Comparative analysis of RNA-seq data from polyA RNAs selection and ribosomal RNAs deletion protocol by strand-specific RNA sequencing technology

Lingjie Fu
Department of Gynaecology and Obstetrics, Shengjing Hospital of China Medical University
Shenyang, China
flina0413@163.com

Meili Chen, Jiayan Wu, Jingfa Xiao, Zhewen Zhang
CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Science
Beijing, China
zhangzw@big.ac.cn

***Abstract*—The conventional non-strand-specific RNA-seq method is widely used for many studies, but it cannot characterize which strand was the transcript originally came from. Strand-specific RNA library construction methods have been developed to overcome this drawback. Here, we compared transcriptomics data from two mainstream RNA enrichment methods (polyA RNAs selection and ribosomal RNAs deletion) by strand-specific RNA sequencing. Using paired-end strategy, we obtained 175 and 149 million high quality reads without ribosomal RNA reads by ribosomal RNAs deletion and poly(A)+ RNAs selection protocol, respectively. From these reads, rmRNA-seq had lower (53.28%) unique mapping rate than the mRNA-seq (73.89%). But, the ribosomal RNAs deletion protocol detected more known non-coding RNAs, particularly lncRNAs, pseudogenes and snoRNAs. Larger proportion (66.7%) of reads mapping to intronic and intergenic regions in ribosomal RNAs deletion method and fewer percentages (33.3%) of reads aligning to exonic regions compared with poly(A)+ RNAs selection method (35.8% and 64.2%). The ribosomal RNAs deletion protocol provides advantages over the poly(A)+ RNAs selection method in sense-antisense pairs detection. In conclusion, the comparison of these two rRNA enrichment methods provides us insight for utility of each protocol. Moreover, we believe that ribosomal RNAs deletion based strand-specific RNA sequencing show us a more comprehensive view of eukaryotic transcriptomes.**

Keywords—rmRNA-seq; mRNA-seq; transcriptomes; strand-specific RNA-seq

I. INTRODUCTION

Due to the rapid development of high throughput next generation sequencing, RNA-seq technology has been widely used in transcriptome analysis [1-5]. With identified all the expressed transcripts, we can understand the transcriptional structure of genes, gain different kinds of RNAs, and quantify expressions of transcripts with different conditions [6]. Conventional non-strand-specific RNA sequencing method cannot tell which strand was the transcript originally came from. Because both strands of cDNA are synthesized by random hexamer primers in library construction process, the calculation of transcripts expression level will occur deviations without knowing the strand information. Recently, this drawback has been overcome by strand-specific RNA library construction protocols [7-12].

There are two mainstream RNA purification methods, one is polyadeylated (poly(A)+) RNAs selection, and the other one is ribosomal RNAs deletion. Numerous researches have been focused on studying the function of poly(A)+ transcripts and enriched RNAs by oligo(dT) selection for transcriptome analysis. However, more and more investigation have indicated that non-polyadeylated (poly(A)-) RNAs also performed important functions [13, 14], and always contain ribosomal RNAs [15], histone mRNAs [16], transfer RNAs and non-coding RNAs (ncRNAs) [17]. If we used ribosomal RNAs deletion method in RNAs extraction for RNA sequencing, it would give us a more comprehensive view of eukaryotic transcriptomes. To address this proposition, we compared transcriptomics data from these two mainstream RNA purification methods by means of strand specific RNA-seq technology.

This work is supported by a grant from the National Science Foundation of China (No. 31271386 and No. 31101063).

978-1-4799-7294-4/14/$31.00 ©2014 IEEE

II. MATERIALS AND METHODS

A. Sample and total RNA isolation

Ovarian cancer sample was collected from Shengjing hospital of china medical university patient at the time of surgery. Total RNA was extracted by means of Trizol method.

selection method and ribosomal RNAs deletion method, respectively. For poly(A)+ RNAs selection, the Illumina TruSeq™ RNA sample preparation kit was used according to the manufacturer's instructions. For ribosomal RNAs deletion, the Life Technologies's RiboMinus™ Eukaryote kit was used following the manufacturer's instructions.

C. Libraries construction and sequencing

For either TruSeqTM or RiboMinusTM based mRNA enrichment, we constructed two strand-specific cDNA libraries with NEXTflexTM Directional RNA-Seq Kit (dUTP Based) according to the manufacturer's instructions. We sequenced the libraries with Illumina HiSeq2000 for 2×101bp pair-end.

D. Reads filtering and alignment

Raw reads were filtered before mapping to the reference genome. Adaptor reads, reads with more than 2% 'N' bases, and reads with low quality (below 15) over half of their lengths were filtered out, respectively. All filtered reads were mapped to 5S, 5.8S, 12S, 16S, 18S and 28S ribosomal RNA sequences by Bowtie[18] (version 0.12.7) with default parameters. The mapped reads were discarded. All the left reads were high quality and mapped to the reference human genome using Tophat[19] (version 2.0.9) with –G option offering the Ensembl GRCh37 release of human gene annotation and other default values. The unmapped reads were remapped with the same methodology by trimming the last 20bp of 101bp reads. We merged these two groups of mapping results using SAMtools [20] (version 0.1.17), and all unique mapped reads were extracted for the following analysis.

E. Libraries assessment

Several criteria are created for assessing RNA-seq libraries， including the percentage of exonic, intronic or intergenic reads, coverage at 5' and 3' ends, evenness of transcript coverage [21]. Here, we used RNA-SeQC [22] (version 1.1.7) to get these metrics. The unique mapped reads were sorted by reference position and marked with duplicated records by picard-tools (version 1.90) before running at RNA-SeQC.

F. Transcripts Identification

We assembled unique mapped reads by Cufflinks [23] (version 2.1.1). The –g option were used for supplying reference annotation to guide RABT assembly [24]. Accoding to Ensembl GRCh37 release of human gene annotation and NONCODE database[25], the known non-coding transcripts The total RNA integrity number was greater than 8, which was analyzed by an Agilent 2100 Bioanalyzer.

B. mRNA Enrichment

Starting from 7ug total RNA of a single human ovarian cancer sample, we purified mRNA with poly(A)+ RNAs

were divided into several groups such as long non-coding RNA (lncRNA), microRNA (miRNA), miscellaneous RNA (miscRNA), small nucleolar RNAs (snoRNA), small nuclear RNA (snRNA), pseudogene, and so on. We also identified the sense and antisense transcripts snoRNA or miRNA. Transcripts without certain transcriptional orientation were excluded. Then we identified sense and antisense transcripts coming from opposite strand and overlapped more than 25nt from the same gene locus.

G. Detection of alternative splicing

We used Tophat to detect alternative splicing (AS) [19] and excluded splice junctions mapped less than two reads. The mapping result were also used to detected five basic AS events, including skipped exon (SE), retained intron (RA), alternative 5' splice site (A5SS), alternative 3' splice site (A3SS) and mutually exclusive exon (MXE) [26] by ASTALAVISTA[27].

III. RESULTS

A. Sequencing and mapping summary of two RNA-seq libraries

To compare ribosomal RNAs deletion based RNA-seq (rmRNA-seq) with poly(A)+ RNAs selection based RNA-seq (mRNA-seq), we constructed two strand specific libraries. Using paired-end strategy, we generated 329.5 million high quality reads by ribosomal RNAs deletion and 149.9 million high quality reads by poly(A)+ RNAs selection. After filtering ribosomal RNA reads, 175 and 149 millions reads were obtained, respectively. The total clean mRNA-seq reads are much higher than the other method. From these reads, 99 million (56.28%) and 126 million (85.16%) reads were mapped to human genome, which referred to rmRNA-seq and mRNA-seq total reads, severally. As to rmRNA-seq, 93 million (53.28%) reads were unique mapped, while 110 million (73.89%) reads were mapped to unique loci in mRNA-seq total reads. In rmRNA-seq unique mapped reads, 33.3% of them were mapped to exonic region, 47.2% to intronic region and 19.5% to intergenic region. Meanwhile, 64.2%, 31% and 4.8% of mRNA-seq unique mapped reads were mapped to exonic, intronic and intergenic regions, respectively (Table.1). It is obvious that poly(A)+ RNAs selection method has greater proportions of reads aligning to exonic region, while reads mapping to intronic and intergenic regions in ribosomal RNAs deletion method are much more.

978-1-4799-7294-4/14/$31.00 ©2014 IEEE

B. *Evenness of transcript coverage*

Evenness of transcript coverage is important for polymorphism detection and transcriptome annotation [28]. To estimate this, we computed the mean coverage for 1000 most highly expressed, 1000 mid-range expressed and bottom 1000 expressed transcripts from 5' to 3' end, respectively, with the lengths of transcripts normalized to 1-100 (Fig. 1a, 1b and 1c).

Table 1 Summary of sequencing and mapping for two RNA-seq libraries.

	rmRNA-seq	mRNA-seq
Total reads	329,478,592	149,921,508
rRNA reads (%)	49.08	6.17
Total clean reads	175,931,452	149,009,166
Mapped reads (%)	56.28	85.16
Unique mapped reads (%)	53.28	73.59
Exonic Rate (%)	33.30	64.20
Intronic Rate (%)	47.20	31.00
Intergenic Rate (%)	19.50	4.80

For rmRNA-seq library, the coverage between 5' and 3' end are very close (83.9% at 5' end and 86.1% at 3' end). On the other hand, we got a significantly increased coverage at 3' end from the mRNA-seq data. This result is consistent with the analysis of transcript coverage evenness as just mentioned above.

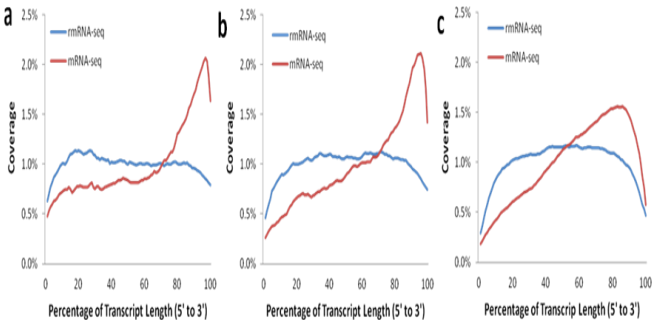

Figure 1 Evenness of transcript coverage. Average relative coverage for 1000 most highly expressed transcripts (a), for 1000 mid-range expressed transcripts (b) and for bottom 1000 expressed transcripts (c) in each library.

D. *Landscape of transcripts*

Using the software Cufflinks, we assembled rmRNA-seq data and mRNA-seq data into 16,897 and 17,247 known protein coding genes, respectively. Of all these genes, 16,362 were expressed in both rmRNA-seq library and mRNA-seq library. Each library also has its own special protein coding genes, 534 for rmRNA-seq and 884 for mRNA-seq (Fig.3A). According to Ensembl GRCh37 release of human gene The coverage of rmRNA-seq data is very even. And there is a significant bias at 3' end in mRNA-seq data.

C. *Coverage at 5' and 3' ends*

To identify full-length transcripts correctly [21], we calculated the ratio of reads covered genes on 5' and 3' ends in each library to estimate the coverage at two ends (Fig. 2).

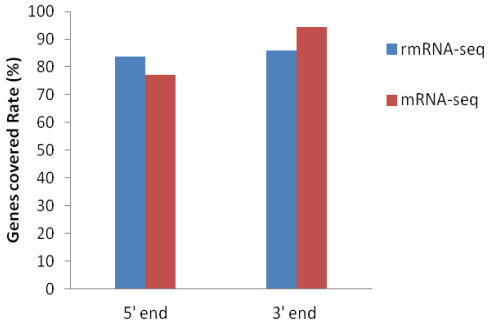

Figure 2 Ratio of reads covered genes on 5' and 3' ends in each library.

annotation and NONCODE database, we identified 27,513 known non-coding transcripts in rmRNA-seq library, which were composed of 23,441 lncRNA, 3,639 pseudogene, 270 misc_RNA, 115 snoRNA, 33 snRNA and 15 miRNA (Table2). In mRNA-seq library, we obtained 26,525 known non-coding transcripts and also divided them into 6 categories (Table2). Seen from Table 2, there are more lncRNAs, pseudogenes and snoRNAs detected in rmRNA-seq library. Comparing non-coding transcripts in two libraries, 18,570 transcripts were both detected, 8,943 transcripts were uniquely found in rmRNA-seq data and 7,955 transcripts were only obtained in mRNA-seq data (Fig 3B). In rmRNA-seq, a total of 59,172 transcripts were identified. Of these, 8,641 putatitive novel transcripts (14.6%) were found. The number of transcripts identified in mRNA-seq came to 57,162 and 5,649 trancripts (9.9%) of these were putatitive novel. To further investigate the association between rmRNA-seq and mRNA-seq, we analyzed the correlation coefficient R among all expressed genes by the Spearman method[29]. FPKM, Fragments Per Kilobase per Million mapped reads [30], was used here to detect the expression level of protein coding genes and non-coding RNAs in each library. Figure 4 showed the scatter plots with expression values, which are normalized to log2 scaled tag counts by R script. As

Table 2 Classification of known non-coding transcripts in rmRNA-seq and mRNA-seq data.

Classification	rmRNA-seq	mRNA-seq
lncRNA	23441 (7323)*	22516 (6398)
pseudogene	3639 (1399)	3480 (1240)
misc_RNA	270 (107)	367 (204)
snoRNA	115 (82)	76 (43)
snRNA	33 (21)	62 (50)
miRNA	15 (11)	24 (20)

*the numbers in brackets refer to specific transcripts in each classification.

978-1-4799-7294-4/14/$31.00 ©2014 IEEE

we anticipated, the correlation between rmRNA-seq and mRNA-seq protein coding genes was moderately high (R=0.71) (Fig 4a). While there was nearly no correlation (R=0.09) between rmRNAseq and mRNA-seq non-coding transcripts (Fig 4b). The correlation of log2 FPKM of our two libraries and data from other lab (SRA: srr926256, Homo sapiens, ES2 cell line, Ovarian cancer) was also calculated. The spearman correlation R is 0.698 (Fig 4c) and 0.591 (Fig 4d), respectively. As these two libraries were constructed by a strand-specific RNA-seq approach, it is convenience to identify the polarity of transcripts. We obtained 6,473 and 6,188 sense-antisense pairs in rmRNA-seq data and mRNA-seq data, respectively, 4,876 of those sense-antisense pairs were shared in these two libraries,

no significant difference were observed between rmRNA-seq and mRNA-seq (Fig 5).

IV. Discussion

In this investigation, we evaluated ribosomal RNAs deletion based RNA-seq and poly(A)+ RNAs selection based RNA-seq with a single human ovarian cancer sample by some

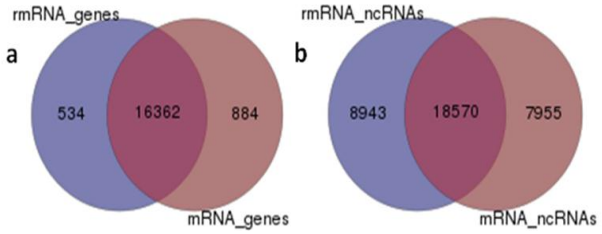

Figure 3 Comparison of annotated genes and non-coding transcripts in each library

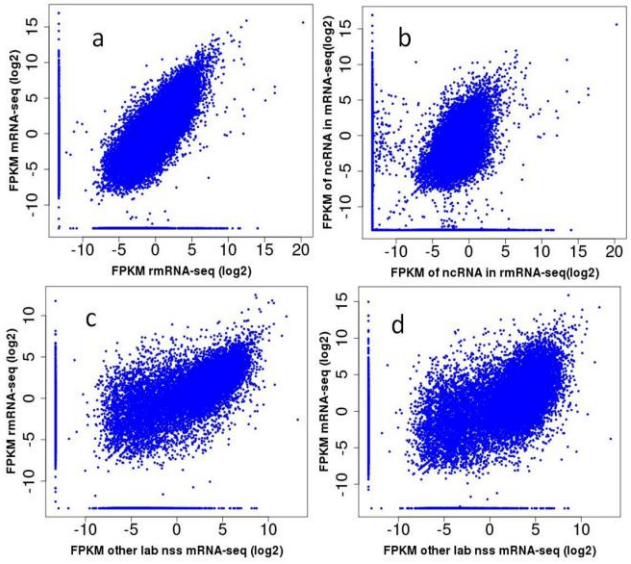

Figure 4 Scatter plots comparing (a) protein coding genes and non-coding transcripts and (b) non-coding transcripts between rmRNA-seq data and mRNA-seq data. Scatter plots comparing protein coding genes in our rmRNA-seq data (c) or our mRNA-seq data (d) with other lab non-strand-specific mRNA-seq data.

1,597 pairs were unique in rmRNA-seq library, and 1,312 pairs were only in mRNA-seq library. Obviously, we had more sense-antisense pairs in rmRNA-seq data.

E. Landscape of alternative splicing

In rmRNA-seq library, 11,440,782 reads were mapped onto 175,817 splice junctions, which belong to 28,257 genes (contains both protein coding and non-coding transcripts). In mRNA-seq library, we identified 152,018 splice junctions for 26,283 genes and 41,416,788 reads were mapped onto these junctions. Comparing with the number of five basic AS events,

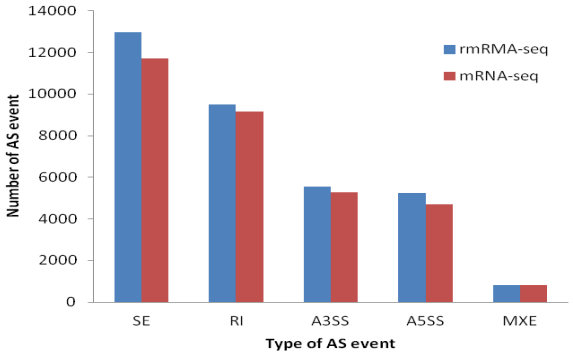

Figure 5 Statistics of alternative splicing events between rmRNA-seq data and mRNA-seq data.

metrics. Ribosomal RNAs deletion method had lower alignment rate than poly(A)+ RNAs selection method. The reasons for this situation are manifold and complex, such as sequencing erroneous, low quality reads contained, and so on, which was also consistent with some previous reports [31, 32]. According to previous studies [33-36], there are only ~20,000 protein-coding genes in human genome, representing <2% of the total genomic sequence (International Human Genome Sequencing Consortium 2004). The major part of the genome is made up of non-coding RNAs [37], which has mostly not been understood their functions yet. We suspected that may cause the difficulty of the reads mapping back. We also found that in ribosomal RNAs deletion method larger proportion of reads mapping to intronic and intergenic regions and fewer percentages of reads aligning to exonic regions compared with poly(A)+ RNAs selection method. For ribosomal RNAs deletion method, we suggest that the greater ratio of reads mapping to intronic and intergenic regions may arise from numerous non-coding RNAs [31, 38, 39]. As to poly(A)+ RNAs selection method, more reads aligning to exonic regions may make the assembly of transcripts simpler [28].

Although enriched mRNAs by oligo(dT) affinity in poly(A)+ RNAs selection procedure, mRNA-seq data still resulted in a obvious coverage bias at 3' end. We presumed that attribute to truncation or degradation of poly(A)+ transcripts 5' end during sample preparation, and the truncated

978-1-4799-7294-4/14/$31.00 ©2014 IEEE

and degraded 5' ends of poly(A)+ transcripts would have not been selected. As a result, missing partial 5' ends of poly(A)+ may influence the assembly of complete transcripts and the accuracy of gene expression profiling. On the other hand, the rmRNA-seq data represented nearly uniform coverage from 5' ends to 3' ends of all transcripts.

We observed that rmRNA-seq had an advantage to detect more known non-coding RNAs, particularly lncRNAs, pseudogenes and snoRNAs. Numerous functional long transcripts are known to lack poly(A) tails. For instance, ribosomal RNAs generated by RNA polymerase Ⅰ and Ⅲ, certain small RNAs generated by RNA polymerase Ⅲ, and replication-dependent histone mRNAs and some lncRNAs synthesized by RNA polymerase Ⅱ [13]. However, mRNAs enriched by oligo(dT) selection approach will miss certain transcripts, which do not have a poly(A) tail. This limitation can be overcome by ribosomal RNAs deletion method. The advantage of ribosomal RNAs deletion method was also reflected in detecting sense antisense transcript pairs. It has been reported that the most prominent form of sense antisense pairs is a non-coding transcript partner of a coding transcript in mammalian genome [40-42]. More non-coding RNAs were detected by ribosomal RNAs deletion method than oligo(dT) selection method, so more sense antisense pairs were found in rmRNA-seq data.

In conclusion, if we just focus on poly(A)+ transcripts, the poly(A)+ RNAs selection based RNA-seq is quite good enough and will require a lower sequencing cost. However, if we want to capture both poly(A)+ and poly(A)- transcripts, ribosomal RNAs deletion based RNA-seq will show us a more comprehensive view of eukaryotic transcriptomes.

ACKNOWLEDGMENT

The authors thank core genomic facility in Beijing Institute of Genomics, Chinese Academy of Science.

REFERENCES

[1] Y. Wu, X. Wang, F. Wu, R. Huang, F. Xue, G. Liang, et al., "Transcriptome profiling of the cancer, adjacent non-tumor and distant normal tissues from a colorectal cancer patient by deep sequencing," PLoS One, vol. 7, p. e41001, 2012.

[2] K. O. Mutz, A. Heilkenbrinker, M. Lonne, J. G. Walter, and F. Stahl, "Transcriptome analysis using next-generation sequencing," Current Opinion in Biotechnology, vol. 24, pp. 22-30, Feb 2013.

[3] J. J. Xu, Y. Y. Li, X. L. Ma, J. F. Ding, K. Wang, S. S. Wang, et al., "Whole transcriptome analysis using next-generation sequencing of model species Setaria viridis to support C-4 photosynthesis research," Plant Molecular Biology, vol. 83, pp. 77-87, Sep 2013.

[4] L. Li, J. Liu, W. Yu, X. Lou, B. Huang, and B. Lin, "Deep transcriptome profiling of ovarian cancer cells using next-generation sequencing approach," Methods Mol Biol, vol. 1049, pp. 139-69, 2013.

[5] S. Ren, Z. Peng, J. H. Mao, Y. Yu, C. Yin, X. Gao, et al., "RNA-seq analysis of prostate cancer in the Chinese population identifies recurrent gene fusions, cancer-associated long noncoding RNAs and aberrant alternative splicings," Cell Res, vol. 22, pp. 806-21, May 2012.

[6] Z. Wang, M. Gerstein, and M. Snyder, "RNA-Seq: a revolutionary tool for transcriptomics," Nat Rev Genet, vol. 10, pp. 57-63, Jan 2009.

[7] D. Parkhomchuk, T. Borodina, V. Amstislavskiy, M. Banaru, L. Hallen, S. Krobitsch, et al., "Transcriptome analysis by strand-specific sequencing of complementary DNA," Nucleic Acids Research, vol. 37, Oct 2009.

[8] Y. He, B. Vogelstein, V. E. Velculescu, N. Papadopoulos, and K. W. Kinzler, "The antisense transcriptomes of human cells," Science, vol. 322, pp. 1855-7, Dec 19 2008.

[9] L. Mamanova, R. M. Andrews, K. D. James, E. M. Sheridan, P. D. Ellis, C. F. Langford, et al., "FRT-seq: amplification-free, strand-specific transcriptome sequencing," Nat Methods, vol. 7, pp. 130-2, Feb 2010.

[10] R. Lister, R. C. O'Malley, J. Tonti-Filippini, B. D. Gregory, C. C. Berry, A. H. Millar, et al., "Highly integrated single-base resolution maps of the epigenome in Arabidopsis," Cell, vol. 133, pp. 523-36, May 2 2008.

[11] Y. Y. Zhu, E. M. Machleder, A. Chenchik, R. Li, and P. D. Siebert, "Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction," Biotechniques, vol. 30, pp. 892-7, Apr 2001.

[12] C. D. Armour, J. C. Castle, R. H. Chen, T. Babak, P. Loerch, S. Jackson, et al., "Digital transcriptome profiling using selective hexamer priming for cDNA synthesis," Nature Methods, vol. 6, pp. 647-U35, Sep 2009.

[13] L. Yang, M. O. Duff, B. R. Graveley, G. G. Carmichael, and L. L. Chen, "Genomewide characterization of non-polyadenylated RNAs," Genome Biol, vol. 12, p. R16, 2011.

[14] I. Livyatan, A. Harikumar, M. Nissim-Rafinia, R. Duttagupta, T. R. Gingeras, and E. Meshorer, "Non-polyadenylated transcription in embryonic stem cells reveals novel non-coding RNA related to pluripotency and differentiation," Nucleic Acids Research, vol. 41, pp. 6300-15, Jul 2013.

[15] I. Grummt, "Regulation of mammalian ribosomal gene transcription by RNA polymerase I," Prog Nucleic Acid Res Mol Biol, vol. 62, pp. 109-54, 1999.

[16] S. Detke, J. L. Stein, and G. S. Stein, "Synthesis of Histone Messenger-Rnas by Rna Polymerase-Ii in Nuclei from S-Phase Hela S3 Cells," Nucleic Acids Research, vol. 5, pp. 1515-1528, 1978.

[17] H. Sunwoo, M. E. Dinger, J. E. Wilusz, P. P. Amaral, J. S. Mattick, and D. L. Spector, "MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles," Genome Res, vol. 19, pp. 347-59, Mar 2009.

[18] B. Langmead, C. Trapnell, M. Pop, and S. L. Salzberg, "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome," Genome Biol, vol. 10, p. R25, 2009.

[19] D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, and S. L. Salzberg, "TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions," Genome Biol, vol. 14, p. R36, Apr 25 2013.

[20] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, et al., "The Sequence Alignment/Map format and SAMtools," Bioinformatics, vol. 25, pp. 2078-9, Aug 15 2009.

[21] J. Z. Levin, M. Yassour, X. Adiconis, C. Nusbaum, D. A. Thompson, N. Friedman, et al., "Comprehensive comparative analysis of strand-specific RNA sequencing methods," Nat Methods, vol. 7, pp. 709-15, Sep 2010.

[22] D. S. DeLuca, J. Z. Levin, A. Sivachenko, T. Fennell, M. D. Nazaire, C. Williams, et al., "RNA-SeQC: RNA-seq metrics for quality control and process optimization," Bioinformatics, vol. 28, pp. 1530-2, Jun 1 2012.

978-1-4799-7294-4/14/$31.00 ©2014 IEEE

[23] C. Trapnell, D. G. Hendrickson, M. Sauvageau, L. Goff, J. L. Rinn, and L. Pachter, "Differential analysis of gene regulation at transcript resolution with RNA-seq," Nat Biotechnol, vol. 31, pp. 46-53, Jan 2013.

[24] A. Roberts, H. Pimentel, C. Trapnell, and L. Pachter, "Identification of novel transcripts in annotated genomes using RNA-Seq," Bioinformatics, vol. 27, pp. 2325-2329, Sep 1 2011.

[25] D. Bu, K. Yu, S. Sun, C. Xie, G. Skogerbo, R. Miao, et al., "NONCODE v3.0: integrative annotation of long noncoding RNAs," Nucleic Acids Research, vol. 40, pp. D210-5, Jan 2012.

[26] E. T. Wang, R. Sandberg, S. Luo, I. Khrebtukova, L. Zhang, C. Mayr, et al., "Alternative isoform regulation in human tissue transcriptomes," Nature, vol. 456, pp. 470-6, Nov 27 2008.

[27] S. Foissac and M. Sammeth, "ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets," Nucleic Acids Research, vol. 35, pp. W297-9, Jul 2007.

[28] X. Adiconis, D. Borges-Rivera, R. Satija, D. S. DeLuca, M. A. Busby, A. M. Berlin, et al., "Comparative analysis of RNA sequencing methods for degraded or low-input samples," Nat Methods, vol. 10, pp. 623-9, Jul 2013.

[29] C. Spearman, "The proof and measurement of association between two things," American Journal of Psychology, vol. 15, pp. 72-101, 1904.

[30] C. Trapnell, B. A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M. J. van Baren, et al., "Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation," Nat Biotechnol, vol. 28, pp. 511-5, May 2010.

[31] P. Cui, Q. A. Lin, F. Ding, C. Q. Xin, W. Gong, L. F. Zhang, et al., "A comparison between ribo-minus RNA-sequencing and polyA-selected RNA-sequencing," Genomics, vol. 96, pp. 259-265, Nov 2010.

[32] M. A. Tariq, H. J. Kim, O. Jejelowo, and N. Pourmand, "Whole-transcriptome RNAseq analysis from minute amount of total RNA," Nucleic Acids Research, vol. 39, p. e120, Oct 2011.

[33] C. P. Ponting, P. L. Oliver, and W. Reik, "Evolution and functions of long noncoding RNAs," Cell, vol. 136, pp. 629-41, Feb 20 2009.

[34] J. E. Wilusz, H. Sunwoo, and D. L. Spector, "Long noncoding RNAs: functional surprises from the RNA world," Genes Dev, vol. 23, pp. 1494-504, Jul 1 2009.

[35] S. Djebali, C. A. Davis, A. Merkel, A. Dobin, T. Lassmann, A. Mortazavi, et al., "Landscape of transcription in human cells," Nature, vol. 489, pp. 101-8, Sep 6 2012.

[36] J. Harrow, A. Frankish, J. M. Gonzalez, E. Tapanari, M. Diekhans, F. Kokocinski, et al., "GENCODE: the reference human genome annotation for The ENCODE Project," Genome Res, vol. 22, pp. 1760-74, Sep 2012.

[37] J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, et al., "The sequence of the human genome," Science, vol. 291, pp. 1304-51, Feb 16 2001.

[38] J. Cheng, P. Kapranov, J. Drenkow, S. Dike, S. Brubaker, S. Patel, et al., "Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution," Science, vol. 308, pp. 1149-1154, May 20 2005.

[39] D. Kampa, J. Cheng, P. Kapranov, M. Yamanaka, S. Brubaker, S. Cawley, et al., "Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22," Genome Research, vol. 14, pp. 331-342, Mar 2004.

[40] T. Lu, C. Zhu, G. Lu, Y. Guo, Y. Zhou, Z. Zhang, et al., "Strand-specific RNA-seq reveals widespread occurrence of novel cis-natural antisense transcripts in rice," Bmc Genomics, vol. 13, p. 721, 2012.

[41] H. Kiyosawa, I. Yamanaka, N. Osato, S. Kondo, and Y. Hayashizaki, "Antisense transcripts with FANTOM2 clone set and their implications for gene regulation," Genome Res, vol. 13, pp. 1324-34, Jun 2003.

[42] S. Katayama, Y. Tomaru, T. Kasukawa, K. Waki, M. Nakanishi, M. Nakamura, et al., "Antisense transcription in the mammalian transcriptome," Science, vol. 309, pp. 1564-6, Sep 2 2005.

978-1-4799-7294-4/14/$31.00 ©2014 IEEE