Int J Clin Exp Med 2018;11(9):8988-9004
www.ijcem.com /ISSN:1940-5901/IJCEM0068703

*Original Article*

# The transcriptome difference between vulnerable and stable carotid atherosclerotic plaque

Tongxun Li[1], Xiaoping Zhang[2], Jinqian Zhang[3], Rui Liu[4], Chengxiong Gu[1]

*Departments of [1]Cardiothoracic Surgery, [2]Beijing Institute of Heart, Lung and Blood Vessel Diseases, Beijing Anzhen Hospital, Capital Medical University, Beijing 100029, China; [3]Department of Laboratory Medicine and Central Laboratories, Guangdong Second Provincial General Hospital, Guangzhou 510317, China; [4]Department of Rehabilitation, Tangdu Hospital, Fourth Military Medical University, Xi'an 710038, Shaanxi, China*

Received July 27, 2017; Accepted April 30, 2018; Epub September 15, 2018; Published September 30, 2018

**Abstract:** In recent years, the instability of atherosclerotic plaques is a hotspot of basic and clinical research. Moreover, it is imperative necessary to accurately identify the vulnerable plaques, and give effectively intervention before acute cardiovascular events. In this work, the stability of atherosclerotic plaques and vulnerable plaques were collected and performed with next-generation of RNA-seq, then data were analyzed. We performed a transcriptome sequence analysis of the vulnerable and stable carotid atherosclerotic plaque by next generation sequencing (NGS) using the Illumina Deep Sequencing technology. We obtained 5,473,799, 3,766,914, 7,713,565 qualified Illumina reads from three vulnerable carotid atherosclerotic plaque, or 5,598,305, 7,249,629, and 5,623,320 qualified Illumina reads from stable carotid atherosclerotic plaque, respectively. Comparative trancriptome analysis differentially revealed 781 genes between vulnerable and stable carotid atherosclerotic plaque, 318 expression of genes was up-regulated and 363 expressions of genes were down-regulated. The volcano plot was used to further identify differential genes expression between the two groups. Cluster analysis was performed, and spearman correlation coefficient was calculated and analyzed. Our work demonstrated differential transcriptomes between vulnerable and stable carotid atherosclerotic plaque. Numbers of genes will serve as a promising resource for revealing the regulatory molecular mechanisms of expression associated with the pathophysiology and pathogenesis of vulnerable atherosclerosis plaque, even their implications in the field of therapy.

**Keywords:** Atherosclerosis (AS), vulnerable plaque, next-generation sequencing (NGS), transcriptome

## Introduction

Atherosclerosis (AS) is a risk factor for cardiovascular disease, which is one of the most serious diseases to harm human health. In recent years, the instability of atherosclerotic plaques is a hotspot of basic and clinical research [1, 2]. However, it is imperative necessary to accurately identify the vulnerable plaques, and give effectively intervention before acute cardiovascular events [3, 4]. In this work, the stability of atherosclerotic plaques and vulnerable plaques were collected and performed with next-generation of RNA-seq, then data were analyzed.

Carotid atherosclerotic disease represents a well-established cause of ischaemic stroke, accounting for up to 20% of strokes or transient ischaemic attacks (TIA) [5]. Stroke constitutes a major cause of acquired disability in adults and the second most frequent cause of mortality in developed nations [6]. Nevertheless, current research has concluded that plaque features other than degree of stenosis contribute to the occurrence of neurologic symptoms, justifying the introduction of the term "vulnerable plaque" [7], responsible for almost half of stroke cases [8].

The plaque surface morphology is among those features related to the risk for embolic stroke and characterising vulnerability. Based on this criterion, carotid plaques are typically classified into smooth, irregular or ulcerated [9, 10]. The presence of ulceration itself is a well-known feature of vulnerability with high clinical significance as entailing increased risk for neurologic symptoms. Beyond grading of stenosis with widely accepted velocity criteria [11].

Moreover, how mRNA expression regulation at translation level in vulnerable plaques remains unclear. This present study is to highlight the role of transcriptomes in the pathophysiology and pathogenesis of vulnerable plaques.

## Materials and methods

### *Patients and Specimens*

*Symptomatic group:* The study consisted of three diagnosed with vulnerable carotid atherosclerotic plaque by pathology companied with symptoms of cerebral ischemia, and diagnosed at vulnerable plaque.

*Asymptomatic group:* Three patients with stable carotid atherosclerotic plaque with pathology were not acompanied with symptoms of cerebral ischemia.

All included patients were referred to Beijing Anzhen Hospital, Capital Medical University from January, 2011 to December, 2016. Their all demographic features, or clinical and carotid artery characteristic were documented and analyzed. All these patients were performed with carotid endarterectomy (CEA) (**Figure 1A**). At the operation day of CEA, the samples of plaque were collected and handled immediately.

This study and use of human specimens in this research were approved by the Ethics Committee of Beijing Anzhen Hospital, Capital Medical University on the basis of the Declaration of Helsinki. We clearly confirmed that informed consents were obtained from all patients. We had record and document participant consent in our hospital. And the ethics committees of our hospital had approved this consent procedure.

### *HE staining*

HE staining was conducted according to routine protocols. Briefly, after deparaffinization and rehydration, 5 μm longitudinal sections were stained with hematoxylin solution for 5 min followed by 5 dips in 1% acid ethanol (1% HCl in 70% ethanol) and then rinsed in distilled water. Then the sections were stained with eosin solution for 3 min and followed by dehydration with graded alcohol and clearing in xylene. The mounted slides were then examined and photographed using an Olympus BX53 fluorescence microscope (Tokyo, Japan). The staining intensity of the trabecular bone was analyzed by Image-Pro Plus 6.0 software and expressed as IOD value.

### *Library preparation*

*Overview of RNA-Seq:* The general steps to prepare a complementary DNA (cDNA) library for sequencing are described below, but often vary between platforms [12, 13].

*RNA isolation:* RNA is isolated from tissue and mixed with deoxyribonuclease (DNase). DNase reduces the amount of genomic DNA. The amount of RNA degradation is checked with gel and capillary electrophoresis and is used to assign an RNA integrity number to the sample. This RNA quality and the total amount of starting RNA are taken into consideration during the subsequent library preparation, sequencing, and analysis steps.

*RNA selection/depletion:* To analyze signals of interest, the isolated RNA can either be kept as is, filtered for RNA with 3' polyadenylated (poly(A)) tails to include only mRNA, depleted ofribosomal RNA (rRNA), and/or filtered for RNA that binds specific sequences. The RNA with 3' poly(A) tails are mature, processed, coding sequences. Poly(A) selection is performed by mixing RNA with poly(T) oligomers covalently attached to a substrate, typically magnetic beads [14, 15], and the website http://www.protocol-online.org/prot/Molecular_Biology/RNA/RNA_Extraction/mRNA_Isolation/index.html, which is the Protocol Online and provides a list of several protocols relating to mRNA isolation Poly(A) selection ignores noncoding RNA and introduces 3' bias [16], which is avoided with the ribosomal depletion strategy. The rRNA is removed because it represents over 90% of the RNA in a cell, which if kept would drown out other data in the transcriptome.

*cDNA synthesis:* DNA sequencing technology is more mature, so the RNA is reverse transcribed to cDNA. Reverse transcription results in loss of strandedness, which can be avoided with chemical labelling. Fragmentation and size selection are performed to purify sequences

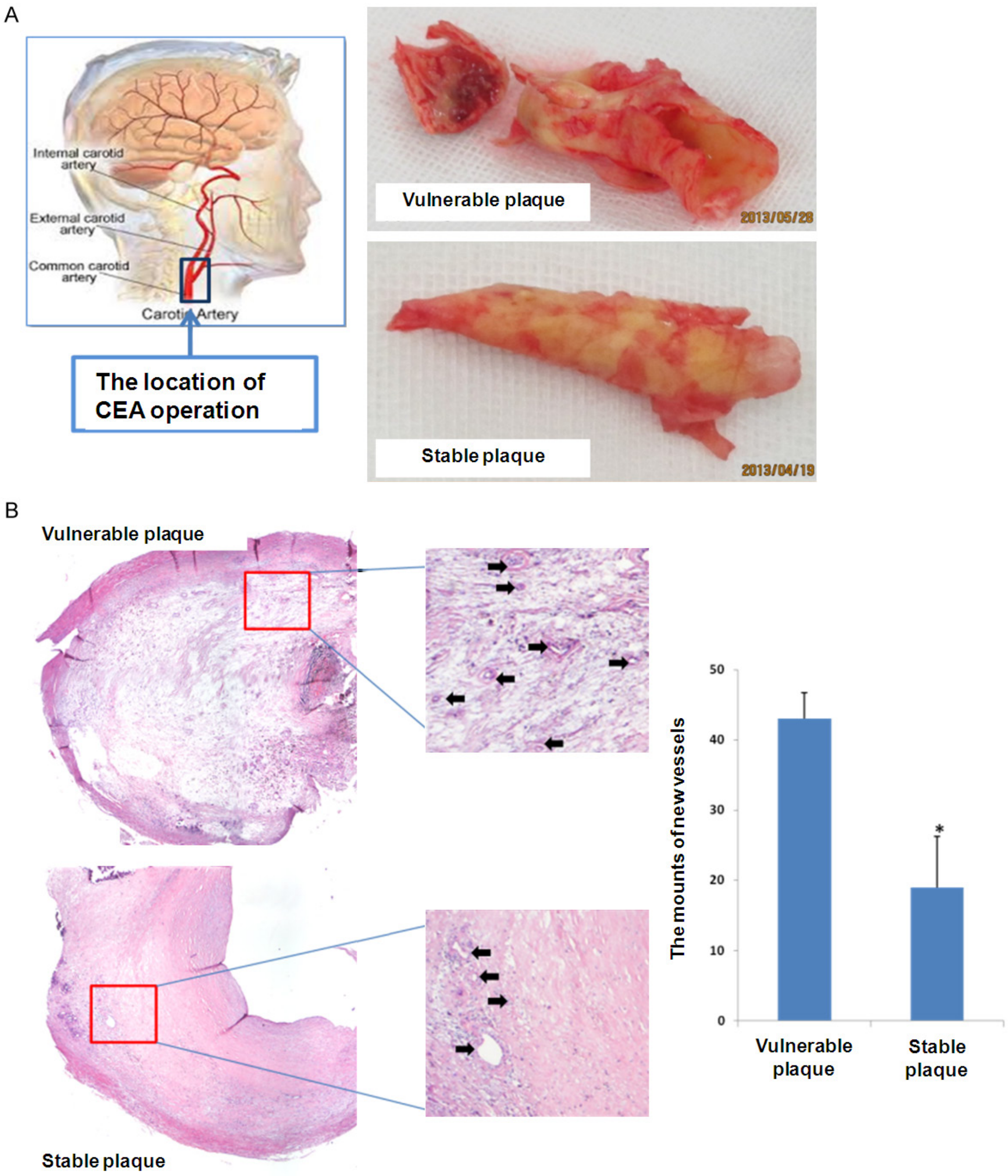

**Figure 1.** Compared to the morphological and pathological characteristics between vulnerable and stable plaque. The patients in this work were all performed with carotid endarterectomy (CEA). The (A) showed the location of CEA operation and the porphology of typical vulnerable and stable plaque. As shown in (B) (×50, and ×200), HE staining demonstrated that the amounts of new vessels in symptomatic patients with vulnerable plaque were much more than those in asymptomatic patients with stable plaque.

that are the appropriate length for the sequencing machine. The RNA, cDNA, or both are fragmented with enzymes, sonication, or nebulizers. Fragmentation of the RNA reduces 5' bias of randomly primed-reverse transcription and the influence of primer binding sites [15], with the downside that the 5' and 3' ends are converted to DNA less efficiently. Fragmentation is followed by size selection, where either small sequences are removed or a tight range of

**Table 1.** Compared to the baseline clinical characteristics of patients between the two groups

| | Symptomatic group (n=13) | Asymptomatic group (n=10) | *P* value |
|---|---|---|---|
| Age (mean ± SD) | 66.23±6.92 | 69.45±9.18 | 0.371 |
| Sex (male/female) | 7/2 | 10/2 | 0.134 |
| BMI (Kg/m$^2$) | 25.5±0.62 | 24±0.73 | 0.382 |
| SP (mmHg, mean ± SD) | 147.14±12.38 | 138.34±11.26 | 0.263 |
| DP (mmHg, mean ± SD) | 81.29±10.57 | 78.38±12.08 | 0.367 |
| TG (mmol/L) | 1.42±0.43 | 1.37±0.51 | 0.886 |
| CHO (mmol/L) | 4.31±0.44 | 3.62±0.87 | 0.119 |
| HDL (mmol/L) | 1.05±0.59 | 1.01±0.63 | 0.675 |
| LDL (mmol/L) | 2.6±0.39 | 2.07±0.74 | 0.098 |
| HCY (µmol/L) | 11.21±0.64 | 22.64±0.18 | 0.033 |

Note: BMI, Body Mass Index; SP, systolic pressure; TG, triglyceride; CHO, cholesterol; HDL, high density lipoprotein; LDL, low density lipoprotein; HCY, homocysteine.

sequence lengths are selected. Because small RNAs like miRNAs are lost, these are analyzed independently. The cDNA for each experiment can be indexed with a hexamer or octamer barcode, so that these experiments can be pooled into a single lane for multiplexed sequencing.

*Direct RNA sequencing*

As converting RNA into cDNA using reverse transcriptase has been shown to introduce biases and artifacts that may interfere with both the proper characterization and quantification of transcripts [17], single molecule Direct RNA Sequencing (DRSTM) technology was under development by Helicos (now bankrupt). DRSTM sequences RNA molecules directly in a massively-parallel manner without RNA conversion tocDNA or other biasing sample manipulations such as ligation and amplification.

*Transcriptome assembly*

Two methods are used to assign raw sequence reads to genomic features (i.e., assemble the transcriptome):

*De novo:* This approach does not require a reference genome to reconstruct the transcriptome, and is typically used if the genome is unknown, incomplete, or substantially altered compared to the reference [18]. Challenges when using short reads for de novo assembly include 1) determining which reads should be joined together into contiguous sequences (contigs), 2) robustness to sequencing errors and other artifacts, and 3) computational efficiency. The primary algorithm used for de novo assembly transitioned from overlap graphs, which identify all pair-wise overlaps between reads, tode Bruijn graphs, which break reads into sequences of length k and collapse all k-mers into a hash table. Overlap graphs were used with Sanger sequencing, but do not scale well to the millions of reads generated with RNA-Seq. Examples of assemblers that use de Bruijn graphs are Velvet [19], Trinity [18], Oases [20], and Bridger [21]. Paired end and long read sequencing of the same sample can mitigate the deficits in short read sequencing by serving as a template or skeleton. Metrics to assess the quality of a de novo assembly include median contig length, number of contigs and N50 [22].

RNA-seq mapping of short reads in exon-exon junctions. The final mRNA is sequenced, which is missing the intronic sections of the pre-mRNA.

*Genome guided:* This approach relies on the same methods used for DNA alignment, with the additional complexity of aligning reads that cover non-continuous portions of the reference genome [23]. These non-continuous reads are the result of sequencing spliced transcripts. Typically, alignment algorithms have two steps: 1) align short portions of the read (i.e., seed the genome), and 2) use dynamic programming to find an optimal alignment, sometimes in combination with known annotations. Software tools that use genome-guided alignment include Bowtie [24], TopHat (which builds on BowTie results to align splice junctions) [25, 26], Subread [27], STAR [23], Sailfish [28], Kallisto [29] and GMAP [30]. The quality of a genome guided assembly can be measured with both 1) de novo assembly metrics (e.g., N50) and 2) comparisons to known transcript, splice junction, genome, and protein sequences using precision, recall, or their combination (e.g., F1 score) [22].

A note on assembly quality: The current consensus is that 1) assembly quality can vary depending on which metric is used, 2) assemblies that scored well in one species do not nec-

**Table 2.** Compared to the carotid arterial characteristics of patients determined by ultrasound between the two groups

| | Symptomatic group (n=13) | Asymptomatic group (n=10) | *P* value |
|---|---|---|---|
| Stroke history n (%) | 13 (100) | 0 (0) | 0.000 |
| Rate of carotid stenosis n (%) | 10 (76.92) | 7 (70.00) | 0.064 |
| Blood flow in the carotid artery (cm/s, mean ± SD) | 378±23.34 | 296±31.08 | 0.297 |
| Carotid artery diameter (cm, mean ± SD) | 0.12±0.03 | 0.19±0.04 | 0.021* |
| Carotid plaque thickness(cm, mean ± SD) | 0.53±0.08 | 0.43±0.09 | 0.918 |

*P<0.05.

**Table 3.** The capacity of sequencing

| Sample | Read (#) | Data (bp) |
|---|---|---|
| Vulnerable plaque 1 | 5,473,799 | 197,056,764 |
| Vulnerable plaque 2 | 3,766,914 | 135,608,904 |
| Vulnerable plaque 3 | 7,713,565 | 277,688,340 |
| Stable plaque 1 | 5,598,305 | 201,538,980 |
| Stable plaque 2 | 7,249,629 | 260,986,644 |
| Stable plaque 3 | 5,623,320 | 202,439,520 |

**Table 4.** The quality of sequencing

| Sample | Q20 (%) | Q30 (%) |
|---|---|---|
| Vulnerable plaque 1 | 99.17 | 97.55 |
| Vulnerable plaque 2 | 99.31 | 97.82 |
| Vulnerable plaque 3 | 99.55 | 98.24 |
| Stable plaque 1 | 99.53 | 98.20 |
| Stable plaque 2 | 99.09 | 97.41 |
| Stable plaque 3 | 99.42 | 97.43 |

essarily perform well in the other species, and 3) combining different approaches might be the most reliable [31, 32].

*Gene expression*

Expression is quantified to study cellular changes in response to external stimuli, differences between healthy and diseased states, and other research questions. Gene expression is often used as a proxy for protein abundance, but these are often not equivalent due to post transcriptional events such as RNA interference and nonsense-mediated decay [33].

Expression is quantified by counting the number of reads that mapped to each locus in the transcriptome assembly step. Expression can be quantified for exons or genes using contigs or reference transcript annotations [12]. These observed RNA-Seq read counts have been robustly validated against older technologies, including expression microarrays and qPCR [33, 34]. Tools that quantify counts are HTSeq [35], FeatureCounts [36], Rcount [37], maxcounts [38], FIXSEQ [39], and Cuffquant. The read counts are then converted into appropriate metrics for hypothesis testing, regressions, and other analyses. Parameters for this conversion are:

*Library size:* Although sequencing depth is prespecified when conducting multiple RNA-Seq experiments, it will still vary widely between experiments [40]. Therefore, the total number of reads generated in a single experiment (library size) is typically adjusted by converting counts to fragments, reads, or counts per million mapped reads (FPM, RPM, or CPM).

*Gene length:* Longer genes will have more fragments/reads/counts than shorter genes if transcript expression is the same. This is adjusted by dividing the FPM by the length of a gene, resulting in the metric fragments per kilobase of transcript per million mapped reads (FPKM) [41].

*Total sample RNA output:* Because the same amount of RNA is extracted from each sample, samples with more total RNA will have less RNA per gene. These genes appear to have decreased expression, resulting in false positives in downstream analyses [40].

Variance for each gene's expression: is modeled to account for sampling error (important for genes with low read counts), increase power, and decrease false positives. Variance can be estimated as a normal, Poisson, or negative binomialdistribution [42-44].

*Differential expression and absolute quantification of transcripts*

RNA-Seq is generally used to compare gene expression between conditions, such as a drug treatment vs. non-treated, and find out which

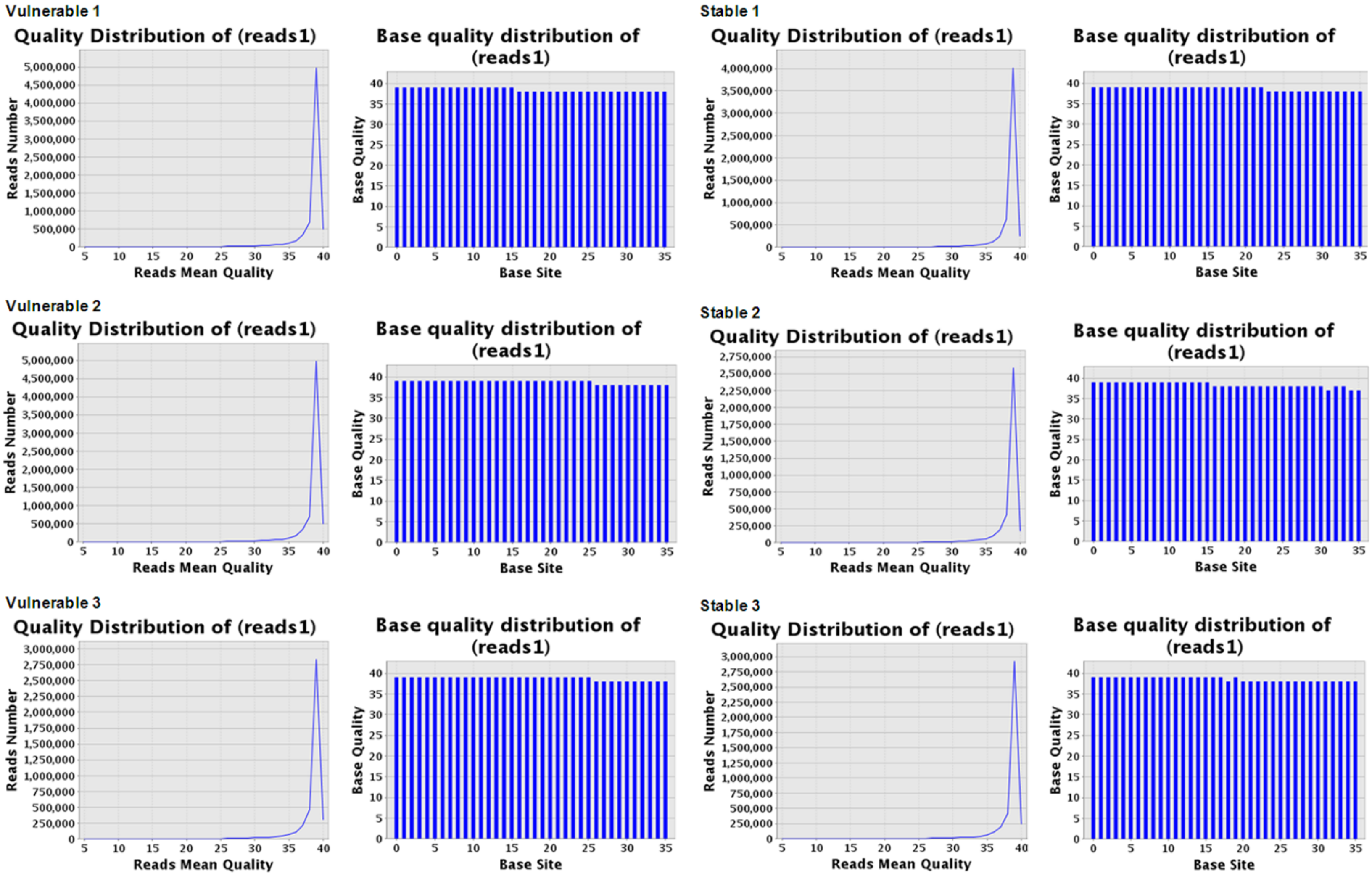

**Figure 2.** Quality of sequencing. The statistical analysis demonstrated that the average quality of every sample was larger than 20, and indicated that the quality was very fine.

**Table 5.** Data Pre-processing of NGS

| Sample | Trim 3' adapter (%) | Short reads |
|---|---|---|
| Vulnerable plaque 1 | 3,623,320 (96.4) | 1,505,942 (40.0) |
| Vulnerable plaque 2 | 5,563,679 (98.9) | 4,481,675 (79.7) |
| Vulnerable plaque 3 | 5,069,495 (92.6) | 2,342,920 (42.8) |
| Stable plaque 1 | 3,956,990 (94.2) | 1,996,908 (47.5) |
| Stable plaque 2 | 2,964,077 (98.3) | 2,465,353 (81.8) |
| Stable plaque 3 | 5,438,345 (97.1) | 2,552,773 (45.6) |

genes are up- or down-regulated in each condition. In principle, RNA-Seq will make it possible to account for all the transcripts in the cell for each condition. Differently expressed genes can be identified using tools that count the sequencing reads per gene and compare them between samples. Many packages are available for this type of analysis [45]; some of the most commonly used tools are DESeq [46] and edger [44], packages from Bioconductor [47, 48]. Both these tools use a model based on the negative binomial distribution [44, 46].

It is not possible to do absolute quantification using the common RNA-Seq pipeline, because it only provides RNA levels relative to all transcripts. If the total amount of RNA in the cell changes between conditions, relative normalization will misrepresent the changes for individual transcripts. Absolute quantification of mRNAs is possible by performing RNA-Seq with added spike ins, samples of RNA at known concentrations. After sequencing, the read count of the spike ins sequences is used to determine the direct correspondence between read count and biological fragments [49, 50]. In developmental studies, this technique has been used in Xenopus tropicalis embryos at a high temporal resolution, to determine transcription kinetics [51].

*Coexpression networks*

Coexpression networks are data-derived representations of genes behaving in a similar way across tissues and experimental conditions [52]. Their main purpose lies in hypothesis generation and guilt-by-association approaches for inferring functions of previously unknown genes [53]. RNASeq data has been recently used to infer genes involved in specific pathways based on Pearson correlation, both in plants [54] and mammals [55]. The main advantage of RNASeq data in this kind of analysis over the microarray platforms is the capability to cover the entire transcriptome, therefore allowing the possibility to unravel more complete representations of the gene regulatory networks. Differential regulation of the splice isoforms of the same gene can be detected and used to predict and their biological functions [56, 57]. Weighted gene co-expression network analysis has been successfully used to identify co-expression modules and intramodular hub genes based on RNA seq data. Co-expression modules may corresponds to cell types or pathways. Highly connected intramodular hubs can be interpreted as representatives of their respective module. Variance-Stabilizing Transformation approaches for estimating correlation coefficients based on RNA seq data have been proposed.

*Statistical analysis*

Welch *t*-test-paired was performed for comparison of transcriptome difference between the two groups. False discovery rate (FDR) was used to evaluate the errors due to multiple comparisons. On the basis of the expression profiles, Hierarchical clustering of selected mRNAs was conducted using Ward's agglomeration method.

## Results

*Compared to the baseline characteristics of patients in the two groups*

The demographic and clinical features of patients in the two groups were compared, including age, sex, BMI, SP, DP, TG, CHO, HDL, LDL, HCY. Moreover, it did not show significantly different of these characteristics between symptomatic group with vulnerable plaque and asymptomatic group with stable plaque (**Table 1**). Furthermore, the carotid arterial characteristic of patients determined by ultrasound were compared. Our results indicated that there were stroke history in symptomatic group, and obviously more than asymptomatic patients (13 vs. 0; *P*=0.000). The carotid artery diameter in asymptomatic group was significantly lower than symptomatic group (0.12±0.03 vs. 0.19±0.04; *P*=0.021). But, there were not obviously different of characteristics, including rate of carotid stenosis, blood flow in the carotid artery, and carotid plaque thickness between

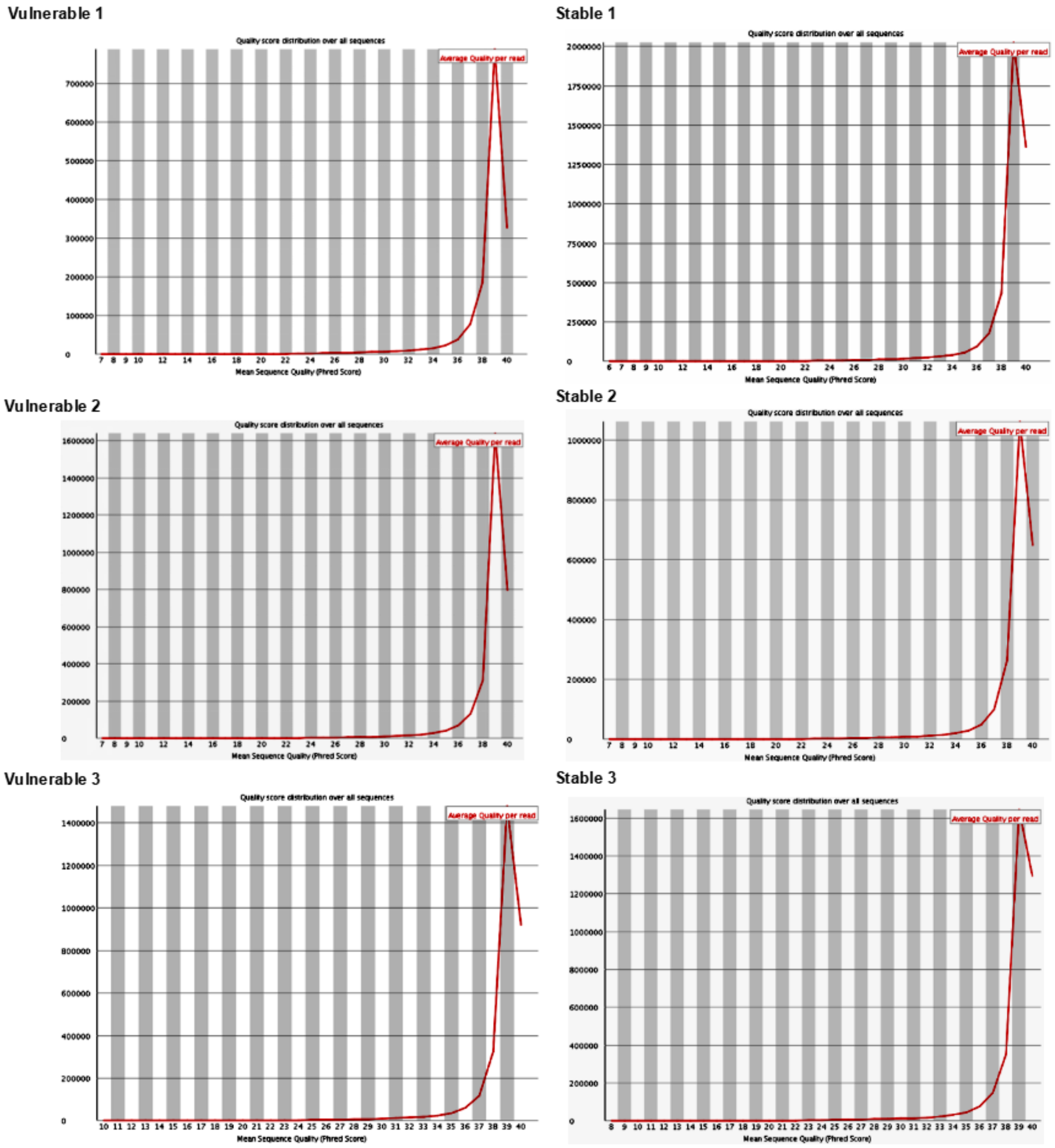

**Figure 3.** Per sequence quality scores. We used fastq_quality_filter program to remove the reads of lower quality, so that the short reads was obtained and the quality score of at least 95% base was not low than 20.

symptomatic group with vulnerable plaque and asymptomatic group with stable plaque (**Table 2**).

*Morphological observation of plaques*

The patients in this work were all performed with carotid endarterectomy (CEA) (**Figure 1A**). The typical plaque of vulnerable and stable were showed in **Figure 1A**. It demonstrated that the vulnerable plaque looks like fragile and uncomplete. As shown in **Figure 1B**, HE staining demonstrated that the amounts of new vessels in symptomatic patients with vulnerable plaque were much more than those in asymptomatic patients with stable plaque.

*Capacity and quality of transcriptome sequencing of plaques*

We obtained 5,473,799, 3,766,914, 7,713,565 qualified Illumina reads from three vulnerable

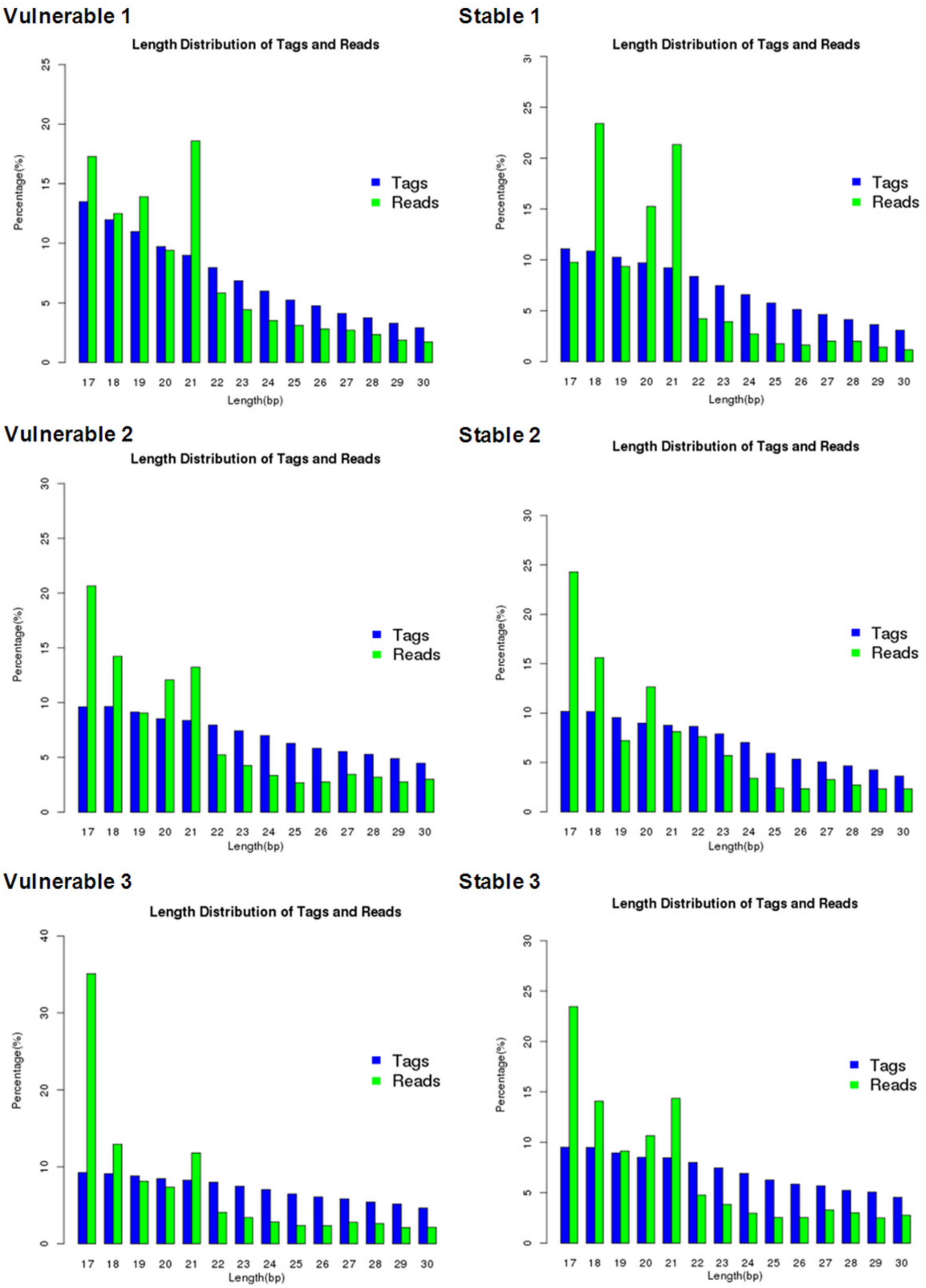

**Figure 4.** The length distribution of tags and reads. The cluster was performed to mark all identical sequences as one tag. The results showed that the length of all the filtered mature transcriptome sequences distributed range from 17-30 nt, and then sequences in the range was analyzed follow up.

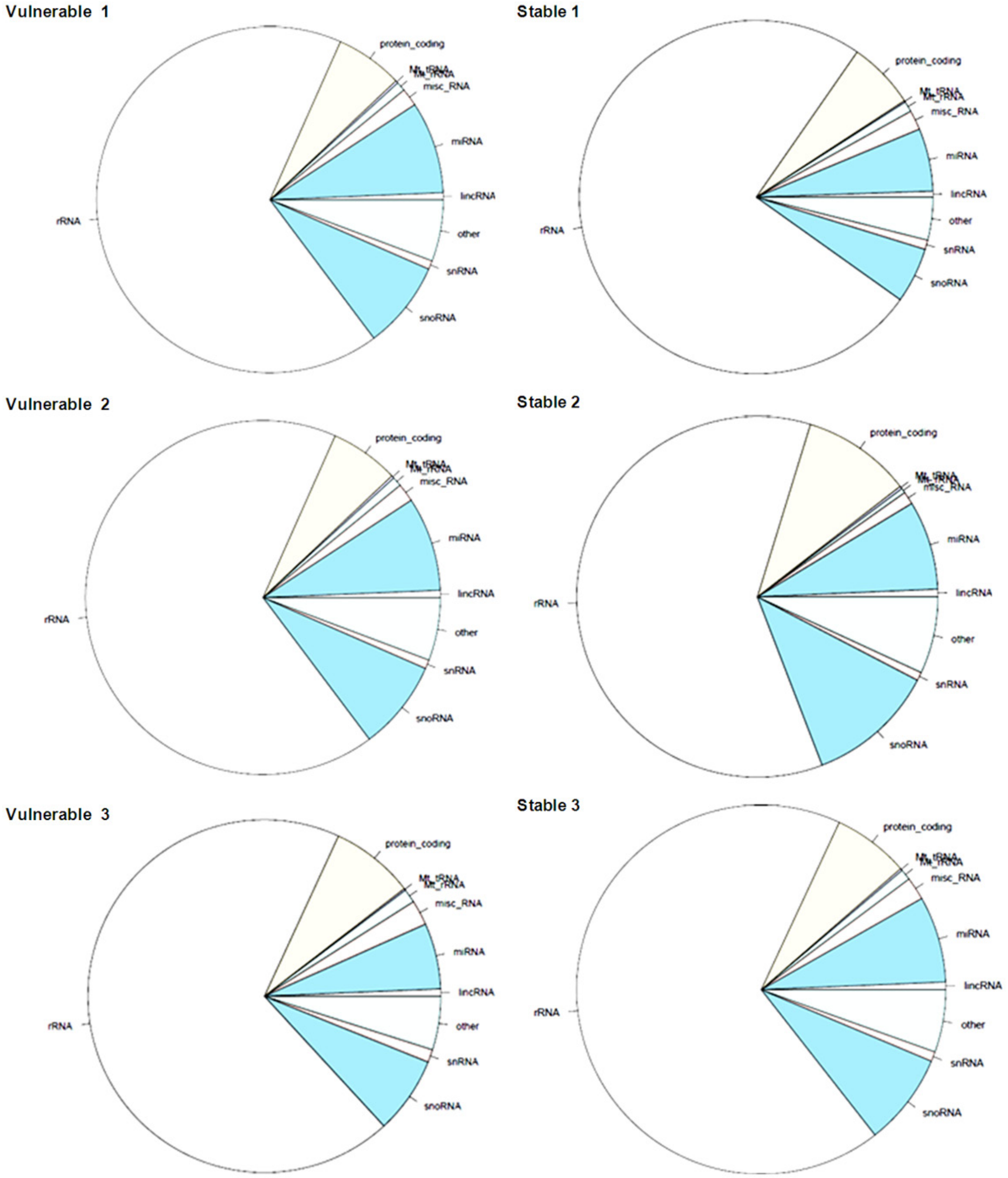

**Figure 5.** The sRNA length distribution. The length distribution of tags and reads were showed in Figure. The Figure indicated that there was no significant difference in length distribution of tags and reads between vulnerable and stable plaques.

carotid atherosclerotic plaque, or 5,598,305, 7,249,629, and 5,623,320 qualified Illumina reads from stable carotid atherosclerotic plaque, respectively (**Table 3**). The statistical analysis demonstrated that the average quality of every sample was larger than 20, and indicated that the quality was very fine (**Table 4** and **Figure 2**).

### *Pre-processing of NGS data of plaques*

Firstly, the 3'-adapter sequence (AGATCGGAAGAGCACACGTCT) was identified from raw data of NGS (**Table 5**) by fastx_cliper program. Then we used fastq_quality_filter program to remove the reads of lower quality, so that the short reads was obtained (**Table 5**) and the quality

**Table 6.** Genome mapped tags and reads

| Sample | Processed Reads | Alignment ($1 \le m^d \le 5$) | Failed | Alignment suppressed ($m^d > 5$) |
|---|---|---|---|---|
| Vulnerable plaque 1 | 4,363,223 | 2,449,906 (56.15) | 440,633 (10.10) | 1,472,684 (33.75) |
| Vulnerable plaque 2 | 3,731,781 | 2,042,342 (54.73) | 337,887 (9.05) | 1,351,552 (36.22) |
| Vulnerable plaque 3 | 2,202,752 | 1,405,721 (63.82) | 245,822 (11.16) | 551,209 (25.02) |
| Stable plaque 1 | 1,141,645 | 653,446 (57.24) | 330,057 (28.91) | 158,142 (13.85) |
| Stable plaque 2 | 3,045,532 | 2,229,938 (73.22) | 414,763 (13.62) | 400,831 (13.16) |
| Stable plaque 3 | 2,260,972 | 1,455,027 (64.35) | 242,198 (10.71) | 563,747 (13.16) |

Note: The bowtie parameter (-m=5, suppress all alignments if > 5 exist).

score of at least 95% base was not low than 20 (**Figure 3**). Then the cluster was performed to mark all identical sequences as one tag. The results showed that the length of all the filtered mature transcriptome sequences distributed range from 17-30 nt, and then sequences in the range was analyzed follow up (**Figure 4**). The length distribution of tags and reads were showed in **Figure 5**.

### *Sequencing data analysis of small RNA with rfam*

All the marked tags were mapped onto sequence of *Homo sapiens* genome in PubMed, by the match software bowtie 0.12.7. **Table 6** displayed the mapped tags and reads computed on account of 17~30 nt reads or tags. **Figure 5** showed the genome-mapped rates of samples. Then we used the database Rfam 11.0 of RNA family to analyse sundry RNAs in the samples.

### *Mirdeep analysis of transcriptome and different transcriptome between the two groups with edgeR*

The Mirdeep (version 2) was utilized to forecast the transcriptome of these samples. The differential transcriptome were analyzed with edgeR software for differential expression analysis (**Table 7**). Then the major different transcriptome between Vulnerable group and Stable control group were analyzed. Comparative trancriptome analysis differentially revealed 781 genes between vulnerable and stable carotid atherosclerotic plaque, 318 expression of genes was up-regulated and 363 expressions of genes were down-regulated. The volcano plot was used to further identify differential transcriptome expression between the two groups (**Figure 6**). Moreover, the cluster analysis was performed (**Figure 7**). Furthermore, the gene ontology (GO) consortium and data was used to analyze the targets gene of different transcriptome (**Figure 8**).

## Discussion

The absolute quantification of next generation sequencing (NGS) is just modestly accurate. The advance of high-throughput sequencing was driven by the high demand of low-cost sequencing, which is known as NGS. Thousands of sequences concurrently manufactured in NGS process. In recent years, the computational analysis of genome-wide scale is increasingly functioned as a backbone to facilitate more novel biomedical discovery. However, owing to the exponential increase of the quantities of sequence data, the analysis bottleneck remains yet to be solved.

In this study, we performed a transcriptome sequencing analysis of the plasma transcriptome of vulnerable and stable plaque by NGS, using the Illumina Deep Sequencing technology. We obtained 5,473,799, 3,766,914, 7,713,565 qualified Illumina reads from three vulnerable carotid atherosclerotic plaque, or 5,598,305, 7,249,629 and 5,623,320 qualified Illumina reads from stable carotid atherosclerotic plaque, respectively. The average quality of more than 99 percent of reads was larger than 20, indicative of good quality of these data. Then we used the fastx cliper program to perform data pre-processing of NGS, and removed the lower quality reads using fastq quality filter program. Subsequently, cluster was carried out; the distribution of the length of mature transcriptome sequence ranged from 18-30 nt and mapped onto the sequence of *Homo sapiens* genome. Then we used the Miranda software to predict the target

**Table 7.** The data of reads by Rfam analysis

| | Vulnerable 1 | Vulnerable 2 | Vulnerable 3 | Stable 1 | Stable 2 | Stable 3 |
|---|---|---|---|---|---|---|
| MiRNA | 579912 | 6931 | 77782 | 76903 | 72272 | 115245 |
| Misc_RNA | 35382 | 3154 | 15126 | 29170 | 22500 | 6023 |
| Protein_coding | 36531 | 10850 | 56657 | 97571 | 77146 | 27970 |
| Mt_tRNA | 9148 | 92 | 2216 | 2223 | 1873 | 1164 |
| rRNA | 203083 | 41636 | 603626 | 874030 | 939094 | 81202 |
| Mt_rRNA | 1435 | 567 | 7832 | 16153 | 12048 | 806 |
| snoRNA | 88520 | 4091 | 74057 | 89938 | 63807 | 39126 |
| Processed_transcript | 48805 | 2199 | 36374 | 41951 | 31649 | 20100 |
| Processed_pseudogene | 2485 | 1250 | 9586 | 11155 | 9933 | 1800 |
| lincRNA | 3896 | 1107 | 5863 | 8388 | 6776 | 2448 |
| snRNA | 7313 | 725 | 7364 | 14921 | 10793 | 2352 |
| Antisense | 2519 | 477 | 3320 | 4763 | 3860 | 1219 |
| Unprocessed_pseudogene | 722 | 123 | 722 | 1307 | 913 | 247 |
| Transcribed_unprocessed_pseudogene | 231 | 83 | 407 | 775 | 538 | 165 |
| TEC | 207 | 63 | 184 | 443 | 295 | 152 |
| Sense_intronic | 336 | 33 | 279 | 419 | 306 | 117 |
| Sense_overlapping | 103 | 56 | 175 | 196 | 169 | 172 |
| Transcribed_processed_pseudogene | 85 | 51 | 275 | 445 | 332 | 64 |
| Unitary_pseudogene | 39 | 15 | 22 | 39 | 37 | 18 |
| Known_ncrna | 23 | 6 | 28 | 61 | 26 | 1 |
| IG_C_gene | 92 | 44 | 252 | 567 | 610 | 144 |
| IG_V_gene | 53 | 18 | 148 | 282 | 336 | 100 |

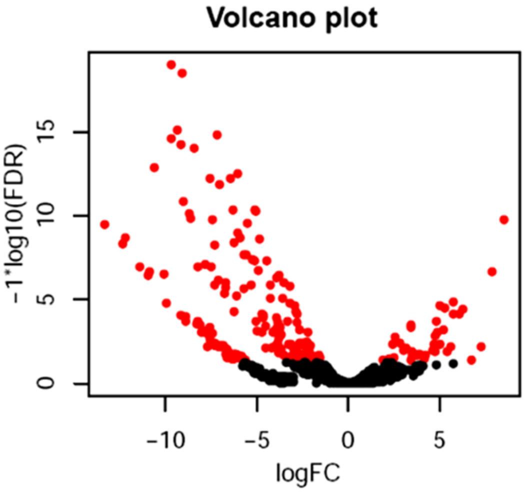

**Figure 6.** The volcano plot. The volcano plot was used to further identify differential transcriptome expression between the two groups. The Figure indicated that there was significant difference in transcriptome expression between vulnerable (red points) and stable plaques (black points).

genes of transcriptome (score > 150, energy < -15).

Rfam is a database of RNA families, and known as a collection of multiple sequence alignments and covariance models. It is available on the US and the UK website. These websites can enable the users to probe a query sequence against the library of covariance models, and to browse family annotation and multiple sequence alignments. The users could download the database in flat file form and search locally by the INFERNAL package (http://infernal.wustl.edu/).

Comparative transcriptome analysis revealed those 781 transcriptome differentially between vulnerable and stable carotid atherosclerotic plaque, including 318 expression of transcriptome was up-regulated and 363 expression of transcriptome were down-regulated. The volcano plot was used to further identify differential transcriptome expression between the two groups. Moreover, the cluster analysis was performed, and the spearman correlation coefficient was calculated and analyzed. Furthermore, the gene ontology (GO) consortium and data was used to analyze the targets gene of differential transcriptome.

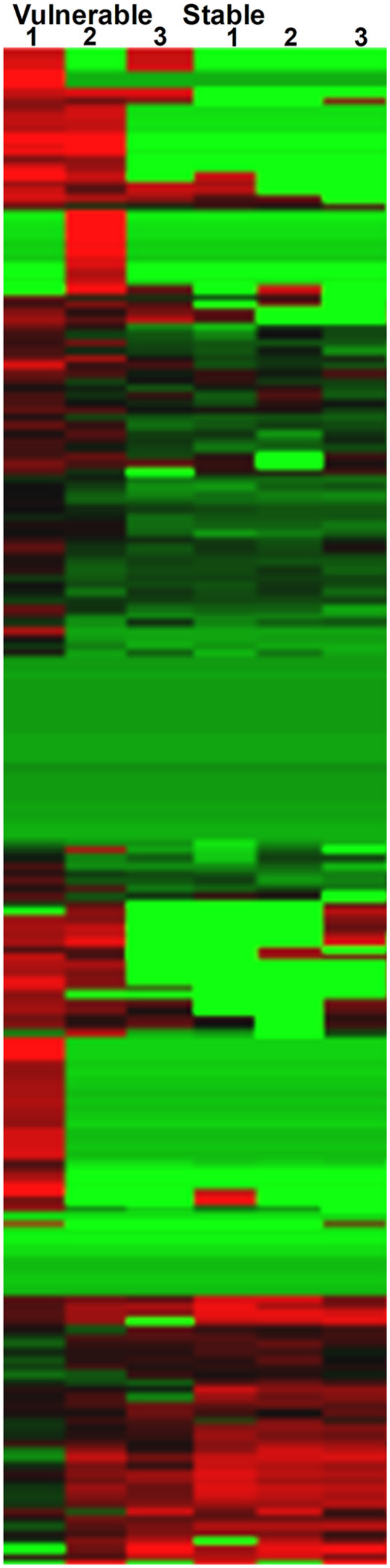

**Figure 7.** The cluster analysis. The cluster analysis was performed. The Figure indicated that there was significant difference in trancriptome expression between vulnerable and stable plaques.

The edger software, an achievement of methology designed by Smyth and Robinson [58, 59], is a bio-conductor package for analysing differential gene expression [60]. The software and methods can also be applied to emerging technologies like RNA-seq [61, 62] to obtain digital expression data. Statistically, a volcano plot can quickly examine differences in large datasets consisted of replicate data [63]. It displays fold-change and significance on the x- and y-axes, respectively. It can combine a measurement of statistical significance from a statistical test (e.g., a *p*-value of ANOVA test) with the magnitude of the change enabling quick visual identification of those data-points (genes, etc.) that display statistically significant large-magnitude changes.

Plenty of studies about transcriptome aberrances primarily focused on the analysis of canonical, reference transcriptome, while isomiRs required more advanced technologies, like NGS and laborious analysis of all the acquired data. Hence, our study employed NGS to acquire transcriptome data come from between pediatric patients and healthy controls by next generation sequencing techniques. Numbers of transcriptome will be pivotal for revealing the molecular regulatory mechanisms of expression relevant to the pathogenesis and pathophysiology of vulnerable carotid atherosclerotic plaque.

Vulnerable plagues have been defined as precursors to lesions that rupture. However, coronary thrombosis may occur from other lesions like plaque erosion and calcified nodules, although to a lesser frequency than rupture. Therefore, the definition of vulnerable plaque should be all-inclusive. Using descriptive terminology, the researches define the precursor lesion of plaque rupture as “thin-cap fibroatheroma” (TCFA). Morphologically, TCFAs have a necrotic core with an overlying thin fibrous cap (< 65 mm) consisting of collagen type I, which is infiltrated by macrophages. These lesions are most frequent in the coronary tree of patients dying with acme myocardial infarction and least common in those with plaque erosion. TCFAs are more common in patients with high serum total cholesterol (TC) and a high TC to high density cholesterol ratio, in women > 50 years, and in those patients with

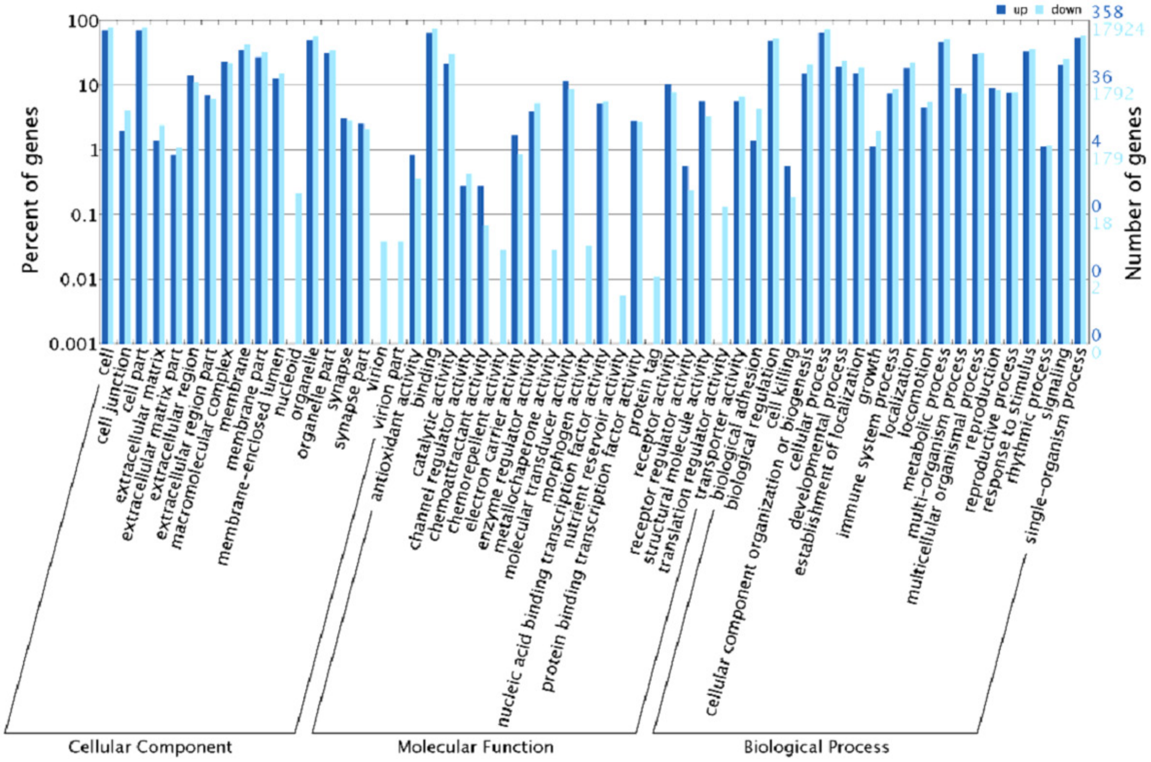

**Figure 8.** The gene ontology (GO) consortium and data. The gene ontology (GO) consortium and data was used to analyze the targets gene of different transcriptome. The Figure indicated that there was significant expression difference of genes in signaling pathway related to the sequencing transcriptome between vulnerable and stable plaques.

elevated levels of high sensitivity C-reactive protein. TCFAs are mostly found in the proximal left anterior descending coronary arteries and less commonly in the proximal right or the proximal left circumflex coronary arteries. In TCFAs, necrotic core length is ~ 2-17 mm (mean 8 mm) and the underlying cross-sectional luminal narrowing in over 75% of cases is < 75% (< 50% diameter stenosis). The area of the necrotic core in at least 75% of cases is ≤ 3 mm$^2$. Clinical studies of TCFAs are limited as angiography and intravascular ultrasound (TVUS) catheters cannot precisely identify these lesions. Newer catheters and other techniques are at various stages of development and will play a significant role in the understanding of plaque progression and the development of symptomatic coronary artery disease [64].

In our work, the demographic and clinical features of patients in the two groups were compared, including age, sex, BMI, SP, DP, TG, CHO, HDL, LDL, HCY. Moreover, it did not show significantly different of these characteristics between symptomatic group with vulnerable plaque and asymptomatic group with stable plaque. Furthermore, the carotid arterial characteristic of patients determined by ultrasound were compared. Our results indicated that there were stroke history in symptomatic group, and obviously more than asymptomatic patients. The carotid artery diameter in asymptomatic group was significantly lower than symptomatic group (0.12±0.03 vs. 0.19±0.04; *P*=0.021) (**Table 2**).

In our study, all patients received CEA operation. The typical plaque of vulnerable and stable were showed in **Figure 1A**. t demonstrated that the vulnerable plaque looks like fragile and uncomplete. As shown in **Figure 1B**, HE staining demonstrated that the amounts of new vessels in symptomatic patients with vulnerable plaque were much more than those in asymptomatic patients with stable plaque.

Ischemic strokes and transient ischemic attacks (TIAs) are frequently caused by cerebral embolism from an atherothrombotic plaque or thrombosis at the site of plaque rupture.

Although the degree of lumen obstruction is a relevant marker of the risk of stroke, the recognition of the role of the vulnerable plaque has opened new avenues in the field of atherothrombotic stroke. The vulnerability is dictated in part by plaque morphology, which, in turn, is influenced by pathophysiologic mechanisms at the cellular and molecular level. A multimodal assessment of plaque vulnerability involving the combination of systemic markers, new imaging methods that target inflammatory and thrombotic components, and the potential of emerging therapies may lead to a new stratification system for atherothrombotic risk and to a better prevention of atherothrombotic stroke [65].

In conclusion, our work demonstrated differential transcriptomes between vulnerable and stable carotid atherosclerotic plaque. Numbers of genes will serve as a promising resource for revealing the regulatory molecular mechanisms of expression associated with the pathophysiology and pathogenesis of vulnerable atherosclerosis plaque, even their implications in the field of therapy.

## Limitations

The major limitation was the small samples due to the complicated technical and expensive costs, which might introduce certain bias. The expression levels of transcriptome will be detected between vulnerable and stable carotid atherosclerotic plaque basing on the results of this research later.

## Acknowledgements

This work was supported by grants from National Natural Science Foundation of China (No. 30600524, No. 81200966 and No. 815-00365), and Beijing Natural Science Foundation (No. 7172040), and Shanxi Province Social Development Science and Technology Attack Project (No. 2016SF-132).

## Disclosure of conflict of interest

None.

**Address correspondence to:** Chengxiong Gu, Department of Cardiothoracic Surgery, Beijing Anzhen Hospital, Capital Medical University, No. 2 Anzhen Road, Chaoyang District, Beijing 100045, China. Tel: 86-010-64412431; 86-010-64432606; E-mail: anzhengu@sina.com

## References

[1] Barquera S, Pedroza-Tobias A, Medina C, Hernandez-Barrera L, Bibbins-Domingo K, Lozano R and Moran AE. Global overview of the epidemiology of atherosclerotic cardiovascular disease. Arch Med Res 2015; 46: 328-338.

[2] Hansson GK and Libby P. The immune response in atherosclerosis: a double-edged sword. Nat Rev Immunol 2006; 6: 508-519.

[3] Zaid M, Fujiyoshi A, Kadota A, Abbott RD and Miura K. Coronary artery calcium and carotid artery intima media thickness and plaque: clinical use in need of clarification. J Atheroscler Thromb 2017; 24: 227-239.

[4] Woo SY, Joh JH, Han SA and Park HC. Prevalence and risk factors for atherosclerotic carotid stenosis and plaque: a population-based screening study. Medicine (Baltimore) 2017; 96: e5999.

[5] Fairhead JF and Rothwell PM. The need for urgency in identification and treatment of symptomatic carotid stenosis is already established. Cerebrovasc Dis 2005; 19: 355-358.

[6] Bonati LH and Nederkoorn PJ. Clinical perspective of carotid plaque imaging. Neuroimaging Clin N Am 2016; 26: 175-182.

[7] Saba L, Anzidei M, Marincola BC, Piga M, Raz E, Bassareo PP, Napoli A, Mannelli L, Catalano C and Wintermark M. Imaging of the carotid artery vulnerable plaque. Cardiovasc Intervent Radiol 2014; 37: 572-585.

[8] Adams HP Jr, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL and Marsh EE 3rd. Classification of subtype of acute ischemic stroke. definitions for use in a multicenter clinical trial. TOAST. Trial of org 10172 in acute stroke treatment. Stroke 1993; 24: 35-41.

[9] Saba L, Anzidei M, Sanfilippo R, Montisci R, Lucatelli P, Catalano C, Passariello R and Mallarini G. Imaging of the carotid artery. Atherosclerosis 2012; 220: 294-309.

[10] Saba L, Caddeo G, Sanfilippo R, Montisci R and Mallarini G. CT and ultrasound in the study of ulcerated carotid plaque compared with surgical results: potentialities and advantages of multidetector row CT angiography. AJNR Am J Neuroradiol 2007; 28: 1061-1066.

[11] Grant EG, Benson CB, Moneta GL, Alexandrov AV, Baker JD, Bluth EI, Carroll BA, Eliasziw M, Gocke J, Hertzberg BS, Katanick S, Needleman L, Pellerito J, Polak JF, Rholl KS, Wooster DL and Zierler RE. Carotid artery stenosis: grayscale and Doppler US diagnosis-society of radiologists in ultrasound consensus conference. Radiology 2003; 229: 340-346.

[12] Griffith M, Walker JR, Spies NC, Ainscough BJ and Griffith OL. Informatics for RNA sequencing: a web resource for analysis on the cloud. PLoS Comput Biol 2015; 11: e1004393.

[13] Wang Z, Gerstein M and Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 2009; 10: 57-63.

[14] Morin R, Bainbridge M, Fejes A, Hirst M, Krzywinski M, Pugh T, McDonald H, Varhol R, Jones S and Marra M. Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. BioTechniques 2008; 45: 81-94.

[15] Mortazavi A, Williams BA, McCue K, Schaeffer L and Wold B. Mapping and quantifying mammalian transcriptomes by RNA-seq. Nat Methods 2008; 5: 621-628.

[16] Chen EA, Souaiaia T, Herstein JS, Evgrafov OV, Spitsyna VN, Rebolini DF, Knowles JA. Effect of RNA integrity on uniquely mapped reads in RNA-Seq. BMC Research Notes 2014; 7: 753.

[17] Liu D and Graber JH. Quantitative comparison of EST libraries requires compensation for systematic biases in cDNA generation. BMC Bioinformatics 2006; 7: 77.

[18] Zhao QY, Wang Y, Kong YM, Luo D, Li X and Hao P. Optimizing de novo transcriptome assembly from short-read RNA-Seq data: a comparative study. BMC Bioinformatics 2011; 12: S2.

[19] Zerbino DR and Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 2008; 18: 821-829.

[20] Oases: a de novo transcriptome assembler for very short reads.

[21] Chang Z, Li G, Liu J, Zhang Y, Ashby C, Liu D, Cramer CL and Huang X. Bridger: a new framework for de novo transcriptome assembly using RNA-seq data. Genome Biol 2015; 16: 30.

[22] Li B, Fillmore N, Bai Y, Collins M, Thomson JA, Stewart R and Dewey CN. Evaluation of de novo transcriptome assemblies from RNA-Seq data. Genome Biol 2014; 15: 553.

[23] Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M and Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013; 29: 15-21.

[24] Langmead B, Trapnell C, Pop M and Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 2009; 10: R25.

[25] Trapnell C, Pachter L and Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 2009; 25: 1105-1111.

[26] Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL and Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. Nat Protoc 2012; 7: 562-578.

[27] Liao Y, Smyth GK and Shi W. The subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Research 2013; 41: e108.

[28] Patro R, Mount SM and Kingsford C. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. Nat Biotechnol 2014; 32: 462-464.

[29] Bray NL, Pimentel H, Melsted P and Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 2016; 34: 525-527.

[30] Wu TD and Watanabe CK. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics (Oxford, England) 2005; 21: 185918-185975.

[31] Lu B, Zeng Z and Shi T. Comparative study of de novo assembly and genome-guided assembly strategies for transcriptome reconstruction based on RNA-Seq. Sci China Life Sci 2013; 56: 143-155.

[32] Bradnam KR, Fass JN, Alexandrov A. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. Gigascience 2013; 2: 10.

[33] Li H, Lovci MT, Kwon YS, Rosenfeld MG, Fu XD and Yeo GW. Determination of tag density required for digital transcriptome analysis: application to an androgen-sensitive prostate cancer model. Proc Natl Acad Sci U S A 2008; 105: 20179-20184.

[34] Zhang ZH, Jhaveri DJ, Marshall VM, Bauer DC, Edson J, Narayanan RK, Robinson GJ, Lundberg AE, Bartlett PF, Wray NR and Zhao QY. A comparative study of techniques for differential expression analysis on RNA-Seq data". PLoS One 2014; 9: e103207.

[35] Anders S, Pyl PT and Huber W. HTSeq--a python framework to work with high-throughput sequencing data. Bioinformatics 2015; 31: 166-169.

[36] Liao Y, Smyth GK and Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 2014; 30: 923-930.

[37] Schmid MW and Grossniklaus U. Rcount: simple and flexible RNA-Seq read counting. Bioinformatics 2015; 31: 436-437.

[38] Finotello F, Lavezzo E, Bianco L, Barzon L, Mazzon P, Fontana P, Toppo S and Di Camillo B. Reducing bias in RNA sequencing data: a novel approach to compute counts. BMC Bioinformatics 2014; 15: S7.

[39] Hashimoto TB, Edwards MD and Gifford DK. Universal count correction for high-throughput sequencing. PLoS Comput Biol 2014; 10: e1003494.

[40] Robinson MD and Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 2010; 11: R25.

[41] Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ and Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 2010; 28: 511-515.

[42] Law CW, Chen YS, Shi W and Smyth GK. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol 2014; 15: R29.

[43] Anders S and Huber W. Differential expression analysis for sequence count data. Genome Biol 2010; 11: R106.

[44] Robinson MD, McCarthy DJ and Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics (Oxford, England) 2010; 26: 139-140.

[45] Soneson C and Delorenzi M. A comparison of methods for differential expression analysis of RNA-seq data. BMC Bioinformatics 2013; 14: 91.

[46] Jiangfeng F, Yuzhu L, Sijiu Y, Yan C, Gengquan X, Libin W, Yangyang P and Honghong H. Transcriptional profiling of two different physiological states of the yak mammary gland using RNA sequencing. PLoS One 2018; 13: e0201628.

[47] Zacher B, Abnaof K, Gade S, Younesi E, Tresch A, Fröhlich H. Joint Bayesian inference of condition-specific miRNA and transcription factor activities from combined gene and microRNA expression data. Bioinformatics 2012; 28: 1714-1720.

[48] Bioconductor-Open source software for Bioinformatics.

[49] Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T, Gottardo R, Hahne F, Hansen KD, Irizarry RA, Lawrence M, Love MI, MacDonald J, Obenchain V, Oleś AK, Pagès H, Reyes A, Shannon P, Smyth GK, Tenenbaum D, Waldron L and Morgan M. Orchestrating high-throughput genomic analysis with bioconductor. Nat Methods 2015; 12: 115-121.

[50] Mortazavi A, Williams BA, McCue K, Schaeffer L and Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 2008; 5: 621-628.

[51] Marguerat S, Schmidt A, Codlin S, Chen W, Aebersold R and Bähler J. Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. Cell 2012; 151: 671-683.

[52] Owens ND, Blitz IL, Lane MA, Patrushev I, Overton JD, Gilchrist MJ, Cho KW and Khokha MK. Measuring absolute RNA copy numbers at high temporal resolution reveals transcriptome kinetics in development. Cell Rep 2016; 14: 632-647.

[53] Marcotte EM, Pellegrini M, Thompson MJ, Yeates TO and Eisenberg D. A combined algorithm for genome-wide prediction of protein function. Nature 1999; 402: 83-86.

[54] Giorgi FM. Comparative study of RNA-seq- and Microarray-derived coexpression networks in Arabidopsis thaliana. Bioinformatics 2013; 29: 717-724.

[55] Iancu OD. Utilizing RNA-Seq data for de novo coexpression network inference. Bioinformatics 2012; 28: 1592-1597.

[56] Eksi R, Li HD, Menon R, Wen Y, Omenn GS, Kretzler M and Guan Y. Systematically differentiating functions for alternatively spliced isoforms through integrating RNA-seq data. PLoS Comput Biol 2013; 9: e1003314.

[57] Li HD, Menon R, Omenn GS and Guan Y. The emerging era of genomic data integration for analyzing splice isoform function. Trends Genet 2014; 30: 340-347.

[58] Robinson MD and Smyth GK. Moderated statistical tests for assessing differences in tag abundance. Bioinformatics 2007; 23: 2881-2887.

[59] Robinson MD and Smyth GK. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. Biostatistics 2008; 9: 321-332.

[60] Fröhlich H. biRte: Bayesian inference of context-specific regulator activities and transcriptional networks. Bioinformatics 2015; 31: 3290-3298.

[61] Li H, Lovci MT, Kwon YS, Rosenfeld MG, Fu XD and Yeo GW. Determination of tag density required for digital transcriptome analysis: application to an androgen-sensitive prostate cancer model. Proc Natl Acad Sci U S A 2008; 105: 20179-20184.

[62] Marioni JC, Mason CE, Mane SM, Stephens M and Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res 2008; 18: 1509-1517.

[63] Cui X and Churchill GA. Statistical tests for differential expression in cDNA microarray experiments. Genome Biol 2003; 4: 210.

[64] Virmani R, Burke AP, Kolodgie FD and Farb A. Vulnerable plaque: the pathology of unstable coronary lesions. J Interv Cardiol 2002; 15: 439-446.

[65] Hermus L, Tielliu IF, Wallis de Vries BM, van den Dungen JJ and Zeebregts CJ. Imaging the vulnerable carotid artery plaque. Acta Chir Belg 2010; 2: 159-64.