LESSeq: local event-based analysis of alternative splicing using RNA-Seq

Jing Leng [1], Sunghee Oh [2], Ekta Khurana [1,3], James P. Noonan [1,4,5], and Mark B. Gerstein [1,3,6*]

[1]Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA
[2]Division of Human Genetics, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA
[3]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA
[4]Department of Genetics, Yale University School of Medicine, New Haven, CT, USA
[5]Kavli Institute for Neuroscience, Yale University School of Medicine, New Haven, CT, USA
[6]Department of Computer Science, Yale University, New Haven, CT, USA
*To whom correspondence should be addressed

ABSTRACT

Summary: Alternative splicing is important in development and evolution, and can be studied genome-wide utilizing RNA-Seq. With comparative RNA-Seq experiments, signatures that distinguish conditions resulting from differential gene regulation, including differential alternative splicing can be detected. However, challenges in statistical inference from short-read technology still preclude reliable identification of alternative splicing signatures that can be prioritized for further biological investigation.][Most published methods do not provide localized, unambiguous regions in genes that undergo differential alternative splicing.]] To enable robust discovery of differential alternative splicing, we developed a pipeline that identifies unambiguous local events of alternative splicing, quantifies their abundance using maximum likelihood estimation, and tests significance of alternative splicing changes between different conditions. We demonstrated the utility of this pipeline through two case studies relevant to human variation and evolution. Using an RNA-Seq dataset of lymphoblastoid cell lines in two human populations and an RNA-Seq dataset of several tissues in human and rhesus macaque, we identified hundreds of population- and lineage-differential alternative splicing events respectively.

Availability: The LESSeq pipeline is implemented in C++ and R, and is available at https://code.google.com/p/lesseq.

Contact: pi@gersteinlab.org

1. INTRODUCTION

Alternative splicing of precursor messenger RNA (pre-mRNA) generates multiple transcripts, or isoforms from a single gene locus that may differ in localization, function or other biological features. Alternative isoform usage is thought to be a major source of biological complexity during development and evolution (Nilsen and Graveley, 2010). In humans, alternative splicing variations have been implicated in differential disease associations and drug responses (Lu, et al., 2012), highlighting the need for deeper understanding of the associations, or even causal relationships between alternative splicing and human biological variations. During evolution, alternative splicing leads to the expansion of transcriptome, and sometimes proteome in organisms through

Jing Leng 1/4/14 1:25 AM
Deleted: Moreover, most

Jing Leng 1/4/14 1:25 AM
Deleted: of

Jing Leng 1/4/14 1:25 AM
Deleted: .

Jing Leng 1/4/14 1:25 AM
Deleted: abundances

Jing Leng 1/4/14 1:25 AM
Deleted: differential splicing. The method uncovers localized candidate regions in genes that exhibit differential

Jing Leng 1/4/14 1:25 AM
Deleted: , facilitating mechanistic studies on the underlying genetic causes

Jing Leng 1/4/14 1:25 AM
Deleted: mark.gerstein@yale.edu

differential inclusion and exclusion of exonic sequences, and could underlie lineage-specific phenotypic traits.

Over the past few years, high-throughput RNA sequencing, or RNA-Seq (Wang, et al., 2009) has dramatically expanded our knowledge of alternative splicing. It was discovered that almost all human multi-exonic pre-mRNAs undergo alternative splicing, and that tissue-specific regulation of alternative splicing maybe pervasive (Wang, et al., 2008), suggesting the functional relevance of alternative isoform usage. However, it has also been shown that noisy products from alternative splicing are extensive (Pickrell, et al., 2010), emphasizing the need to distinguish biologically important alternative splicing events from those of no functional consequences, which could be achieved in part through comparative RNA-Seq experimental designs that include multiple biological conditions.

On the informatics side of RNA-Seq research, many computational methods have been developed to assemble and quantify transcripts utilizing RNA-Seq data. Since the transciptome of a given condition is unlikely to be fully captured by any reference annotation, it is desirable to assemble the transcripts for a specific study leveraging the RNA-Seq data. Expression levels of the newly assembled transcripts can then be calculated, followed by downstream analysis (e.g. differential transcript usage detection between two conditions). However, there are many challenges in such transcript-based inference. First, assembling correct isoforms utilizing short-read data for the sample(s) of study is very difficult, especially for mammalian genomes such as the human (Steijger, et al., 2013). In most transcript assembly methods designed for mammalian genomes, RNA-Seq reads are first mapped to the reference genome. The resulting exonic and spliced reads are subsequently used (sometimes in conjunction with a reference transcriptome annotation) to construct a splicing graph for each gene locus, which is then used to derive isoform structures according to a specific graph-traversing algorithm. Because short-read technology does not provide full connectivity of different regions in the splicing graph, the strategy for traversing such graphs varies wildly across methods (i.e. some generate the most parsimonious set, some output all possible ones, and the others lie between these two extremes), and none was shown to yield satisfactory full-transcript annotation in human (Steijger, et al., 2013). Second, even if the “correct” transcriptome annotation is provided, it is not straightforward to calculate transcript expression levels (Pachter, 2011), as short-read technology necessitates probabilistic estimation of the transcript abundances. Most transcript quantification tools calculate the Maximum Likelihood Estimate (MLE) of transcript abundances based on a specific objective function, whose form and complexity differ across methods and are dependent on the modeling of RNA-Seq process. In reality, transcript quantification algorithms based on similar ideas can generate different results (Steijger, et al., 2013), with the agreement generally decreasing as the number of isoforms of a gene increases (Du, et al., 2012). Moreover, the assembly and quantification steps are tightly linked, as incorrect transcript annotation exacerbates the quantification problem (Du, et al., 2012). Combined, the inaccuracies and uncertainties in both transcript assembly and quantification make transcript-based comparative study extremely challenging.

We thus propose to step back from the transcript-based RNA-Seq inference problem, and devised a local event-based analytical approach that focuses on localized regions in genes where isoform structures diverge (e.g. one skipped exon encompassed by two constitutive exons) (Katz, et al., 2010; Wang, et al., 2008). By assessing local alternative splicing events, we can bypass several aspects of uncertainty in transcript-based analysis and generate more robust results. Local alternative splicing events are essentially local parts of splicing graphs that contain diverging paths, and analyzing such regions abrogates the need to assemble full-length transcripts. Since transcript assembly methods output different transcripts even with the same underlying splicing graph, circumventing the step of whole transcript assembly bypasses the errors produced from it. In addition, the number of local events in a given gene is never greater than that of all the isoforms of a gene, thus yielding more robust quantification results - as has been shown previously (Du, et al., 2012), fewer isoforms per gene lead to more consistent quantification results between methods. Moreover, in the implementation of our pipeline, focusing on defined simple patterns of local events yield regions that are guaranteed to be computationally identifiable (Hiller, et al., 2009), which is not always true for transcript-level inference.

2. METHODS

Below we describe the four major steps of the LESSeq pipeline.

2.1 Refine gene models using RNA-Seq

This first step of the pipeline aims to derive comprehensive splicing annotations for the specific sample(s) of study. For species with a reference annotation, this step is optional but strongly recommended, because alternative splicing is highly tissue-specific and there may exist splicing events in the condition of interest which are not annotated in a reference transcriptome. The current pipeline employs Cufflinks (Trapnell, et al., 2010) for this purpose, and we recommend using reference annotation based transcript (RABT) assembly for well annotated organisms such as human. In the RABT method, faux-reads that tile the reference transcripts are used together with the RNA-Seq reads to assemble transcripts. However, alternatives to Cufflinks can also be used. For species that do not have a reference genome and/or reference transcriptome annotation, *de novo* methods to build gene models are available (Garber, et al., 2011), and users should substitute their own preferred approach for the current implementation in LESSeq, while the remainder of the pipeline still applies.

2.2 Build splicing graphs and define local events

In this step, the pipeline builds a splicing graph for each gene locus utilizing assembled transcripts from the previous step. The locally diverged parts in the graphs are identified, and termed as "local events" (Figure XXX). As illustrated in the two scenarios in XXX, the definition of local events is conservative, and in strict terms means that such local graphs should have the numbers of both in-edges to the leftmost node and out-edges from the rightmost node equal to that of the total number of isoforms in this gene. In the implementation of the pipeline, the shortest local graphs that satisfy such criteria are taken. Because the local events for some genes can still be very complex, and suffer similar quantification difficulties as those in transcript-based analysis, the pipeline provides filtering steps that generate pre-defined simple events for which all isoforms of

Jing Leng 1/4/14 1:25 AM
Deleted: options

a gene can be grouped into either of the two forms as shown in Figure XXX. Of note, the events selected according to such criteria are restrictive, so that the downstream analyses yield robust results. [[Mention we provide a resource for human GENCODE annotation]]

2.3 Count reads compatible with local events and estimate their relative expression levels

A metric to quantify isoform usage is the relative expression level of an isoform, with each isoform's expression level divided by the total expression from all isoforms of a given gene. As such, the relative expression level represents isoform abundance relative to other isoforms of a given gene, with the sum of all isoforms' relative expression levels for a gene being 1. This metric is useful if one aims to compare alternative isoform usage independent of gene expression level changes. For the local events identified from the previous step, such concept leads to the natural definition of relative expression levels of each local event (Figure XXX). For each event type shown in Figure XXX, there are two fractional values representing the relative expression level of either of the two possible forms of a local event. Such metric provides a quantitative measurement of the extent of alternative splicing at each given locus, and the values concatenated for all local events can be used as a feature vector for each sample, which can then be used to perform downstream analysis such as clustering (Figure XXX).

To calculate the relative expression levels of each local event identified from the previous step, the pipeline counts reads that are compatible with either of the two local events (Figure XXX) at each locus, and derives the Maximum Likelihood Estimates using the read counts and local event annotation. To calculate the MLE, LESSeq uses the same strategy as a method that was developed to estimate relative transcript expression levels (Du, et al., 2012). In this process, RNA-Seq is modeled as a probabilistic partial sampling process, assuming uniform sampling of short-reads from each form of local event, and an Expectation-Maximization (EM) algorithm is used to infer the MLE. Hence, this step of LESSeq outputs the raw number of read counts compatible with each local event as well as the estimated relative expression levels.

2.4 Test differential alternative splicing

In this final step, the pipeline determines the statistical significance of differential alternative splicing between conditions. We provide parametric tests that are capable of handling situations when very few replicates (e.g. three replicates or fewer) are generated for each condition. We also provide a non-parametric test that can be utilized when there are many replicates per condition. The non-parametric test can be used to supplement the parametric tests for two purposes: to compare the abundance of significant differential alternative splicing events, and to derive the most confident candidates by taking the intersection of results from different tests.

When only one replicate is generated for each condition, a parametric test based on two-sided Fisher exact test is used. In this case, a two-way contingency table is constructed for each local event, with each cell's value being the raw read-count for one local event form in one condition, as shown in XXX (Wang, et al., 2008). When more than one replicate is available in each condition, the parametric test is based on a log-linear model

Jing Leng 1/4/14 1:25 AM
Deleted: a

Jing Leng 1/4/14 1:25 AM
Deleted: test

Jing Leng 1/4/14 1:25 AM
Deleted: is applicable in all

Jing Leng 1/4/14 1:25 AM
Deleted: , and it is the only applicable test

Jing Leng 1/4/14 1:25 AM
Deleted: fewer than three

Jing Leng 1/4/14 1:25 AM
Deleted: per

Jing Leng 1/4/14 1:25 AM
Deleted: are available.

Jing Leng 1/4/14 1:25 AM
Deleted: provided two

Jing Leng 1/4/14 1:25 AM
Deleted: tests

Jing Leng 1/4/14 1:25 AM
Deleted: equal to or greater than three

Jing Leng 1/4/14 1:25 AM
Deleted: two

Jing Leng 1/4/14 1:25 AM
Deleted: tests

Jing Leng 1/4/14 1:25 AM
Deleted: test

Jing Leng 1/4/14 1:25 AM
Deleted: The

with a Poisson link and a likelihood ratio test based on model fit, using the raw read-counts for one local event form in one condition (Bullard, et al., 2010; Cotney, et al., 2012; Cotney, et al., 2013). The number of reads compatible with one form p of a local event in sample i is denoted as X_{pi}, and is modeled as $\log\left(E\left[X_{pi}\middle|X_{.i}\right]\right) = \log X_{.i} + \lambda_{pj(i)} + \theta_{pi}$, where $X_{.i}$ is the total number of reads mapped in the local event for sample i, $\lambda_{pj(i)}$ is the condition-specific splicing level for condition j, and θ_{pi} is the replicate error term. For each event, p can be either of the two forms in XXX, so there are two tests, and therefore two p-values for each event. In this test, the raw read counts compatible with either of the two events in a locus are compared between conditions, normalizing for the total raw read counts in the entire locus, so that the effect of differential alternative splicing is tested, independent of total expression level changes. When many replicates exist in each condition, a non-parametric test can also be applied. The non-parametric test takes as input the relative expression levels estimated from the previous step, as opposed to raw read counts in the parametric test. It is to perform Wilcoxon rank sum test on the of relative expression level values between two conditions.

3. APPLICATIONS

3.1 Within-species variation

Comparative RNA-Seq experiments under different conditions in a single species can be used to uncover alternative splicing signatures important in various aspects of the biology for the species of interest. For example, data from different time points during organismal development can yield insights to events driving developmental progression, sampling from different organs may identify signatures underlying tissue differentiation, and comparison between healthy versus disease samples facilitates discovery of aberrant splicing in a specific disease.

One important question in human biology is the difference between individuals and populations, and we studied human differences in alternative splicing using a dataset generated by the Geuvadis Consortium (Lappalainen, et al., 2013). Messenger RNA-Seq data of lymphoblastoid cell lines (LCLs) in five human populations was produced (Lappalainen, et al., 2013). Mapped reads were downloaded (http://www.ebi.ac.uk/arrayexpress/files/E-GEUV-1/processed/) for two human populations – CEU and YRI, with 91 and 89 samples respectively. After gene annotation refinement using RNA-Seq reads and Ensembl (V67) human annotation, 2948 local events were identified (each event was required to have 80nt-long exons and 50nt-long introns). Using the relative expression levels for all local events, individuals were clustered, revealing that individuals do not segregate by population with regard to alternative splicing (Figure XXX). Statistical tests for alternative splicing changes yield between 8% to 10% significant differential events between the two populations, with 174 events detected by both parametric and non-parametric methods (BHP cutoff at 0.05, Figure XXX).

The original research paper showed that the populations cluster by genotype, but not exon-level expression (Lappalainen, et al., 2013). Exon-level expression value is a

Jing Leng 1/4/14 1:25 AM
Deleted: generated by the previous step

Jing Leng 1/4/14 1:25 AM
Deleted: ,

Jing Leng 1/4/14 1:25 AM
Deleted: The two non-parametric tests take

Jing Leng 1/4/14 1:25 AM
Deleted: One test is to perform Wilcoxon rank sum test on the vectors of relative expression level values across samples; the other is to conduct a permutation-based test using the SAMr packge(Tusher, et al., 2001).

Jing Leng 1/4/14 1:25 AM
Deleted: Messenger RNA-Seq data of lymphoblastoid cell lines in multiple human populations generated by the Geuvadis Consortium was used

Jing Leng 1/4/14 1:25 AM
Deleted: were

Jing Leng 1/4/14 1:25 AM
Deleted: differential

Jing Leng 1/4/14 1:25 AM
Deleted: show good overlap, yielding 152

Jing Leng 1/4/14 1:25 AM
Deleted: local

Jing Leng 1/4/14 1:25 AM
Deleted: all three

Jing Leng 1/4/14 1:25 AM
Deleted: [[The original research paper shows population clustering by genotype, but not exon-level expression. Transcript-based (FluxCapacitor+Wilcoxn) and exon-based (DEXSeq) differential splicing analyses were conducted in the original research paper, and thousands of significant alternative isoform usage cases were discovered…]]

combined product of gene expression level and alternative splicing, and is not informative for assessing the two aspects of gene regulation separately. Our clustering result using local event relative expression levels revealed that, when the effects of alternative splicing alone is examined, the individuals do not cluster by population (Figure XXX). The original research paper also attempted to identify population-differential alternative splicing events. Using Gencode annotation, Transcript-based (FluxCapacitor for transcript quantification and Wilcoxn for significance testing) and exon-based (DEXSeq) analyses were conducted, yielding around 20% and 50% significant genes in each case. The methodologies bear several shortages. In the first place, using a reference annotation while deep transcriptome sequencing data is available ignores the specific splicing structures present in this large number of LCL samples. As such, the downstream calculations are not reliable since the correct exon and splicing annotation are missed. The problem with transcript-based approach is further exacerbated by the transcript quantification step – as discussed in the introduction section, and the exon-based method compare less favorably to the local event-based approach – as mentioned in the discussion section.

3.2 Cross-species variation

Comparative RNA-Seq experiments across species can help to better understand organismal evolution in terms of whole transcriptome variation, and to identify candidates exhibiting differential gene regulation that could drive phenotypic evolution (i.e. differential expression and/or differential alternative splicing).

To study the evolution of alternative splicing between human and other primates, we utilized a messenger RNA-Seq dataset that profiled six organs (brain, cerebellum, heart, kidney, liver and testis) in ten species (Brawand, et al., 2011). Fastq files were downloaded for all six organs for human and rhesus macaque. LiftOver tool (http://genome.ucsc.edu/) was used to match orthologous exon coordinates between the two species (Ensembl V67 annotation for both species were used), yielding 1683 "skipped exon" local events at 0.9 reciprocal mapping rates. Using relative expression levels as well as total expression levels (measured by RPKM values) at all events, hierarchical clustering was performed, revealing that alternative splicing patterns cluster by species whereas total expression levels cluster by tissues. Lineage-differential alternative splicing events were also identified (XXX).

The original research paper revealed that there is strong selection pressure for expression level in organs (Brawand, et al., 2011). Our study confirms this by showing that the expression levels at local events cluster by organs. Additionally, we find that the relative expression levels cluster by species. This observation agrees with two recent studies that also found faster evolutionary changes in alternative splicing compared to expression level (Barbosa-Morais, et al., 2012; Merkin, et al., 2012). However, the extent to which this is due to neutral evolution or selection is unknown. In a more recent analysis (Reyes, et al., 2013), it was shown that most exon expression level exhibit weak cross-tissue difference and large interspecies variability, indicating neutral drift; while only a minority show conserved tissue-specific usage patterns. Our study tackles the problem from a

Jing Leng 1/4/14 1:25 AM
Deleted: Messenger

Jing Leng 1/4/14 1:25 AM
Deleted: data

Jing Leng 1/4/14 1:25 AM
Deleted: was used

Jing Leng 1/4/14 1:25 AM
Deleted: derive

Jing Leng 1/4/14 1:25 AM
Deleted: annotation

Jing Leng 1/4/14 1:25 AM
Deleted: ,

Jing Leng 1/4/14 1:25 AM
Deleted: few thousand exons

different perspective – by identifying lineage-differential alternative splicing events, we are the first to identify candidates for lineage-specific phenotypes.

4. DISCUSSION

We developed a pipeline for comparative alternative splicing study with RNA-seq. Our method aims to provide robust differential alternative splicing detection by using a local-event based approach, and it uncovers localized candidate regions in genes that exhibit differential alternative splicing. By pinpointing specific loci of interest, the method can ease the design of mechanistic studies such as mini-gene assay. It also allows unambiguous design of PCR primers and microarray probes for large-scale applications (e.g. healthy versus disease state biomarkers). We applied the pipeline to two RNA-Seq dataset for studying human variation and evolution, and were able to identify population- and lineage-differential alternative splicing events.

LESSeq employs a local event-based analysis strategy, and is thus more robust to transcript annotation errors compared to transcript-based methods. A few other methods are also built around similar "local event" ideas. For example, DEXSeq (Anders, et al., 2012) is a method that tests differential exon usage. However, compared to LESSeq, DEXSeq loses information on local connectivity of exons, and does not define the differential usage of exons as results of different mechanisms, as shown in XXX. DEXSeq also does not provide the relative expression level estimation of each event, or exon as a metric to assess the degree of alternative splicing. Most importantly, the estimation of each exon in DEXSeq maybe confounded by other alternative splicing events in the same gene – unlike LESSeq, in which the estimation is not affected by other events due to the strict definition of what type of events should be included XXX. DiffSplice is a method that both quantifies and tests for differential local events, or "alternative splicing modules"(Hu, et al., 2013). Compared to LESSeq, DiffSplice does not filter out very complex local events. The results of analyzing these complex events could be very unreliable. For instance, some complex local events are still unidentifiable using short-read data (Hiller, et al., 2009). Furthermore, DiffSplice only provides a permutation-based test and is not applicable when there are fewer than three replicates per condition. In such situations, LESSeq provides a parametric test for studies that have very few replicates. When many replicates are available, LESSeq's parametric and non-parametric tests can both be applied, and the most confident candidates can be identified as the intersection (Soneson and Delorenzi, 2013). [[The statistical tests are conducted in the R framework (http://cran.us.r-project.org/), and can be further supplemented with tests from other R packages.]]

Jing Leng 1/4/14 1:25 AM
Deleted: DISCUSSIONS

Jing Leng 1/4/14 1:25 AM
Deleted: .

Jing Leng 1/4/14 1:25 AM
Deleted: .

Jing Leng 1/4/14 1:25 AM
Deleted: three or more

Jing Leng 1/4/14 1:25 AM
Deleted: of all methods

ACKNOWLEDGEMENTS

Funding:

REFERENCES

Anders, S., Reyes, A. and Huber, W. (2012) Detecting differential usage of exons from RNA-seq data, *Genome research*, **22**, 2008-2017.

Barbosa-Morais, N.L., *et al.* (2012) The evolutionary landscape of alternative splicing in vertebrate species, *Science*, **338**, 1587-1593.

Brawand, D., *et al.* (2011) The evolution of gene expression levels in mammalian organs, *Nature*, **478**, 343-348.
Bullard, J., *et al.* (2010) Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments, *BMC bioinformatics*, **11**, 94.
Cotney, J., *et al.* (2012) Chromatin state signatures associated with tissue-specific gene expression and enhancer activity in the embryonic limb, *Genome research*, **22**, 1069-1080.
Cotney, J., *et al.* (2013) The evolution of lineage-specific regulatory activities in the human embryonic limb, *Cell*, **154**, 185-196.
Du, J., *et al.* (2012) IQSeq: integrated isoform quantification analysis based on next-generation sequencing, *PloS one*, **7**, e29175.
Garber, M., *et al.* (2011) Computational methods for transcriptome annotation and quantification using RNA-seq, *Nature methods*, **8**, 469-477.
Hiller, D., *et al.* (2009) Identifiability of isoform deconvolution from junction arrays and RNA-Seq, *Bioinformatics*, **25**, 3056-3059.
Hu, Y., *et al.* (2013) DiffSplice: the genome-wide detection of differential splicing events with RNA-seq, *Nucleic acids research*, **41**, e39-e39.
Katz, Y., *et al.* (2010) Analysis and design of RNA sequencing experiments for identifying isoform regulation, *Nature methods*, **7**, 1009-1015.
Lappalainen, T., *et al.* (2013) Transcriptome and genome sequencing uncovers functional variation in humans, *Nature*.
Lu, Z.X., Jiang, P. and Xing, Y. (2012) Genetic variation of pre‐mRNA alternative splicing in human populations, *Wiley Interdisciplinary Reviews: RNA*, **3**, 581-592.
Merkin, J., *et al.* (2012) Evolutionary dynamics of gene and isoform regulation in mammalian tissues, *Science*, **338**, 1593-1599.
Nilsen, T.W. and Graveley, B.R. (2010) Expansion of the eukaryotic proteome by alternative splicing, *Nature*, **463**, 457-463.
Pachter, L. (2011) Models for transcript quantification from RNA-Seq, *arXiv preprint arXiv:1104.3889*.
Pickrell, J.K., *et al.* (2010) Noisy splicing drives mRNA isoform diversity in human cells, *PLoS genetics*, **6**, e1001236.
Reyes, A., *et al.* (2013) Drift and conservation of differential exon usage across tissues in primate species, *Proceedings of the National Academy of Sciences*, **110**, 15377-15382.
Soneson, C. and Delorenzi, M. (2013) A comparison of methods for differential expression analysis of RNA-seq data, *BMC bioinformatics*, **14**, 91.
Steijger, T., *et al.* (2013) Assessment of transcript reconstruction methods for RNA-seq, *Nature methods*.
Trapnell, C., *et al.* (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation, *Nature biotechnology*, **28**, 511-515.
Wang, E.T., *et al.* (2008) Alternative isoform regulation in human tissue transcriptomes, *Nature*, **456**, 470-476.
Wang, Z., Gerstein, M. and Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics, *Nature Reviews Genetics*, **10**, 57-63.

Jing Leng 1/4/14 1:25 AM

Deleted: Tusher, V.G., Tibshirani, R. and Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response, *Proceedings of the National Academy of Sciences*, **98**, 5116-5121.