



Published in final edited form as:

Methods Mol Biol. 2012 ; 825: 75–92. doi:10.1007/978-1-61779-436-0_7.

Revealing Transcriptome Landscape of Mouse Spermatogonial Cells by Tiling Microarray

Tin-Lap. Lee¹, Owen M. Rennert¹, and Wai-Yee. Chan^{1,2}

¹Laboratory of Clinical and Developmental Genomics, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892

²School of Biomedical Sciences, The Chinese University of Hong Kong, Hong Kong SAR, China

Summary

Spermatogenesis is a highly regulated developmental process by which spermatogonia develop into mature spermatozoa. This process involves many testis- or male germ cell-specific events through tightly regulated gene expression programs. In the past decade the advent of microarray technologies has allowed functional genomic studies of male germ cell development, resulting in the identification of genes governing various processes. A major limitation with conventional gene expression microarray is that there is a bias from gene probe design. The gene probes for expression microarrays are usually represented by a small number probes located at the 3' end of a transcript. Tiling microarrays eliminate such issue by interrogating the genome in an unbiased fashion through probes tiled for the entire genome. These arrays provide a higher genomic resolution and allow identification of novel transcripts. To reveal the complexity of the genomic landscape of developing male germ cells, we applied tiling microarray to evaluate the transcriptome in spermatogonial cells. Over 50% of the mouse and rat genome are expressed during testicular development. More than 47% of transcripts are uncharacterized. The results suggested the transcription machinery in spermatogonial cells are more complex than previously envisioned.

Keywords

Spermatogenesis; male germ cells; development; tiling microarray; expression; profiling

1. Introduction

Spermatogenesis is a key process in mammalian reproduction. This highly ordered process requires well-controlled programs governed by dynamic patterns of gene expression. Some genes are exclusive to spermatogenic cells, while others are closely related to genes expressed in somatic cells. Although key genes in male germ cell development have been identified, the biological mechanisms and transcripts that govern the programs of

*Corresponding Author: Tin-Lap Lee, Ph.D., Laboratory of Clinical and Developmental Genomics, Eunice Kennedy Shriver National Institute of Child Health and Human Development, 49 Convent Drive, MSC4429, Bethesda, MD 20892-4429, USA, Tel:(301) 451-8368, Fax: (301) 480-4700, leetl@mail.nih.gov.

spermatogonial stem cell renewal, germ cell differentiation during spermatogenesis or fertilization remain largely unknown.

The advent of microarray technologies provides solutions to study the gene expression profile in germ cell development. A major observation from these microarray studies is that the genome is actively transcribed during germ cell development (1–3). Importantly, many differentially expressed transcripts were unknown or uncharacterized (4–6). In addition, the germ cell transcriptome exhibits dynamic changes in conjunction with specific developmental regulation. Male germ cell development happens in three phases. The first phase, from day 0 to postnatal day 8, is mitosis when spermatogonial proliferation predominates. The second phase occurred at the initiation of meiosis, on day 14, during early pachytene spermatocytes development. This was followed by entry into spermiogenesis on day 20 when round spermatids first appear (7). The dynamic and specific nature of the germ cell transcriptome was shown to associate with specific development and regulatory programs. Ontology analysis of germ cell transcriptome revealed different biological processes distinctively associated with mitotic, meiotic and post-meiotic male germ cells respectively (2,7–9).

Although microarrays prove to be an invaluable tool in the functional genomics study of germ cell development, it has some limitations. Traditional microarrays represent only a selected set of transcripts at the time of fabrication, and only a small number of probe sets determine the expression level of a given transcript. Transcripts with alternative splicing isoforms may be misinterpreted, if the probes are all located in the common region shared by the variants. The probes in traditional microarray platforms are designed based on sequences located at the 3' end region of a transcript. Therefore, it does not provide a complete and accurate picture of transcript expression data, which presents a challenge in downstream validation experiments.

Tiling microarrays addressed these issues by designing probes throughout the entire genome or contigs of the genome in an unbiased fashion with higher resolution (10–14). The whole genome coverage feature allows a wide range of genomic applications, including whole genome transcriptome analysis (14), detection of transcription factor binding sites (13), chromatin modifications and epigenetic modifications such as DNA or histone methylation (15). In this chapter, we apply Affymetrix mouse tiling 1.1R microarray sets (GeneChips) to examine the transcriptional landscape in the spermatogonial cells. Overall, we found more than 45 percent of transcripts were not annotated. Current annotation accounts for only about 30 percent of the data set. Majority of data set is in the form of expressed sequence tags (ESTs) or non-coding RNAs.

2. Materials

2.1. Spermatogonial 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 have spermatocytes or spermatids, type A spermatogonia are typically >99% pure after separation from contaminating Sertoli cells by differential plating.

2. For Transcriptome tiling array application in triplicates, at least 15 μg of total RNA is required. The minimum concentration is 0.24 $\mu\text{g}/\mu\text{l}$.
3. RNA quality is critical to the overall success of experiment. Minimize the number of freeze and thaw cycle if possible and make sure the experiment is performed in RNAase-free environment.

2.2 Tiling Microarrays

1. Affymetrix GeneChip® Mouse Tiling 1.1R Array Set designed for whole transcriptome analysis.
2. The chips are based on Affymetrix 49 format. Each chip contains more than three million probe pairs (perfect and mismatch) to cover whole mouse genome at an average 35 base pair resolution. The average gap between probes is 10 base pair. All repetitive elements derived from RepeatMasker were removed.
3. Sequences used in probe set design were based on NCBI mouse genome assembly (Build33). The results could be updated to other build version using genome coordinate conversion tools (see methods).

2.3 rRNA reduction

1. RiboMinus Human/Mouse Transcriptome Isolation Kit. Resuspend the RiboMinus Magnetic Beads completely before use.
2. Magna-Sep Magnetic Particle Separator
3. Betaine, 5M

Synthesis, Amplification and Labeling of cDNA and cRNA—GeneChip WT Amplified Double-Stranded cDNA Synthesis Kit containing:

- T7 primers, 2.5 $\mu\text{g}/\mu\text{l}$
- 5X 1st Strand Buffer
- DTT, 0.1M
- dNTP, 10mM
- RNase Inhibitor
- SuperScript II
- MgCl_2 , 1M
- DNA Polymerase I
- RNase H
- Random Primers, 3 $\mu\text{g}/\mu\text{l}$
- dNTP+dUTP, 10mM
- RNase-free Water

GeneChip WT cDNA Amplification Kit containing:

- 10X IVT Buffer
- IVT NTP Mix
- IVT Enzyme Mix
- IVT Control

GeneChip WT Double-Strand DNA terminal Labeling Kit containing:

- 10X cDNA Fragmentation Buffer
- UDG, 10U/ μ l
- APE1, 100U/ μ l
- 5X TdT Buffer
- TdT, 30U/ μ l
- GeneChip DNA Labeling Reagent, 5mM
- RNase-free Water

cDNA, cRNA Clean-up and Fragmentation—GeneChip Sample Cleanup Module containing:

- cDNA Cleanup Spin Columns
- cDNA Binding Buffer
- cDNA Wash Buffer, 6ml concentrate
- cDNA Elution Buffer
- In-vitro Transcription (IVT) cRNA Cleanup Spin Columns
- IVT cRNA Binding Buffer
- IVT cRNA Wash Buffer, 5ml concentrate
- 5X Fragmentation Buffer

Gel-Shift Assay—RNA 6000 Nano LabChip Reagents and Supplies. (Agilent) TBE Gel, 4–20%, 1mm, 12 well. (Invitrogen)

GeneChip Hybridization, Wash, and Staining—GeneChip Hybridization, Wash, and Stain kit containing:

- Pre-Hybridization Mix
- 2X Hybridization Mix
- DMSO, store at room temperature
- Stain Cocktail 1, light sensitive, store in dark bottle
- Stain Cocktail 2

- Array Holding Buffer
- Wash Buffer A, store at room temperature to prevent formation of salt particles
- Wash Buffer B, light sensitive, store in dark bottle
- Control Oligonucleotide B2, 3nM
- 20X Eukaryotic Hybridization Controls (bioB, bioC, BioD, cre), frozen stocks needs to be heated to 65°C for 5 minutes to resuspend the cRNA completely before use

Instrumentation

1. GeneChip Scanner 3000 7G. Previous versions of 3000 series scanner can be upgraded to 7G by through Affymetrix scanner upgrade program.
2. GeneChip Hybridization Oven 640.
3. GeneChip AutoLoader (Optional)
4. GeneChip Fluidics Station 400 or 450 series. This protocol in this chapter is based on Fluidics 450 station.
5. NanoDrop ND-1000 (Ambion)
6. Bioanalyzer 2100 (Agilent)

Software

1. Affymetrix GeneChip Command Console Software (AGCC) or GeneChip Operating Software (GCOS) version 1.3 or higher.
2. Affymetrix Tiling Analysis Software (TAS)
3. Integrated Genome Brower from web start (http://www.affymetrix.com/partners_programs/programs/developer/tools/download_igb.affx) or through local installation.
4. The softwares are available for free for Affymetrix users, downloadable at <http://www.affymetrix.com>.

3 Methods

The procedure from capturing polyadenylated transcripts to microarray scanning is illustrated in figure 1. Basically, mRNA population in the total RNA is first enriched by RiboMinus kit. The enriched mRNA population is then converted to cDNA by cDNA synthesis. The cDNA product serves as a template for subsequent *in vitro* transcription to cRNA. Contrary to conventional 3'IVT GeneChip microarrays, the cRNA will be amplified for another round to cDNA before fragmentation and labeling. Once the fragmented products are labeled with biotin, it will be mixed with eukaryotic RNA and oligo B2 controls for array hybridization.

Input RNA quality is critical for the overall success of the experiment. Therefore, the quality and quantities of cRNA generated are very dependent on the O.D_{260/280} ratio and amount of input total RNA. The intensity of signals is sensitive to artifacts contained in the poor quality samples. According to our experience, we recommend the input total RNA quality has an O.D_{260/280} ratio of at least 2. Three micrograms of total RNA is sufficient for up to nine GeneChip hybridization reactions. To study non-polyadenylated transcripts or transcripts unique to cytosol or nuclear fraction, additional separation protocols are required before proceeding to cDNA synthesis step.

3.1 Enriching mRNA Fraction from Total RNA using RiboMinus Kit

1. Prepare RiboMinus buffer with betaine by mixing 120 μl of betaine (5M) and 280 μl of RiboMinus hybridization buffer.
2. Prepare RiboMinus probe hybridization mix in a 0.2 ml strip tube according to the table below and incubate at 70°C for 5 minutes in a thermal cycler.

	Concentration of total RNA input	
	0.4 $\mu\text{g}/\mu\text{l}$	0.4 $\mu\text{g}/\mu\text{l}$ and 0.24 $\mu\text{g}/\mu\text{l}$
Total RNA	3 μg	3 μg
RiboMinus Probe	2.4 μl	2.4 μl
Hybridization Buffer with Betaine	60 μl	90 μl
RNase-free Water	up to 70 μl	up to 105 μl
Total Volume	70 μl	105 μl

3. Meanwhile, pre-heat a 37°C heat block and a 50°C heat block for mRNA capture using magnetic beads.
4. Re-suspend the RiboMinus magnetic beads by pipetting several times until no deposit is observed at the bottom. Transfer 150 μl of beads to a 1.5 ml non-stick RNase-free tube (*see* Note 1).
5. Put the tube on a magnetic stand. Aspirate and discard the clear supernatant after one minute.
6. Wash the magnetic bead pellet with 150 μl RNase-free water by resuspending the pellet. Spin briefly and put the tube back to the magnetic stand for 1 minute. Repeat the step for one more time.
7. Wash with 150 μl of hybridization buffer with betaine. Aspirate and discard the supernatant.
8. Depending on the concentration of the input RNA, add 90 (input 0.4 $\mu\text{g}/\mu\text{l}$) or 60 (less than 0.4 $\mu\text{g}/\mu\text{l}$) μl of hybridization buffer with betaine to resuspend the magnetic bead pellet.

¹Do not allow the magnetic beads to dry out. Dried beads will reduce the recovery of RNA. Resuspend the beads immediately.

9. Incubate the mix in the 37°C heat block for 5 minutes, resuspend the mix and incubate for 5 more minutes. Place the tube in the magnetic stand for 1 minute.
10. Transfer the clear supernatant containing the purified RNA to a 1.5 ml non-stock RNase-free tube and leave on ice.
11. Wash the magnetic beads with 50 µl of hybridization buffer with betaine, incubate in the 50°C heat block for 5 minutes. Place the tube back to magnetic stand for 1 minute.
12. Aspirate the clear supernatant and transfer to the tube containing the purified RNA in step 10

3.2 RNA cleanup

1. Add 735 µl of cRNA binding buffer from the GeneChip Sample Cleanup Module to each sample and vortex for 3 seconds. Then add 525 µl of 100% ethanol to each sample. Mix by flicking.
2. Apply the sample to a IVT cRNA column in a 2 ml collection tube.
3. Centrifuge for 15 seconds at 8000 g. Discard the flow-through.
4. Using the same column, repeat step 2 and 3 for the remaining samples.
5. Discard the collection tube and combine the column in a new 2 ml collection tube. Add 500 µl of cRNA wash buffer and centrifuge at 8000 g for 15 seconds. Discard the flow through.
6. Add 500 µl of 80% ethanol and centrifuge at 8000 g for 15 seconds. Discard the flow through.
7. Open the column cap and centrifuge at maximum speed for 5 minutes.
8. Transfer the column to a new 1.5 ml collection tube. Add 11 µl of RNase-free water directly to the center of the membrane. Centrifuge at maximum speed for 1 minutes.
9. The eluted RNA should be around 9.8 µl. Analyze the quality of purified RNA using a Bioanalyzer (optional, *see* Note 2).

3.3 First cycle Double-Stranded cDNA Synthesis

1. Prepare the RNA/T7 primer mix by combining 4 µl of RNA and 1 µl of 500 ng/µl T7 primer (*see* Note 3).
2. Incubate at 70°C for 5 minutes. Place the sample on ice.
3. Prepare the first cycle, first strand master mix as follows:

²To check if rRNA depletion is efficient, load 1 µl of the purified RNA to a RNA 6000 Nano LabChip for RNA profile analysis.

³Use SpeedVac to reduce the volume of RNA sample. Working T7 primer is diluted in 1:5 from stock.

Component	Volume in 1 reaction
5× 1 st Strand Buffer	2 µl
DTT, 0.1M	1 µl
dNTP mix, 10 mM	0.5 µl
RNase Inhibitor	0.5 µl
SuperScript II	1 µl
Total Volume	5 µl

4. Add 5 µl of master mix to each sample to make a total volume of 10 µl.
5. Incubate the samples in a thermal cycler as follows: 25°C for 10 minutes; 42°C for 1 hour; 70°C for 10 minutes and 4°C for 10 minutes.
6. Prepare the second cycle, second strand master mix as follows:

Component	Volume in 1 reaction
MgCl ₂ , 17.5 mM	4 µl
dNTP mix, 10 mM	0.4 µl
DNA Polymerase I	0.6 µl
RNase H	0.2 µl
RNase-free Water	4.8 µl
Total Volume	10 µl

7. Combine the first cycle, first strand mix from step 5 to the mix from step 6 to make a total volume of 20 µl.
8. Incubate the samples in a thermal cycler as follows: 16°C for 2 hours without heated lid; 75°C for 10 minutes with heated lid and 4°C for 10 minutes.

3.3 In-vitro Transcription (IVT) Amplification

1. Prepare the IVT master mix as follows:

Component	Volume in 1 reaction
10× IVT Buffer	5 µl
IVT NTP Mix	20 µl
IVT Enzyme Mix	5 µl
Total Volume	30 µl

2. Add 30 µl of IVT master mix to each sample to make a total volume of 50 µl. Mix and briefly spin down the mix.
3. Incubate the samples in a thermal cycler as follows: 37°C for 16 hours and 4°C at hold position.

4. On day 2, clean up the cRNA from step 3 using the cRNA cleanup spin columns from the cleanup module.
5. Add 50 μ l of RNase-free water to each sample to make a total volume of 100 μ l.
6. Add 350 μ l of cRNA binding buffer to each sample and vortex for 3 seconds.
7. Add 250 μ l of 100% ethanol to each sample and flick mix.
8. Transfer the sample to the IVT cRNA column in a 2 ml collection tube.
9. Centrifuge at 8000g for 15 seconds. Discard the flow through.
10. Discard the collection tube and combine the column to a new 2 ml collection tube. Add 500 μ l of cRNA wash buffer and centrifuge at 8000 g for 15 seconds. Discard the flow through.
11. Add 500 μ l of 80% ethanol and centrifuge at 8000 g for 15 seconds. Discard the flow through.
12. Open the column cap and centrifuge at maximum speed for 5 minutes.
13. Transfer the IVT cRNA column to a new 1.5 ml collection tube. Add 12 μ l of RNase-free water directly to the center of the membrane. Centrifuge at maximum speed for 1 minutes.
14. Repeat step 13. The eluted cRNA will be about 21 μ l. Analyze the quality of purified RNA by measuring the readings at 260nm, 280nm and 320nm using spectrophotometer. The concentration of cRNA is calculated by this formula (see Note 4): $[A_{260} - A_{320}] \times 0.04 \times \text{dilution factor}$

3.4 Second Cycle Double-strand cDNA Synthesis

1. Divide each sample into three separate reactions to be used for hybridization with three tiling microarrays by preparing the second cycle, cRNA/Random primer mix as follows:

Component	Volume in 1 reaction
cRNA, 7 μ g	Variable
Random Primers (3 μ g/ μ l)	1 μ l
RNase-free water	up to 8 μ l
Total Volume	8 μ l

2. Flick mix and spin down the tubes.
3. Incubate the samples in a thermal cycler as follows: 70°C for 5 minutes; 25°C for 5 minutes and 4°C for 10 minutes.
4. Prepare second cycle, first strand cDNA synthesis master mix as follows:

⁴Make sure the total cRNA is equal to or above 21 μ g. Poor amplification efficiency or RNA input quality could lead to low cRNA yield.

Component	Volume in 1 reaction
5× 1 st Strand Buffer	4 µl
DTT, 0.1M	2 µl
dNTP+dUTP, 10 mM	1.25 µl
SuperScript II	4.75 µl
Total Volume	12 µl

5. Combine the first cycle, first strand cDNA synthesis master mix with cRNA/ Random primer mix from step to make a total volume of 20 µl.
6. Incubate the samples in a thermal cycler as follows: 25°C for 10 minutes; 42°C for 90 minutes; 70°C for 10 minutes and 4°C for 10 minutes.
7. Prepare second cycle, second strand cDNA synthesis master mix as follows:

Component	Volume in 1 reaction
MgCl ₂ , 17.5 mM	8 µl
dNTP+dUTP, 10 mM	1 µl
DNA Polymerase I	1.2 µl
RNase H	0.5 µl
RNase-free water	9.3 µl
Total Volume	20 µl

8. Combine the second cycle, second strand cDNA synthesis master mix with second cycle, first strand cDNA reaction from step to make a total volume of 40 µl.
9. Incubate the samples in a thermal cycler as follows: 16°C for 2 hours without heated lid; 75°C for 10 minutes with heated lid and 4°C for 10 minutes.
10. Clean up the cDNA using cDNA clean up spin columns from the GeneChip Sample Cleanup Module.
11. Add 60 µl of RNase-free water and 370 µl of cDNA binding buffer to each cDNA sample. Vortex for 3 seconds.
12. Transfer the mix to the cDNA binding column with 2 ml collection tube and centrifuge at 8000 g for 15 seconds. Discard the flow through.
13. Add 500 µl of 80% ethanol and centrifuge at 8000 g for 15 seconds. Discard the flow through.
14. Discard the collection tube and combine the cDNA column to a new 2 ml collection tube. Add 750 µl of cDNA wash buffer and centrifuge at 8000 g for 15 seconds. Discard the flow through.
15. Open the column cap and centrifuge at maximum speed for 5 minutes.

16. Transfer the cDNA column to a new 1.5 ml collection tube. Add 15 μ l of RNase-free water directly to the center of the membrane. Centrifuge at maximum speed for 1 minutes.
17. Repeat step 16. The eluted cDNA will be about 28 μ l. Analyze the quality of purified cDNA by measuring the readings at 260nm, 280nm and 320nm using spectrophotometer. The concentration of cDNA is calculated by this formula:
 $[A_{260} - A_{320}] \times 0.05 \times \text{dilution factor}$
18. Each sample tube should contain at least 7.5 μ g of cDNA. Pooling the cDNA sample is recommended to minimize variability across reactions.

3.5 Fragmentation and Labeling

1. Fragment the cDNA by preparing the fragmentation mix as follows:

Component	Volume in 1 reaction
cDNA	7.5 μ g
10 \times Fragmentation Buffer	4.8 μ l
UDG, 10U/ μ l	1.5 μ l
APE1, 100U/ μ l	2.25 μ l
RNase-free water	up to 48 μ l
Total Volume	48 μ l

2. Incubate the samples in a thermal cycler as follows: 37°C for 1 hour; 93°C for 2 minutes and 4°C for 10 minutes.
3. Transfer 45 μ l of fragmented sample to a new tube and use the remaining for fragmentation analysis using RNA 6000 Nano LabChip.
4. Label the fragmented cDNA by preparing the labeling mix as follows:

Component	Volume in 1 reaction
Fragmented cDNA	45 μ l
5 \times TdT Buffer	4.8 μ l
TdT, 30U/ μ l	2 μ l
DNA Labeling Reagent, 5mM	1 μ l
Total Volume	60 μ l

5. Incubate the samples in a thermal cycler as follows: 37°C for 1 hour; 70°C for 10 minutes and 4°C for 10 minutes.
6. Take 4 μ l for Gel-shift analysis (optional, *see* Note 5)

⁵Details on Gel-Shift assay can be found in Appendix A of GeneChip Whole Transcript Double-Stranded Target Assay Manual.

3.6 Hybridization and Scanning

1. Prepare hybridization cocktail as follows:

Component	Volume in 1 reaction
Labeled target	60 μ l
Control Oligonucleotide B2	4.17 μ l
2 \times Hybridization Mix	125 μ l
DMSO	17.5 μ l
RNase-free water	up to 250 μ l
Total Volume	250 μ l

2. Heat the cocktail at 99°C for 5 minutes. Cool to 45°C for 5 minutes and centrifuge at maximum speed for 1 minute.
3. Inject 200 μ l of cocktail into the tiling array chip (*see* Note 6).
4. Place the chips in the hybridization oven at 45°C. Set to rotate at 60 rpm. Incubate for 16 hours.
5. Remove and save the hybridization cocktail after incubation. Refill the chip with 250 μ l of Wash A buffer.
6. Set up the Fluidic Station 450 and apply FS450_0001 protocol. Place 600 μ l of Stain Cocktail 1 in sample holder 1; Place 600 μ l of Stain Cocktail 2 in sample holder 2 and 800 μ l of Array Holding Buffer in sample holder 3 (*see* Note 7).
7. Load the chip to scanner when the fluidic procedure is finished (*see* Note 8).

3.7 Data Analysis

The data analysis workflow in this section is based on Tiling Array Analysis (TAS) software provided by Affymetrix. It provides basic analysis functions including raw data processing to generate normalized intensity values and p-values for the probes, genomic interval computation and visualization for evaluating the quality of array data. The analysis can be further processed in downstream software pipeline, such as IGB and UCSC genome browser. Here we will show how to process and analyze the scanned raw intensity data in CEL format.

In most cases, TAS is capable for analyzing simple sample setup like single transcriptome analysis discussed in this chapter. However, it is important to note that TAS has certain limitations. For example, TAS could be used to compare paired samples only; the output file does not include any genomic annotations. To perform multi-sample comparisons and generate mapped annotation data to the significant intervals requires third party software

⁶Do not discard the remaining cocktail mix. The mix is to be re-used to hybridize chips up to 3 rounds. Store the cocktail at -20°C.

⁷Refer to fluidic station operation manual for operation procedure.

⁸The scanning will take about 30 minutes for each chip. Refer to the GCOS operation manual for details on setting up experiment.

such as Partek Genomic Suite (<http://www.partek.com>) or TileMapper (<http://nichd.nih.gov/TileMapper>).

Define analysis group

1. Create a folder for tiling array analysis (e.g C:\TMA-workflow) and put the CEL intensity data file in this folder (*see* Note 9). Then define the sample set in TAS software through generation of Tiling Analysis Group (TAG) file (*see* Note 10). A total of fourteen TAG files (Chip A to Chip N) will be generated to represent a complete transcriptome.
2. Identify the CEL replicates from the same chip. To associate the probe set intensity to corresponding genomic position, the user needs to supply genomic location (bmap) file.
3. Select “Edit → default” and press “paths” tab in TAS to locate the bmap file. Make sure the file is properly uncompressed.

Data Normalization

1. Normalization is required for replicate data. It helps mitigating signal intensity differences caused by variations in biological samples and target preparation process.
2. This method is applied in our transcriptome data analysis (*see* Note 11).
3. Select “Edit → default” and press “paths” tab in TAS to 1

Intensity Analysis

1. Once data are normalized, the intensity values signal and p-values for each probe are calculated. Calculate the window size by $2 \times \text{bandwidth} + 1$ (*see* Note 12).
2. For transcriptome analysis, we will use a window size of 151bp (*see* Note 13). Select “Edit → default” and press “Probe Analysis” and put 75 in the bandwidth window.
3. Apply one sample analysis to calculate the p-values based on signals from PM and MM probe sets (*see* Note 14). The p-values will be presented in $-10 \log_{10}(\text{p-value})$, which means a p-value of 0.1 will be equal to 10 and a p-value of 0.01 will be

⁹Biological instead of technical replicates are highly recommended. To minimize batch-to-batch variations, all replicates should be processed and analyzed together if possible.

¹⁰Name the TAG file in format like (sample name)-(date)-(Chip number), (e.g. Spga-010309-A1) will enable better tracking and prevent confusion in subsequent data handling on the samples.

¹¹Two normalization methods are available in TAS. They are linear scaling and quantile normalization. In linear scaling, all array samples were scaled to the same value by multiplying signal intensities of each feature in an array by a factor. For quantile normalization, the signal intensity distribution is adjusted equally, so that the median signal intensities are equal across the experiments.

¹²To lower the noise and susceptibility to outliers, the intensity signal and p-values are calculate based on a cluster of oligonucleotide probes (data points), not by individual calculation. In TAS such cluster of data points are known as window, which is determined by genomic distance (bandwidth) from the probe set of interest.

¹³The optimal size of bandwidth depends on the nature of biological samples and experimental application. For transcriptome analysis, we determined the windows based on average exon size of a transcript. Although a larger bandwidth has more statistical power, it could result in diluted signal values. It is important to test against known controls to come up with a bandwidth size that gives the best data representation.

equal to 20. This is done by selecting “Edit → default” in TAS, press “Probe Analysis” and select “One side Upper” under “Test Type” field and “PM/MM” under “intensity” field.

4. To generate the intensity output files consisted of signal intensity and p-values, select “Edit → default” and press “Export” tab in TAS. Under “Results section”, select “Save signal and p-values”, the files will be saved separately. Save the files in .bar file format by selecting “Save to BAR format” in “Signal/p-values Output File” section (*see* Note 15).
5. Proceed to “Scale” in the “Default Properties” window to select scale of the data. Select “Log2” under “Signal Scale” and select “ $-10\log_{10}$ ” under “p-value Scale”, click “OK” to finish.
6. To run intensity analysis, select “Analyze → Intensities” in the menu of TAS. Open the TAG file created in define analysis group step to start analysis.

Assessing Intensity Data

1. To access the quality of transcriptome data, start IGB browser (*see* Note 16). Depending on the scale of analysis, select the memory requirement of IGB. “Pres ???”
2. Select the genome and build version corresponding to the bmap file applied in TAS by accessing the “Data Access” tab in the lower window of IGB, the Refseq gene annotation and associated genome coordinates will be retrieved in the upper window.
3. Choose the chromosome corresponding to the intensity or p-value data file by highlighting the sequence under “Current genome”. The table below shows the chromosome information contained in the mouse tiling 1.1R microarray.
4. To navigate the genome, use the scroll bar on the bottom of the viewer to move in either direction along the genome. Use the arrows on both sides for fine movements. To zoom in or out at a specific position, slide the scroll bar on the top window of IGB.
5. The location of a gene can be retrieved by entering genomic coordinate or gene symbol under the “Pattern Search” tab in the lower window (*see* Note 17).

¹⁴Signal calculation in one sample analysis is based on Hodges-Lehmann estimator defined by the bandwidth size. The median of all pair-wise averages derived from the differences of PM and MM signals in every window are calculated. Signals are expressed in log base 2 scale. P-values are determined by Wilcoxon signed rank test using PM and MM probe value differences within bandwidth area from all replicates. The test is based on the null hypothesis that no signal difference between PM and MM probes.

¹⁵The high-resolution data created by tiling microarray is best visualized in genome browser rather than in spreadsheet (text) format. IGB can be used for viewing whole transcriptome intensity analysis result files.

¹⁶IGB can be started from web or through local installation. To run IGB from web, make sure Java Web Start (www.java.com) is installed. IGB requires java at least version 1.4.2. However, version 1.5 is recommended and any more recent version can also be used. The minimum memory requirement for IGB alone is 256mb of system memory, therefore a system of at least 512mb is required. However, we recommend to run IGB in a system with at least 2gb of system memory. To load the intensity data from all the chips in triplicate setting, start IGB with 1.5gb memory requirement. Use lower memory requirement options will cause slow down or program crash.

¹⁷Other than known gene/position search, IGB also offers pattern search functionality. Regular expression pattern search is very useful for looking for specific sequence features or finding instances of a sequence containing unknowns. A complete list of regular expression syntax can be found at <http://java.sun.com/j2se/1.5.0/docs/api/java/util/regex/Pattern.html>.

6. Open an intensity file from intensity analysis by selecting “File → Open file” in the IGB menu. Open single or multiple (by holding shift key) intensity files and then click “open” to finish. The intensity analysis result will be loaded into IGB and visible above the Refseq annotation track (*see* Note 18).
7. By default, the background of IGB is in black and the annotations are in green. For publication or printing purpose, the background and signal color can be changed. Go to “Graph Adjuster” tab in the lower window to change the desired color.
8. IGB does not filter the intensity data. It has to be manually filtered by value or by percentile. Key in “0%” under “Min” and “95%” under “Max” in the “Graph Adjuster” tab to analyze the top 5% percentile of data*. The selection can be also adjusted dynamically by using the adjacent slide bar (*see* Note 19). An example showing expression profile output for heat shock protein 1 (Hspe1) is illustrated in Figure 2.

Calculate Intervals

1. Define the thresholds for significant transcribed regions in IGB by highlighting the result track from intensity analysis (*see* Note 20).
2. Select “Graph Adjuster” tab in the lower window of IGB and click “Thresholding” to open graph thresholds option.
3. Turn “Visibility” to “On” and select the threshold value at 95% percentile. Put 230 in Max Gap and 60 in Min Run windows. Region that meet the threshold criteria will be visible under the signal track as solid blocks (*see* Note 21).
4. To create an independent track, click “Make Track” in the graph thresholds window.
5. Right-click the track and select “Save Annotations” to export the track data in .bed file format. The .bed files can be viewed in IGB or UCSC genome browser.
6. Alternatively, the intervals can be calculated in TAS, select “Edit → defaults”, open “Interval Analysis” tab and put the corresponding numbers from step 3 in each field. Then go to “Analysis option in the menu and select “Intervals” to open the intensity files to be analyzed, press “Open” to start intervals analysis.

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

¹⁸In addition to Refseq annotation, IGB also provides access to other annotation information, which can be retrieved from NetAffx or DAS/2 server.

¹⁹The percentile cut-off is very subjective to application. A low percentile cutoff will give more data point at the cost of false-positivity. High percentile cutoff give better success rate in down-stream validation at the cost of data coverage.

²⁰Interval analysis is performed based on intensity analysis results (bar or chp files). The goal is to define regions of transcripts in the genome based on the signal or p-values from intensity analysis. These significant transcribed regions known as transcribed fragments (Transfrags) will be revealed as blocks in IGB.

²¹Max_gap is defined as maximum distance between positive positions while Min_run is defined as minimum length of positive region. The values for these two parameters is dependent on type of application and adjustment based on known positive controls.

References

1. Schultz N, Hamra FK, Garbers DL. A multitude of genes expressed solely in meiotic or postmeiotic spermatogenic cells offers a myriad of contraceptive targets. *Proc Natl Acad Sci U S A.* 2003; 100:12201–12206. [PubMed: 14526100]
2. Shima JE, McLean DJ, McCarrey JR, Griswold MD. The murine testicular transcriptome: characterizing gene expression in the testis during the progression of spermatogenesis. *Biol Reprod.* 2004; 71:319–330. [PubMed: 15028632]
3. Johnston DS, Wright WW, Dicandeloro P, Wilson E, Kopf GS, Jelinsky SA. Stage-specific gene expression is a fundamental characteristic of rat spermatogenic cells and Sertoli cells. *Proc Natl Acad Sci U S A.* 2008; 105:8315–8320. [PubMed: 18544648]
4. 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]
5. 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]
6. 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]
7. Wrobel G, Primig M. Mammalian male germ cells are fertile ground for expression profiling of sexual reproduction. *Reproduction.* 2005; 129:1–7. [PubMed: 15615893]
8. Chalmel F, Rolland AD, Niederhauser-Wiederkehr C, Chung SS, Demougin P, Gattiker A, Moore J, Patard JJ, Wolgemuth DJ, Jegou B, Primig M. The conserved transcriptome in human and rodent male gametogenesis. *Proc Natl Acad Sci U S A.* 2007; 104:8346–8351. [PubMed: 17483452]
9. Pang AL, Taylor HC, Johnson W, Alexander S, Chen Y, Su YA, Li X, Ravindranath N, Dym M, Rennert OM, Chan WY. Identification of differentially expressed genes in mouse spermatogenesis. *J Androl.* 2003; 24:899–911. [PubMed: 14581517]
10. Bertone P, Gerstein M, Snyder M. Applications of DNA tiling arrays to experimental genome annotation and regulatory pathway discovery. *Chromosome Res.* 2005; 13:259–274. [PubMed: 15868420]
11. Johnson JM, Edwards S, Shoemaker D, Schadt EE. Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments. *Trends Genet.* 2005; 21:93–102. [PubMed: 15661355]
12. Mockler TC, Chan S, Sundaresan A, Chen H, Jacobsen SE, Ecker JR. Applications of DNA tiling arrays for whole-genome analysis. *Genomics.* 2005; 85:1–15. [PubMed: 15607417]
13. Cowell JK, Hawthorn L. The application of microarray technology to the analysis of the cancer genome. *Curr Mol Med.* 2007; 7:103–120. [PubMed: 17311536]
14. Yazaki J, Gregory BD, Ecker JR. Mapping the genome landscape using tiling array technology. *Curr Opin Plant Biol.* 2007; 10:534–542. [PubMed: 17703988]
15. Yagi S, Hirabayashi K, Sato S, Li W, Takahashi Y, Hirakawa T, Wu G, Hattori N, Ohgane J, Tanaka S, Liu XS, Shiota K. DNA methylation profile of tissue-dependent and differentially methylated regions (T-DMRs) in mouse promoter regions demonstrating tissue-specific gene expression. *Genome Res.* 2008

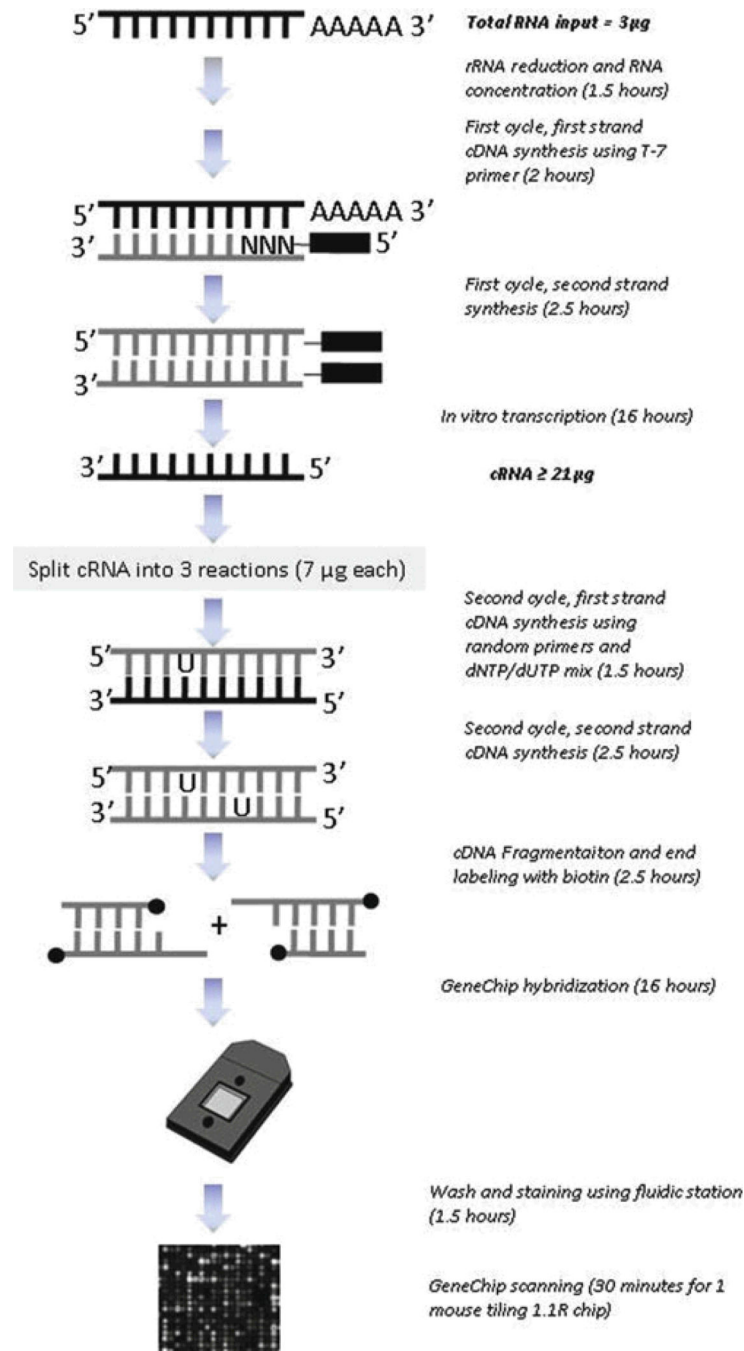


Figure 1.

Overview of tiling microarray protocol for Affymetrix Mouse 1.1R GeneChip. The mRNA population from input total RNA is first enriched by rRNA reduction. Isolated mRNA is then reverse transcribed to cDNA and amplified by *in vitro* transcription into cRNA. The cRNA population will be split into three reactions for another round of cDNA synthesis. The cDNA is subsequently fragmented and labeled, and ready for hybridization with GeneChip. Hybridization mix can be reused for up to three times. Therefore, the protocol is sufficient for nine GeneChip hybridization. The protocol takes three days to complete.

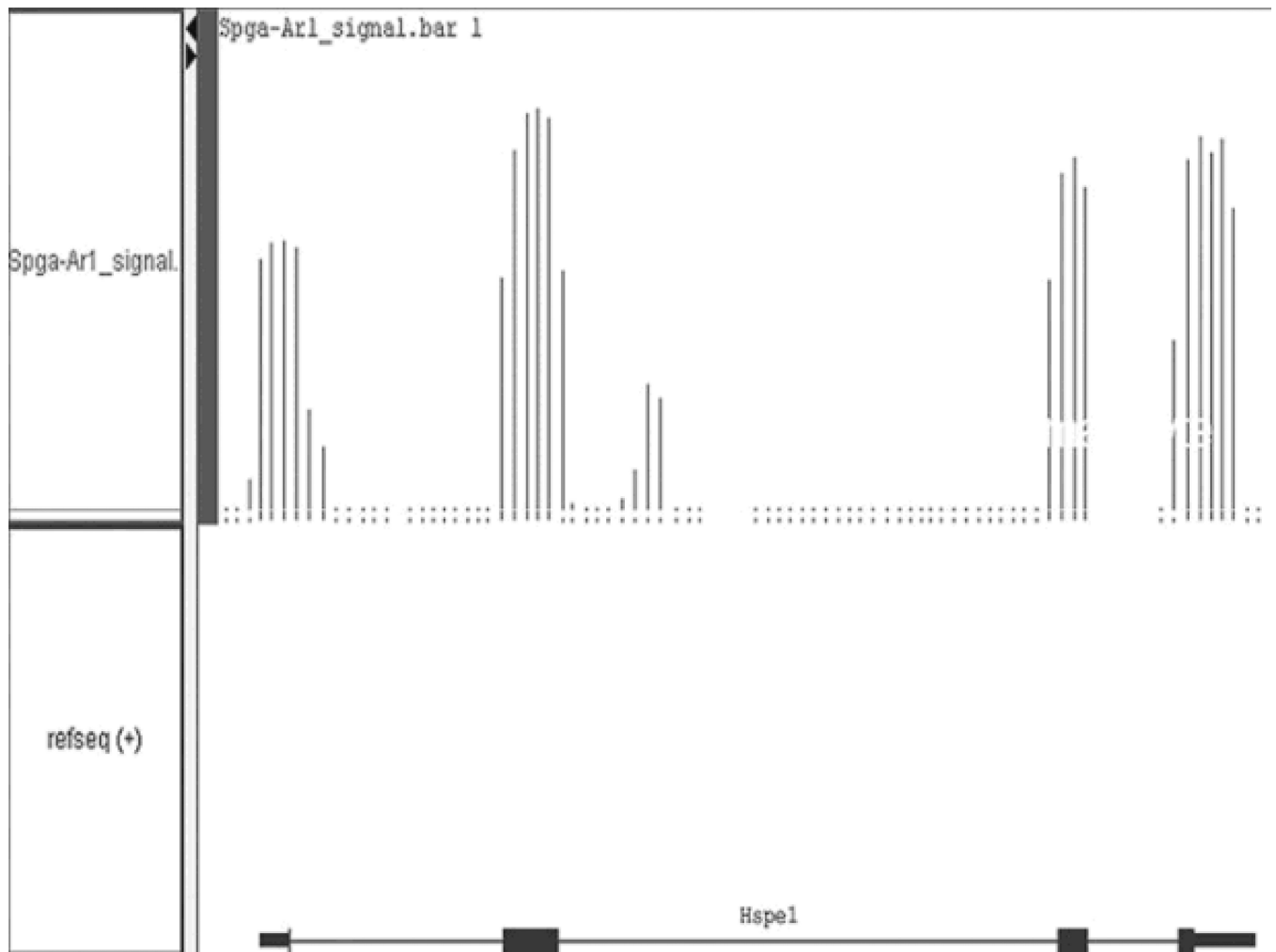


Figure 2. Expression profiling of mouse heat shock protein 1 (*Hspe1*) in mouse spermatogonia. The top panel represents filtered intensity signal from triplicate array set cut off at 95% percentile. The lower panel represents Refseq annotation of *Hspe1* from Genbank. Note that the intensity signal aligns perfectly with the exon annotation of *Hspe1*. The tiling array signals also indicate a genomic region between exon 2 and 3 of *Hspe1* is expressed, which was not annotated by any public database.