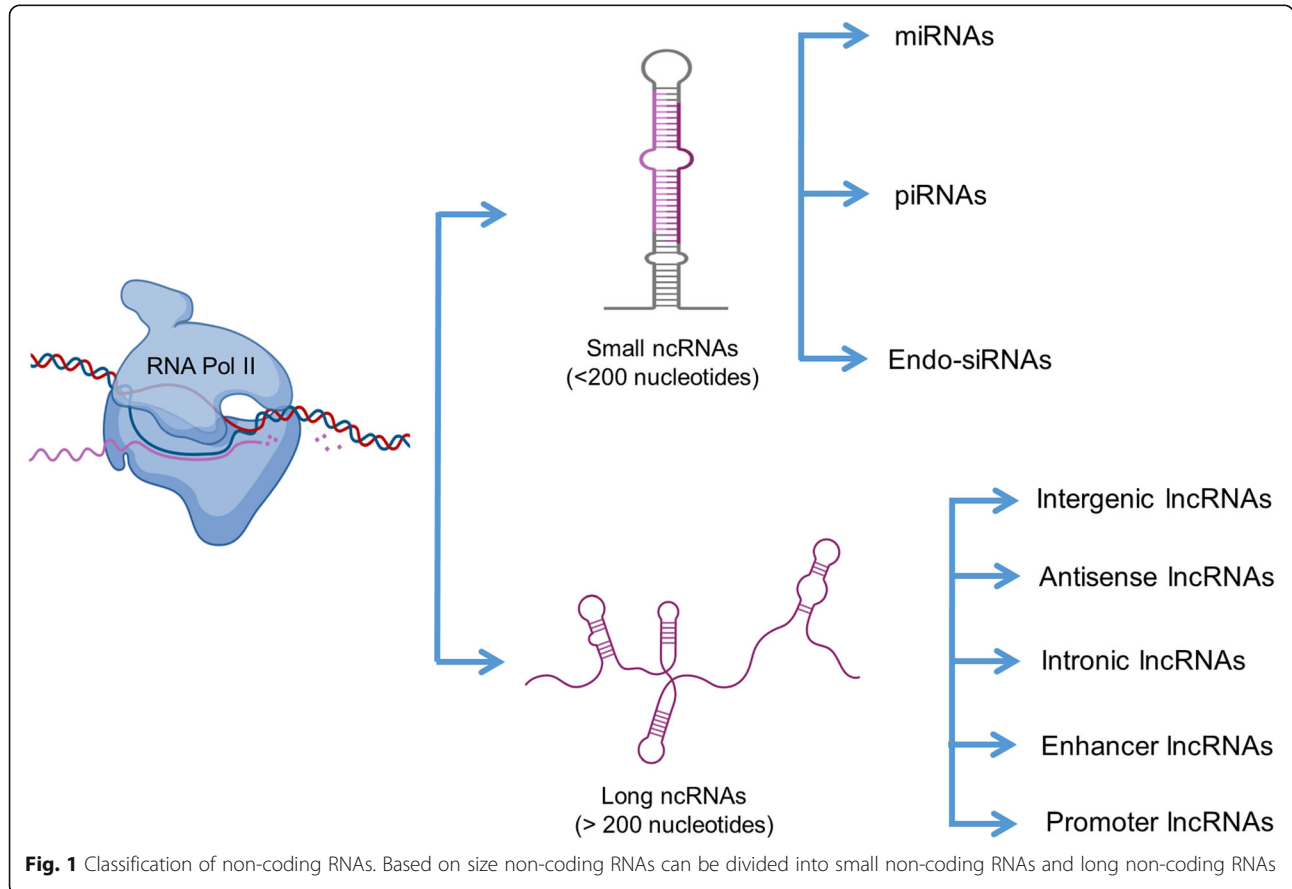


Fig. 1 Classification of non-coding RNAs. Based on size non-coding RNAs can be divided into small non-coding RNAs and long non-coding RNAs

tract [11]. Further, it was found that at the embryonic stage E13.5, the developing lungs of the KO mouse had smaller size with globular and disorganised lobes. These KO mice die soon after birth due to breathing problems [12]. Several lncRNAs are reported to play crucial role in stem cell maintenance and differentiation. For example, RMST [13], Miat [14], Tunar [15] are reported to have roles in neural differentiation. LncRNAs like Braveheart [16], Alien [17], SENCER [18] and hoxBln1 [19] are required for cardiac, endoderm, endothelial and hematopoietic differentiation, respectively. A large number of lncRNAs are involved in a variety of cancers, such as ANRIL [20], CARLo-5 [21], *FALI* [22], *GAS5* [23], HOTAIR [24], HOTTIP [25], PCGEM1 [26], PCA3 [27]. The diversity of lncRNAs and their biological functions are taking giant leap at present.

Till date many potential lncRNAs have been reported in spermatogenesis and male infertility, however, very few of them have been validated and functionally characterized. A comparison of the findings across these studies can lead to the identification of lncRNAs important for spermatogenesis which could also be exploited as markers of male infertility. Therefore, in this article, we have reviewed the long noncoding RNAs (lncRNAs), their genomic locations,

databases for lncRNA prediction and analysis, recent discoveries made with respect to spermatogenesis and male fertility, and discuss important lncRNAs known to play a functional role in spermatogenesis and male fertility.

Long non-coding (lnc) RNAs

LncRNAs (long non-coding RNAs) are defined as the non-protein-coding RNAs, which are at least 200 nucleotides in length [28, 29]. Like mRNAs, lncRNAs are transcribed by RNA polymerase II and sometimes processed like mRNAs i.e. have 5mG cap, spliced and polyadenylated [30]. With the advent of sensitive, high throughput technologies like next-generation sequencing (NGS) and microarray, several lncRNAs have been discovered, but till date only few have been functionally characterized by gene specific studies.

Genomic contexts of lncRNAs

LncRNAs can be classified on the basis of their genomic location i.e. depending upon where from these RNAs are transcribed. These can be grouped into five broad classes: stand-alone lncRNAs, natural antisense transcripts, pseudogenes, long intronic ncRNAs, promoter-associated transcripts, and enhancer transcripts (Fig. 2).

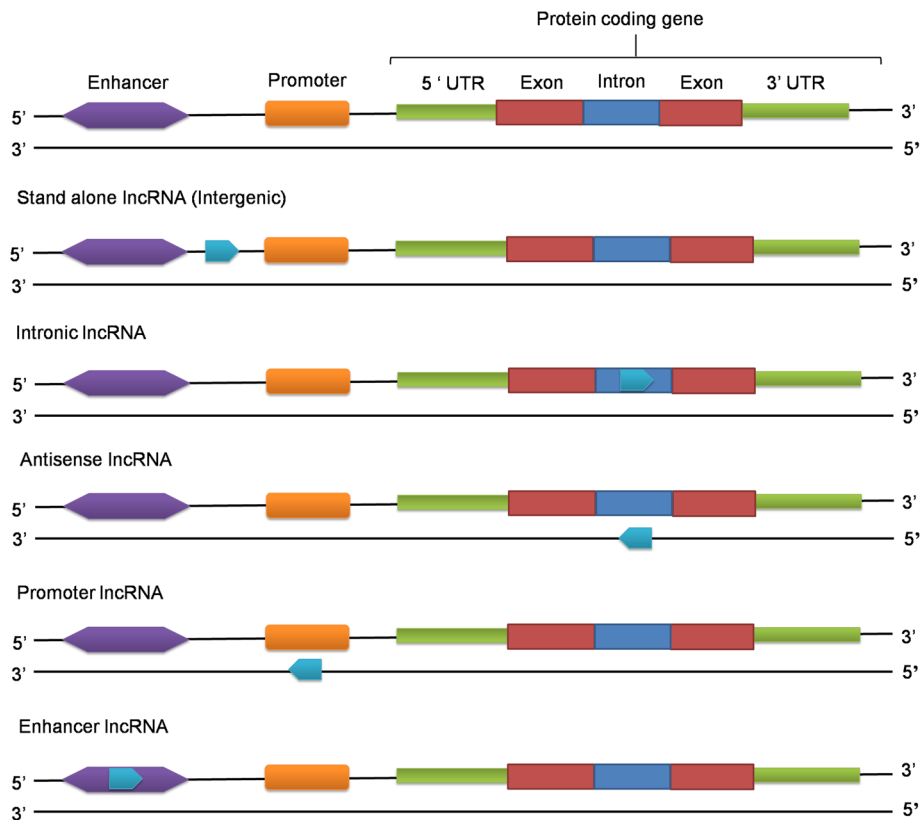


Fig. 2 Genomic contexts of lncRNAs. LncRNAs (blue) may be stand-alone transcription units, intronic, antisense to other genes, or transcribed from promoters or enhancers

Stand-alone lncRNAs

Stand-alone lncRNAs (also known as lincRNAs or large intergenic noncoding RNAs) are transcribed from a distinct genomic loci and do not overlap with protein-coding genes [31–33]. They have an average length of 1 kb and are transcribed by RNA Pol II. These transcripts are usually polyadenylated and spliced. In humans, more than half of the lncRNAs constitute lincRNAs [34]. LincRNAs have been reported to have functional roles in a wide-variety of species. Till date, these lncRNAs have been reported in human, mouse, rat, pig, bovine, bull, zebrafish, *C. elegans*, yeast, *Arabidopsis thaliana* [34–37]. Xist [38], HOTAIR [39], H19 [40] and MALA T1 [41] are some known examples of stand-alone lncRNAs.

Natural antisense transcripts (NATs)

Natural antisense transcripts are transcribed from the opposite sense strand [41, 42]. Generally, NATs are transcribed from both the ends of sense genes and are complementary to the 5' and 3' ends of sense mRNAs [42–44]. As NATs are complementary to the protein-coding mRNAs, these regulate gene expression by RNA:RNA interaction. The best example is the X-inactive-specific transcript Xist/Tsix transcript pair, which is involved in X-chromosome silencing [45]. Just like lincRNAs, this class of lncRNA has been reported in a wide variety of species, including humans, mouse, rat, cattle, chicken, pigs and yeast [35, 37, 46, 47].

Pseudogenes

Pseudogenes are 'relics' of genes, which are non-functional and do not encode for any protein [46]. They are the result of duplication of a parent gene or retrotransposition of the mRNA sequence into a different genomic locus. These are non-functional as a consequence of non-sense, frameshift and other mutations. Approximately, 2–20% of the pseudogenes are transcribed into RNAs, which come under this category of lncRNAs. Pseudogenes derived lncRNAs have been reported in both human and mouse. Lethe, a pseudogene derived lncRNA, was induced by pro-inflammatory cytokines via TNF α signaling in mouse [47]. Pseudogenes derived lncRNAs have also been reported to be critical for cancer development and progression. For instance, DUXAP10 was found to be up-regulated in colorectal cancer [48]. Many studies have revealed that pseudogenes regulate gene expression of their parental gene by producing natural siRNAs or antisense transcripts [49].

Long intronic ncRNAs

It has been known that introns harbour small ncRNAs, such as snoRNAs and miRNAs. Several transcriptome studies have found that many long transcripts are encoded by the introns of annotated genes [50, 51]. Till date only

few have been studied, but it has been seen that these are dysregulated in cancer and many have differential expression pattern in disease conditions [52]. The intronic lncRNAs have been reported in human, mouse, rat, cattle, chicken, pig and *Arabidopsis* [35, 37, 53, 54]. An lncRNA, ANRASSF1, epigenetically regulates the expression of a tumor-suppressor gene, RASSF1, in tumorigenesis [53]. Similarly, an lncRNA SYISL, which is highly expressed in muscles in mouse, has been reported to recruit epigenetic machinery (EZH2, PRC2) to the promoters of p21, MyoG, MCK and Myh4, leading to epigenetic silencing [54]. In *Arabidopsis thaliana*, COLDAIR, an intron present in the FLC (flowering repressor locus) gene, is responsible for plant vernalization [55].

Promoter-associated long transcripts and enhancer RNAs

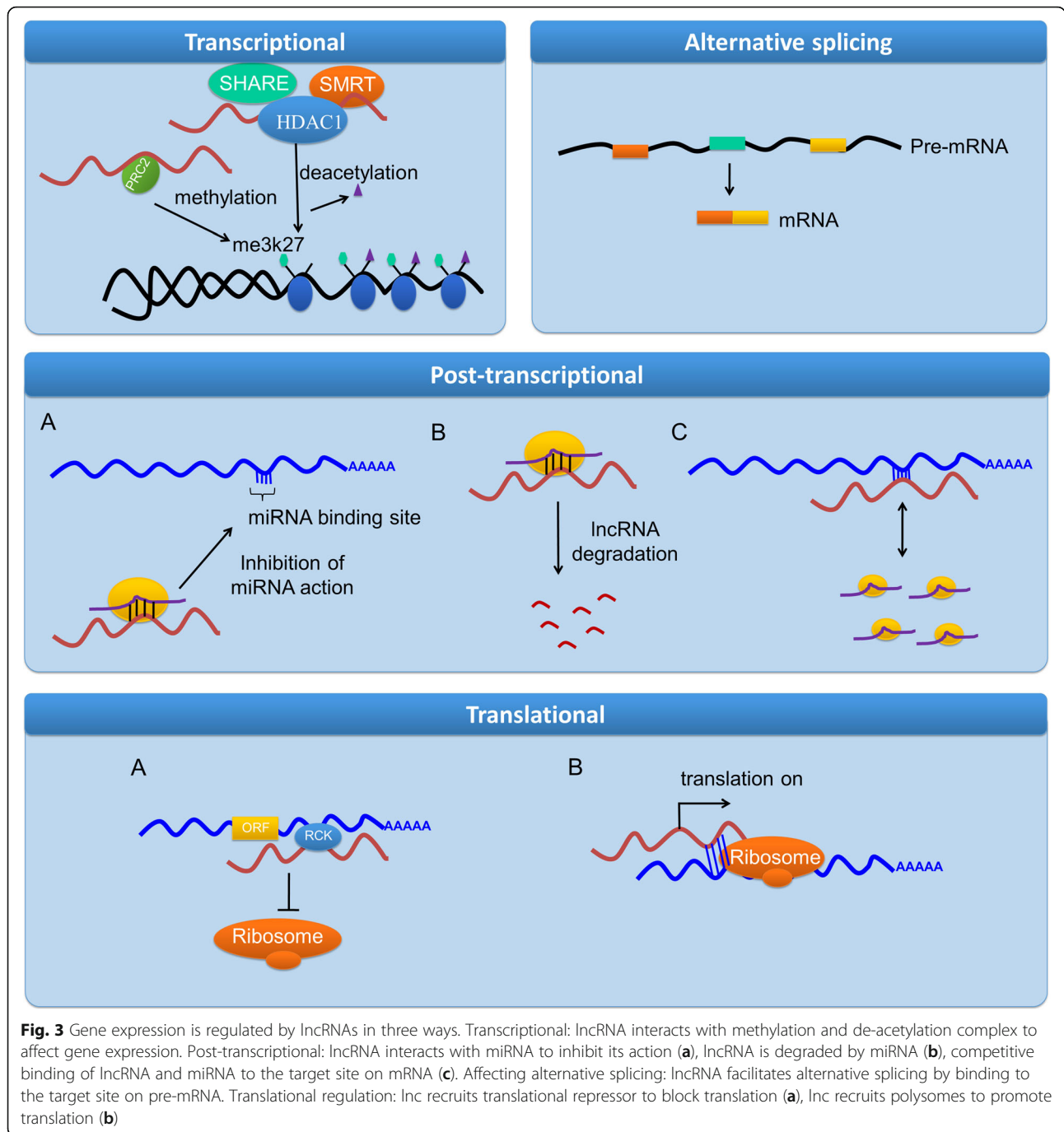
Promoter-associated long RNAs (PALPs) and enhancer-associated RNAs (eRNAs) are lncRNAs, which are transcribed from promoters and enhancers by RNA Pol II [56]. Both promoter and enhancer-associated lncRNAs are known in human and mouse [57, 58]. Promoter associated RNAs are transcribed from the transcription start sites (TSS) of protein coding or non-coding genes. These are cis-acting elements which regulate the expression of their neighbouring genes. The PALPs have recently emerged as an important epigenetic regulator of genes involved in a variety of diseases [59]. Myoparr is a promoter associated lncRNAs, which is transcribed from the promoter region of human and mouse myogenin gene. It activates the expression of myogenin gene by interacting with Ddx17 and PCAF and inhibits proliferation by activating myogenic microRNA expression, thereby promoting myogenesis [57]. Enhancer associated RNAs (eRNAs) are approximately 2 kb long and are generally non-polyadenylated RNAs [60]. It has been reported that eRNAs are involved in chromatin opening and promoting RNA Pol II binding at the transcription site of the target gene [61, 62]. An enhancer associated RNA, CARMEN, was found to be highly expressed in human cardiac precursor cells (CPC) where it interacts with SUZ12 and EZH2 (components of PRC2 complex) and promotes cardiac specification and differentiation in CPC [58].

Regulation by long non-coding RNAs

Gene regulation is a complex process, which takes place at transcriptional, post-transcriptional and translational levels (Fig. 3). Long non-coding RNAs in association with other molecules regulate gene expression in the following ways:

Transcriptional level

LncRNAs are involved in epigenetic gene regulation and genomic imprinting, such as X-inactive specific transcript



(XIST), AIR, H19 [63]. In females, during early embryonic stage one X-chromosome is inactivated to ensure dosage compensation with regard to hemizygous males [64]. This inactivation of X-chromosome is initiated by the recruitment of PRC2 (polycomb repressive 2) complex to specific sites, resulting in H3K27me3 (Histone H3 lysine K27 trimethylation). H19, Airn, Kcnq1ot1, Meg3, and Meg8 are the examples of lncRNAs that are involved in controlling genomic imprinting [65]. In this process, gene expression is regulated by the recruitment of DNA and histone

methyl transferases for DNA methylation and histone modifications, respectively [66]. This type of regulation has been reported in human, mouse and yeast [66, 67].

Post-transcriptional level

After transcription, long non-coding RNAs can regulate the expression of miRNAs by working as a miRNA sponge. It can also interfere with the function of miRNA by competing for its target site on mRNA. Moreover, miRNA can also regulate the expression of lncRNAs by

binding to its 3'UTR region. This type of regulation has been reported in human, mouse and rat [68–71].

- A. There are several non-coding RNAs, which regulate gene expression by binding to the target mRNA through sequence complementarity. There are certain lncRNAs which work as competitive endogenous RNAs (ceRNAs), also called miRNA sponges because they contain binding sequences for miRNAs [72]. These ceRNAs interact with miRNAs and impair the interaction between miRNA and its target mRNA. For example, lnc-MG blocks miR-125b expression and controls IGF2 (Insulin growth factor 2) levels [73].
- B. MicroRNAs can also regulate the expression of lncRNAs [68]. For example, miR-101 and miR-217 negatively regulate lncRNA MALAT1 (metastasis associated with lung adeno-carcinoma transcript1) [69] and miR-449a blocks lncRNA NEAT1 (nuclear enriched abundant transcript 1) expression in cancer [74].
- C. Long non-coding RNAs can interrupt in miRNA function by competing for its target site on mRNA. A very good example of this is the binding of lncRNA BACE1-AS (β -site amyloid precursor protein cleaving enzyme 1- antisense transcript) to its sense partner, where miR-485-5p binds, thus impairing the function of miR-485-5p in Alzheimer's disease [75].

Alternative splicing

The process of alternative splicing takes place in both animals and plants to generate different mRNA isoforms [76]. It was recently found that lncRNAs regulate the process of alternative splicing by interacting with the splicing factors or by forming RNA-RNA duplexes with pre-mRNAs [70]. Till date, this type of regulation by lncRNA has been reported in human and mouse [77]. Recently, it has been reported that lncRNAs NEAT1 and MALAT1 co-localize with nuclear speckles containing the splicing factor SC35, suggesting their involvement in the process of alternative splicing [71].

Translation

- A. lncRNAs play important roles in regulating the translation process as well. lncRNAs regulating target genes at the translational level have been reported in human and mouse [68]. lncRNA-p21 was reported to suppress the translation of JunB and β -catenin mRNA by recruiting translational repressors [68].
- B. The translation of Uchl1 (ubiquitin carboxyl-terminal hydrolase L1) is facilitated by the lncRNA

generated from its antisense strand AS-Uchl1, which recruits polysomes with the help of a repetitive domain SINEB2 to promote a cap- independent translation [68].

Post-translational modifications

lncRNAs can also modulate the post-translation modifications (PTM) of a protein by masking the PTM sites, which has been reported in both human and mouse [78, 79]. lnc-DM is a cytosolic lncRNA exclusively expressed in human conventional dendritic cells (DCs). This lncRNA regulates the differentiation of dendritic cells by phosphorylating STAT3 on Tyr205, thereby preventing the binding of the protein tyrosine phosphatase SHP1 [80]. Similarly, a NF- κ B interacting lncRNA NKILA interacts with I κ B and inhibits its phosphorylation, leading to the activation of NF- κ B and suppression of breast cancer metastasis [81].

Databases for lncRNA prediction and analysis

High throughput studies, such as, microarray and deep sequencing, have identified a huge number of non-coding RNAs [77]. Around 90,000 lncRNAs have been reported by high throughput transcriptome studies [78], but only 16,000 are annotated in GENCODE [79]. Many new lncRNA databases have been developed to annotate such a large number of lncRNAs (Table 1). The number of lncRNAs stored in the databases ranges from < 2000 to > 70,000 transcripts, Noncode v3.0 being the largest database, storing > 73,000 transcripts. DIANA-lncRNA is the database which contains the largest number of computationally predicted lncRNAs (> 56,000 transcripts) as well as the largest number of experimentally validated lncRNAs (2958 transcripts). These databases automatically generate results as lists or tables according to the query. Additionally, most of them also provide graphical visualizations as diagrams or plots (Noncode v3.0, DIANA-lncBase, CHIPBase, LNCipedia, and lncRNome) [96]. There is at least one database which predicts the targets for lncRNAs [94]. This database predicts lncRNA-lncRNA and lncRNA-mRNA interaction in human and mouse. Another database enlists targets for lncRNA in disease-specific conditions, such as cancer, atherosclerosis, cardiac fibrosis and pulmonary hypertension [95]. The information regarding stored lncRNAs and their biological annotations are taken from the literature, computational predictions or primary data repositories. GENCODE project, which is a part of the ENCODE project [97] is a primary data repository providing accurate annotations of human genome, including non-coding RNAs [98]. On the other hand, database like lncRNADisease and functional lncRNA databases rely on manually identified, literature extracted annotations.

Table 1 List of databases for lncRNA annotation and characterization

Database	Description	Organism	Website	Reference
CHIPBase v2.0	It explores the transcriptional regulatory networks of non-coding RNAs (ncRNAs) and protein coding genes (PCGs).	Human and mouse	Rna.sysu.edu.cn/chipbase/	[82]
DIANA-LncBase	This database contains information for both predicted and experimentally verified miRNA-lncRNA interactions.	Human and mouse	Diana.imis.athena-innovation.gr	[83]
LNCipedia	It is a public database for long non-coding RNA (lncRNA) sequences and annotation. It contains 127,802 transcripts and 56,946 genes.	Human	www.lncipedia.org	[84]
lncRNadb v2.0	This database provides a comprehensive annotation of eukaryotic long non-coding RNAs.	71 different organisms	www.lncrnadb.org	[85]
lncRNADisease	Database containing experimentally validated lncRNAs-disease associations.	Human	Cmbi.bjmu.edu.cn/lncrnadisease	[86]
lncRNome	Database containing comprehensive list of lncRNA in humans. It contains information for over 17,000 long non-coding RNAs in humans.	Human	Genome.igib.res.in/lncRNome	[87]
Noncode v5.0	Integrated knowledge database of non-coding RNAs containing non-coding RNAs expressing in 17 different species.	Human, mouse, rat, chicken, fruit fly, zebrafish, <i>C. elegans</i> , yeast, <i>Arabidopsis</i> , chimpanzee, gorilla, orangutan, rhesus macaque, opossum platypus and pig	www.bioinfo.org/noncode/	[88]
The Functional lncRNA database	Database having information of long non-protein-coding transcripts that have been experimentally shown to be both non-coding and functional. Currently the database contains lncRNAs from Human, Mouse and Rat.	Human, mouse and rat	www.valadkhanlab.org	[89]
Encyclopedia of DNA Elements (ENCODE) project	It contains comprehensive list of functional elements in the human genome, including elements that act at the protein and RNA levels.	Human	www.encode.project.org	[90]
StarBase v3.0	Contains information of more than 1.1 million miRNA-ncRNA, 2.5 million miRNA-mRNA, 2.1 million RBP-RNA and 1.5 million RNA-RNA interactions from sequencing data.	Human, mouse and <i>C. elegans</i>	www.starbase.sysu.edu.cn	[91]
RNA-Protein Interaction Prediction (RPI Seq)	This database helps in predicting any possible interaction between lncRNA of known sequence and proteins.	Human, mouse, yeast, fruit fly and <i>E. coli</i>	Prisdb.gdcb.iastate.edu/RPISeq/	[92]
lnc2Catlas	It gives information about lncRNAs and their association with cancer risk. It provides association between 33 different cancers and 27,670 lncRNA transcripts.	Human	lnc2catlas.bioinfotech.org	[93]
lncRIsearch	It is a web server which predicts human and mouse lncRNA-lncRNA and lncRNA-mRNA interactions. It uses RI blast to predict RNA-RNA interaction. It also provides information of tissue-specific expression and subcellular localization data for the lncRNAs.	Human and mouse	rtools.cbrc.jp/lncRIsearch/	[94]
lncTarD	It provides information about disease-related lncRNAs for humans. This database provides a comprehensive resource of 2822 key lncRNA-target regulations, lncRNA influenced functions, and lncRNA-mediated regulatory mechanisms in human diseases.	Human	Biocc.hrbmu.edu.cn/lncTarD/	[95]

All the databases provide information about lncRNAs identified in humans. Databases which provide information of mouse lncRNAs are DIANA-LncBase, CHIPBase, lncRNadb, Noncode v3.0, and the Functional lncRNA

Database. lncRNadb and Noncode v3.0 databases provide information on lncRNAs expressed in different species, from yeast to plants. There are certain databases that provide information about the cell or tissue specific

lncRNAs; for example, lncRNadb, CHIPBase, Noncode v3.0, DIANA-LncBase, and lncRNome. Only lncRNadb and Noncode v3.0 provide information about the cellular localization of the lncRNAs. There are certain databases which provide information about lncRNAs and their association with diseases: Noncode v3.0, lncRNadb, DIANA-lncBase, lncRNADisease, and lncRNome. Few databases provide information about the association of lncRNAs with other non-coding RNAs. For instance, databases like DIANA-LncBase, LNCipedia, the Functional lncRNA Database, and lncRNome give information about the association between miRNAs and lncRNAs.

Differential expression of lncRNAs in spermatogenesis and male infertility

Microarray

Microarrays with probes specific to lncRNAs detection became available about a decade back. These include Arraystar Mouse Stringent and Affymetrix Genechip 2.0 ST series and lncRNA microarray, which are based on updated lncRNA genomic annotation resources such as NONCODE, ENCODE, RefSeq, UCSC Genes, NRED, and fRNadb. These platforms allow direct measurement of the levels of known lncRNA candidates. Only three microarray studies have investigated lncRNAs in relation to spermatogenesis [99–101] (Table 2). Sun et al. (2013) conducted a microarray study on neonatal (6 days) and adult mouse testis (8 weeks old) and identified 8265 lncRNAs, of which 3025 showed differential expression [99]. On comparing the two groups, they found that

1062 lncRNAs were significantly down-regulated and 1963 were significantly up-regulated in the adult mouse testis as compared to the neonatal testis. Some of the male germ cell specific lncRNAs, such as Aldoart2, Speer5-ps1, and Speer9-ps1 were enriched in adult testis and some of the imprinted lncRNAs, such as H19, Meg3, Airn were down-regulated in adult testis. The study also identified that there is a correlation between the lncRNAs expression and epigenetic modification marks, H3K4me3 and H3K27me3 [99]. However, no microarray study has analysed differential lncRNA expression in humans.

Bao et al. (2013) conducted another microarray study to identify lncRNAs expression at different developmental stages: embryonic day (E) 12.5, E 15.5, postnatal day 7 (PND 7), 14, 21, and adult [100]. They found that during fetal testicular development, 2593 lncRNAs were up-regulated and 1125 were down-regulated in E12.5 stage as compared to E15.5 stage. Further, 1976 lncRNAs were up-regulated whereas 3096 were down-regulated in P21 as compared to adult testes. In the other three time points, over 2000 and over 1500 lncRNAs were down- and up-regulated [100]. They selected 28 lncRNAs from microarray data for validation by real-time polymerase chain reaction (RT-PCR) and found a consistent expression of selected lncRNAs. These lncRNAs were AK005782, AK052477, BC049716, ENSMUS T00000136860, ENSMUST00000145068, ENSMUST00000 161511, ENSMUST00000172055, NR_027848, uc008hqc.1, uc009fxb.1, uc009rzf.1, AK020521, AK132382, ENSMUS T0000098955, ENSMUST00000142338, ENSMUST000001

Table 2 List of transcriptome studies in spermatogenesis

Cell type/Tissue	Technique	Organism	Reference
Pre-natal day testes (E12.5 and E15.5) and post-natal day testes PND 7, 14, 21 and adult testes	Microarray	Mouse	[100]
Post-natal day testes PND 7, 14, 17, 21 and 28	RNA sequencing	Mouse	[102]
6 day and 8 week old testes	Microarray	Mouse	[99]
Spermatogonia (type A and B), pre-leptotene, combined leptotene/zygotene, pachytene spermatocytes and round spermatids	RNA sequencing	Mouse	[103]
Sperm and round spermatids	Strand-specific RNA sequencing	Mouse	[104]
Germ cells at mitotic, meiotic and post-meiotic stages	Single cell RNA sequencing	Mouse	[105]
Testes, brain, kidney, heart and liver	Microarray	Mouse	[101]
Sperm	RNA sequencing	Human	[106]
Spermatogonia, spermatocyte and spermatids	RNA sequencing	Human	[107]
A _{pale} and A _{dark} spermatogonia, leptotene/zygotene, early pachytene and late pachytene spermatocyte and round spermatids	RNA sequencing	Human	[108]
Testicular cells	RNA sequencing	Human	[109]
Motile and immotile (asthenozoospermic) sperm	RNA sequencing	Human	[110]
Sperm	RNA sequencing	Bovine	[111]
High-feed (HG) and hay testes	RNA sequencing	Sheep	[112]
High and low motility sperm	RNA sequencing	Bull	[113]

53774, ENSMUST00000161823, NR_003270, NR_028123, uc009cxn.1, uc009ipj.1, uc009til.1, ENSMUST00000119676, ENSMUST00000131790, uc007kqi.1, uc007pgs.1, Tsx. In both the above studies, most of the differentially expressed lncRNAs overlapped with small RNAs, suggesting a possible precursor or regulatory relationship. Also, both these studies identified a close association between lncRNAs and epigenetic regulation. Bao et al. (2013) also identified by RNA immunoprecipitation experiment that 11 lncRNAs interact with one or many factors of chromatin-remodelling complexes, such as EZH2 and LSD1 [100].

In a recent study, microarray analysis of transcripts from mouse brain, kidney, heart, liver and testis was performed [101] (Table 2). It was found that the lncRNAs were most abundant in testes as compared to other tissues. A total of 14,256 lncRNAs were detected in testis, out of which 1607 lncRNAs were specifically expressed in testis. They selected 26 testis-specific abundant lncRNAs for validation by reverse transcription polymerase chain reaction (RT-PCR) and found a pattern of expression similar to the microarray results. These testis-specific lncRNAs were AK018981, Uc007xmk.2, NR_01596, AK005929, NR_040424, ENSMUST0000086914, NR_102286, Uc008vvo.1, NR_028107, NR_027704, NR_038002, ENSMUST00000135865, NR_040326, Uc009dqq.1, NR_038180, ENSMUST00000131403, NR_045045, NR_033583, NR_003953, ENSMUST00000132787, NR_033788, NR_015571, NR_015547, ENSMUST00000136906, Uc029xya.1, and AK018904. The microarray studies on lncRNAs have been undertaken only on different developmental stages and between neonatal and adult testis to identify RNAs which participate in spermatogenesis. No study on purified germ cells or Sertoli cells has been undertaken. Studies on humans lack completely, providing a good opportunity to score the expression of lncRNAs between testicular tissues from azoospermic and normozoospermic individuals and in sperm from oligozoospermic infertile, normozoospermic fertile and normozoospermic infertile individuals.

RNA sequencing (RNA seq) studies

Because of the falling cost of the Next Generation Sequencing (NGS) technology, RNA seq has become a popular approach and is commonly used for transcriptome analysis. Its advantage over microarray technique is its sensitivity, coverage, resolution and the capability to identify new transcripts. Thousands of lncRNAs have been identified in human and mouse transcriptome sequencing projects [98]. RNA sequencing generates billions of reads having size in the range of 30–400 bp, depending upon the technology used. In RNA seq, the number of reads mapping to a region directly reflects the expression level of transcripts, facilitating comparison between two or more samples. Till date, several RNA sequencing studies have been conducted on testis

or sperm in order to investigate their roles in male germ cell development. Nevertheless, due to huge data output from these studies, thousands of transcripts have been reported in relation to spermatogenesis.

Studies in humans At least three studies have investigated lncRNAs in human germ cells (Table 2). Zhu et al. (2016) undertook sequencing and identified a total of 1800 lncRNAs, out of which 157 showed differential expression between different population of testicular cells [107]. In a similar study conducted by Jan et al. (2017) found a total of 137 lncRNAs were differentially expressed between different population of testicular cells [108]. Recently, Rolland et al. (2019) undertook RNA sequencing in human testicular cells to identify lncRNAs, and compared them with rat germ cell RNA profile [109]. They identified 113 known lncRNAs and 20 novel unannotated transcripts (NUTs) were conserved between human and rodents, suggesting them to be important for spermatogenesis. We compared the data across the above three studies and found 15 lncRNAs (LINC00635, LINC00521, LINC00174, LINC00654, LINC00710, LINC00226, LINC00326, LINC00494, LINC00535, LINC00616, LINC00662, LINC00668, LINC00467, LINC00608, LINC00658) to be common across them. We predicted their potential targets using lncRsearch database, and found some crucial spermatogenic genes such as CENPB, FAM98B, GOLGA6 family, RPGR, TPM2, GNB5 and KCNQ10T1 (Table 3). In yet another search for targets using disease-specific target database (lncTarD), we identified TAZ, LIN28A, CDKN2B, CDKN2A, CDKN1A, CDKN1B, CDKN1C, EZH2, SUZ12, VEGFA as targets relevant to spermatogenesis for some of the lncRNAs (Table 3). This opens up new avenues of research in establishing regulatory connections to the signaling involved in spermatogenesis.

lncRNA profiling of human sperm is very limited (Table 2). Sendler et al. (2013) sequenced sperm transcriptome of fertile men and identified 155 lncRNAs in human sperm [106]. Most of these lncRNAs were antisense or lincRNAs. Only one study analysed lncRNA profile in infertility. Zhang et al. (2019) undertook sequencing to identify lncRNAs that differ in motile and immotile human sperm [110]. While 9879 lncRNA genes (13,819 lncRNA transcripts) showed differential expression in between motile and immotile sperm, three lncRNAs i.e. lnc32058, lnc09522, and lnc98497 showed specific and high expression in immotile sperm in comparison to normal motile sperm. Gene ontology of lncRNAs targets revealed that most of the targets were involved in the process of spermatogenesis and sperm function.

Studies in mouse Most of the RNA seq studies have been undertaken on mouse testis or sperm (Table 2). Laiho et al. (2013) undertook RNA sequencing on whole testis of mouse on post-natal days 7, 14, 17, 21, and 28

Table 3 Predicted and experimentally validated targets of selected spermatogenesis-associated lncRNAs

lncRNA gene ID	Global targets	Functions	Disease specific targets	Reference
LINC00635	CENPB^a, FAM98B, GOLGA6L1, GOLGA6L22, GOLGA6L6, MLYCD, MUC2, RP11-573D15.8, RPGR, SNX8	Cenpb ^{-/-} male mice were fertile but have low sperm content and testis weight. FAM98A, FAM98B along with DDX1 and C14orf166 are required for PRMT1 expression (which is important for spermatogenesis). The whole GOLGA6 family is highly expressed in testis. Overexpression of RpgR in transgenic mice has led to increased flagellar defects and reduced sperm count.	-	[114–117]
LINC00174	RP11-573D15.8, CTD-2095E4.4, GOLGA6L22, GOLGA6L6, RPGR, SNX8	-	TAZ-involved in cancer progression.	-
LINC00654	ABI2, FAM98B, GOLGA6L22, GOLGA6L6, RP11-573D15.8, RP3-323A16.1, RPGR, SNX8	-	-	-
LINC00710	ABI2, CCDC85C, FAM98B, RP11-573D15.8, RPGR, SNX8	-	-	-
LINC00226	CCDC85C, FAM98B, SNX8, TPM2, C00717C00720L.1	TPM2 was found to be up-regulated in cauda region as compared to caput/corpus region of human epididymis.	-	[118]
LINC00326	C9orf152, FLJ35934, GNB5, LINC00940,	GNB5 could play a role in acrosome biogenesis and vesicle trafficking.	-	[119]
LINC00494	ABI2, EMC10, FAM98B, GOLGA6L22, GOLGA6L6, IGFN1, RP11-573D15.8, RPGR, SNX8	-	-	-
LINC00535	ABI2, FAM98B, SNX8	-	-	-
LINC00616	ABI2, CCDC85C, EMC10, FAM98B, MLYCD, PARVG, RP11-573D15.8, RP3-323A16.1, RPGR, SNX8	-	-	-
LINC00662	RP11-573D15.8	-	LIN28A- involved in cell metastasis and stemness in lung cancer.	-
LINC00668	ABI2, EMC10, KCNQ1OT1, MIR6820, RP11-573D15.8, RP3-323A16.1, RPGR, SNX8	KCNQ1OT1 is paternally expressed and plays a role in silencing of six genes <i>Ascl2, Kcnq1, Cdkn1c, Slc22a18, Phlda2,</i> and <i>Osbp15.</i>	CDKN2B, CDKN2A, CDKN1A, CDKN1B, CDKN1C, EZH2, SUZ12, VEGFA- involved in cancer progression and cell proliferation.	[120]
LINC00467	RP3-323A16.1, RPGR	-	-	-
LINC00608	ABI2, EMC10, FAM98B, RP11-573D15.8, RP3-323A16.1, RPGR, SNX8	-	-	-
LINC00658	ABI2, AP006621.9, B3GALNT2, CCDC85C, FAM98B, RP11-573D15.8, RPGR, SNX8	-	-	-

^agene names in bold have known functions in spermatogenesis

[102]. When compared between two adjacent developmental stages, they found that 17–46 lncRNAs were down-regulated and 16–605 lncRNAs were up-regulated in each transition. Wichman et al. (2017) analysed the expression of lncRNAs at specific stages of spermatogenesis in mouse by purifying key stages (combined type A and type B spermatogonia, pre-leptotene, combined leptotene/zygotene, pachytene spermatocytes and round spermatids) of spermatogenesis by STA-PUT method

[103]. They found that 1630 lncRNAs were differentially expressed in spermatogonia versus pachytene spermatocytes with Gm14244, 1700026D11Rik, 1700066O22Rik, RP24-503F14.1, Gm1698, 1700110I01Rik, Gm20621, Gm10619, RP23-52 K8.2, 1700006J14Rik, H19, Gm26751, Gm13054, Gm26656, Gm26835, 5930412G12Rik, Gm26775, D830 015G02Rik, Gm26673, Gm5532 being the top 20. 1534 lncRNAs were differentially expressed between pachytene spermatocytes and round spermatids with 1810059H22Rik,

4930412B13Rik, 4930467K11Rik, Gm14249, Axtx6, RP23-245G9.1, Gm11782, Gm13481, 9130204K15Rik, Gm13175, Gm27032, Gm16278, 1700040D17Rik, Gm10619, Gm14244, 4930477E14Rik, RP23-46B12.8, 1520401A03Rik, A730090 N16Rik, Gm13589 being the top 20. They also identified a subset of lncRNAs that escape meiotic sex chromosome inactivation at the pachytene stage of meiosis and others with strong testicular expression pattern, suggesting their importance in spermatogenesis. Lastly, they generated a knock out mouse having the deletion of one X-linked lncRNA, Tslrn1 (testis-specific long noncoding RNA1) and found that males carrying this deletion displayed normal fertility, but a significant reduction in sperm count [103]. This shows its participation in the regulation of spermatogenesis, though the molecular mechanisms of its action remain to be investigated.

Zhang et al. (2017) undertook strand-specific RNA sequencing that revealed a large repertoire of lncRNAs in mouse mature sperm [104]. They identified 20,907 known and 4088 novel lncRNAs transcripts in sperm. Chen et al. (2018) undertook single-cell-RNA sequencing on mouse germ cells at mitotic, meiotic and post-meiotic stages in order to identify the molecular events occurring during male germ cell development and to reveal several novel crucial regulators of mammalian spermatogenesis [105]. A large proportion of known protein-coding genes (18,037 out of 20,088, 89.8%) were identified in the spermatogenic cells. They also identified 9431 out of 11,962 (78.8%) annotated lncRNAs to be expressed in the spermatogenic cells, which had different spatio-temporal expression. The highest numbers and the expression level of lncRNAs were found in the diplo-tene, MI spermatocytes and steps 6–8 spermatids.

Studies in other animals Selvaraju et al. (2017) undertook sequencing to identify the spermatozoal transcripts in fresh bull semen [111] (Table 2). They found that spermatozoa retain a variety of RNAs, including lncRNAs. The most abundant lncRNAs reported in this study were NONBTAT002138.2, NONBTAT029740.1, NONBTAT026069.2, NONBTAT026075.2, NONBTAT015718.2, NONBTAT015717.2, NONBTAT017665.2, NONBTAT027794.1, NONBTAT007060.2, NONBTAT029554.1, NONBTAT008960.2, NONBTAT022624.2, NONBTAT030080.1, NONBTAT030770.1, NONBTAT030342.1, NONBTAT030589.1, NONBTAT031209.1, NONBTAT022014.2, NONBTAT025041.2, and NONBTAT001981.2. It remains to be seen if some of these lncRNAs match with human or rodent sperm.

Yanli et al. (2017) conducted RNA sequencing to identify lncRNAs and their role in the testes development of high-grain (HG) diet fed sheep [112]. A total of 6460 lncRNAs were reported. Further, on comparing the lncRNAs expressed in testes of hay and HG diet groups, researchers identified 28 up-regulated and 31 down-

regulated lncRNAs differentially expressed between the two groups. Ten differentially expressed lncRNAs were randomly selected to validate the RNA sequencing results by using qRT-PCR. These lncRNAs were XR_001435329.1, XR_001025228.2, XR_0010443569.1, XR_001042765.2, XR_001045031.2, XR_001040775.1, XR_001434819.1, XR_001044933.2, XR_001044935.2, XR_001044859.2. This study apart from providing lncRNA profile of sheep testis also suggested a significant role of 59 lncRNAs in regulating the effect of nutrition on fertility.

In another study, Wang et al. (2019) undertook an integrated analysis of lncRNAs and mRNAs in high and low motility Holstein bull sperm, in order to identify mRNAs regulated by lncRNAs and their impact on sperm motility [113] (Table 2). They reported 11,561 lncRNAs in sperm, of which 2517 were differentially expressed between high and low motility sperm groups. The top 20 significantly differentially expressed lncRNAs reported were TCONS_00000538, TCONS_00003123, TCONS_00005494, TCONS_00007217, TCONS_00008747, TCONS_00009226, TCONS_00011237, TCONS_00012048, TCONS_00012181, TCONS_00012775, TCONS_00013706, TCONS_00014744, TCONS_00014816, TCONS_00015123, TCONS_00019053, TCONS_00020130, TCONS_00021031, TCONS_00021657, TCONS_00024090 and TCONS_00025506. They also found that out of the 20,875 protein coding genes detected in semen, 19 were differentially expressed between high and low motility groups. Further, they found that lncRNA TCONS_000417333 targets a gene EFNA1 (ephrin A1), which is involved in male reproductive physiology. Several other lncRNAs and mRNAs identified in this study are good targets for further studies. This is a landmark study in identifying bull sperm motility and fertility lncRNA markers, which can find potential application in animal husbandry.

Functional lncRNAs in spermatogenesis

Many potential lncRNAs have been identified in male germ cell development, some of them showing differential expression between fertile and infertile sperm. However, only a few have been functionally annotated and characterized [1] (Fig. 4). Below are some lncRNAs, which have been shown to participate in spermatogenesis.

Humans

Narcolepsy candidate-region 1 gene (NLC1-C) Narcolepsy candidate-region 1 gene (NLC1-C) is an lncRNA expressing in the cytoplasm of spermatogonia and early spermatocytes. It was found that the over-expression of NLC1-C promoted cell growth whereas its loss inhibited cell growth and promoted apoptosis [121]. Microarray analysis revealed that NLC1-C had lower expression in MA (maturation arrest) patients than those with normal spermatogenesis. In another study, it was reported that

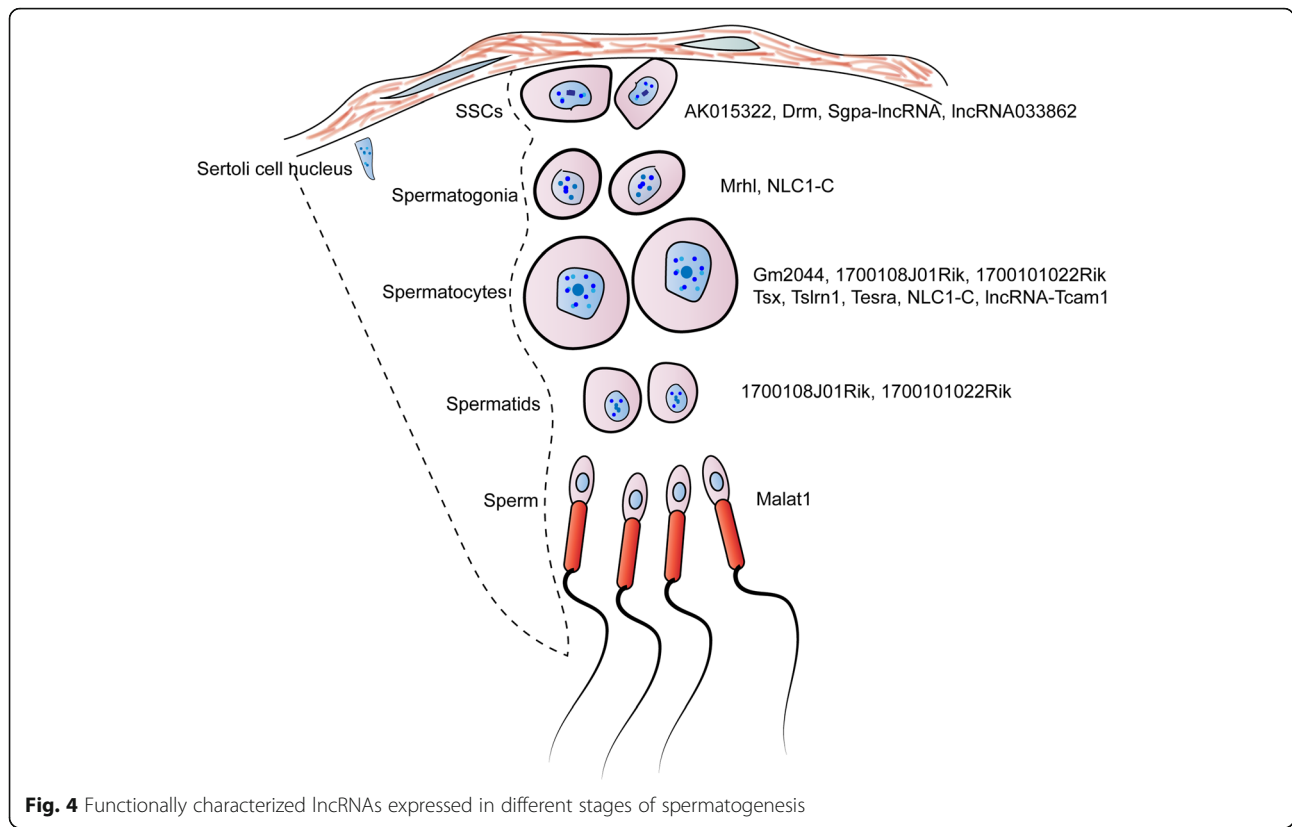


Fig. 4 Functionally characterized lncRNAs expressed in different stages of spermatogenesis

NLC1-C inhibited the transcription of miR-320a and miR-383 by binding to the RNA-binding domain of nucleolin and promoted the proliferation of spermatogonia and spermatocytes in MA patients [121].

Rodents

Meiotic recombination hot spot locus (Mrhl) Meiotic recombination hot spot locus (Mrhl) RNA is a 2.4 monoexonic residing in the nucleus. Mrhl was first identified by Nishant et al. (2004) while working on a 17.2 kb fragment of a novel meiotic recombination hot spot in mouse chromosome 8 [122]. Later, the same research group showed that Mrhl RNA was cleaved by Drosha to form an 80-nucleotide RNA intermediate [123]. Further evidence showed that both 2.4 kb full length and cleaved 80 nucleotide RNAs were localized in the nucleoli of GC1 cell line, suggesting their possible role in some specific chromatin domains [123]. Further studies by the same research group showed that Mrhl interacts with p68 and inhibits the Wnt signaling pathway [124]. Recently, they identified that the transcription factor Sox8 is a regulatory link between mrhl RNA expression, Wnt signaling activation and meiotic progression. In this study, Mrhl was shown to regulate Sox8 by binding to its promoter. Mrhl RNA interacts with other regulatory factors like Myc-Max-Mad, co-repressor Sin3a and co-

activator Pcaf at the Sox8 promoter site and regulates spermatogonial differentiation [124].

HongrES2 HongrES2 is a 1588 bp lncRNA, which is co-transcribed by chromosomes 5 and 9 in rats. Its expression is seen in the cauda region of epididymis [125]. The expression of HongrES2 was found to increase constantly till the first wave of spermatogenesis and become constant around day 450 in rats. This expression pattern of hongrES2 perfectly matches with its reported role in regulating sperm maturation. In the nucleus, HongrES2 is processed to a 23 bp RNA called as mil-HongrES2. This processed transcript inhibits the expression of CES7, an epididymis specific protein. Overexpression of mil-HongrES2 resulted in decreased sperm capacitation [125], indicating that low HongrES2 expression promotes sperm maturation process in the epididymis.

Testis-specific X-linked (Tsx) Testis specific X-linked (Tsx) is an lncRNA, which is located within the X-inactivation center. It is known to be highly expressed in the pachytene spermatocytes and not in spermatogonia or round spermatids, suggesting its role in regulating the meiotic division. An increased number of abnormal pachytene spermatocytes undergoing apoptosis were seen in Tsx knockout mice, suggesting its role in the progression of meiosis in mice [126].

Dmrt1-related gene (Drm) Similarly, Dmrt1-related gene (Drm) is an exclusively expressing lncRNA in testis. It was discovered by Zhang et al. (2010) in an attempt to clone the Dmrt1 gene when they found another RNA isoform of different size and sequence on the 3' half [127]. Later, it was found that this 3' half was encoded by a gene present on chromosome 5, whereas Dmrt1 gene is present on chromosome 19. This chimeric form of mRNA was the result of disrupted coding region of Dmrt1 gene and the replacement of its 3'UTR region, which resulted in decreased expression of normal Dmrt1 protein. Dmrt1 is known to play a crucial role in spermatogonial development by upregulating *Sohlh1* and prevents early meiosis of spermatogonia by repressing *Stra8* [128, 129]. Therefore, the regulation of Dmrt1 by Drm could be involved in the switching between mitosis and meiosis during germ cell development.

Spga-lncRNAs Chan et al. (2006) used two different techniques, SAGE and whole genome tiling microarray, in three different germ cell stages: type A spermatogonia, pachytene spermatocytes, and round spermatids to identify the protein coding genes [130]. Later, Lee et al. (2012) used this published SAGE and tiling microarray data to identify lncRNAs [131]. With this approach, they found 50, 35, and 24 single-exonic lncRNAs with expression in type A spermatogonia, pachytene spermatocyte and round spermatids, respectively. After validation by qPCR, northern blotting and RACE, two lncRNAs Spga-lncRNA1 and Spga-lncRNA2 were found to be highly expressed in spermatogonia. Further, in vitro experiments showed that they may play important roles in maintaining stemness of spermatogonia.

lncRNA- testicular cell adhesion molecule 1 (lncRNA-Tcam1) lncRNA- testicular cell adhesion molecule 1 (lncRNA-Tcam1) is a 2.4 kb nucleotide long nuclear lncRNA. It is exclusively expressed in mouse male germ cells and is assumed to play a crucial role in meiosis. In a recent report, it was found that lncRNA-Tcam1 regulates immune related genes in mouse germ cells [132]. The induction of lncRNA-tcam1 transcription using the tet-off system in GC-2 cells upregulated six genes, *Trim30a*, *Ifit3*, *Tgtp2*, *Ifi47*, *Oas1g*, and *Gbp3*. All these genes were found to be associated with immune response, suggesting that lncRNA-Tcam1 regulates immune related genes.

Tug1 lncRNA Tug1 locus resides on chromosome 11 in murine and is among the most conserved lncRNAs between mouse and humans [133]. The expression of this lncRNA was seen not only in adult tissues but also in a number of embryonic tissues during the embryonic development in both human and mouse. Lewandowski

et al. (2019) reported that the loss of Tug1 locus leads to male sterility [133]. They found a significant decrease in sperm number in Tug1^{-/-} males as compared to the wild-type mice (oligospermia), they also found morphological defects in the head and mid-piece region of sperm (teratospermia). Thus, these findings revealed an essential role of the Tug1 locus in male fertility. Further investigations showed that Tug1 locus harbours three distinct regulatory activities: cis-acting repressive DNA function, trans-acting Tug1 lncRNA and peptide (TUG1-BOAT) having role in maintaining mitochondrial membrane potential [133].

Tesra Tesra is a 4435 nucleotide long testis-specific lncRNA, which is present at the *Prss/Tessp* gene cluster on chromosome 9 [134]. The expression of tesra was detected in the cytoplasm, and the nucleus of germ cells and only in the cytoplasm of Leydig cells while no expression was seen in the Sertoli cells. Satoh et al. (2019) reported a nuclear function of this lncRNA where it acts as a transcriptional activator of the *Prss/Tessp* gene. The *Prss/Tessp* gene cluster specifically activates in the primary spermatocytes at the late pachytene stage and has been found to be critical for meiotic progression of the germ cells. The molecular mechanism by which tesra enhances the activity of *Prss/Tessp* gene promoter still needs to be explored [134]. However, previous studies have shown nuclear lncRNAs to recruit histone modifiers or transcription factors to chromatin regions of target genes or work as a scaffold that interacts with multiple protein complexes to regulate the target gene [134].

lncRNA033862 This lncRNA resides on chromosome 19 and is transcribed from the antisense strand of the GDNF family receptor alpha (*Gfra1*) gene. lncRNA033862 was found to be highly dysregulated in response to GDNF (a ligand of GFRA1) in mouse SSCs [135]. GFRA1 has been considered as a cell surface marker of SSCs and the loss of *Gfra1* has been reported to disrupt SSC self-renewal in neonatal mouse testis. It was found that this lncRNA is predominantly expressed in spermatogonia and SSCs and regulates SSC self-renewal by acting as a transcriptional activator of *Gfra1* [135].

AK015322 Hu et al. (2016) performed a microarray study to investigate the expression profile of lncRNA in different germ cell populations (spermatogonial stem cell, type A spermatogonia, pachytene spermatocyte and round spermatids) [136]. They found lncRNA AK015322 expression to be highest in the spermatogonial stem cells as compared to other germ cells. Knockdown of AK015322 resulted in decreased proliferation, suggesting its role in

maintaining SSC self-renewal capacity. AK015322 regulates the expression of ETV5, important for SSC proliferation, by working as a decoy of miR-19b-3p [136]. Therefore, this study revealed that lncRNA AK01522 suppresses the effect of miR-19b-3p on ETV5 expression, thus promoting the proliferation of SSCs.

Gm2044 In a microarray study, the expression of Gm2044 was found to be highest in the pachytene spermatocytes of mouse testis [137]. Further, the expression of Gm2044 was found to be more at day 12 to day 20, which represents spermatocyte formation and meiotic phase during the first wave of spermatogenesis, suggesting its role in meiotic progression. RNA pull down experiment revealed that Gm2044 interacts with Utl1 and the overexpression of Gm2044 resulted in decreased protein level of UTL1, suggesting Utl1 to be a direct target of Gm2044. The study identified Gm2044 to be critical for germ cell transition and meiotic progression by inhibiting Utl1 translation [137].

Conclusion and future perspective

Once thought to be junk, non-genic regions of DNA encode several non-coding RNAs, which play important roles in gene regulation. The advent of next generation sequencing has accelerated the discovery of lncRNA along with other types of non-coding RNAs. Investigations on lncRNAs expression in the embryonic and post-natal testis identified about 1000 lncRNAs that show change in expression with the onset and progression of spermatogenesis. At least one study investigated lncRNAs in different germ cells in adult testis and identified about 1500 such RNAs showing change in expression during spermatogenesis. Specific investigations on particular lncRNAs identified their impact on sperm count (Tslrn1, Tug1), sperm motility (lnc32058, lnc09522, and lnc98487), sperm capacitation (mil-hongrES2) and meiosis (Drm, Tsx, Gm2044). There are certain lncRNAs involved in maintaining the balance between stemness of spermatogonia and their differentiation. For example, Spga-lncRNA, AK015322, LncRNA033862 are involved in maintaining the stemness of spermatogonia while Mrhl lncRNA binds to the Sox8 promoter to regulate spermatogonial differentiation in mouse. Comparison across three studies on human germ cells identified 15 (LINC00635, LINC00521, LINC00174, LINC00654, LINC00710, LINC00226, LINC00326, LINC00494, LINC00535, LINC00616, LINC00662, LINC00668, LINC00467, LINC00608, LINC00658) common lncRNAs that can be great candidates for further analysis for their role in spermatogenesis and their differential expression in infertility.

A number of databases and analysis tools have now become available to facilitate the data analysis and comparison. Transcriptome sequencing offers the identification of novel RNAs along with known RNAs; therefore, sequencing studies should provide new candidate RNAs for further investigation in spermatogenesis and male infertility. Several lncRNAs have been identified which play roles in different developmental stages and adult mouse testes. Hong et al. (2018) identified testis-specific lncRNAs in mouse, of which 26 most abundant (AK018981, Uc007xmk.2, NR_01596, AK005929, NR_040424, ENSMUST00000086914, NR_102286, Uc008vvo.1, NR_028107, NR_027704, NR_038002, ENSMUST00000135865, NR_040326, Uc009dqq.1, NR_038180, ENSMUST00000131403, NR_045045, NR_033583, NR_003953, ENSMUST00000132787, NR_033788, NR_015571, NR_015547, ENSMUST00000136906, Uc029xya.1, AK018904) are strong candidates for further investigations. Bao et al. (2013) reported 28 candidate lncRNAs (AK005782, AK052477, BC049716, ENSMUST00000136860, ENSMUST00000145068, ENSMUST00000161511, ENSMUST00000172055, NR_027848, uc008hqc.1, uc009fxb.1, uc009rzf.1, AK020521, AK132382, ENSMUST0000098955, ENSMUST00000142338, ENSMUST00000153774, ENSMUST00000161823, NR_003270, NR_028123, uc009cxn.1, uc009ipj.1, uc009til.1, ENSMUST00000119676, ENSMUST00000131790, uc007kgi.1, uc007pgs.1, Tsx) which were differentially expressed across different developmental stages in mouse.

Although many potential lncRNAs have been discovered till date with the help of advanced technologies and public lncRNA annotation resources, yet the functional annotation of lncRNAs in spermatogenesis is in infancy and has tremendous possibilities. Long non-coding RNAs involved in spermatogenesis can be identified by using public lncRNA annotation, but their specific localization and biological function in testis need experimental evidence. A number of studies have undertaken RNA sequencing in testis, germ cells and sperm [108, 138–140]. Some of these have compared RNA sequencing data between fertile and infertile men. However, most of these studies did not pay attention to lncRNA. One of the reasons could be the lack of tools available to analyse lncRNAs. The availability of such tools now provides an opportunity to analyse their transcriptome data for the analysis of lncRNAs. lncRNAs may be used as biomarkers or/and targets to diagnose or/and treat infertility in the future when more high throughput data, more refined bioinformatics tools, protein-RNA binding models, and regulatory evidence from ENCODE become available.

New techniques have been developed to identify the function of lncRNAs through their interaction with other molecular species. These techniques involve the identification of interacting partners like RNA, protein

and/or DNA and isolation of a component of interest (RNA/protein). For identification of RNAs linked to proteins, RIP (RNA immunoprecipitation) and HITS/ PAR-CLIP (High-Throughput Sequencing of RNA/ PhotoActivatable-Ribonucleoside-Cross Linking and Immunoprecipitation) techniques can be used. Other techniques like RNA pull down and/or ChIRP/CHART would be useful to identify proteins and DNA sequences associated with a particular lncRNA.

Acknowledgements

Meghali Joshi would like to thank the Council of Scientific and Industrial Research (CSIR), Govt. of India for graduate fellowship (31/004(1285)/2015-EMR-I). The authors would like to thank the Council of Scientific and Industrial Research (CSIR), Govt. of India for financial support under MLP2026. Meghali Joshi would like to thank Poonam Mehta for help in drawing figures.

Authors' contributions

MJ and SR conceived, prepared and finalized the manuscript. The author(s) read and approved the final manuscript.

Funding

This work was supported by financial support from the Council of Scientific and Industrial Research (CSIR), Govt. of India, under project code MLP2026.

Availability of data and materials

No associated data is available.

Ethics approval and consent to participate

Not applicable.

Consent for publication

All authors have provided consent for publication.

Competing interests

All authors have declared to have no competing interests.

Received: 13 July 2020 Accepted: 21 October 2020

Published online: 30 October 2020

References

- Sahlu BW, Zhao S, Wang X, Umer S, Zou H, Huang J, et al. Long noncoding RNAs: new insights in modulating mammalian spermatogenesis. *J Anim Sci Biotechnol*. 2020;11(1):1–12.
- Richard Boland C. Non-coding RNA: It's Not Junk. *Dig Dis Sci*. 2017;62:1107–9.
- Palazzo AF, Lee ES. Non-coding RNA: what is functional and what is junk? *Front Genet*. 2015;6:2.
- Hombach S, Kretz M. Non-coding RNAs: Classification, Biology and Functioning. In: Slaby O, Calin G. (eds) *Non-coding RNAs in Colorectal Cancer*. *Advances in Experimental Medicine and Biology*. Cham Springer; 2016;937.
- He Z, Kokkinaki M, Pant D, Gallicano GI, Dym M. Small RNA molecules in the regulation of spermatogenesis. *Reproduction*. 2009;137(6):901–11.
- Rajender S, Meador C, Agarwal A. Small RNA in spermatogenesis and male infertility. *Front Biosci (Schol Ed)*. 2012;4:1266–74.
- Dinger ME, Amaral PP, Mercer TR, Pang KC, Bruce SJ, Gardiner BB, et al. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res*. 2008;18(9):1433–45.
- Lee S, Seo H-H, Lee CY, Lee J, Shin S, Kim SW, et al. Human long noncoding RNA regulation of stem cell potency and differentiation. *Stem Cells Int*. 2017;2017(6374504):10.
- Fang Y, Fullwood MJ. Roles, functions, and mechanisms of long non-coding RNAs in cancer. *Genomics Proteomics Bioinformatics*. 2016;14(1):42–54.
- Bond AM, VanGompel MJW, Sametsky EA, Clark MF, Savage JC, Disterhoft JF, et al. Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry. *Nat Neurosci*. 2009;12(8):1020.
- Sauvageau M, Goff LA, Lodato S, Bonev B, Groff AF, Gerhardinger C, et al. Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *elife*. 2013;2:e01749.
- Lai K-MV, Gong G, Atanasio A, Rojas J, Quispe J, Posca J, et al. Diverse phenotypes and specific transcription patterns in twenty mouse lines with ablated lincRNAs. *Plos One*. 2015;10(4):e0125522.
- Ng S-Y, Bogu GK, Soh BS, Stanton LW. The long noncoding RNA RMST interacts with SOX2 to regulate neurogenesis. *Mol Cell*. 2013;51(3):349–59.
- Apra J, Prenninger S, Dori M, Ghosh T, Monasor LS, Wessendorf E, et al. Transcriptome sequencing during mouse brain development identifies long non-coding RNAs functionally involved in neurogenic commitment. *EMBO J*. 2013;32(24):3145–60.
- Lin N, Chang K-Y, Li Z, Gates K, Rana ZA, Dang J, et al. An evolutionarily conserved long noncoding RNA TUNA controls pluripotency and neural lineage commitment. *Mol Cell*. 2014;53(6):1005–19.
- Klattenhoff CA, Scheuermann JC, Surface LE, Bradley RK, Fields PA, Steinhauser ML, et al. Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell*. 2013;152(3):570–83.
- Kurian L, Aguirre A, Sancho-Martinez I, Benner C, Hishida T, Nguyen TB, et al. Identification of novel long noncoding RNAs underlying vertebrate cardiovascular development. *Circulation*. 2015;131(14):1278–90.
- Boulberdaa M, Scott E, Ballantyne M, Garcia R, Descamps B, Angelini GD, et al. A role for the long noncoding RNA SENCN in commitment and function of endothelial cells. *Mol Ther*. 2016;24(5):978–90.
- Deng C, Li Y, Zhou L, Cho J, Patel B, Terada N, et al. HoxBlinc RNA recruits Set1/MLL complexes to activate Hox gene expression patterns and mesoderm lineage development. *Cell Rep*. 2016;14(1):103–14.
- Kong Y, Hsieh C-H, Alonso LC. ANRIL: a lincRNA at the CDKN2A/B locus with roles in cancer and metabolic disease. *Front Endocrinol*. 2018;9:405.
- Dou C, Sun L, Jin X, Han M, Zhang B, Jiang X, et al. Long non-coding RNA CARLo-5 promotes tumor progression in hepatocellular carcinoma via suppressing miR-200b expression. *Oncotarget*. 2017;8(41):70172.
- Liu T, Wang Z, Zhou R, Liang W. Focally amplified lincRNA on chromosome 1 regulates apoptosis of esophageal cancer cells via DRP1 and mitochondrial dynamics. *IUBMB Life*. 2019;71(2):254–60.
- Ji J, Dai X, Yeung S-CJ, He X. The role of long non-coding RNA GAS5 in cancers. *Cancer Manag Res*. 2019;11:2729.
- Tang Q, Hann SS. HOTAIR: an oncogenic long non-coding RNA in human cancer. *Cell Physiol Biochem*. 2018;47(3):893–913.
- Yang B, Gao G, Wang Z, Sun D, Wei X, Ma Y, et al. Long non-coding RNA HOTTIP promotes prostate cancer cells proliferation and migration by sponging miR-216a-5p. *Biosci Rep*. 2018;38(5):BSR20180566.
- Yanga W, Hea F, Lia Y, Zhaib Y, Tana B, Wuha H. Long non-coding RNA PCGE M1 as a biomarker for prostate cancer. *Scienceasia*. 2016;42(3):201–6.
- Zhang G, He X, Ren C, Lin J, Wang Q. Long noncoding RNA PCA3 regulates prostate cancer through sponging miR-218-5p and modulating high mobility group box 1. *J Cell Physiol*. 2019;234(8):13097–109.
- Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet*. 2009;10(3):155.
- Batista PJ, Chang HY. Long noncoding RNAs: cellular address codes in development and disease. *Cell*. 2013;152(6):1298–307.
- Rutenberg-Schoenberg M, Sexton AN, Simon MD. The properties of long noncoding RNAs that regulate chromatin. *Annu Rev Genomics Hum Genet*. 2016;17:69–94.
- Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*. 2009;458(7235):223.
- Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev*. 2011;25(18):1915–27.
- Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell*. 2011;147(7):1537–50.
- Ransohoff JD, Wei Y, Khavari PA. The functions and unique features of long intergenic non-coding RNA. *Nat Rev Mol Cell Biol*. 2018;19(3):143.
- Ji H, Niu C, Zhan X, Xu J, Lian S, Xu B, et al. Identification, functional prediction, and key lincRNA verification of cold stress-related lincRNAs in rats liver. *Sci Rep*. 2020;10(1):1–14.
- Billerey C, Boussaha M, Esquerré D, Rebours E, Djari A, Meersseman C, et al. Identification of large intergenic non-coding RNAs in bovine muscle using next-generation transcriptomic sequencing. *BMC Genomics*. 2014;15(1):499.

37. Kern C, Wang Y, Chitwood J, Korf I, Delany M, Cheng H, et al. Genome-wide identification of tissue-specific long non-coding RNA in three farm animal species. *BMC Genomics*. 2018;19(1):684.
38. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Bruggmann SA, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*. 2007;129(7):1311–23.
39. Brannan CI, Dees EC, Ingram RS, Tilghman SM. The product of the H19 gene may function as an RNA. *Mol Cell Biol*. 1990;10(1):28–36.
40. Ji P, Diederichs S, Wang W, Böing S, Metzger R, Schneider PM, et al. MALAT-1, a novel noncoding RNA, and thymosin β 4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene*. 2003;22(39):8031.
41. Magistri M, Faghihi MA, Wahlestedt C. Regulation of chromatin structure by long noncoding RNAs: focus on natural antisense transcripts. *Trends Genet*. 2012;28(8):389–96.
42. Finocchiaro G, Carro MS, Francois S, Parise P, DiNinni V, Muller H. Localizing hotspots of antisense transcription. *Nucleic Acids Res*. 2007;35(5):1488–500.
43. He Y, Vogelstein B, Velculescu VE, Papadopoulos N, Kinzler KW. The antisense transcriptomes of human cells. *Science*. 2008;322(5909):1855–7.
44. Baranello L, Bertozzi D, Fogli MV, Pommier Y, Capranico G. DNA topoisomerase I inhibition by camptothecin induces escape of RNA polymerase II from promoter-proximal pause site, antisense transcription and histone acetylation at the human HIF-1 α gene locus. *Nucleic Acids Res*. 2009;38(1):159–71.
45. Werner A, Carlile M, Swan D. What do natural antisense transcripts regulate? *RNA Biol*. 2009;6(1):43–8.
46. Balakirev ES, Ayala FJ. Pseudogenes: are they “junk” or functional DNA? *Annu Rev Genet*. 2003;37(1):123–51.
47. Rapicavoli NA, Qu K, Zhang J, Mikhail M, Laberge R-M, Chang HY. A mammalian pseudogene lncRNA at the interface of inflammation and anti-inflammatory therapeutics. *elife*. 2013;2:e00762.
48. Lian Y, Xu Y, Xiao C, Xia R, Gong H, Yang P, et al. The pseudogene derived from long non-coding RNA DUXAP10 promotes colorectal cancer cell growth through epigenetically silencing of p21 and PTEN. *Sci Rep*. 2017;7(1):1–11.
49. Delpu Y, Larri u D, Gayral M, Arvanitis D, Dufresne M, Cordelier P, et al. Noncoding RNAs: clinical and therapeutic applications. In: Egger G, Arimondo P. (eds) *Drug Discovery in Cancer Epigenetics*. Elsevier; 2016. p. 305–26.
50. Louro R, Smirnova AS, Verjovski-Almeida S. Long intronic noncoding RNA transcription: expression noise or expression choice? *Genomics*. 2009;93(4):291–8.
51. Rearick D, Prakash A, McSweeney A, Shepard SS, Fedorova L, Fedorov A. Critical association of ncRNA with introns. *Nucleic Acids Res*. 2010;39(6):2357–66.
52. Guil S, Soler M, Portela A, Carr re J, Fonalleras E, G mez A, et al. Intronic RNAs mediate EZH2 regulation of epigenetic targets. *Nat Struct Mol Biol*. 2012;19(7):664.
53. Beckedorff FC, Ayupe AC, Crocci-Souza R, Amaral MS, Nakaya HI, Soltys DT, et al. The intronic long noncoding RNA ANRASSF1 recruits PRC2 to the RASSF1A promoter, reducing the expression of RASSF1A and increasing cell proliferation. *Plos Genet*. 2013;9(8):e1003705.
54. Jin JJ, Lv W, Xia P, Xu ZY, Dai Zheng A, Wang XJ, et al. Long noncoding RNA SYSL regulates myogenesis by interacting with polycomb repressive complex 2. *Proc Natl Acad Sci*. 2018;115(42):E9802–11.
55. Heo JB, Sung S. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science*. 2011;331(6013):76–9.
56. Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science*. 2007;316(5830):1484–8.
57. Hitachi K, Nakatani M, Takasaki A, Ouchi Y, Uezumi A, Ageta H, et al. Myogenin promoter-associated lnc RNA Myoparr is essential for myogenic differentiation. *EMBO Rep*. 2019;20(3):e47468.
58. Ounzain S, Micheletti R, Arnan C, Plaisance I, Cecchi D, Schroen B, et al. CARMEN, a human super enhancer-associated long noncoding RNA controlling cardiac specification, differentiation and homeostasis. *J Mol Cell Cardiol*. 2015;89:98–112.
59. Chellini L, Frezza V, Paronetto MP. Dissecting the transcriptional regulatory networks of promoter-associated noncoding RNAs in development and cancer. *J Exp Clin Cancer Res*. 2020;39(1):1–10.
60.  rom UA, Shiekhattar R. Long noncoding RNAs usher in a new era in the biology of enhancers. *Cell*. 2013;154(6):1190–3.
61. Cesana M, Cacchiarelli D, Legnini I, Santini T, Sthandier O, Chinappi M, et al. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell*. 2011;147(2):358–69.
62. Melo CA, Drost J, Wijchers PJ, van de Werken H, de Wit E, Vrieling JAFO, et al. eRNAs are required for p53-dependent enhancer activity and gene transcription. *Mol Cell*. 2013;49(3):524–35.
63. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. *Annu Rev Biochem*. 2012;81:145–66.
64. Yang C, Chapman AG, Kelsey AD, Minks J, Cotton AM, Brown CJ. X-chromosome inactivation: molecular mechanisms from the human perspective. *Hum Genet*. 2011;130(2):175–85.
65. Kanduri C. Long noncoding RNAs: Lessons from genomic imprinting. *Biochim Biophys Acta (BBA)-Gene Regul Mech*. 2016;1859(1):102–11.
66. Mohammad F, Mondal T, Guseva N, Pandey GK, Kanduri C. Kcnq1ot1 noncoding RNA mediates transcriptional gene silencing by interacting with Dnmt1. *Development*. 2010;137(15):2493–9.
67. Senmatsu S, Asada R, Abe T, Hoffman CS, Ohta K, Hirota K. lncRNA transcriptional initiation induces chromatin remodeling within a limited range in the fission yeast *fbp1* promoter. *Sci Rep*. 2019;9(1):1–8.
68. Yoon JH, Abdelmohsen K, Gorospe M. Functional interactions among microRNAs and long noncoding RNAs. In: *Seminars in cell & developmental biology*. Academic Press; 2014;34:9–14.
69. Wang X, Li M, Wang Z, Han S, Tang X, Ge Y, et al. Silencing of long noncoding RNA MALAT1 by miR-101 and miR-217 inhibits proliferation, migration, and invasion of esophageal squamous cell carcinoma cells. *J Biol Chem*. 2015;290(7):3925–35.
70. Li Z, Zhao W, Wang M, Zhou X. The role of long noncoding RNAs in gene expression regulation. In: Vlachakis D.(eds). *Gene Expression Profiling in Cancer*. IntechOpen; 2019.
71. Romero-Barrios N, Legascue MF, Benhamed M, Ariel F, Crespi M. Splicing regulation by long noncoding RNAs. *Nucleic Acids Res*. 2018;46(5):2169–84.
72. Furi -Tari P, Tarazona S, Gabald n T, Enright AJ, Conesa A. spongeScan: a web for detecting microRNA binding elements in lncRNA sequences. *Nucleic Acids Res*. 2016;44(W1):W176–80.
73. Zhu M, Liu J, Xiao J, Yang L, Cai M, Shen H, et al. lnc-mg is a long non-coding RNA that promotes myogenesis. *Nat Commun*. 2017;8:14718.
74. You J, Zhang Y, Liu B, Li Y, Fang N, Zu L, et al. MicroRNA-449a inhibits cell growth in lung cancer and regulates long noncoding RNA nuclear enriched abundant transcript 1. *Indian J Cancer*. 2014;51(7):77.
75. Faghihi MA, Zhang M, Huang J, Modarresi F, Van der Brug MP, Nalls MA, et al. Evidence for natural antisense transcript-mediated inhibition of microRNA function. *Genome Biol*. 2010;11(5):R56.
76. Kan Z, Rouchka EC, Gish WR. Gene structure prediction and alternative splicing analysis using genomically aligned ESTs. *Genome Res*. 2001;11(5):889–900.
77. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, et al. The transcriptional landscape of the mammalian genome. *Science*. 2005;309(5740):1559–63.
78. Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, et al. The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet*. 2015;47(3):199.
79. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: the reference human genome annotation for the ENCODE project. *Genome Res*. 2012;22(9):1760–74.
80. Wang P, Xue Y, Han Y, Lin L, Wu C, Xu S, et al. The STAT3-binding long noncoding RNA lnc-DC controls human dendritic cell differentiation. *Science*. 2014;344(6181):310–3.
81. Liu B, Sun L, Liu Q, Gong C, Yao Y, Lv X, et al. A cytoplasmic NF- κ B interacting long noncoding RNA blocks I κ B phosphorylation and suppresses breast cancer metastasis. *Cancer Cell*. 2015;27(3):370–81.
82. Zhou K-R, Liu S, Sun W-J, Zheng L-L, Zhou H, Yang J-H, et al. ChIPBase v2.0: decoding transcriptional regulatory networks of non-coding RNAs and protein-coding genes from ChIP-seq data. *Nucleic Acids Res*. 2016;45(D1):D43–50.
83. Paraskevopoulou MD, Georgakilas G, Kostoulas N, Reczko M, Maragkakis M, Dalamagas TM, et al. DIANA-lncBase: experimentally verified and computationally predicted microRNA targets on long non-coding RNAs. *Nucleic Acids Res*. 2012;41(D1):D239–45.
84. Volders P-J, Helsens K, Wang X, Menten B, Martens L, Gevaert K, et al. LNCipedia: a database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Res*. 2012;41(D1):D246–51.

85. Quek XC, Thomson DW, Maag JLV, Bartonicek N, Signal B, Clark MB, et al. lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Res.* 2014;43(D1):D168–73.
86. Chen G, Wang Z, Wang D, Qiu C, Liu M, Chen X, et al. lncRNADisease: a database for long-non-coding RNA-associated diseases. *Nucleic Acids Res.* 2012;41(D1):D983–6.
87. Bhartiya D, Pal K, Ghosh S, Kapoor S, Jalali S, Panwar B, et al. lncRNome: a comprehensive knowledgebase of human long noncoding RNAs. *Database.* 2013;2013:bat034.
88. Fang S, Zhang L, Guo J, Niu Y, Wu Y, Li H, et al. NONCODEV5: a comprehensive annotation database for long non-coding RNAs. *Nucleic Acids Res.* 2017;46(D1):D308–14.
89. Niazi F, Valadkhan S. Computational analysis of functional long noncoding RNAs reveals lack of peptide-coding capacity and parallels with 3' UTRs. *Rna.* 2012;18(4):825–43.
90. Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, et al. The encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res.* 2017;46(D1):D794–801.
91. Yang J-H, Li J-H, Shao P, Zhou H, Chen Y-Q, Qu L-H. starBase: a database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. *Nucleic Acids Res.* 2010;39(suppl_1):D202–9.
92. Muppurala UK, Honavar VG, Dobbs D. Predicting RNA-protein interactions using only sequence information. *BMC Bioinformatics.* 2011;12(1):489.
93. Ren C, An G, Zhao C, Ouyang Z, Bo X, Shu W. Lnc2Catlas: an atlas of long noncoding RNAs associated with risk of cancers. *Sci Rep.* 2018;8(1):1909.
94. Fukunaga T, Iwakiri J, Ono Y, Hamada M. lncRRsearch: a web server for lncRNA-RNA interaction prediction integrated with tissue-specific expression and subcellular localization data. *Front Genet.* 2019;10:462.
95. Zhao H, Shi J, Zhang Y, Xie A, Yu L, Zhang C, et al. LncTarD: a manually-curated database of experimentally-supported functional lncRNA-target regulations in human diseases. *Nucleic Acids Res.* 2020;48(D1):D118–26.
96. Fritah S, Niclou SP, Azuaje F. Databases for lncRNAs: a comparative evaluation of emerging tools. *Rna.* 2014;20(11):1655–65.
97. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012;489(7414):57.
98. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 2012;22(9):1775–89.
99. Sun J, Lin Y, Wu J. Long non-coding RNA expression profiling of mouse testis during postnatal development. *Plos One.* 2013;8(10):e75750.
100. Bao J, Wu J, Schuster AS, Hennig GW, Yan W. Expression profiling reveals developmentally regulated lncRNA repertoire in the mouse male germline. *Biol Reprod.* 2013;89(5):101–7.
101. Hong SH, Kwon JT, Kim J, Jeong J, Kim J, Lee S, et al. Profiling of testis-specific long noncoding RNAs in mice. *BMC Genomics.* 2018;19(1):539.
102. Laiho A, Kotaja N, Gyenesei A, Sironen A. Transcriptome profiling of the murine testis during the first wave of spermatogenesis. *Plos One.* 2013;8(4):e61558.
103. Wichman L, Somasundaram S, Breindel C, Valerio DM, McCarrey JR, Hodges CA, et al. Dynamic expression of long noncoding RNAs reveals their potential roles in spermatogenesis and fertility. *Biol Reprod.* 2017;97(2):313–23.
104. Zhang X, Gao F, Fu J, Zhang P, Wang Y, Zeng X. Systematic identification and characterization of long non-coding RNAs in mouse mature sperm. *Plos One.* 2017;12(3):e0173402.
105. Chen Y, Zheng Y, Gao Y, Lin Z, Yang S, Wang T, et al. Single-cell RNA-seq uncovers dynamic processes and critical regulators in mouse spermatogenesis. *Cell Res.* 2018;28(9):879.
106. Sendler E, Johnson GD, Mao S, Goodrich RJ, Diamond MP, Hauser R, et al. Stability, delivery and functions of human sperm RNAs at fertilization. *Nucleic Acids Res.* 2013;41(7):4104–17.
107. Zhu Z, Li C, Yang S, Tian R, Wang J, Yuan Q, et al. Dynamics of the transcriptome during human spermatogenesis: predicting the potential key genes regulating male gametes generation. *Sci Rep.* 2016;6:19069.
108. Jan SZ, Vormer TL, Jongejan A, Röling MD, Silber SJ, de Rooij DG, et al. Unraveling transcriptome dynamics in human spermatogenesis. *Development.* 2017;144(20):3659–73.
109. Rolland AD, Evrard B, Darde TA, Le Béguet C, Le Bras Y, Bensalah K, et al. RNA profiling of human testicular cells identifies syntenic lncRNAs associated with spermatogenesis. *Human Reprod.* 2019;34(7):1278–90.
110. Zhang X, Zhang P, Song D, Xiong S, Zhang H, Fu J, et al. Expression profiles and characteristics of human lncRNA in normal and asthenozoospermia sperm. *Biol Reprod.* 2018;100(4):982–93.
111. Selvaraju S, Parthipan S, Somashekar L, Kolte AP, Binsila BK, Arangasamy A, et al. Occurrence and functional significance of the transcriptome in bovine (*Bos taurus*) spermatozoa. *Sci Rep.* 2017;7:42392.
112. Zhang Y, Yang H, Han L, Li F, Zhang T, Pang J, et al. Long noncoding RNA expression profile changes associated with dietary energy in the sheep testis during sexual maturation. *Sci Rep.* 2017;7(1):5180.
113. Wang X, Yang C, Guo F, Zhang Y, Ju Z, Jiang Q, et al. Integrated analysis of mRNAs and long noncoding RNAs in the semen from Holstein bulls with high and low sperm motility. *Sci Rep.* 2019;9(1):2092.
114. Hudson DF, Fowler KJ, Earle E, Saffery R, Kalitsis P, Trowell H, et al. Centromere protein B null mice are mitotically and meiotically normal but have lower body and testis weights. *J Cell Biol.* 1998;141(2):309–19.
115. Akter KA, Mansour MA, Hyodo T, Senga T. FAM98A associates with DDX1-C14orf166-FAM98B in a novel complex involved in colorectal cancer progression. *Int J Biochem Cell Biol.* 2017;84:1–13.
116. Bekpen C, Tautz D. Human core duplication gene families: game changers or game players? *Brief Funct Genomics.* 2019;18(6):402–11.
117. Brunner S, Colman D, Travis AJ, Luhmann UFO, Shi W, Feil S, et al. Overexpression of RPGR leads to male infertility in mice due to defects in flagellar assembly. *Biol Reprod.* 2008;79(4):608–17.
118. Légaré C, Sullivan R. Differential gene expression profiles of human efferent ducts and proximal epididymis. *Andrology.* 2020;8(3):625–36.
119. Teves ME, Roldan ERS, Krapf D, Strauss JF III, Bhagat V, Sapao P. Sperm differentiation: the role of trafficking of proteins. *Int J Mol Sci.* 2020;21(10):3702.
120. Mancini-DiNardo D, Steele SJS, Levorske JM, Ingram RS, Tilghman SM. Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. *Genes Dev.* 2006;20(10):1268–82.
121. Lü M, Tian H, Cao YX, He X, Chen L, Song X, et al. Downregulation of miR-320a/383-sponge-like long non-coding RNA NLC1-C (narcolepsy candidate-region 1 genes) is associated with male infertility and promotes testicular embryonal carcinoma cell proliferation. *Cell Death Dis.* 2015;6(11):e1960.
122. Nishant KT, Ravishanker H, Rao MRS. Characterization of a mouse recombination hot spot locus encoding a novel non-protein-coding RNA. *Mol Cell Biol.* 2004;24(12):5620–34.
123. Ganesan G, Rao SMR. A novel noncoding RNA processed by Drosha is restricted to nucleus in mouse. *Rna.* 2008;14(7):1399–410.
124. Arun G, Akhade VS, Donakonda S, Rao MRS. Mrhl RNA, a long noncoding RNA, negatively regulates Wnt signaling through its protein partner Ddx5/p68 in mouse spermatogonial cells. *Mol Cell Biol.* 2012;32(15):3140–52.
125. Ni M-J, Hu Z-H, Liu Q, Liu M-F, Lu M, Zhang J-S, et al. Identification and characterization of a novel non-coding RNA involved in sperm maturation. *PLoS One.* 2011;6(10):e26053.
126. Anguera MC, Ma W, Clift D, Namekawa S, Kelleher RJ III, Lee JT. Tsx produces a long noncoding RNA and has general functions in the germline, stem cells, and brain. *Plos Genet.* 2011;7(9):e1002248.
127. Zhang L, Lu H, Xin D, Cheng H, Zhou R. A novel ncRNA gene from mouse chromosome 5 trans-splices with Dmrt1 on chromosome 19. *Biochem Biophys Res Commun.* 2010;400(4):696–700.
128. Ottolenghi C, Veitia R, Barbieri M, Fellous M, McElreavey K. The human doublesex-related gene, DMRT2, is homologous to a gene involved in somitogenesis and encodes a potential bicistronic transcript. *Genomics.* 2000;64(2):179–86.
129. Agbor VA, Tao S, Lei N, Heckert LL. A Wt1-Dmrt1 transgene restores DMRT1 to Sertoli cells of Dmrt1^{-/-} testes: a novel model of DMRT1-deficient germ cells. *Biol Reprod.* 2013;88(2):51.
130. Chan W-Y, Lee T-L, Wu S-M, Ruszczyk L, Alba D, Baxendale V, et al. Transcriptome analyses of male germ cells with serial analysis of gene expression (SAGE). *Mol Cell Endocrinol.* 2006;250(1–2):8–19.
131. Lee TL, Xiao A, Rennett OM. Identification of Novel Long Noncoding RNA Transcripts in Male Germ Cells. In: Chan WY, Blomberg L (eds) *Germline Development. Methods in Molecular Biology (Methods and Protocols)*. New York: Springer; 2012;825.
132. Kurihara M, Otsuka K, Matsubara S, Shiraishi A, Satake H, Kimura AP. A testis-specific long non-coding RNA, lncRNA-Tcam1, regulates immune-related genes in mouse male germ cells. *Front Endocrinol.* 2017;8:299.
133. Lewandowski JP, Dumbović G, Watson AR, et al. The Tug1 lncRNA locus is essential for male fertility. *Genome Biol.* 2020;21:237.
134. Satoh Y, Takei N, Kawamura S, Takahashi N, Kotani T, Kimura AP. A novel testis-specific long noncoding RNA, Tesra, activates the Prss42/Tessp-2 gene during mouse spermatogenesis. *Biol Reprod.* 2019;100(3):833–48.

135. Li L, Wang M, Wu X, Geng L, Xue Y, Wei X, et al. A long non-coding RNA interacts with Gfra1 and maintains survival of mouse spermatogonial stem cells. *Cell Death Dis.* 2016;7(3):e2140.
136. Hu K, Zhang J, Liang M. LncRNA AK015322 promotes proliferation of spermatogonial stem cell C18-4 by acting as a decoy for microRNA-19b-3p. *In Vitro Cell Dev Biol Anim.* 2017;53(3):277–84.
137. Hu K, Li L, Liao Y, Liang M. LncRNA Gm2044 highly expresses in spermatocyte and inhibits Utf1 translation by interacting with Utf1 mRNA. *Genes Genomics.* 2018;40(7):781–7.
138. Pereira N, Cheung S, Parrella A, O'Neill C, Nikprelevic N, Rosenwaks Z, et al. Investigating the role of sperm-specific RNA to screen men with unexplained infertility. *Fertil Steril.* 2017;108(3):e46.
139. Sohni A, Tan K, Song H-W, Burow D, de Rooij DG, Laurent L, et al. The neonatal and adult human testis defined at the single-cell level. *Cell Rep.* 2019;26(6):1501–17.
140. Guo J, Grow EJ, Carrell DT, Hotaling JM, Cairns B. The adult human testis cell atlas via single cell RNA sequencing. *Fertil Steril.* 2018;110(4):e165–6.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

