Article

Chinese Science Bulletin

Bioinformatics

August 2012 Vol.57 No.22: 2920–2928
doi: 10.1007/s11434-012-5077-3

A comparative analysis of tissue gene expression data from high-throughput studies

PING Jie[1], WANG YaJun[1], YU Yao[2], LI YiXue[1,2], LI Xuan[2*] & HAO Pei[2*]

[1] *College of Life Science and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China;*
[2] *Key Laboratory of Systems Biology/Key Laboratory of Synthetic Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China*

Received November 10, 2011; accepted January 16, 2012; published online May 30, 2012

High-throughput technologies were employed over the past decade to study the expression profiles of cells and tissues. There are large collections of accumulated data from public databases and numerous research articles were published on these data. In the current study, we performed meta-analysis on the gene expression data from human liver and kidney tissues produced from five different technologies: EST, SAGE, MPSS, microarray, and RNA-Seq. We found RNA-Seq was the most sensitive in the number of genes it detected while SAGE and MPSS were the least sensitive. For the genes detected by all the platforms, there were generally good correlations to the measured expression levels of corresponding genes. We further compared detected genes to liver/kidney proteomics data from the Human Protein Atlas, and found 960 of the 8764 genes only detected by RNA-Seq were validated by proteomics results. In conclusion, RNA-Seq is a more sensitive and consistent technology compared to the other four high-throughput platforms, though their results are in general agreement. Average coverage was determined to be the preferred measurement to represent gene expression levels by RNA-Seq data and will be used in future works.

high-throughput sequencing, tissue transcriptome, comparative analysis

Citation: Ping J, Wang Y J, Yu Y, et al. A comparative analysis of tissue gene expression data from high-throughput studies. Chin Sci Bull, 2012, 57: 2920–2928, doi: 10.1007/s11434-012-5077-3

The transcriptome is the collection of all RNA molecules in one or a group of cells. Studying the transcriptome is vital to understanding the functions of unknown genes, to revealing the regulating mechanism of specific genes, and also to researching the diagnosis and treatment of some diseases [1–3]. In the last few years, with the development of microarray and sequencing technologies, transcriptome research has become faster and more accurate.

Several technologies have been developed to study the transcriptome, namely EST, SAGE [4], MPSS [5], microarray, and RNA-seq. ESTs are short sequences of cDNA fragments (usually 200–800 nucleotides long), which have been used in gene discovery [6] and gene sequence determination [7]. SAGE allows fast and detailed analysis of thousands of transcripts. As it does not require a known template clone, SAGE can be used to recognize and quantitate new genes, and it has been successfully used to describe the transcriptome of various diseases and most organisms [8,9]. MPSS is a gene expression quantification approach and has high sensitivity and absolute gene expression although it is confined to restriction enzyme recognition site [10]. Microarray-based gene expression profiling can be used to measure changes in gene expression levels and to genotype or re-sequence mutant genes, etc [11].

RNA-Seq [12], uses next-generation sequencing technologies to sequence all cDNAs from a sample. Because of its high throughput, RNA-Seq can provide more comprehensive information on the transcriptome than the four methods above [1]. RNA-Seq can be used to detect the overall transcriptional activities of many species at the level of single nucleotides [13], discover unknown or rare transcripts [14,15], and recognize alternative splicing sites [16]

*Corresponding authors (email: lixuan@sippe.ac.cn; phao@sibs.ac.cn)

or cSNP [17] accurately. Furthermore, RNA-Seq can be performed with different platforms to deal with different problems, as they have different advantages. For example, using the Illumina Genome Analyzer platform, recent applications include sequencing mammalian transcriptomes [18], ABI Solid Sequencing to profile stem cell transcriptomes [13] or Life Science's 454 Sequencing to discover SNPs [19]. Even though each platform has its technical individualities, the information gathered from each is of the same nature.

Early technologies were limited by the size of reads, such as ESTs, or the cost, like MPSS and SAGE, or the throughput, as microarray. However, these limiting features are precisely the advantage of RNA-Seq. In this work, we compared these five technologies in a study of transcriptomes from two tissues, liver and kidney, to assess the sensitivity and reproducibility. Comparative analysis was performed with the historic research data to investigate how data from these five technologies are correlated.

1 Methods

1.1 RNA-Seq

The raw RNA-Seq data were obtained from the Sequence Read Archive (SRA) of NCBI with accession number SRA000299. Data for each tissue contains three datasets, SRR002321, SRR002322 and SRR002323 for liver, SRR002320, SRR002324 and SRR002325 for kidney. The original tissue samples were from the liver and kidney tissues of one normal human male dead less than 6 h. All cDNA samples were sequenced by the Illumina Solexa platform [20].

The length of each read is 36 bp. The reads were mapped to the UniGene data, in which alternatively spliced transcripts were removed, by using the alignment software MAQ [21]. Two or less mismatches were allowed in the mapping process. The genes that mapped to at least one read were considered as potentially expressed. After mapping to the reference data, the amount of coverage at each position of a gene can be calculated through the alignment result.

In this work, we used two standards, average coverage and max depth for the measurement of gene expression. The average coverage indicates the normalized coverage based on gene length, and the max depth shows the largest depth position of one gene.

1.2 Microarray

Microarray data on liver tissue were from the Chinese Human Liver Proteome Project (CNHLPP) [22]. The raw RNA samples from ten liver tissues were hybridized to HG-U133plus 2.0 high-density oligonucleotide arrays in two technical replicates. Then, using GENECHIP 3.2 to perform the primary image analysis of the arrays, we collected three series of human liver-related gene expression data (platform: GPL570; Series: GSE11045, GSE7117, GSE7741). Microarray data on kidney tissue was acquired from the GEO database of NCBI (platform: GPL570; Series: GSE11045, GSE11151, GSE12606).

The data from livers were then filtered according to the Present (P) versus Absent (A) call percentage using MAS 5.0 algorithm [23] and the statistical software R (http://www.r-project.org/). The genes that were mapped to at least one present probe set were considered as expressed. The probe sets from kidneys were mapped to Ensembl genes. If only a single probe set was mapped to a gene then the corresponding intensities were used in all future analyses. If multiple probe sets mapped to the same gene then the probe set that was most often called as present was considered. If two or more probe sets were called present in the same number of hybridizations then a probe set at random was chosen and used in all further analyses.

1.3 Other datasets

The EST data were all acquired from dbEST. The SAGE data were a subset of Human SAGE Genie data from the CGAP website (http://cgap.nci.nih.gov/). The MPSS data were produced from experiments performed by TaKaRa Co., Japan using CNHLPP samples, and the data can be downloaded from the website (http://202.127.18.238/hepatocytes/) [24]. The expression data of proteins in normal human tissues were obtained from the Human Protein Atlas database [25].

In our work, we determined the overriding significance of RNA-Seq by comparing results from different methods with sampling errors. RNA-Seq datasets have biological repeats, some microarray data have technical repeats, and others are from public databases. The details of the data used in this work can be found in Supporting Information S5.

1.4 Correlation coefficient

The strength of the linear association between two variables is quantified by the correlation coefficient. Given a set of observations (x_1, y_1), (x_2, y_2), …, (x_n, y_n), the formula for computing the correlation coefficient is given by

$$r = \frac{1}{n-1}\sum\left(\frac{x-\bar{x}}{s_x}\right)\left(\frac{y-\bar{y}}{s_y}\right). \quad (1)$$

2 Results

2.1 Study design and data collections: EST, SAGE, MPSS, microarray, and RNA-Seq

In this study, we compare the gene expression data of liver

and kidney tissues from five different platforms, RNA-Seq, microarray, EST, SAGE, and MPSS. All the original data were acquired from public databases. The raw data of RNA-Seq consist of 88055898 reads from liver with 18301710 reads mapped to the reference, and 83696940 reads from kidney with 17968282 reads mapped to the reference (Table 1). In the MPSS, SAGE or EST dataset, the gene which contains at least one scanned sequence was considered as expressed while in the microarray dataset, the gene should contain at least one present probe. In total, we found 24272 genes expressed in liver tissue and 24694 genes expressed in kidney tissue. In the 24272 candidate genes of liver, 21210 genes came from RNA-Seq data, 12821 genes from microarray data, 12152 genes from EST data, 6413 genes from MPSS data, and 6780 genes from SAGE data. In the 24694 candidate genes of kidney, 21759 genes were derived from RNA-Seq data, 12900 genes from microarray data, 14093 genes from EST data, 5821 genes from MPSS data, and 6621 genes from SAGE data (Figure 1 and Table 1).

There are in total 3047 unique genes detected from all five datasets from liver and 2517 unique genes from kidney. Interestingly, over 5000 genes were detected as expressed from the RNA-Seq dataset and, in comparison, less than 300 genes were detected from the other datasets (Figure 2).

Table 1 The number of genes of two tissues, liver and kidney, detected by five high-throughput transcriptome technologies

	Liver	Kidney
RNA-Seq	21210	21759
Microarray	12821	12900
EST	12152	14093
MPSS	6413	5821
SAGE	6780	6621
Total	24272	24694

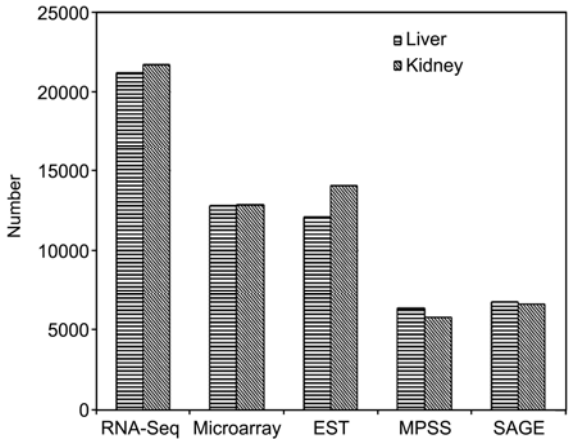

Figure 1 Number of genes of two tissues, liver and kidney, detected by five high-throughput transcriptome technologies. There are in total 25905 genes in the reference data after removing alternatively-spliced transcripts. From RNA-Seq data, 21210 genes for liver and 21759 genes for kidney were detected as expressed, while less than 15000 genes were detected by the other four technologies.

2.2 Comparison of gene expression data between microarray and RNA-Seq data

We compared RNA-Seq with microarray datasets, as they are currently the two most popular technologies for transcriptome studies. We compared commonly and differently detected genes between microarray and RNA-Seq. Almost 97% of the genes identified by microarray can also be detected by RNA-Seq (Figure 3), and about 9000 genes from the RNA-Seq dataset cannot be detected by microarray. Therefore, we mapped the sequences of microarray probes to the UniGene reference data. Of the 25905 genes in the reference data, 21883 genes can be mapped by microarray

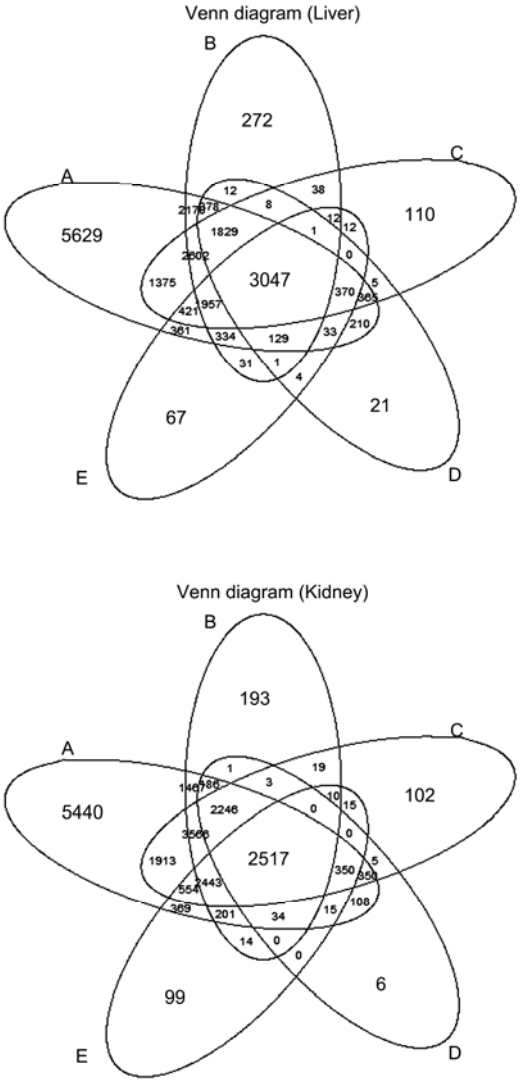

Figure 2 Venn diagrams describing the number of expressed genes detected by five high-throughput transcriptome technologies from liver and kidney. A stands for RNA-Seq, B for microarray, C for EST, D for MPSS, and E for SAGE.

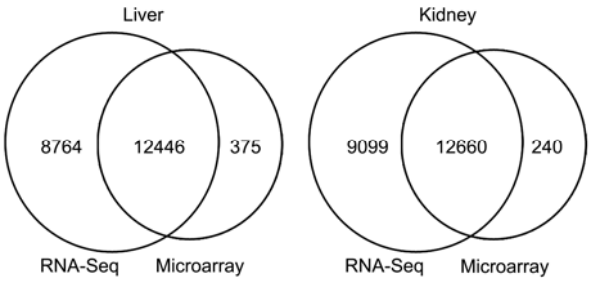

Figure 3 Venn diagrams representing the number of genes detected by RNA-Seq and microarray from liver and kidney tissues.

probes. Respectively, 2093 genes of liver and 2309 genes of kidney cannot be mapped by probes. According to the results, those un-detected genes from microarray data, including non-probe-mapped genes and probe-mapped genes without expression signals, can be detected by the RNA-Seq method and have relatively lower expression signals (Figure 4). The detecting range of RNA-Seq is significantly larger than the microarray technique.

Then expression intensities of each gene were compared, using two different standards of RNA-Seq data and absolute intensities of microarray data (Figure 5). These two independent measures of gene expression are highly correlated, especially the highly expressed genes. We compared the distribution and characteristics of commonly and differently detected genes between microarray and RNA-Seq data (Figure 6). The genes called by both methods expressed significantly more highly than those detected only from RNA-Seq data. The expression values of the genes called by both methods are symmetrically distributed while most of the expression values of the genes only detected by RNA-Seq were lower. Moreover, from Figures 5 and 6, it can be found that average coverage is more appropriate than max depth as a standard, for the expression intensities are more correlative and the distribution is more symmetrical.

2.3 Comparison of gene expression data between MPSS and RNA-Seq, ESTs and RNA-Seq, and SAGE and RNA-Seq data

The gene expression data were analyzed and compared between MPSS and RNA-Seq, between ESTs and RNA-Seq, and between SAGE and RNA-Seq (Figures S1–S9). There were 9244 and 7820 more genes detected by RNA-Seq than by ESTs from liver and kidney, respectively (Figure S1).

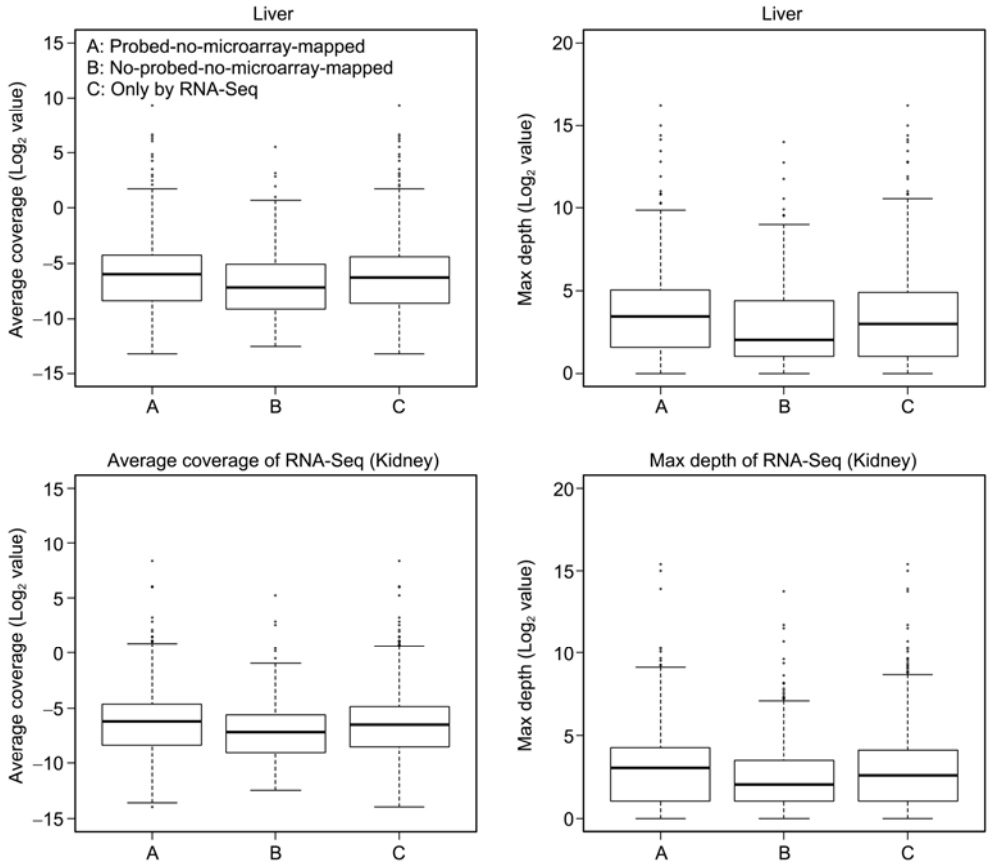

Figure 4 Boxplots summarizing the distribution of the expression values from RNA-Seq data. A represents the genes which cannot be detected from microarray data but can be mapped by probe sequences and can be called from RNA-Seq data. B represents the genes which cannot be detected from microarray data or mapped by probe sequences but can be called from RNA-Seq data. C represents the non-probe-mapped genes that can be detected only from RNA-Seq.

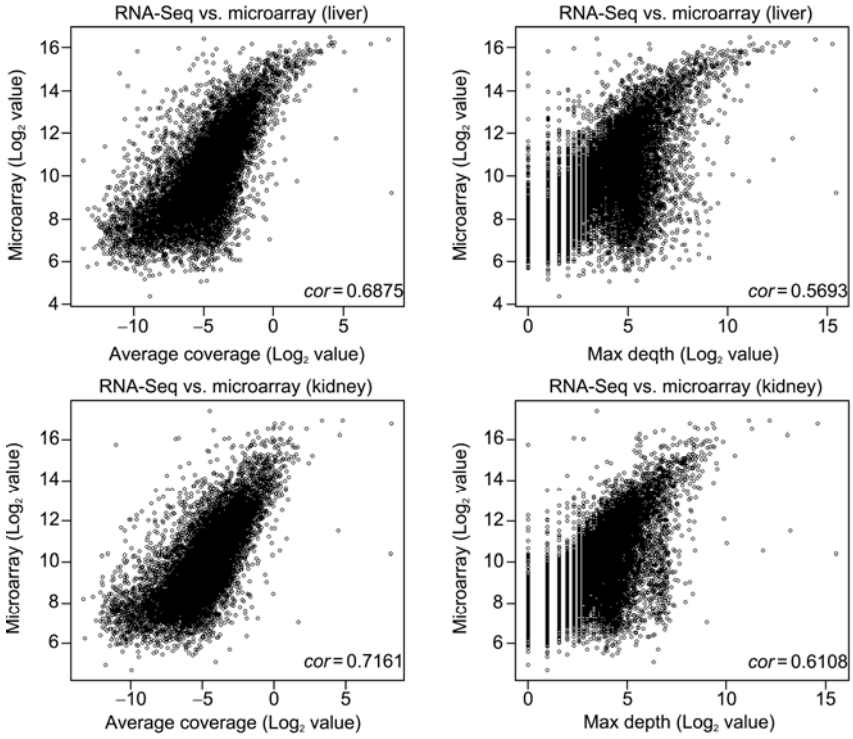

Figure 5 Comparing the gene expression values from RNA-Seq data and microarray data of liver and kidney tissues. In the top two diagrams, the *X*-axis indicates the Log_2 transformed expression values using average coverage as the standard. The corresponding Log_2 transformed expression values from microarray data of the each gene is plotted on the *Y*-axis. In the lower two diagrams, the *X*-axis indicates the Log_2 transformed expression values from RNA-Seq data use max depth as the standard.

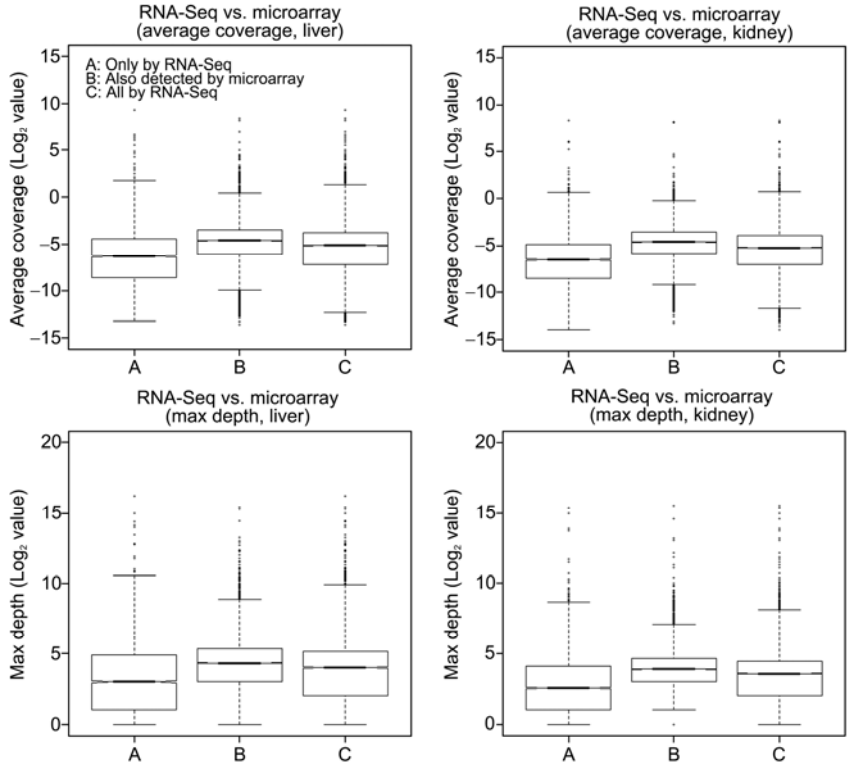

Figure 6 Boxplots summarizing the distribution of the gene expression values from RNA-Seq and microarray data. In each panel, the left box shows the expression data of genes which were only detected from RNA-Seq data, not microarray; the data detected from both RNA-Seq and microarray data is presented in the middle box, and the right box shows all gene expression values detected from RNA-Seq data. In the top two diagrams, the *Y*-axis indicates the expression values (Log_2) using average coverage as the standard and in the bottom two diagrams using max depth as the standard.

The high sensitivity of RNA-Seq was even more obvious when compared to SAGE and MPSS (Figures S4 and S7). There are significant correlations in the measured expression levels of corresponding genes between RNA-Seq and the other platforms. They are; however, lower than the correlation between microarray and RNA-Seq.

2.4 Relationship between transcriptome and proteome

To investigate the relationship between the transcriptome and proteome, we used the Human Protein Atlas dataset to study the correlation between the genes expressed at the protein level and those expressed at the mRNA level in both liver and kidney. All genes with strong, high, medium or moderate intensity in the Human Protein Atlas dataset were labeled as "protein expressed", while those with negative, low or weak intensity were labeled as "protein unexpressed". There were 4304 and 3227 proteins detected from liver and kidney, respectively, in the Human Protein Atlas (Figure 7). Of those, 3964 (92.1%) and 3011 (93.3%) from liver and kidney, respectively, were found in RNA-Seq with corresponding gene sequences. More interestingly, only 960 proteins (22.3%) found in the liver proteome were detected by RNA-Seq. For the kidney, only 703 (21.8%) proteins were detected by RNA-Seq.

3 Discussion

In this study, we present a comparative analysis of gene expression in the transcriptome of liver and kidney tissues by utilizing data from five different technologies, ESTs, MPSS, SAGE, microarray and RNA-Seq. In our study, 3047 genes in liver and 2517 genes in kidney were detected as expressed genes using all five technologies (Figure 2). RNA-Seq could identify an additional 5629 genes in liver and 5440 genes in kidney over the 15581 liver- and 16319 kidney-expressed genes detected by all of the other four technologies. Each of the other four platforms could detect very few unique genes.

In previous studies, only one-third or one-half of the genes were detected as expressed in one tissue. The number of genes detected by microarray, ESTs, MPSS and SAGE was limited by the detection sensitivity of these traditional

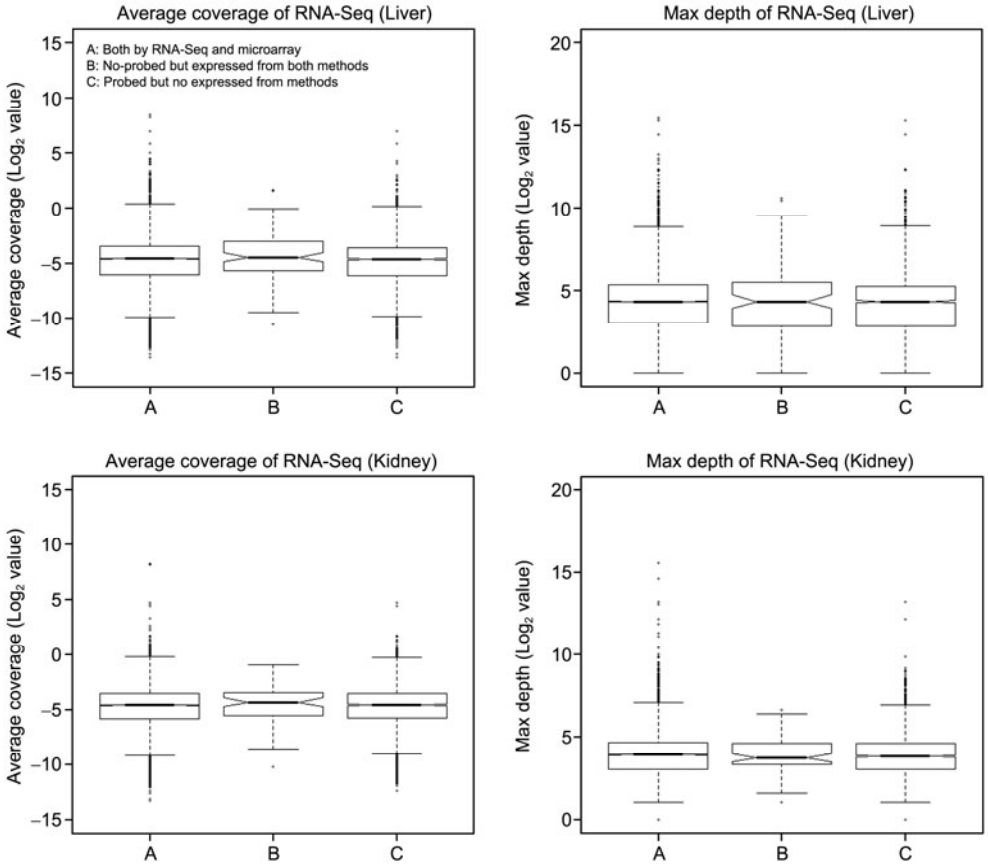

Figure 7 Boxplots summarizing the distribution of protein expression data. The *Y*-axis indicates the Log_2 transformed expression intensity of average coverage and max depth. A represents the genes which can be detected from both RNA-Seq and microarray data. B represents the genes which are expressed in the Human Protein Atlas data but cannot be mapped by probe sequences from those detected from both RNA-Seq and microarray data. C represents the genes which can be mapped by probe sequences from those detected from both RNA-Seq and microarray data but are not expressed in the Human Protein Atlas data.

methods. For example, HG-U133plus 2.0 high-density oligonucleotide arrays contain 54675 probe sets representing only 19164 human genes. However, the RNA-Seq technique provides a whole genome scanning method for gene expression including unknown genes and transcripts. The high sensitivity and accuracy (Figures 2 and 3) in detecting those genes that have very low expression levels is one of the greatest advantages of RNA-Seq.

We compared five different technologies in the study of tissue transcriptomes. One important challenge for four of the traditional technologies is to acquire enough tags from the overall transcripts. ESTs are small segments from 200 to 800 bp, SAGE uses short sequences tags (10–14 bp) to identify transcripts [4], MPSS uses signature sequence tags (about 16–20 bp) [5], and microarray probes are special sequences of particular known genes. Deep sequencing of RNAs (RNA-Seq) is a useful tool to characterize and quantify transcriptomes [1]. Compared with traditional techniques, the advantages of RNA-Seq include the following: owing to the high-throughput of next-generation sequencing, the coverage and depth of RNA-Seq data are significant higher than ESTs, SAGE and MPSS sequencing; furthermore, the ability to identify *novo* transcripts should be highlighted when compared with the microarray technique. The number of genes detected by each technology provides evidence. When considering the probe sequences and gene expression data at the protein level, the probe sequences of the microarray in use might not have wide coverage. The non-probe-mapped genes also can be detected as expressed at the protein level (Figure 7).

In addition, we evaluated the performance of two measurement standards for the expression of genes detected by RNA-Seq: average coverage and max depth. The results indicate that the measurement of average coverage was more continuous than that of max depth and had stronger correlations with the results of the other four methods. Therefore, we will use average coverage as the standard to measure gene expression levels from RNA-Seq data in future studies. We also compared the widely used index of expression level, RPKM [26], with our two indexes. The results using RPKM were highly consistent with each of our two indexes (Supporting Information S6). Considering the good performance of our index with low expressing genes and their low computational complexity, these two measurement indexes are more suitable for biologists to perform a rough estimate of gene expression levels from RNA-Seq data.

Using the genes detected by different technologies and their expression values, we found RNA-Seq was the most sensitive in terms of the number of genes it detected while SAGE and MPSS were the least sensitive. For the genes detected by all other platforms, there were generally good correlations in the measured expression levels of corresponding genes. It is likely that sample differences were responsible for the few genes that were not detectable by RNA-Seq. In our work, we have used different experimental data or data from different public databases to compare five different platforms for tissue gene expression. As the experimental samples are not the same and the experiments are not performed by same lab at the same time, we have reason to infer that there is sample error. In addition, samples from RNA-Seq datasets are biological repeats while microarray data contains technical repeats from the same library (Supporting Information S5). For this reason, our results contain sampling errors and have overriding significance.

Our findings highlight that many liver and/or kidney expressed genes, which had not yet been identified in traditional techniques, were detected by the RNA-Seq technique. This is particularly important since most of them are highly associated with biological functions of the liver and/or kidney.

FLT3, detected only by RNA-Seq in the liver, encodes a class III receptor tyrosine kinase that regulates hematopoiesis [27]. FLT3-ligand (FL) is important for the proliferation and differentiation of human hematopoietic progenitors both *in vivo* and *in vitro*, and FL administration significantly induces an antitumor effect and inhibits liver metastases [28]. ENTPD8, also known as NTPDase8, detected only by RNA-Seq, also plays an important role in liver function [29]. It is the liver canalicular ecto-ATPase/ATPDase and is responsible for the main NTPDase activity in liver [29]. CDH16 is a member of the cadherin superfamily and is expressed exclusively in the kidney [30] was also detected by RNA-Seq only. The connection of CDH16 to the cytoskeleton is important for maintaining tissue integrity in the kidney, which relies on the interaction of CDH16 with alpha B-crystallin [31]. RhCG is a member of an ammonia transporter family and plays a critical role in ammonium handling and pH homeostasis in the human kidney [32]. According to Brown's work, under normal conditions RhCG is the major putative ammonia transporter expressed in the human kidney [33]. The expression of RhCG was also only detected with RNA-Seq data.

4 Conclusion

The major significance of our work is that, by performing a meta-comparison analysis of five different technologies, we found RNA-Seq had the highest number of genes detected as expressed in two tissues and that it is sufficiently substitutable for the other four technologies. Our work provided a technical framework for the analysis of expressed genes, the correlation of gene expression and provided a catalog of expressed genes in the liver and kidney. The integrated tissue transcriptome data should provide a valuable resource for the in-depth understanding of human tissues and diseases.

This work was supported by National Basic Research Program of China (2012CB316501 and 2012CB517900), the National Natural Science Foundation of China (90913009), and in part by Shanghai Pujiang Scholarship Program (10PJ1408000). The author gratefully acknowledges the support of SA-SIBS Scholarship Program.

1 Wang Z, Gerstein M, Snyder M. RNA-Seq: A revolutionary tool for transcriptomics. Nat Rev Genet, 2009, 10: 57–63
2 Attardo G M, Ribeiro J M, Wu Y, et al. Transcriptome analysis of reproductive tissue and intrauterine developmental stages of the tsetse fly (*Glossina morsitans morsitans*). BMC Genomics, 2010, 11: 160
3 Lincecum J M, Vieira F G, Wang M Z, et al. From transcriptome analysis to therapeutic anti-CD40L treatment in the SOD1 model of amyotrophic lateral sclerosis. Nat Genet, 2010, 42: 392–399
4 Velculescu V E, Zhang L, Vogelstein B, et al. Serial analysis of gene expression. Science, 1995, 270: 484–487
5 Brenner S, Johnson M, Bridgham J, et al. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol, 2000, 18: 630–634
6 Sterky F, Regan S, Karlsson J, et al. Gene discovery in the wood-forming tissues of poplar: Analysis of 5,692 expressed sequence tags. Proc Natl Acad Sci USA, 1998, 95: 13330–13335
7 Adams M D, Dubnick M, Kerlavage A R, et al. Sequence identification of 2,375 human brain genes. Nature, 1992, 355: 632–634
8 Ye S Q, Usher D C, Zhang L Q. Gene expression profiling of human diseases by serial analysis of gene expression. J Biomed Sci, 2002, 9: 384–394
9 George A J, Gordon L, Beissbarth T, et al. A serial analysis of gene expression profile of the Alzheimer's disease Tg2576 mouse model. Neurotox Res, 2010, 17: 360–379
10 Reinartz J, Bruyns E, Lin J Z, et al. Massively parallel signature sequencing (MPSS) as a tool for in-depth quantitative gene expression profiling in all organisms. Brief Funct Genomic Proteomic, 2002, 1: 95–104
11 Schena M, Shalon D, Davis R W, et al. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science, 1995, 270: 467–470
12 Morin R, Bainbridge M, Fejes A, et al. Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. Biotechniques, 2008, 45: 81–94
13 Cloonan N, Forrest A R, Kolle G, et al. Stem cell transcriptome profiling via massive-scale mRNA sequencing. Nat Methods, 2008, 5: 613–619
14 Bertone P, Stolc V, Royce T E, et al. Global identification of human transcribed sequences with genome tiling arrays. Science, 2004, 306: 2242–2246
15 David L, Huber W, Granovskaia M, et al. A high-resolution map of transcription in the yeast genome. Proc Natl Acad Sci USA, 2006, 103: 5320–5325
16 Sultan M, Schulz M H, Richard H, et al. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science, 2008, 321: 956–960
17 Chepelev I, Wei G, Tang Q, et al. Detection of single nucleotide variations in expressed exons of the human genome using RNA-Seq. Nucleic Acids Res, 2009, 37: e106
18 Mortazavi A, Williams B A, McCue K, et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods, 2008, 5: 621–628
19 Barbazuk W B, Emrich S J, Chen H D, et al. SNP discovery via 454 transcriptome sequencing. Plant J, 2007, 51: 910–918
20 Marioni J C, Mason C E, Mane S M, et al. RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. Genome Res, 2008, 18: 1509–1517
21 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 2009, 25: 1754–1760
22 He F. Human liver proteome project: Plan, progress, and perspectives. Mol Cell Prot, 2005, 4: 1841–1848
23 Gautier L, Cope L, Bolstad B M, et al. Affy-analysis of Affymetrix GeneChip data at the probe level. Bioinformatics, 2004, 20: 307–315
24 Yu Y, Ping J, Chen H, et al. A comparative analysis of liver transcriptome suggests divergent liver function among human, mouse and rat. Genomics, 2010, 96: 281–289
25 Uhlen M, Bjorling E, Agaton C, et al. A human protein atlas for normal and cancer tissues based on antibody proteomics. Mol Cell Prot, 2005, 4: 1920–1932
26 Mortazavi A, Williams A B, McCue K, et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods, 2008, 5: 621–628
27 Gilliland D G, Griffin J D. The roles of FLT3 in hematopoiesis and leukemia. Blood, 2002, 100: 1532–1542
28 Péron J M, Esche C, Subbotin V M, et al. FLT3-ligand administration inhibits liver metastases: Role of NK cells. J Immunol, 1998, 161: 6164–6170
29 Fausther M, Lecka J, Kukulski F, et al. Cloning, purification, and identification of the liver canalicular ecto-ATPase as NTPDase8. Am J Physiol Gastrointest Liver Physiol, 2007, 292: G785–795
30 Thedieck C, Kuczyk M, Klingel K, et al. Expression of Ksp-cadherin during kidney development and in renal cell carcinoma. Br J Cancer, 2005, 92: 2010–2017
31 Thedieck C, Kalbacher H, Kratzer U, et al. Alpha B-crystallin is a cytoplasmic interaction partner of the kidney-specific cadherin-16. J Mol Biol, 2008, 378: 145–153
32 Planelles G. Ammonium homeostasis and human rhesus lycoproteins. Nephron Physiol, 2006, 105: 11–17
33 Brown A C, Hallouane D, Mawby W J, et al. RhCG is the major putative ammonia transporter expressed in the human kidney, and RhBG is not expressed at detectable levels. Am J Physiol Renal Physiol, 2009, 296: F1279–1290

Supporting Information

S1 Comparison of gene expression data between EST and RNA-Seq data

Figure S1 Venn diagrams representing the number of genes detected by RNA-Seq data (left cycle) and EST data (right cycle) for liver and kidney.

Figure S2 Comparing the gene expression values from RNA-Seq data and EST for liver and kidney.

Figure S3 Boxplots summarizing the non-parametric distribution of the expression values from RNA-Seq data comparing with that from the EST data for liver and kidney.

S2 Comparison of gene expression data between MPSS and RNA-Seq data

Figure S4 Venn diagrams representing the number of genes detected by RNA-Seq data (left cycle) and MPSS data (right cycle) for liver and kidney.

Figure S5 Comparing the gene expression values from RNA-Seq data and MPSS for liver and kidney.