Published in final edited form as:
Methods Mol Biol. 2012 ; 825: 105–114. doi:10.1007/978-1-61779-436-0_9.

Identification of Novel Long Noncoding RNA Transcripts in Male Germ Cells

Tin-Lap Lee, **Amy Xiao**, and **Owen M. Rennert**

Abstract

Emerging evidence from these studies suggested that the male germ cell transcriptome is more complex than previously envisioned. In addition to protein-coding genes, the transcriptome also encodes a significant number of nonprotein-coding transcripts. These noncoding (nc) RNAs appear to be involved in a variety of cellular activities, ranging from simple housekeeping to complex regulatory functions. A class of ncR-NAs known as long ncRNAs (lncRNAs) were recently shown to be expressed in a developmentally regulated manner during brain and embryonic stem cell development. This protocol aims to predict and identify potential lncRNA candidates using Serial Analysis of Gene Expression (SAGE) data. We also illustrate how to validate the potential lncRNAs by expression analyses using real-time PCR and Northern Blot. Potential lncRNA candidates in male germ cells are identified using our previously established male germ cell SAGE database (GermSAGE).

Keywords

Long noncoding RNA; Male germ cells; Development; SAGE

1. Introduction

An important observation emerged from various genome-wide transcriptome studies is that the majority of transcripts do not code for proteins. They are referred to as noncoding RNAs (ncRNAs) (1). Mammalian cells produce thousands of ncRNAs of unknown function. These nonprotein-coding portions of the genome were often considered “junk,” but recent research has shown that ncRNAs have a wide range of regulatory functions. Initial discoveries identified small ncRNAs, such as microRNA (miRNA), small interfering RNA (siRNAs), and Piwi-interacting RNA (piRNA); these have been widely reported to function in various regulatory processes, including male germ cell development (2). Recently, a new class of ncRNAs known as long ncRNAs (lncRNAs), defined as ncRNA species at least 200 nucleotides in length, has also been demonstrated to function in developmental regulations in brain and mouse embryonic stem cells (ESCs) (3, 4). The precise functional role of lncRNA in male germ cell development remains largely unknown. Its functional importance in developmental regulations suggests that lncRNAs may also be indispensable in male germ cell development.

The presence of ncRNAs was functionally implicated in our previous male germ cell transcriptome studies employing Serial Analysis of Gene Expression (SAGE) (5–10). This sequence-based genomic assay provides an unbiased interrogation of all polyadenylated transcripts in male germ cell transcriptome. Up to 75% of tags were not properly annotated in our male germ cell SAGE libraries. Therefore, we hypothesize that a subset of

unannotated SAGE tags may be related to ncRNAs. In this protocol, we demonstrate how to identify lncRNAs in spermatids by using the GermSAGE database (11), a publicly available SAGE database for male germ cell transcriptome data.

2. Materials

2.1. Male Germ Cell Samples

1. Type A spermatogonia were isolated from 6-day-old Balb/c mice (university stocks, Georgetown University). Since 6-day-old mice testis do not contain spermatocyte or spermatid, type A spermatogonia are typically >99% pure after separation by differential plating from contaminating Sertoli cells.
2. RNA quality is critical to the overall success of the experiment. Minimize the number of freeze and thaw cycle if possible and make sure that the experiment is performed in an RNAase-free environment.

2.2. Serial Analysis of Gene Expression (Optional)

1. I-SAGE kit module (Invitrogen).
2. I-SAGE Ditag PCR Module (Invitrogen).
3. Cloning kit (Qiagen).
4. ABI 3730xl sequencer.
5. SAGE2000 software.

2.3. Quantitative Real-Time PCR Validation

1. Primers are designed by Primer3 and show no significant homology with other genomic regions by NCBI Blast.
2. Taqman PCR Reagent Kit: Ampli*Taq*Gold DNA polymerase, dUTP, dATP, dCTP, dGTP, 1.2 ml 10× Taqman buffer A, and 2 × 1.5 ml 25 mM $MgCl_2$.
3. Bovine serum albumin (BSA).
4. 1× Tris–EDTA buffer (10 mM Tris–HCl containing 1 mM EDTA·Na_2).
5. ABI PRISM Sequence Detection System.

2.4. Northern Blot Analysis

1. Digoxigenin (DIG) Northern Starter Kit (Roche).
2. DIG Easy Hyb Granules.
3. Anti-DIG-AP, Fab fragments.
4. Chemiluminescent substrate.
5. Actin RNA Probe, DIG labeled.
6. DIG-11-UTP.
7. DIG wash and block buffer set.
8. Nylon membrane, positively charged.
9. Hybridization bags.
10. 20× SSC buffer (pH 7.0).

11. Immunological detection solutions.
12. Stripping buffer:50% deionized formamide, 5% SDS, 50 mM Tris–HCl, adjust to pH 7.5.

Solution	Preparation
Washing buffer	0.1 M maleic acid, 0.15 M NaCl, and 0.3% Tween 20 (pH 7.5)
Maleic acid buffer	0.1 M maleic acid, 0.1 M NaCl (pH 7.5)
Detection buffer	0.1 M Tris–HCl, 0.1 M NaCl (pH 9.5)
1× blocking solution	Dilute 10× stock with maleic acid (1:10)
Antibody solution	Centrifuge anti-DIG-AP for 5 min at 13,750×*g* (10,000 rpm)and dilute in blocking solution (1:10,000)

13. Stripping solution (optional): DEPC water, stripping buffer, 20× SSC, and 2× SSC.

2.5. Bioinformatics Tools and Databases

1. SAGE2000 software (Invitrogen).
2. Primer3 (http://frodo.wi.mit.edu/primer3/).
3. NCBI Blast (http://blast.ncbi.nlm.nih.gov).
4. RNAfold (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi).
5. SAGEmap (http://www.ncbi.nlm.nih.gov/projects/SAGE/).
6. GermSAGE (http://germsage.nichd.nih.gov/).
7. UCSC genome browser (http://genome.ucsc.edu/).
8. Galaxy (http://galaxy.psu.edu/).
9. fRNAdb (http://www.ncrna.org/frnadb/).

3. Methods

3.1. Construction of SAGE Libraries (Optional, see Note 1)

1. Mix 5 μg of total RNA with oligo dT magnetic beads.
2. Synthesize double-strand cDNA.
3. Digest with NlaIII to form one end of the tag.
4. Divide in half and ligate 40-bp adapters containing the recognition sequence for the type-II restriction enzyme BsmF1.
5. Cleave with BsmF1 to form ~50 bp tag (40 bp adaptor/13 bp tag).
6. Fill in 5′ overhangs and ligate to form an ~100-bp ditag.
7. PCR amplify using ditag primers 1 and 2.
8. Cut 40-bp adapters with Nla III to release the 26 bp ditag.

[1]Please refer the outline of a SAGE experiment to manufacturer's protocol. This is optional in this chapter. We applied data from GermSAGE database in this protocol. The instruction for data downloading in GermSAGE is indicated in the help section of the Web site.

9. Ligate ditags to form concatemers.
10. Clone into pZErO®-1 and sequence.
11. Annotate the tag by SAGEmap.

3.2. Identification of Spermatid-Specific SAGE Tags from GermSAGE

1. The data from SAGE libraries have been deposited in the publicly available GermSAGE database (http://germsage.nichd.nih.gov/). Download the raw data files from the three main germ cell stages (type A spermatogonia, pachytene spermatocytes, round spermatids) from GermSAGE.
2. To screen for abundantly expressed tags, filter the tag count by equal or greater than five.
3. Transcripts represented by the tags in step 2 may be expressed in more than one male germ cell stage. To identify stage-specific expressed tags, apply Venn diagram analysis on the three datasets (see Note 2).
4. Identify genomic coordinates of Sptd tags by BLAST (see Note 3).
5. To increase the reliability of transcript prediction, align the coordinate results with polyadenylation (PolyA) signal data (see Note 4).
6. Identify stand orientation of tags, and calculate the relative distance of polyA signal to the 3′ end of each tag. Tags without PolyA evidence are eliminated.
7. Download mouse long noncoding transcript track data (fRNAdb::uc-ncRNA) from fRNAdb (see Note 5).
8. Import the long noncoding transcript data from fRNAdb and the filtered tag data from step 6 to Galaxy.
9. Use “Operate on Genomic Intervals” function in Galaxy to align both datasets. Set to detect the presence of tags around 500 bp to both ends. Inconsistent orientations between the transcript and the tag are removed.
10. Determine the stability of transcripts by RNAfold by uploading the sequence in FASTA format. Select “minimum free energy (MFE) and partition function” and check “avoid isolated base pairs” under “Fold algorithms and basic options.” Check all boxes in the “Output options” section to generate interactive RNA secondary structure plot, RNA secondary structure plots with reliability annotation, and mountain plot.
11. Rank the transcripts based on the distance of matching SAGE tag to the 3′ end of the lncRNA transcripts. This is calculated by subtracting the genomic coordinate at the 3′ end of the transcript from the start position of SAGE tag.
12. Use “Intersect” function of Galaxy to find the association of Sptd lncRNAs to the existing Refseq annotation. Classify the lncRNAs into one of the following categories: (a) Promoter (7), (b) Intronic (3), (c) 3′UTR (8), and (d) Intergenic (6) (see Note 6).

---

[2] Venn diagram can be generated by using BioVenn (http://www.cmbi.ru.nl/cdd/biovenn/).
[3] Local Blast is more fl exible. The window size is set to 7.
[4] As transcripts identified by SAGE should be polyadenylated, including nearby polyadenylated signal evidence increases the reliability of prediction. Polyadenylated database can be found at PolyA_DB (http://polya.umdnj.edu/polyadb/).
[5] fRNAdb is a database for comprehensive noncoding RNA sequences. fRNAdb::uc-ncRNA track contains unannotated lncRNAs with only one exon. This eliminates the complexity of ORF predication and follow-up validations.
[6] The workfl ow details of Galaxy can be found at http://main.g2.bx.psu.edu/screencast.

13. Examine the Sptd lncRNA and tag annotation in UCSC genome browser (optional). Sptd lncRNAs with high conversion evidence (MultiZ) score are selected first.

3.3. Validate Stage-Specific Expression by Quantitative Real-Time PCR

1. Design primer sets targeting the predicted lncRNA transcripts by Primer3. Include 18 s rRNA actin or tubulin as reference control (see Note 7).
2. Reverse transcribe RNA into cDNA by using random hexamers as follows:

Reagent	Volume for each reaction
10× Taqman RT buffer	10 μl
25 mM $MgCl_2$	22 μl
Random hexamers (50 ng/ml)	5 μl
500 μM dNTP mix	20 μl
Reverse transcriptase (50 U/μl)	2.5 μl
DEPC water	Variable
Final volume	100 μl

3. Mix and incubate in a PCR thermal cycler as follows: At 25°C for 10 min, 48°C for 30 min, and 95°C for 5 min.
4. Prepare qPCR component as follows:

Reagent	Volume for each reaction
2× SYBR green mix	15 μl
Primers 10 μM (forward)	1 μl
Primers 10 μM (reverse)	1 μl
cDNA template	1 μl
DEPC water	Variable
Final volume	25 μl

5. Mix and incubate in a real-time PCR machine. The conditions for the amplification are as follows: At 50°C for 2 min, 95°C for 10 min followed by 95°C for 15 s, 60°C for 30 s, 72°C for 30 s for 40 cycles, and end the protocol at 72°C for 10 min.
6. Analyze the data according to manufacturer's protocol.

3.4. Confirmation of Predicted ORF by Northern Blots

1. Reverse transcribe input RNA (>1 μg) using a standard reverse transcriptase system.
2. Prepare DNA template from total RNA according to the table below:

[7] Parameters to be measured during primer design are as follows: melting temperature, free energy for the formation of hairpins, self-dimerization or heterodimer formation should be below $\Delta G \leq -9$ kcal/mol. Verify that the amplicon with values of $\Delta G \leq -9$ kcal/mol using the DINAMelt Server (http://dinamelt.bioinfo.rpi.edu) to avoid problematic secondary structures.

Reagent	Volume
DEPC-treated water	Variable
10× expand buffer	5 μl
10 mM d(ACGT)P	1 μl each
Primer (sense)	1 μl
Primer (antisense)	1 μl
Expand high fidelity	0.75 μl
cDNA	2 μl
Final volume	50 μl

3. Incubate the samples in a thermocycler as follows: 94°C for 45 s, 60°C for 45 s, and 72°C for 90 s. Repeat the reaction for 30 cycles.
4. Mix 4 μl of PCR product and 6 μl of DEPC-treated water. Prepare the DIG-labeling mix as follows (see Note 8):

Reagent	Volume
5× labeling mix	4 μl
5× transcription buffer	4 μl
RNA polymerase (SP6, T7, or T3)	2 μl

5. Mix the PCR product and the DIG-labeled mix. Incubate at 42°C or 1 h.
6. Add 2 μl DNA to remove template DNA. Incubate at 37°C for 15 min.
7. Stop the reaction by adding 2 μl of 0.2 M (pH 8.0) EDTA.
8. To determine labeling efficiency, apply 1 μl of labeled probes and controls to the nylon membrane (see Note 9).
9. Fix the probes on membrane by baking at 120°C for 30 min.
10. Wash the membrane with 20 ml of washing buffer and incubate under shaking at room temperature for 2 min (see Note 10).
11. Remove the washing buffer and replace with 10 ml of blocking solution. Incubate for 30 min.
12. Remove the blocking solution and replace with 10 ml of antibody solution. Incubate for 30 min.
13. Remove the antibody solution and wash with 20 ml of washing buffer for 15 min.
14. Place the membrane on a development folder and apply four drops of CDP-star. Incubate at room temperature for 5 min.

[8] DIG-labeled single-stranded RNA probes are constructed by in vitro transcription of input template DNA. DNA templates have to be linearized at a restriction site of cloned insert. The transcribed sequence should be between 200- and 1,000-bp long. High concentration of EDTA (>0.1 mM) inhibits transcription reaction.

[9] The yield of DIG-labeled RNA is critical for hybridization results. High probe concentration causes significant background while too low concentration causes compromised signals.

[10] The volume is for small membrane (3 × 5 cm), which can be processed in a plastic container or petri dish.

15. Expose the membrane to an X-ray film at room temperature for 20 min.
16. Compare the spot intensity of labeling reaction to control probe to evaluate the amount of DIG-labeled product.
17. To prepare for gel electrophoresis, add 20 μl of loading buffer to the RNA probe, denature at 65°C for 15 min, and cool down on ice for 1 min.
18. Run the product on a formaldehyde gel for 2 h (see Note 11).
19. Evaluate the quality of target RNA by staining in 0.25 μg/ml of ethidium bromide briefly. Examine the gel under UV light.
20. Transfer the RNAs using standard transfer protocol (see Note 12). Rinse gel in 20× SSC for 15 min twice.
21. Fix the RNAs to the membrane by UV-cross-linking or baking. For UV cross-linking, place the membrane on a Whatman paper soaked with 2× SSC. For baking, briefly rinse membrane in 2× SSC twice. Bake at 80°C for 2 h (see Note 13).
22. Prepare hybridization solution by reconstituting granules in 64 ml of DEPC-treated water at 37°C.
23. Prewarm 15 ml of DIG Easy Hyb solution to 68°C. Prehybridize the membrane in prewarmed DIG Easy Hyb for 30 min with gentle agitation.
24. Denature DIG-labeled RNA probe (100 ng/ml) at 99°C in a thermal cycler for 5 min. Place the probe on ice immediately.
25. Mix the denatured DIG-labeled RNA probe with prewarmed DIG Easy Hyb (see Note 14).
26. Remove prehybridization solution and replace with hybridization solution. Incubate overnight with gentle agitation.
27. Remove the hybridization mix and wash in 2× SSC with 0.1% SDS at room temperature under constant agitation. Repeat the wash twice.
28. Wash twice with 0.1× SSC in 0.1% SDS prewarmed at 68°C under constant agitation.
29. Rinse the membrane with washing buffer for 5 min.
30. Remove washing buffer and incubate with 100 ml of blocking solution for 30 min.
31. Remove blocking solution and incubate with 50 ml of antibody solution for 30 min.
32. Remove antibody solution and wash with 100 ml of washing buffer for 15 min. Repeat twice.
33. Remove washing buffer and incubate with 100 ml of detection buffer for 5 min.
34. Put the membrane in a hybridization bag, add 1 ml of CDP-star solution to soak the membrane. Incubate at room temperature for 5 min (see Note 15).

---

11 Run the gel at 4 V/cm in RNase-free gel for at least 2 h. Overnight run is preferred.

12 The details of transfer procedure can be found at the following URL: http://www.nature.com/nmeth/journal/v2/n12/full/nmeth1205-997.html. Use only positively charged nylon membrane. Usually, gels blotted by capillary transfer in 20× SSC overnight give the best results. Alkali transfer is not recommended.

13 We found that baking the membrane at 80°C gave the most consistent results. Baking at 120°C is another option, but not recommended.

14 The homology between the probe and target RNA is a critical factor for determining hybridization condition. Hybridization stringency is dependent on temperature. High temperature increases stringency, and low temperature decreases stringency. The optimal temperature is 68°C for DIG-labeled probes.

35. Put the membrane to an image cassette with an X-ray film. Expose at room temperature for 15–25 min (see Note 16).

3.5. Stripping and Reprobing of RNA Blots (Optional)

Stripping and reprobing allow signal optimization and reuse of the existing membrane without tedious RNA loading and membrane preparation procedure. It is important to note that only membranes without drying at any time during the hybridization and detection procedure could be used.

1. Prewarm water bath to 80°C.
2. Place the membrane in a hybridization bag and rinse membrane in DEPC-treated water.
3. Replace DEPC-treated water with stripping solution and incubate at 80°C for 60 min to remove the DIG-labeled probe.
4. Aspirate the stripping buffer. Wash in 2× SSC for 5 min twice.
5. Prehybridize and hybridize with another probe as mentioned in the previous section. Store membrane in a sealed bag at 4°C if not used immediately.

Acknowledgments

This work was supported by the Intramural Research Program of the National Institutes of Health (NIH), Eunice Kennedy Shriver National Institute of Child Health and Human Development.

References

1. Mattick JS. Non-coding RNAs: the architects of eukaryotic complexity. EMBO Rep. 2001; 2:986–991. [PubMed: 11713189]
2. Mattick JS, Makunin IV. Small regulatory RNAs in mammals. Hum Mol Genet. 2005; 14(Spec No 1):R121–132. [PubMed: 15809264]
3. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. Nat Rev Genet. 2009; 10:155–159. [PubMed: 19188922]
4. Taft RJ, Pang KC, Mercer TR, Dinger M, Mattick JS. Non-coding RNAs: regulators of disease. J Pathol. 2010; 220:126–139. [PubMed: 19882673]
5. Wu SM, Baxendale V, Chen Y, Pang AL, Stitely T, Munson PJ, Leung MY, Ravindranath N, Dym M, Rennert OM, et al. Analysis of mouse germ-cell transcriptome at different stages of spermatogenesis by SAGE: biological significance. Genomics. 2004; 84:971–981. [PubMed: 15533714]
6. Chan WY, Lee TL, Wu SM, Ruszczyk L, Alba D, Baxendale V, Rennert OM. Transcriptome analyses of male germ cells with serial analysis of gene expression (SAGE). Mol Cell Endocrinol. 2006; 250:8–19. [PubMed: 16413108]
7. Chan WY, Wu SM, Ruszczyk L, Law E, Lee TL, Baxendale V, Lap-Yin Pang A, Rennert OM. The complexity of antisense transcription revealed by the study of developing male germ cells. Genomics. 2006; 87:681–692. [PubMed: 16458478]
8. Lee TL, Li Y, Alba D, Vong QP, Wu SM, Baxendale V, Rennert OM, Lau YF, Chan WY. Developmental staging of male murine embryonic gonad by SAGE analysis. J Genet Genomics. 2009; 36:215–227. [PubMed: 19376482]

[15] Do not allow membrane to dry and make sure that enough solution covers the blotting area to avoid “halo” effect in developed image. Make sure to spread the substrate evenly to prevent formation of air bubbles over the membrane. Avoid adding too much solution or use large-size hybridization bag. Squeeze out excess solution if necessary. Unevenly soaked membrane results in dark background.

[16] Luminescence could last up to 24 h. Therefore, multiple exposure could be made within this period.

9. Lee TL, Li Y, Cheung HH, Claus J, Singh S, Sastry C, Rennert OM, Lau YF, Chan WY. GonadSAGE: a comprehensive SAGE database for transcript discovery on male embryonic gonad development. Bioinformatics. 2010; 26:585–586. [PubMed: 20028690]

10. Lee TL, Pang AL, Rennert OM, Chan WY. Genomic landscape of developing male germ cells. Birth Defects Res C Embryo Today. 2009; 87:43–63. [PubMed: 19306351]

11. Lee TL, Cheung HH, Claus J, Sastry C, Singh S, Vu L, Rennert O, Chan WY. GermSAGE: a comprehensive SAGE database for transcript discovery on male germ cell development. Nucleic Acids Res. 2009; 37:D891–897. [PubMed: 18832368]