Long non-coding RNAs: new players in mammalian female gametogenesis

Rosalia Battaglia
Cinzia Di Pietro

Department of Biomedical and Biotechnological Sciences, Unit of Biology and Genetics - G. Sichel, University of Catania, Italy

Address for correspondence:
Rosalia Battaglia
Department of Biomedical and Biotechnological Sciences, Unit of Biology and Genetics - G. Sichel, University of Catania
Via S. Sofia 87
95123 Catania, Italy
E-mail: rosaliabattaglia04@gmail.com

Summary

New advances in high-throughput sequencing technologies have revealed long non-coding RNAs (lncRNAs) as functional RNA molecules with a crucial role during human early embryo development. Although lncRNAs lack protein-coding function, they have emerged as key regulators of protein-coding genes, acting at different steps of gene expression, from chromatin architecture and epigenetic to transcription, RNA splicing, editing, and translation. Here we review the existing lncRNA classification, nomenclature and molecular function. Then, we discuss the current understanding of lncRNAs in reproductive biology. LncRNAs could be novel biomarkers of oocyte quality in reproductive medicine or ART and could be potential new molecular targets for new drugs for many infertility disorders in the future.

KEY WORDS: non-coding transcriptome, long non-coding RNA, ovarian follicle, reproduction.

Introduction

Long non-coding RNAs (lncRNAs) represent a heterogeneous class of non-protein-coding RNAs whose length is greater than 200 nucleotides and are basically involved in the regulation of gene expression. Unlike microRNAs (miRNAs) that represent negative regulators of gene expression at the post-transcriptional level, lncRNAs may be both positive and negative regulators of protein-coding genes, acting at different steps of gene expression, from chromatin architecture and epigenetic to transcription, RNA splicing, editing, and translation (1, 2) (Figure 1). They have been classified according to their genomic position into long intergenic non-coding RNAs (lincRNAs), long intragenic non-coding RNAs or bidirectional, if located less than 1000 base pairs away from a protein coding gene and transcribed in the opposite direction. The long intragenic non-coding RNAs, exonic or intronic, are classified as sense lncRNA if their sequence overlaps with the sense strand of the protein coding gene, antisense if the sequence overlaps with the antisense strand of the protein coding gene (1, 2) (Figure 2). Another class of lncRNA molecules has recently been described, circular RNAs (circRNAs). CircRNAs can arise from exons (exonic circRNA) or introns (intronic circRNA) with independent modes of generation. They can act as natural miRNA sponges to lower miRNA levels and reduce their regulatory effect on mRNAs (3). Unlike protein encoding genes and miRNAs, whose sequences are conserved in evolution, most lncRNAs show low sequence conservation. In spite of this, their function, among different species, is preserved. Probably, their function might not require nucleotide sequence conservation, but just the maintenance of the secondary structure needed to bind DNA, other RNAs and proteins. Moreover, the promoter regions of lncRNAs are often conserved as the promoters of many protein-coding genes (4).
they may be involved in cell-cycle regulation, stemness, differentiation, and apoptosis. LncRNA expression profiles are altered in several types of cancers, including human prostate cancer, renal cell carcinomas, breast cancer, ovarian cancer, and human lung adenocarcinomas (5). Moreover, specific lncRNAs have also been shown to be involved in other human diseases (6). In addition, it was proposed that, similar to miRNAs, lncRNA expression profiling may be an informative biomarker in cancer diagnoses (7).

The discovery that most of the mammalian transcriptome include long molecules that do not code for proteins dates back to 2002, even if in 1990 the first two lncRNAs were identified (8-10). Xist (X inactive specific transcript) controls X chromosome inactivation during early female embryonic development. Xist is transcribed from the inactive X chromosome, binds to chromatin and recruits chromatin regulatory complexes associated with gene repression. In differentiated cells, X inactivation is mediated through DNA cytosine methylation by DNA methyltransferases (DNMTs) and in this stage the role of Xist is unknown (11). The lncRNA H19, located in an imprinting region of chromosome 11 near the insulin-like growth factor 2 (IGF2) gene, is a maternally expressed and paternally imprinted gene. It has been shown that it controls genome expression at multiple levels and: I) acts on chromatin organization; II) sequesters miR-106 and miR-let7; III) is the precursor of miR-675. Moreover, interacting with p53, H19 inhibits tumor suppressor protein action and checks genome stability. It is expressed during embryonic life, down-regulated at birth and its altered expression is involved in different steps of tumorigenesis. Mutations in this gene have been associated with Beckwith-Wiedemann Syndrome and Wilms tumor (12). To date, the number of lncRNAs in Homo Sapiens is unknown, however, it is estimated that several thousand genes are present in our genome, even if only a few of them have been examined for their biological function. Different
online databases are available to explore the IncRNA world, the most known are: Lncrnadb.org, LNCipedia and LncRNAWiki (13-15). Moreover, exhaustive reviews, about IncRNA function and their involvement in human diseases, have been published (16).

**LncRNAs and Reproduction**

LncRNAs were recently identified in granulosa and cumulus cells, oocytes and early embryos, and their role in oocyte and early embryo development has been suggested (17-20). By using RNA sequencing techniques, 89 differentially expressed IncRNAs between compact and expanded cumulus cells have been found. Some of them were located in introns of genes known to be involved in GC physiology (17). Another paper, using microarray analysis, compared the IncRNA profile in cumulus cells from mature oocytes producing high quality embryos with those from oocytes producing poor quality embryos. The Authors found about 200 IncRNAs differentially expressed and suggested that IncRNAs may contribute to the processes of oocyte and early embryo development (19). The same groups described the up-regulation of IncRNA AK124742, in cumulus cells related to high quality embryos. AK124742 is antisense to PSMD6, a gene that encodes a subunit of 26S proteasome. AK124742 colocalizes with DNA damage foci and is involved in the ATP-dependent degradation of ubiquitinated proteins. The Authors suggested that this IncRNA could be considered as a potential biomarker for embryo quality (21).

An interesting class of IncRNAs, involved in development, is represented by bidirectional IncRNAs that lead to the upregulation of their oppositely transcribed partner gene. Probably, they induce DNA demethylation in promoter regions in a sequence-specific manner (18). 618 of these non-coding RNAs, named promoter-associated...

---

*Figure 2 - Classification of LncRNAs according to genomic location.*
Long non-coding RNAs (PancRNAs), have been identified in mouse MII oocytes and more than 1000 in 2-cell embryos. Moreover, the pancRNA regulating interleukin 17d (Il17d) seems to be essential for embryonic development (18).

Even if the research of IncRNAs in reproduction and development is still at the beginning, there are evidences to support their relevant role, and a comprehensive review in mammals and in different species has been recently published (22-24). In 2016, we characterized human oocytes miRNome and more recently, by a computational approach, we identified 41 IncRNAs able to target 9 of the miRNAs identified in oocytes (25, 26). The predicted IncRNAs are expressed in embryonic cells, in the ovary or the placenta and are involved in several human pathologies. Among them, we found Xist and another two well characterized nuclear IncRNAs: MALAT1 (Metastasis Associated Lung Adenocarcinoma Transcript 1) and NEAT1 (Nuclear Enriched Abundant Transcript 1). MALAT1 up-regulation was initially described in aggressive non-small cell lung cancer (27). MALAT1 localizes to nuclear bodies, named nuclear speckles, involved in pre-mRNA splicing. It has been described that its down-regulation may impair proliferation, cell cycle, apoptosis and migration of trophoblast cells involved in preeclampsia (28). NEAT1 was found exclusively localized to paraspeckles and is a core component of these nuclear bodies involved in the nuclear retention of specific mRNAs (29). Moreover, NEAT1 seems to be essential for the formation of the corpus luteum and the establishment of pregnancy in mice, although its precise molecular mechanism remains to be investigated (30). Recently, increased levels of NEAT1 were found associated with placental dysfunction in idiopathic intrauterine Growth Restriction (IUGR) fetuses (31). Paraspeckles have not been described in human oocytes, in spite of this, in female germ cells, NEAT1 could contribute to regulate mRNA stability. Other than being described for cytoplasmic polyadenylation, this further mechanism allowed for the storage of maternal RNAs and their use during early development, before the activation of the embryonic genome (32). We also identified GAS5 (growth arrest specific 5) a circRNA that targets miR-136-5p. GAS5 is the host gene of 11 human small nucleolar RNAs (snoRNA) involved in the 2’-O-methylation of rRNA. This IncRNA seems to be involved in embryogenesis, in fact, it increases OCT4, NANOG and SOX2 expression by Nodal regulation and is directly regulated by these stemness factors in hESCs, forming a circuit that promotes pluripotency (33). It has been demonstrated that ncRNAs are able to control different points of embryogenesis such as maternal-zygotic transition, maintenance of pluripotency the patterning of the body axes, specification and differentiation of cell types and organ morphogenesis (34). Another predicted IncRNA that could perform its role in human oocytes is OIP5-AS1, an antisense transcript of OIP5 (Opa interacting protein 5) gene. The protein encoded by this gene localizes to centromeres, where it is essential for the recruitment of CENP-A, and it is required for centromeric heterochromatin organization (35). It is known that IncRNAs perform an important role in the control of chromatin structure and function (Figure 1). They can directly interact with many histone- and DNA-modifying enzymes to participate in covalent modifications of histones or DNA (36). Chromatin remodeling occurs at the end of oocyte growth, before meiosis resumption, and induces transcriptional silencing. In fact, it has been demonstrated that in this phase of development, in maturing oocytes, the abundance of several genes decreases as chromatin compaction increases, but the molecular mechanisms are not completely understood. It is possible to suppose that in these processes different IncRNAs could play important roles (37). Understanding the regulation of gene expression inside the ovarian follicle is important in basic reproductive research and could also be useful for clinical applications. In fact, the characterization of non-coding RNAs in ovarian follicles could improve reproductive disease diagnosis, provide biomarkers of oocyte quality in Assisted Reproductive Treatment, and allow the development of therapies for infertility disorders.

References

4. Johnsson P, Lipovich L, Grandé D, Morris KV. Evolu-