BMC Biotechnology

METHODOLOGY ARTICLE   **Open Access**

CrossMark

Evaluating whole transcriptome amplification for gene profiling experiments using RNA-Seq

Sheena L Faherty*, C Ryan Campbell, Peter A Larsen and Anne D Yoder

Abstract

Background: RNA-Seq has enabled high-throughput gene expression profiling to provide insight into the functional link between genotype and phenotype. Low quantities of starting RNA can be a severe hindrance for studies that aim to utilize RNA-Seq. To mitigate this bottleneck, whole transcriptome amplification (WTA) technologies have been developed to generate sufficient sequencing targets from minute amounts of RNA. Successful WTA requires accurate replication of transcript abundance without the loss or distortion of specific mRNAs. Here, we test the efficacy of NuGEN's Ovation RNA-Seq V2 system, which uses linear isothermal amplification with a unique chimeric primer for amplification, using white adipose tissue from standard laboratory rats (*Rattus norvegicus*). Our goal was to investigate potential biological artifacts introduced through WTA approaches by establishing comparisons between matched raw and amplified RNA libraries derived from biological replicates.

Results: We found that 93% of expressed genes were identical between all unamplified versus matched amplified comparisons, also finding that gene density is similar across all comparisons. Our sequencing experiment and downstream bioinformatic analyses using the Tuxedo analysis pipeline resulted in the assembly of 25,543 high-quality transcripts. Libraries constructed from raw RNA and WTA samples averaged 15,298 and 15,253 expressed genes, respectively. Although significant differentially expressed genes ($P < 0.05$) were identified in all matched samples, each of these represents less than 0.15% of all shared genes for each comparison.

Conclusions: Transcriptome amplification is efficient at maintaining relative transcript frequencies with no significant bias when using this NuGEN linear isothermal amplification kit under ideal laboratory conditions as presented in this study. This methodology has broad applications, from clinical and diagnostic, to field-based studies when sample acquisition, or sample preservation, methods prove challenging.

Keywords: RNA-Seq, Whole transcriptome amplification, White adipose tissue, Rattus norvegicus

Background

RNA-Seq is quickly becoming the preferred method for comprehensively characterizing global transcriptome activity. This approach has emerged as a powerful tool for determining the link between genotype and phenotype given that the transcriptomes of specific tissue types and individual cells reflect functionality [1-4]. Monitoring changes in gene expression across thousands of genes simultaneously provides an untargeted view of the molecular workings that contribute to adaptive responses to extrinsic and intrinsic stimuli [5,6]. When conducting RNA-Seq studies, one potential obstacle encountered by researchers can result from suboptimal starting quantities

of RNA that the use of NGS platforms often requires. This can be especially problematic for certain tissue types wherein starting RNA is in low quantities, such a white adipose tissue, or when using minimally invasive techniques to acquire samples [7-10]. Many of these approaches yield considerably less than a microgram of total RNA [11-16] though current library preparation protocols are typically optimized for a minimum of one microgram of total RNA.

Technologies are emerging to circumvent these challenges. With whole transcriptome amplification (WTA), RNA can be synthetically enriched for gene profiling experiments with both research and diagnostic applications [12,17-19]. WTA amplifies the transcriptome, even in the face of low starting material and/or when samples are heavily degraded due to insufficient preservation [5].

* Correspondence: sheena.faherty@duke.edu
Department of Biology, Duke University, Durham, NC 27708, USA

Faherty *et al. BMC Biotechnology* (2015) 15:65

Page 2 of 10

The development and refinement of technologies for high fidelity RNA amplification, therefore, have tremendous appeal for numerous applications spanning the disciplines of field ecology to biomedical diagnostics [12,17,18].

An obvious concern when using WTA for sensitive RNA-Seq is that the replication of transcript abundance occurs without the loss or distortion of transcript expression. Using technologies such as quantitative polymerase chain reaction (qPCR) or microarrays to measure expression levels investigators have compared expression profiles from i.) amplification of varying amounts of starting total RNA [5,20,21]; ii.) amplified and unamplified RNA from the same source [5,14,22,23]; iii.) different commercially available kits or lab-based techniques [20,21,24-26]. These studies are valuable proxies for assessing amplification bias, but these technologies can be limited in their detection capabilities [12,14]. NGS approaches have greater resolution over qPCR and microarrays, and can detect changes in gene expression on a finer scale with digital precision in which one digital unit represents a single mapped sequence read [3,27]. It is therefore critical to assess the accuracy of transcript amplification with the highest resolution technology, particularly when investigators are choosing NGS over other techniques in current experimental methodologies.

Commercially available kits are becoming increasingly prevalent for seamless incorporation of amplified RNA into the NGS pipeline. These are offered by companies such as Clontech, Sigma, Miltenyi Biotec, and NuGEN, and use a variety of techniques to amplify RNA. When considering WTA for RNA-Seq studies, it is important to carefully consider the differences in chemistry and approaches that each commercial kit utilizes. Variations in priming strategies, cDNA synthesis, and amplification of newly converted cDNA can have impacts on length of cDNA products, 3′ bias, amplification efficiency, and fidelity of maintaining relative transcript abundance [1]. All of these factors can potentially introduce unwanted bias in expression studies using RNA-Seq. A researcher limited to using samples collected from the field or from heavily degraded clinical specimens, for example, might consider choosing a kit that uses both oligo (dT) and random hexamers primers for reverse transcription. This might alleviate the potential mis-amplification due to the loss of the poly-A tail from RNA degradation. NuGEN's Ovation RNA-Seq V2 kit is based on the linear isothermal amplification of double-stranded cDNA that encompasses a unique RNA/DNA heteroduplex at one end using the RNA-dependent DNA polymerase activity [5]. Amplification is initiated at the 3′ end as well as randomly throughout the whole transcriptome in the sample, using both oligoDT and random hexamer primers. This approach will putatively yield a uniform and accurate representation of the transcriptome, an assumption that we directly test in this study.

As commercially available WTA kits continue to improve, investigators should continually reassess their efficiency and compatibility with new sequencing platforms, to determine the best practices as methods evolve. In previous work, Tariq et al. (2011) performed an exploration of the efficacy of the NuGEN Ovation RNA-Seq system by using two different library preparation protocols to reduce ribosomal RNA contamination (i.e. poly-A enrichment and rRNA depletion), and ran their sequencing experiment on two separate sequencing platforms, the SOLiD platform and the Illumina Genome Analyzer IIX, to assess sequencing fidelity and platform-specific biases. Notably, this study attempted to identify the best combination of approaches (i.e. which combination of library preparation protocols and sequencing platforms) will result in the largest amount of high quality data generated from a WTA sample for RNA-Seq characterization. As comprehensive as this investigation was, the authors neglected to determine how WTA technologies, such as the Ovation kit, influence gene expression profiles. Their investigation was instead limited to evaluating differential gene expression between the two sequencing platforms. In addition, since the publication of that study, NuGEN has released an enhanced version of their Ovation kit and Illumina sequencing technology has become more advanced with the advent of their HiSeq platform. In particular, the HiSeq boasts lower error rates, better precision, and longer reads than the GAIIX [7]. Moreover, Shanker et al. (2015) recently conducted a broad-scale investigation of the reproducibility of four commercially-available kits, including NuGEN's Ovation RNA-Seq V2 system, at multiple laboratories using three varying amounts of starting concentrations on the Illumina HiSeq platform. They find that NuGEN performs remarkably well across multiple laboratories and across technical replicates as low as 500 pg [28]. With RNA-Seq data, however, the largest variability is known to come from biological stochasticity rather than from technical inaccuracies [29,30].

It remains unknown whether WTA preferentially amplifies specific gene products to the exclusion of others. Therefore, our study integrated the design of matched amplified and unamplified RNA from six biological replicates to empirically test this question using NGS transcriptomic sequencing on the Illumina HiSeq 2000 platform, which has become the gold standard for next-generation sequencing experiments [11,13,15,16]. In the present study, we aim to investigate the degree to which NuGEN's Ovation RNA-Seq V2 system influences gene expression profiles in white adipose tissue excised from standard laboratory rats (*Rattus norvegicus*), building upon and enhancing knowledge regarding current WTA

Faherty *et al. BMC Biotechnology* (2015) 15:65

Page 3 of 10

methodologies. To that end, we used the Tuxedo analysis pipeline [2-4,31] to assemble our RNA-Seq reads using the *R. norvegicus* genome, in addition to providing annotation and an exploration of differentially expressed genes. Our study was designed to sequence samples at the greatest depth possible, given the resources at hand, in order to identify any subtle differences between raw RNA and amplified RNA in biological replicates. Ultimately, our goal is to provide foundational insight for future studies to build upon regarding the utility of WTA for clinical and field-based studies that are investigating the functional link between genotype and phenotype via gene expression and RNA-Seq.

Results

We sequenced over 829 million paired-end reads, 2 ×100 base pairs in length (Lane 1 = 368,942,612, Lane 2 = 460,712,194 reads). Quality filtering resulted in the retention of approximately 90% and 86% of reads from Lanes 1 and 2, respectively (Table 1). Approximately 62% of all reads successfully mapped to the *R. norvegicus* genome (Table 1; Additional file 1: Table S1). Transcript assembly using Cufflinks resulted in 25,543 assembled transcripts.

We identified strong correlations in gene expression within each of the six pairwise comparisons of matched raw RNA and amplified RNA, and FPKM (fragments per kilobase of exon per million fragments mapped) values averaged slightly higher in libraries constructed from raw RNA samples (Figure 1). Transcripts with FPKM values greater than 0.05 were identified as being expressed (i.e. above the expected false discovery rate following Trapnell *et al.* 2010), and this threshold was used to determine the number of expressed genes for each sample. Libraries constructed from raw RNA and WTA samples averaged 15,298 and 15,253 expressed genes, respectively (Table 1; Additional file 1: Table S1). Approximately 93% of expressed genes were identical between each of the six matched raw RNA and WTA comparisons (Table 1, Figure 2) and gene density was similar across all samples (Figure 3). Using Cuffdiff, significant differentially expressed genes (P < 0.05) were

Table 1 Summary of sequencing results, quality filtering, and transcript assembly from pooled raw RNA samples vs. pooled WTA samples

	Raw RNA	WTA
Total sequenced reads	368,942,612	460,712,194
Filtered reads	334,297,804	395,807,422
Aligned reads	208,496,473	241,691,534
Uniquely aligned reads	173,011,711	221,886,287
Multiple aligned reads	35,484,762	19,805,247
Unmapped reads	121,884,321	152,473,973
Avg. expressed genes	15,298	15,253

identified in four of the six comparisons (Table 2), with each of these representing less than 0.15% of all shared genes for each comparison. This percentage is representative of 54 genes, in total, that show differences in expression levels when comparing unamplified vs. amplified. Of the 54 differentially expressed genes, 19 are identified as putative in the Ensembl database as pseudo-genes, while 35 represent functional protein-coding genes, and of those that are protein-coding, 33% percent have homology to human, while the others are putatively rat-specific.

As a complimentary bioinformatic approach to investigate potential differential gene expression, we used HT-Seq [32] to quantify reads by gene, and DESeq [33] to compare these quantities across biological replicates. We find less than 0.14% of genes expressed display differential expression when comparing raw vs. amplified RNA from biological replicates (Table 2). This percentage is representative of 31 genes that show differential expression between unamplified and amplified matched pairs.

We next assessed potential biases in transcript representation (i.e. coverage at 3′ and 5′ end) by determining the average coverage at each percentile of length from 3′ to 5′ end of the known transcripts (Figure 4). The coverage depth analysis at the extreme 3′ and 5′ ends of the transcripts confirms 3′ bias for all of our samples independent of WTA treatment and is likely an artifact of library creation, which were prepared using a poly-A-based enrichment method.
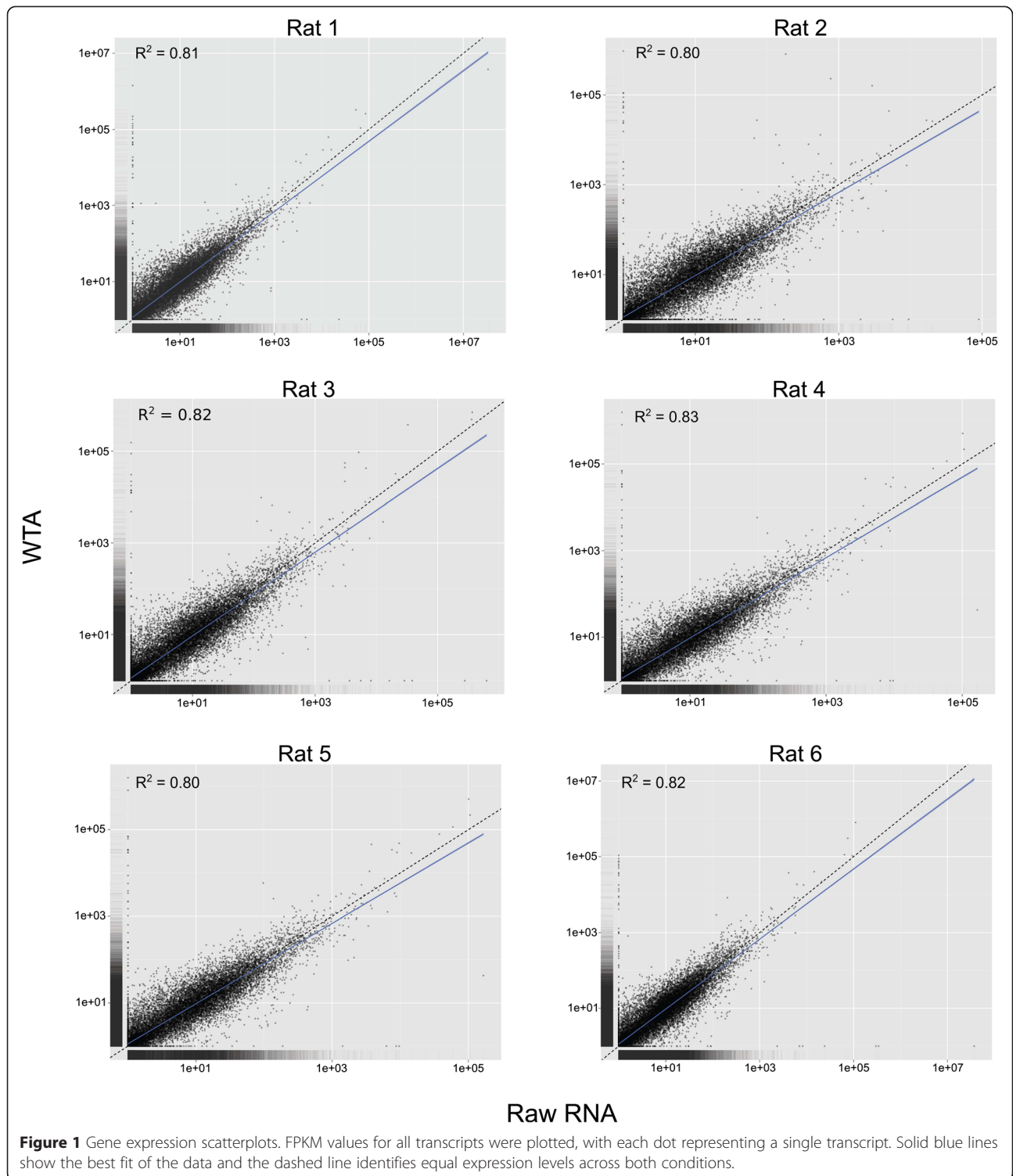
Discussion

RNA-Seq has enabled the sequencing of transcriptomes to allow both identification and quantification of transcript levels. The digital precision and sensitivity of RNA-Seq is well-suited to the analysis of low-input or degraded samples, especially with respect to older methods such as qPCR or microarrays, yet many available protocols require total RNA around 1.0 microgram, presenting a potential obstacle for clinical or field-based studies. Here, we have tested the efficacy of NuGEN's Ovation RNA-Seq V2 system for whole transcriptome amplification by comparing matched amplified and unamplified samples to identify if there exists any subtle changes in gene expression after WTA among biological replicates. We were specifically interested in investigating the degree to which accurate representations of gene diversity and transcript abundance levels were achieved when sequenced on an Illumina HiSeq.

Gene expression patterns in unamplified versus amplified libraries

We identified consistent patterns of gene presence and absence across each of the six comparisons, with an average overlap of approximately 93%. We found no evidence that
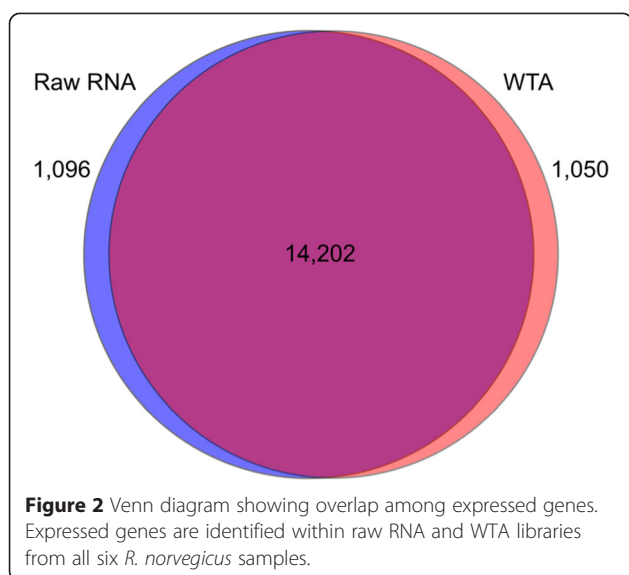
Faherty *et al. BMC Biotechnology* (2015) 15:65

Page 4 of 10

Figure 1 Gene expression scatterplots. FPKM values for all transcripts were plotted, with each dot representing a single transcript. Solid blue lines show the best fit of the data and the dashed line identifies equal expression levels across both conditions.

one or more specific gene products were differentially expressed in matched unamplified vs. amplified RNA in this study. Using Cuffdiff, the only gene found to be differentially expressed in as many as four of the six sample pairs was classified as a pseudo-gene and one protein-coding gene, heterogeneous nuclear ribonucleoprotein A0

(*Hnrnpa0*), a DNA binding protein that is thought to influence pre-mRNA processing, mRNA metabolism and transport, was found to be differentially expressed in three out of the six sample pairs. The remaining differentially expressed genes were only found in two or fewer comparisons, and no pairwise comparison shared more than five

Faherty *et al. BMC Biotechnology* (2015) 15:65

Page 5 of 10

Figure 2 Venn diagram showing overlap among expressed genes. Expressed genes are identified within raw RNA and WTA libraries from all six *R. norvegicus* samples.

genes. Our results from DESeq also indicate that there is no consistent biological distortion, as 48% (15 out of 31) of differentially expressed genes were found in five or fewer of our pairwise comparisons. This result is remarkably consistent with the differential expression results produced using Cuffdiff. Moreover, we observe no overlap with the genes that Cuffdiff identified as differentially expressed, indicative of the small proportion of differentially expressed genes being a product of background noise from stochastic events observed with typical RNA-Seq experiments. This result thus lends strong support to our hypothesis that there is no consistent biological distortion when using WTA.

Testing the efficacy of NuGEN's Ovation RNA-Seq V2 kit
Only a few studies have thus far investigated potential biases in gene expression introduced by WTA approaches using NuGEN's WTA technology [5,21,28,34]. These studies have been foundational in determining that the WTA chemistry tested herein performs well across technical replicates of varying amounts of starting material down to the limits of 50–100 pg. However, to our knowledge, no previous work has included biological replicates in their analyses. Thus, the strength of our investigation is that it incorporates the experimental design of matched amplified and unamplified RNA from six biological replicates to test for preferential amplification of WTA approaches. In the present study, we do not find that WTA selectively amplifies specific gene products and the small variability observed among replicates is most likely attributed to background noise. Biological replicates are essential for identifying potential distortion from gene expression stochasticity inherent in any biological system.

We chose NuGEN's Ovation kit based upon the single primer isothermal amplification (SPIA) technology for achieving WTA. This technology incorporates a combination of 3'-Ribo-SPIA (i.e. polydT) primers, which prime synthesis of cDNA at the 3' polyA tail, and whole transcript (WT)-Ribo-SPIA (i.e. random hexamer) primers to prime cDNA synthesis across the full length of the transcripts, providing amplification independent of sample degradation at the 3' end [5,8-10,35]. Ideally, this feature makes the Ovation RNA-Seq System V2 kit useful for NGS platforms, such as the Illumina HiSeq which function by distributing reads across the entire length of the transcript. Previous studies have also found that commercial kits show high reproducibility in WTA procedures completed by researchers in different laboratories (correlation coefficient of ~0.95), thus providing encouraging support for the use of WTA prior to sequencing when working with challenging samples [12,14,21].

Utility of whole transcriptome amplification for non-model species or field-based studies
Studies that investigate the relationship between genotype and phenotype have largely focused on model organisms for which a genome is publicly available, and have been performed using captive populations, or with cell culture methods under controlled laboratory settings for which sample material is relatively easy to obtain [12,17,18,36-46]. Although these studies provide a wealth of vital information on the drivers of physiology, behavior, and disease dynamics, comparable studies on non-model organisms or studies that use crucial pathological specimens (e.g. laser capture microdissection, FFPE, or biopsied tissue) or samples collected in the field remain scarce. This is due to the challenging nature of working with these sample types, often adding another layer of complexity (e.g. potential degradation) to the already present challenge of low RNA input [5,12,20,21,47-50]. Potential degradation is due to the fact that RNA is highly sensitive to endogenous and exogenous RN*ases*, which rapidly degrade RNA molecules at ambient conditions [5,14,22,23,51,52]. Degradation can result in immediate changes in expression profiles or a prospective loss of rare transcripts immediately after sample collection [20,21,24-26,53,54]. Although, this issue may be unavoidable for certain experiments, the priming strategy of the NuGEN kit may alleviate potential downstream issues when using partially degraded samples.

Conclusions
Our study builds upon and enhances existing knowledge regarding the NuGEN Ovation RNA-Seq V2 system's performance in studies that use WTA for RNA-Seq experiments. Previous studies have demonstrated that this
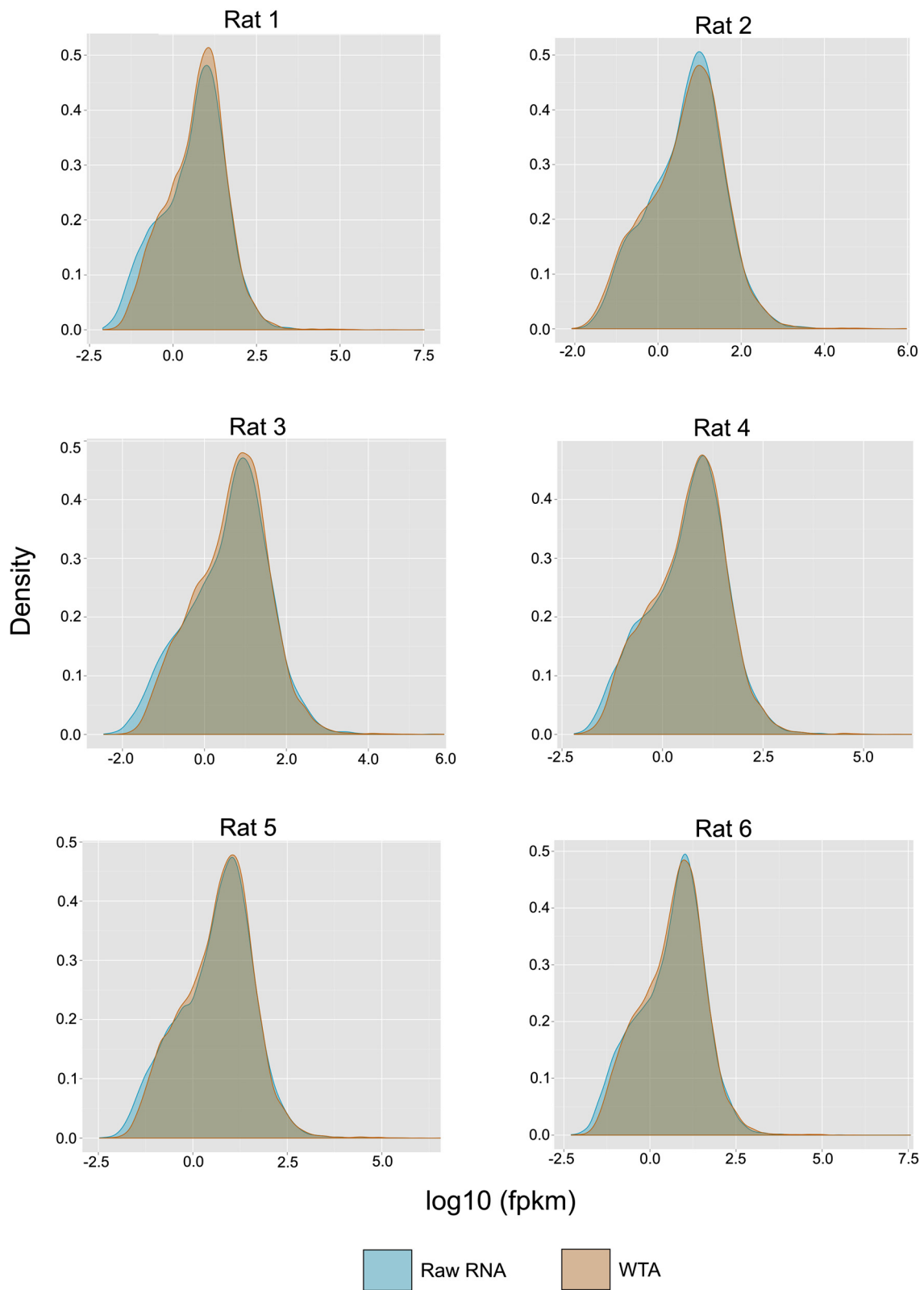
Faherty *et al. BMC Biotechnology* (2015) 15:65

Page 6 of 10

Figure 3 Distribution of gene expression levels across each sample/condition. Comparative distribution analysis of reads correlated to gene density in rat white adipose tissue.

Faherty et al. BMC Biotechnology (2015) 15:65

Page 7 of 10

Table 2 Number of significant differentially expressed genes (P < 0.05) per individual identified using Cuffdiff and DESeq

	Cuffdiff		DESeq	
	Differentially expressed genes (Raw RNA vs. WTA)	Percent of total shared genes between conditions	Differentially expressed genes (Raw RNA vs. WTA)	Percent of total shared genes between conditions
Rat 1	0	-	20	0.11%
Rat 2	16	0.11%	15	0.08%
Rat 3	19	0.14%	23	0.13%
Rat 4	18	0.13%	22	0.12%
Rat 5	0	-	22	0.12%
Rat 6	17	0.12%	25	0.14%

chemistry performs well across technical replicates of varying starting amounts of input RNA and across multiple laboratories. The novelty of our study is that it empirically tests biological variation, by incorporating the design of matched amplified and unamplified RNA from six biological replicates. The Ovation RNA-Seq V2 kit is able to amplify small quantities of RNA from tissue samples while maintaining transcript levels nearly identical to those from matched unamplified input RNA, demonstrating an overlap of 93% when comparing each matched pair individually. Despite this high fidelity of transcript representation, future studies should anticipate the slight discrepancy revealed here and take measures to account for the potential for small differences in downstream analyses. Additionally, future work could benefit from testing this methodology using more realistic scenarios, such as repeating the experiment with smaller input concentrations, or using partially degraded samples that may be more representative of RNA that may be extracted from clinical or from field-based studies. Nonetheless, this study provides the essential first step into investigations into the putative alteration of gene expression profiles due to WTA among biological replicates.

Methods

Tissue sampling

Animal handling was carried out in strict accordance with the approval of Duke University's Institutional Animal Care and Use Committee (IACUC protocol #A096-13-04). Gonadal white adipose tissue was excised from standard laboratory rats (*Rattus norvegicus*, n = 6) and snap frozen in liquid nitrogen. Samples were transferred to –80°C and stored there until sample processing.

RNA extraction

Approximately 100 mg of white adipose tissue was used for total RNA extraction. RNA was purified using an optimized TRIzol protocol in conjunction with the Microarray Tissue Kit (Qiagen; Valencia, CA, USA). Briefly, frozen tissues were homogenized in 1 ml TRI Reagent (Ambion; Grand Island, NY, USA) using a hand-held rotor-stator homogenizer to provide efficient disruption

and homogenization of samples. BCP (1-bromo-3-chloropropane) extraction was performed using 100 ul of BCP, followed by centrifugation at $12,000 \times g$ for 15 min at 4°C. Aqueous layer containing total RNA was transferred to a fresh 1.5 ml centrifuge tube and one volume of 70% EtOH was added to sample. The entire sample volume was loaded onto Qiagen filter columns and kit protocol was followed according to manufacturer's instructions, including an on-column DN*ase* step to remove residual contaminating DNA before amplification and downstream sequencing.

RNA integrity and concentration was assessed using the 2100 Bioanalyzer (Agilent Technologies; Santa Clara, CA, USA) at the Duke Institute for Genome Science and Policy's Microarray Facility. RNA integrity is provided by a RNA Integrity Number (RIN), which is calculated based on the comparison of the areas of 18S rRNA and 28 s rRNA [12,14,55]. Values range from 1 being the most degraded to 10 being the most intact. RIN values for our total RNA extractions ranged from 9.2 – 9.4, a measure of high quality (Additional file 2: Figure S1). Total RNA concentrations ranged from 200–300 ng/ul.

Whole transcriptome amplification

Total RNA extractions were divided into two tubes for each rat, for a total of 12 tubes. Six tubes were normalized to 100 ng/μl in a 5 μl volume, for a total of 500 ng of RNA, and subjected to whole transcriptome amplification using NuGEN's Ovation RNA-Seq V2 kit (San Carlos, CA, USA) according to manufacturer's instructions. This kit provides a rapid method for preparing amplified cDNA from total RNA for downstream RNA-Seq applications. It employs a single primer isothermal amplification (SPIA) method to amplify total RNA into double stranded cDNA and depletes rRNA without preselecting mRNA. Amplified cDNA samples were then purified using the MinElute Reaction Cleanup Kit (Qiagen; Valencia, CA, USA) according to manufacturer's protocol. After the amplification procedure, cDNA concentrations ranged from 421–575 ng/ul (Additional file 1: Table S2). Paired amplified cDNA (n = 6) and unamplified raw RNA samples (n = 6) were then sent to Duke GCB Genome

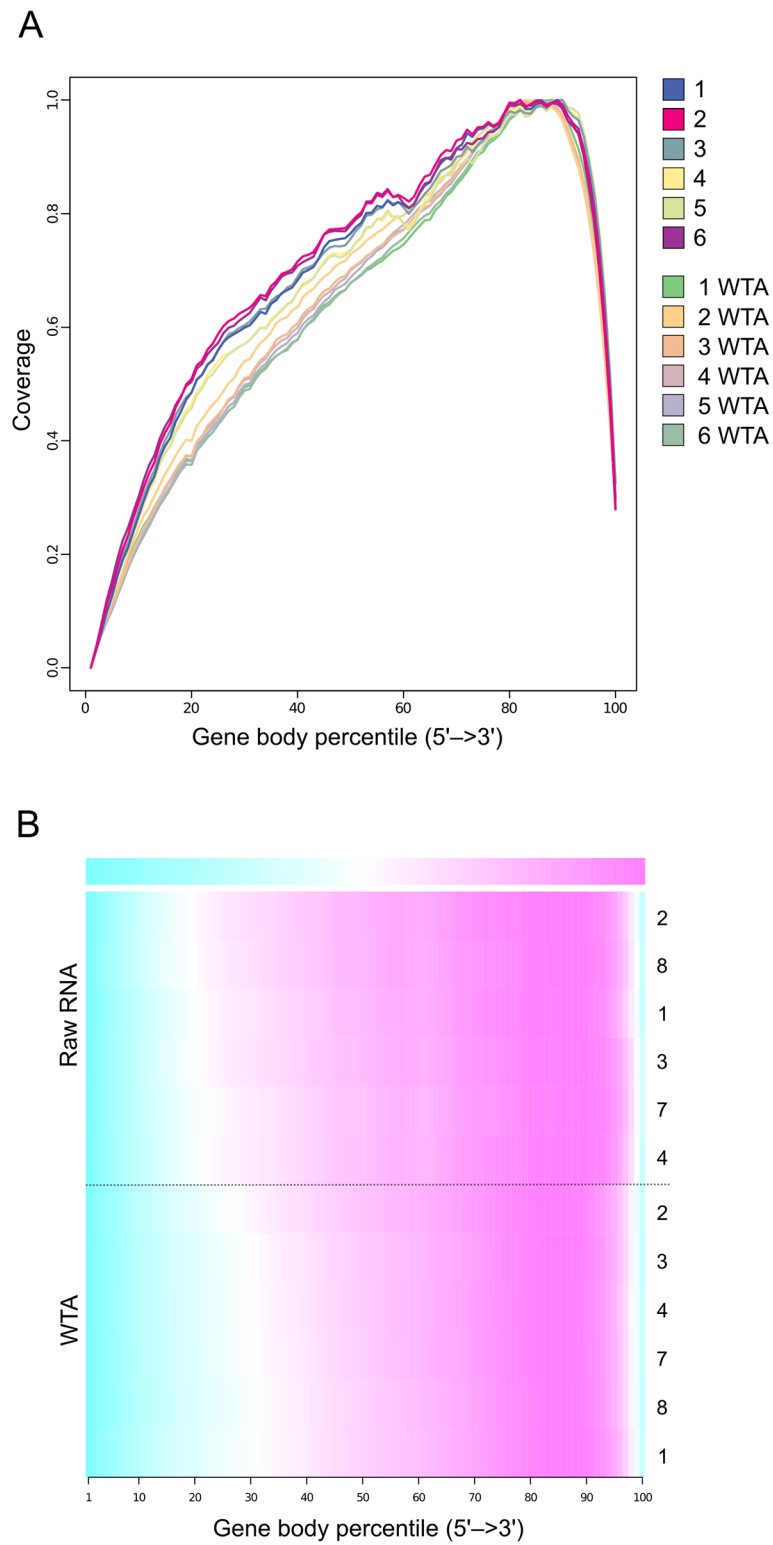Faherty *et al. BMC Biotechnology* (2015) 15:65

Page 8 of 10

Figure 4 RNA-Seq read coverage of rat gene models. **(A)** Coverage across all transcripts based on mapping of transcriptome reads to the *Rattus norvegicus* genome. All samples showed similar 3′ bias. **(B)** Heat map showing read coverage across all rat genes. Samples are ranked according to Pearson's skewness coefficients.

Faherty *et al. BMC Biotechnology* (2015) 15:65

Page 9 of 10

Sequencing Shared Resource for library preparation and sequencing.

Library preparation and sequencing

Total RNA samples were converted to cDNA libraries using Illumina's TruSeq RNA Sample Preparation Kit (San Diego, CA, USA). In summary, 1 μg of total RNA was enriched for mRNA using oligo-dT coated magnetic beads, fragmented, and reverse transcribed into cDNA. The cDNA was fragmented into smaller pieces, blunt-ended, and ligated to indexed (barcoded) adaptors and amplified using PCR. Previously amplified cDNA libraries were prepared for sequencing using Illumina's TruSeq DNA Sample Preparation Kit according to manufacturer's protocol. Final library size distribution was determined using Agilent Bioanalyzer 2100. Insert size was between 120–130 base pairs (bp) and average library size was 230 bp. Twelve libraries were multiplexed, pooled and sequenced on two lanes of the Illumina HiSeq2000 platform (San Diego, CA, USA) using 100 bp paired-end reads. Specifically, six multiplexed non-amplified samples were sequenced on Lane 1, while six multiplexed amplified samples were sequenced on Lane 2. The data set supporting the results of this article are available in the NCBI Short Read Archive repository [accession number SRP049463; http://www.ncbi.nlm.nih.gov/sra/?term=SRP049463].

Read mapping and transcriptome assembly

RNA-Seq reads were filtered using the Trimmomatic program [56] and aligned to the *R. norvegicus* genome (Rnor_5.0; downloaded from Ensembl.org on 3/2014) using TopHat2 (v 2.0.5) [3,27,57]. The resulting alignment data from TopHat2 were then fed into Cufflinks (v 2.0.2) to assemble aligned RNA-Seq reads into transcripts for each rat individually. Annotated transcripts were obtained from Ensembl database (Rnor_5.0.75.gtf; downloaded from Ensembl.org on 3/2014). Transcript abundances estimates were measured in FPKM. Cuffdiff was used to determine differential gene expression profiles between amplified and unamplified matched samples for each biological replicate included in our study [58]. HTSeq was used under default parameters to produce gene counts for each of our twelve samples [32]. To estimate differential expression, we fed the matrix of read counts generated by HTSeq into DESeq [33]. Since we were working with biological, and not technical replicates, we reduced the read counts to two conditions for each sample (amplified and raw) and estimated dispersion by ignoring condition and treating all samples as if they were replicates of the same condition. We used the option sharingMode = "fit-only" to retain outliers, accounting for biological replications in our experimental design. The geneBody_coverage.py script within the RseqC package [59] was used to examine read coverage across gene models and test for 5′ and 3′ bias in transcript coverage.

Additional files

Additional file 1: Table S1. Summary of sequencing results, quality filtering, and transcript assembly from biological replicates included in our analysis. **Table S2.** Summary of pre- and post-amplification RNA concentrations.

Additional file 2: Figure S1. Agilent Bioanalyzer traces demonstrate total RNA extractions are high-quality.

Abbreviations
WTA: Whole transcriptome amplification; qPCR: Quantitative polymerase chain reaction; FPKM: Fragments per kilobase of exon per million fragments mapped; BCP: 1-bromo-3-chloropropane; RIN: RNA integrity number; SPIA: Single primer isothermal amplification.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
SLF designed the study; SLF, PAL, and CRC completed the bioinformatic work and drafted the manuscript. ADY participated in all aspects of the experimental design. All authors read and approved the final manuscript.

References
1. Nygaard V, Løland A, Holden M, Langaas M, Rue H, Liu F, et al. Effects of mRNA amplification on gene expression ratios in cDNA experiments estimated by analysis of variance. BMC Genomics. 2003;4:11.
2. Mortazavi A, Williams BA, Mccue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Meth. 2008;5:621–8.
3. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10:57–63.
4. Streelman JT, Kocher TD. From phenotype to genotype. Evol Dev. 2000;2:166–73.
5. Kurn N. Novel isothermal, linear nucleic acid amplification systems for highly multiplexed applications. Clin Chem. 2005;51:1973–81.
6. López-Maury L, Marguerat S, Bähler J. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. Nat Rev Genet. 2008;9:583–93.
7. Minoche AE, Dohm JC, Himmelbauer H. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. Genome Biol. 2011;12:R112.
8. Fietz J, Tataruch F, Dausmann K, Ganzhorn J. White adipose tissue composition in the free-ranging fat-tailed dwarf lemur (Cheirogaleus medius; Primates), a tropical hibernator. J Comp Physiol B. 2003;173:1–10.
9. Feeney EJ, Austin S, Chien Y-H, Mandel H, Schoser B, Prater S, et al. The value of muscle biopsies in Pompe disease: identifying lipofuscin inclusions in juvenile-and adult-onset patients. Acta Neuropathol Commun. 2014;2:2.
10. Hennessey JV, Chromiak JA, Dellaventura S, Guertin J, Maclean DB. Increase in percutaneous muscle biopsy yield with a suction-enhancement technique. J Appl Physiol. 1997;82:1739–42.

Faherty et al. BMC Biotechnology (2015) 15:65

Page 10 of 10

11. Barber LJ, Sandhu S, Chen L, Campbell J, Kozarewa I, Fenwick K, et al. Secondary mutations in BRCA2associated with clinical resistance to a PARP inhibitor. J Pathol. 2013;229:422–9.

12. Wang E. RNA amplification for successful gene profiling analysis. J Transl Med. 2005;3:28.

13. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 2012;6:1621–4.

14. Wang E, Miller LD, Ohnmacht GA, Liu ET, Marincola FM. High-fidelity mRNA amplification for gene profiling. Nat Biotechnol. 2000;18:457–9.

15. Sausen M, Leary RJ, Jones S, Wu J, Reynolds CP, Liu X, et al. Integrated genomic analyses identify ARID1A and ARID1B alterations in the childhood cancer neuroblastoma. Nat Genet. 2012;45:12–7.

16. Liu P, Morrison C, Wang L, Xiong D, Vedell P, Cui P, et al. Identification of somatic mutations in non-small cell lung carcinomas using whole-exome sequencing. Carcinogenesis. 2012;33:1270–6.

17. Eberwine J, Yeh H, Miyashiro K, Cao Y, Nair S, Finnell R, et al. Analysis of gene expression in single live neurons. Proc Natl Acad Sci U S A. 1992;89:3010–4.

18. Van Gelder RN, von Zastrow ME, Yool A, Dement WC, Barchas JD, Eberwine JH. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. Proc Natl Acad Sci U S A. 1990;87:1663–7.

19. Macaulay IC, Voet T. Single cell genomics: advances and future perspectives. PLoS Genet. 2014;10:e1004126.

20. Baugh LR, Hill AA, Brown EL, Hunter CP. Quantitative analysis of mRNA amplification by in vitro transcription. Nuc Acids Res. 2001;29:E29.

21. Clément-Ziza M, Gentien D, Lyonnet S, Thiery J-P, Besmond C, Decraene C. Evaluation of methods for amplification of picogram amounts of total RNA for whole genome expression profiling. BMC Genomics. 2009;10:246.

22. Rudnicki M, Eder S, Schratzberger G, Mayer B, Meyer TW, Tonko M, et al. Reliability of T7-based mRNA linear amplification validated by gene expression analysis of human kidney cells using cDNA microarrays. Nephron Exp Nephrol. 2004;97:e86–95.

23. Scheidl SJ, Nilsson S, Kalén M, Hellström M, Takemoto M, Håkansson J, et al. mRNA expression profiling of laser microbeam microdissected cells from slender embryonic structures. Am J Pathol. 2002;160:801–13.

24. Klur S, Toy K, Williams MP, Certa U. Evaluation of procedures for amplification of small-size samples for hybridization on microarrays. Genomics. 2004;83:508–17.

25. Viale A, Li J, Tiesman J, Hester S, Massimi A, Griffin C, et al. Big results from small samples: evaluation of amplification protocols for gene expression profiling. J Biomol Tech. 2007;18:150–61.

26. Gold D, Coombes K, Medhane D, Ramaswamy A, Ju Z, Strong L, et al. A comparative analysis of data generated using two different target preparation methods for hybridization to high-density oligonucleotide microarrays. BMC Genomics. 2004;5:2.

27. Zheng W, Chung LM, Zhao H. Bias detection and correction in RNA-Sequencing data. BMC Bioinformatics. 2011;12:290.

28. Shanker S, Paulson A, Edenberg HJ, Peak A, Perera A, Alekseyev YO, Beckloff N, Bivens NJ, Donnelly R, Gillaspy AF, Grove D, Gu W, Jafari N, Kerley-Hamilton JS, Lyons RH, Tepper C, Nicolet CM: Evaluation of Commercially Available RNA Amplification Kits for RNA Sequencing Using Very Low Input Amounts of Total RNA. J Biomol Tech. 2015 Apr;26(1):4-18. doi: 10.7171/jbt.15-2601-001

29. Hansen KD, Wu Z, Irizarry RA, Leek JT. Sequencing technology does not eliminate biological variability. Nature Biotechnol. 2011;29:572–3.

30. Liu Y, Zhou J, White KP. RNA-seq differential expression studies: more sequence or more replication? Bioinformatics. 2014;30:301–4.

31. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28:516–20.

32. Anders S, Pyl PT, Huber W. HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics. 2014;31:166–9.

33. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11:R106.

34. Tariq MA, Kim HJ, Jejelowo O, Pourmand N. Whole-transcriptome RNAseq analysis from minute amount of total RNA. Nuc Acids Res. 2011;39:e120–e120.

35. Dafforn A, Chen P, Deng G, Herrler M, Iglehart D, Koritala S, et al. Linear mRNA amplification from as little as 5 ng total RNA for global gene expression analysis. Biotechniques. 2004;37:854–7.

36. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. 2008;18:1509–17.

37. Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, et al. The developmental transcriptome ofDrosophila melanogaster. Nature. 2012;471:473–9.

38. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, et al. Highly integrated single-base resolution maps of the Epigenome in Arabidopsis. Nuc Acids Res. 2009;133:e123–e123.

39. Cloonan N, Forrest ARR, Kolle G, Gardiner BBA, Faulkner GJ, Brown MK, et al. Stem cell transcriptome profiling via massive-scale mRNA sequencing. Nat Meth. 2008;5:613–9.

40. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet. 2008;40:1413–5.

41. Morozova O, Marra MA: Applications of next-generation sequencing technologies in functional genomics. Genomics 2008;92:255–264. doi:10.1016/j.ygeno.2008.07.001

42. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science. 2008;320:1344–9.

43. Rowley JW, Oler AJ, Tolley ND, Hunter BN, Low EN, Nix DA, et al. Genome-wide RNA-seq analysis of human and mouse platelet transcriptomes. Blood. 2011;118:e101–11.

44. Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, et al. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature. 2008;453:1239–43.

45. Hillier LW, Reinke V, Green P, Hirst M, Marra MA, Waterston RH. Massively parallel sequencing of the polyadenylated transcriptome of C. elegans. Genome Res. 2009;19:657–66.

46. Daines B, Wang H, Wang L, Li Y, Han Y, Emmert D, et al. The Drosophila melanogaster transcriptome by paired-end RNA sequencing. Genome Res. 2011;21:315–24.

47. Oshlack A, Robinson MD, Young MD. From RNA-seq reads to differential expression results. Genome Biol. 2010;11:220.

48. Ekblom R, Galindo J. Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity. 2010;107:1–15.

49. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29:644–52.

50. Cahais V, Gayral P, Tsagkogeorga G, Melo-Ferreira J, Ballenghien M, Weinert L, et al. Reference-free transcriptome assembly in non-model animals from next-generation sequencing data. Mol Ecol Resour. 2012;12:834–45.

51. Deutscher MP. The metabolic role of RNases. Trends Biochem Sci. 1988;13:136–9.

52. Stevens A. Ribonucleic acids: biosynthesis and degradation. Annu Rev Plant Physiol Plant Mol Biol. 1963;32:15–42.

53. Gopee NV, Howard PC. A time course study demonstrating RNA stability in postmortem skin. Exp Mol Pathol. 2007;83:4–10.

54. Ohashi Y, Creek KE, Pirisi L, Kalus R, Young SR. RNA degradation in human breast tissue after surgical removal: a time-course study. Exp Mol Pathol. 2004;77:98–103.

55. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, et al. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Mol Biol. 2006;7:3.

56. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.

57. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes inthe presence of insertions, deletions and gene fusions. Genome Biol. 2013;14:R36.

58. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nature Biotechnol. 2013;31:46–53.

59. Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. Bioinformatics. 2012;28:2184–5.