TRANSCRIPT LEVEL ANALYSIS IMPROVES THE UNDERSTANDING OF BLADDER CANCER

Xiang Ao and Shuaicheng Li

Department of Computer Science, City University of Hong Kong, Hong Kong

ABSTRACT

Bladder cancer (BC) is one of the most globally prevalent diseases, attracting various studies on BC relevant topics. High-throughput sequencing renders it convenient to extensively explore genetic changes, like the variation in gene expression, in the development of BC. In this study, we did differential analysis on gene and transcript expression (DGE and DTE) and differential transcript usage (DTU) analysis in an RNA-seq dataset of 42 bladder cancer patients. DGE analysis reported 8543 significantly differentially expressed (DE) genes. In contrast, DTE analysis detected 14350 significantly DE transcripts from 8371 genes, and DTU analysis detected 27914 significantly differentially used (DU) transcripts from 8072 genes. Analysis of the top 5 DE genes demonstrated that DTE and DTU analysis provided the source of changes in gene expression at the transcript level. The transcript-level analysis also identified some DE and DU transcripts from previously reported mutated genes that related to BC, like ERBB2, ESPL1, and STAG2, suggesting an intrinsic connection between gene mutation and alternative splicing. Hence, the transcript-level analysis may help disclose the underlying pathological mechanism of BC and further guide the design of personal treatment.

KEYWORDS

Bladder Cancer, Differential Gene Expression, Differential Transcript Expression, Differential Transcript Usage

1. INTRODUCTION

Bladder cancer (BC) is the 10th most common malignant carcinoma worldwide, with about an estimate of 549,000 new cases and 200,000 deaths in 2018 [1]. Depending on the invasion state of tumor cells in the muscle layer, bladder cancer is clinically sorted out into two distinct subtypes. One is non-muscle invasive bladder cancer (NMIBC), and the other is muscle-invasive bladder cancer (MIBC). NMIBCs are rarely muscle-invasive and incline recurrence, reporting a rate as high as 70%. Its five-year survival rate is about 90% [2]. MIBCs are, in contrast, frequently metastasize, with a five-year survival rate of less than 50%. Currently, the main treatments for NMIBCs are transurethral resection and postoperative intravesical chemotherapy. In comparison, the treatments of MIBCs involve radical cystectomy or radiotherapy [3-6]. A systemic therapy, such as agents targeting dysfunctional or mutational genes or agents acting at the molecular level for BC treatment, is not available for BC patients at present and is in an urgent need [7-9].

The advance of the treatment of BC requires a comprehensive understanding of its pathogenesis. So far, researchers have paid substantial effort in investigating the potential mechanisms of BC [8, 10-14]. For example, elaborate differential gene expression (DGE) analyses of BC expression data advance our understanding of BC and are expected to improve its current treatments and therapies [15,16]. However, differential expression analysis at the gene level unable to reveal the

David C. Wyld et al. (Eds): ITCSE, NLCA, ICAIT, CAIML, ICDIPV, CRYPIS, WiMo - 2020
pp. 29-40, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.100503

details and changes in the composition of gene expression when gene generates more than one transcripts. Changes in the population and proportion of the entire transcripts from the same gene are vague from the result of DGE analysis. Therefore, transcript-level analysis is requisite and necessary to discover variant transcripts that lead to changes in gene expression and contribute to abnormality or phenotype of interest [17].

Consequently, differential transcript expression (DTE) analysis and differential transcript usage (DTU) analysis rose to respond to the needs. They are aiming at detecting the transcript that presents variance in its expression level or abundance. Here, we combine gene-level analysis (DGE analysis) and transcript-level analysis (DTE analysis and DTU analysis) to reveal the potential critical biomarkers contributing to the development of BC.

2. MATERIALS AND METHODS

2.1. Data Preparation

Guo G et al. presented an RNA-seq dataset of 42 bladder cancer patients to study the genetic basis of transitional cell carcinoma in [18]. We utilized such a dataset throughout this manuscript. The cohort contains 42 patients of bladder tumor, 16 of which contained paired morphological normal bladder tissue. Among the 42 patients, 6 were females, and 36 were males. The range of age of the cohort was 25 to 87 years old at the time when recruited the patients. The overall mean and median age was 62.3 and 64.5 years, respectively. Besides, 25 of 42 samples were MIBC, while 17 were NMIBC. Table 1 manifests the details about the clinical characteristics of the cohort.

We collected the data (accession code SRA063495) from the Sequence Read Archive (SRA). The mRNA libraries were generated from the TruSeq RNA Sample Preparation kit (Illumina), and the sequencing platform was the HiSeq 2000. We refer to [18] for the procedures of reads sequencing. We downloaded all the raw ".sra" files from SRA and used the fastq-dump program from the SRA toolkit (version 2.9.6) to obtain clean fastq files by decompressing such ".sra" files. We then treated those fastq files as the input to our analysis pipeline.

2.2. Expression Quantification

Our analysis pipeline contains three main steps, including expression quantification, differential analysis, and gene enrichment analysis. Figure 1 shows the flowchart of our pipeline. In expression quantification, we quantified the genome-wide expression level of genomic features, i.e., the expression level of genes and transcripts. In the differential analysis, we did differential expression analysis on genes and transcripts and differential usage analysis on transcripts. Finally, we did gene enrichment analysis on the results from the differential analysis.

In the quantification step, the expression level of each genomic feature (either gene or transcript) was represented by the number of reads originated from the feature. We adopted two distinct programs to quantify expression levels. More precisely, we used featureCounts from the Subread package (version 1.6.4) [19] to measure gene-level expression while applied Salmon (version 0.13.1) [20] to estimate transcript-level expression. For gene-level expression quantification, we first applied STAR (version 2.5.3a) [21] to align reads to human genome reference GRCh38.p2, and then adopted featureCounts for each gene to calculate the number of reads mapped to the gene (genes that shared overlapping regions were merged into one unique gene). For transcript-level expression quantification, we incorporated the Salmon to estimate the counts of reads for each transcript. The reference human gene annotation utilized in this step was from Ensembl

release 79, which consisted of 65217 records of genes and 213622 records of transcripts. After quantification, we obtained counts of reads for all the genes as well as the transcripts. We then used them in the differential expression analysis.

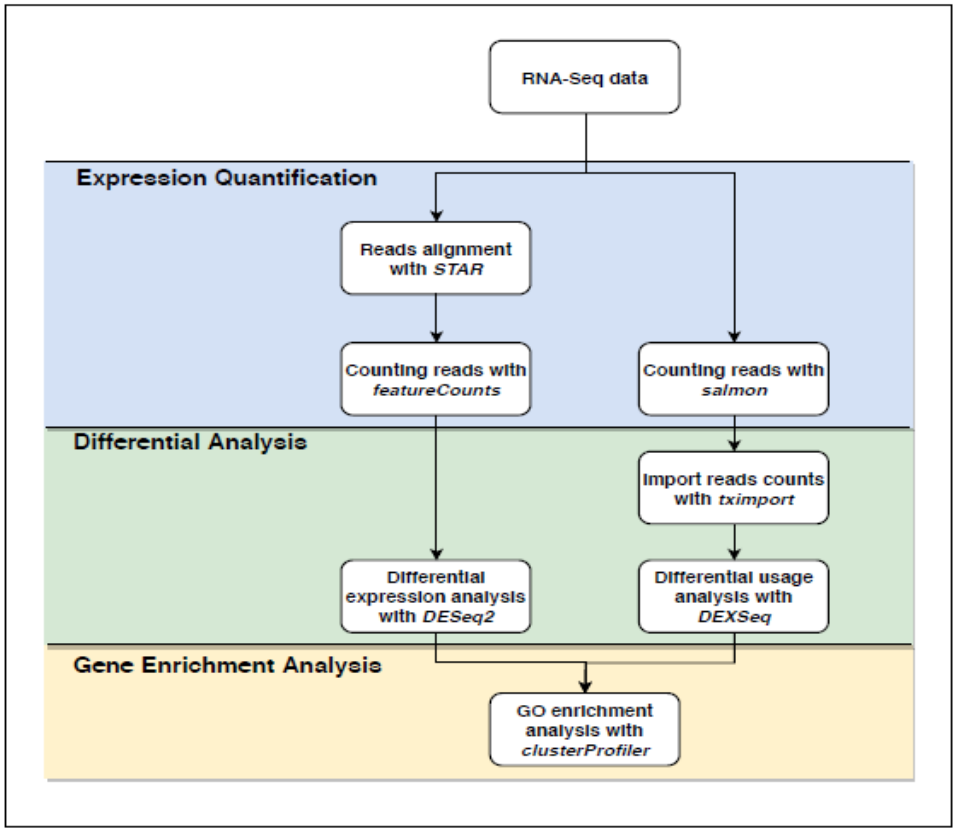

Figure 1. Flowcharts of the analysis pipeline. Our analysis pipeline contains three main parts: expression quantification, differential analysis, and gene enrichment analysis.

2.3. Differential Analysis

In step 2, we did differential expression analysis at both gene-level and transcript-level to compute the variance of expression between tumor and normal samples. Because of alternative splicing, major human genes generate more than one transcript. Consequently, we can present transcripts' expression levels in both absolute values (reads counts) and relative values (transcript usage, defined as (number of a transcript) / (amount number of transcripts from the same gene)). Transcripts from a single gene (i.e., isoforms) may hold comparable usages between conditions but increase dramatically in absolute expression, and reverse cases exist as well. Differential transcript usage analysis thus complements differential gene/transcript expression analysis. We did DGE, DTE, and DTU analysis on a BC cohort, to extensively explore the differences between BC and normal tissue and further discover the potential mechanisms behind the development of BC.

We made use of an R package DESeq2 (version: 1.24.0) [22] to do differential expression (DE) analysis. For each genomic feature and the counts of reads aligned to it, DESeq2 adopts a generalized linear model to fit the counts to a negative binomial distribution to detect differentially expressed features. Depending on the distinct read-counting programs for genes and transcripts, we took diverse data importing methods. For DGE analysis, we directly import reads

counts data from the featureCounts. As to DTE analysis, we used an R package tximport (version: 1.12.3) [17] to import counts data from the Salmon by setting the argument countsFromAbundance to "scaledTPM". Non-expressed genes/transcripts in all samples (defined as genes/transcripts with a sum of counts across all samples less than 10) were filtered out to reduce computational burdens. Samples' sex, age, bladder cancer subtype (muscle / non-muscle invasive), cancer grade, cancer recurrence status, and condition (cancer/normal) served as independent variables. We did differential comparison between cancer and normal samples. We selected significantly DE feature according to its log2 fold change (log2(FC)) and false discovery rate (FDR). The threshold was $|\log2(FC) \geq 1|$ and $FDR < 0.05$.

As a complementary analysis method, we did DTU analysis by taking advantage of an R package DEXSeq (version: 1.30.0) [23, 24]. Anders, et al., incipiently designed it for the exon-level differential usage analysis. To infer changes in exon usage, it compared the number of reads mapping to a certain exon to the number mapping to any other exons generated from the same gene. We used it for the transcript-level differential usage analysis. We applied the same method (i.e., tximport) to import reads counts for each transcript and also excluded non-expressed transcripts from analysis. The same independent variables used in DE analysis were adopted in DTU analysis as well. Besides, the same criteria were applied to choose significant DU transcripts.

2.4. Gene Ontology Enrichment Analysis

We used the R package clusterProfiler (version: 3.12.0) [25] to do gene ontology (GO) enrichment analysis.Significant genes were employed as input and converted to ENTREZ identifiers. A threshold was set to p-value < 0.05 to select significantly enriched GO terms.

3. RESULTS

From differential gene expression analysis, we found 8543 significantly differentially expressed genes between bladder cancer samples and normal bladder samples from 65065 tested genes. Of all the significant DE genes, 5293 genes were down-regulated with a mean log2(Fold Change) equals to -2.52 while 3250 were up-regulated with a mean log2(Fold Change) equals to 1.76. Principle component analysis (PCA) with the gene expression data shows that a normal sample B77_Normal was more likely to be a cancer sample (Figure 2). However, including this sample in DE analysis did not affect its result as DESeq2 sets aside outliers from the analysis. Figure 3 shows the volcano plot of DE genes, with the gene names of the top 5 over-expressed and suppressed genes labeled (names of merged genes contain a plus symbol). We used the expression level of the top 500 significant DE genes to do hierarchical clustering (Figure 4). In Figure 4, most cancer samples and normal samples were grouped correctly except the sample B77_Normal, which was more likely a cancer sample and illustrated by PCA. Among the 37 significantly mutated genes reported in [18], 10 of them (27%) were found differentially expressed (Table 1).

Gene ontology (GO) enrichment analysis was done for DE genes to identify the GO terms that were activated by DE genes. GO enrichment analysis demonstrated that extracellular matrix (GO:0031012), collagen-containing extracellular matrix (GO:0062023), muscle system process (GO:0003012), muscle contraction (GO:0006936), extracellular matrix organization (GO:0030198), extracellular structure organization (GO:0043062), extracellular matrix structural constituent (GO:0005201), muscle organ development (GO:0007517), regulation of leukocyte activation (GO:0002694), external side of plasma membrane (GO:0009897) were the top 10 significantly enriched terms (Table 2).

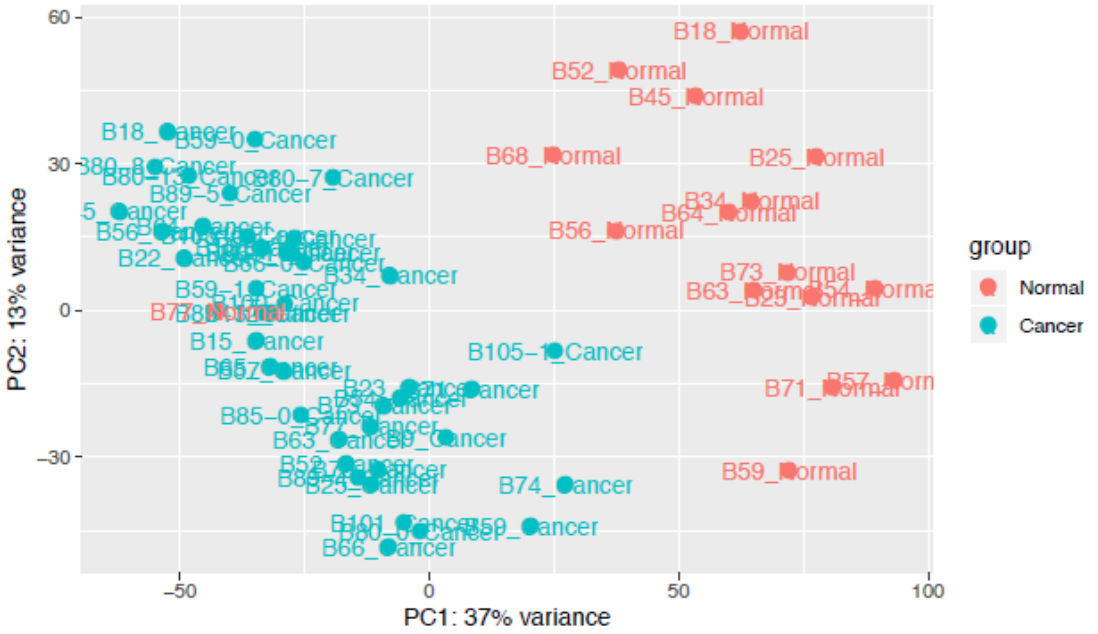

Figure 2. PCA plot on gene expression data with the first two principal components. Normal samples were colored with orange while cancer samples were colored with cyan. Although sample B77_Normal was a normal sample, it was closer to cancer samples than normal samples.

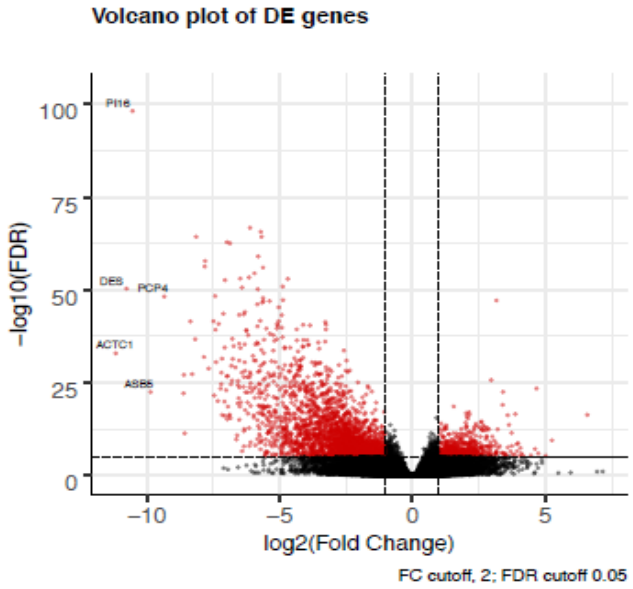

Figure 3. Volcano plot of DE genes. The x axis shows the log2(Fold change) between the cancer samples and the normal samples, the y-axis shows the –log10(FDR). Labels in the figure stand for the top 5 significantly down/up-regulated genes.

Table 1. The 37 significantly mutated genes and the distribution in DGE, DTE and DTU.

Gene set	Genes
Significantly Mutated genes	ANK2, ANK3, ARID1A, ATM, CHD6, CREBBP, CSMD3, ELF3, EP300, ERBB2, ERBB3, ERCC2, ESPL1, FAT4, FGFR3, HRAS, KALRN, KRAS, LAMA4, LRP2, MLL, MLL3, NCOR1, NF1, NFE2L3, PDZD2, PIK3CA, PIK3R4, RB1, STAG2, SYNE1, SYNE2, TP53, TRAK1, TRRAP, TSC1, UTX
DE genes	ANK2, CSMD3, ERBB2, ESPL1, FAT4, FGFR3, HRAS, LAMA4, SYNE1, TRAK1
DTE genes	ANK2, ARID1A, ATM, CREBBP, ELF3, ERBB2, ERBB3, ESPL1, FAT4, FGFR3, HRAS, KRAS, LAMA4, NF1, NFE2L3, STAG2, SYNE1, TP53, TRAK1
DTU genes	ANK2, ANK3, ARID1A, ATM, CREBBP, CSMD3, ELF3, ERBB2, ERBB3, ERCC2, ESPL1, FGFR3, HRAS, KALRN, LAMA4, NCOR1, NF1, PDZD2, PIK3R4, RB1, STAG2, SYNE2, TP53, TRAK1, TRRAP

Differential transcript expression (DTE) analysis and differential transcript usage (DTU) analysis provide details of the change in gene expression. From 213622 transcripts that were generated by all 65065 genes, DTE analysis identified 14350 significant DE transcripts, in which 6053 were up-regulated and 8297 were down-regulated. Table 3 lists the top 10 differentially expressed transcripts. In contrast, DTU analysis detected 27914 significant DU transcripts where contained 15012 over-expressed transcripts and 12902 down-regulated transcripts. Table 4 presents the top 10 differentially used transcripts. To discover the contrast of the results from DTE analysis and DTU analysis, we separated both results in positive (over-expressed) and negative (down-regulated) subgroups and showed their intersections in Figure 5a. The figure shows that 5502 transcripts changed their absolute expression level as well as their proportions in gene expression simultaneously. Among such 5502 transcripts, most of them showed the same direction of changes. That is, 2623 (46.7%) transcripts' absolute expression level and relative usage level obtained increase in BC samples, and 2742 (49.8%) transcripts got both types of levels decreased. However, some transcripts displayed opposite directions of changes. For example, there were 9 (0.2%) transcripts whose absolute expression level got promoted while usage level reduced in BC samples. Besides, there were 128 (2.3%) transcripts had pure expression level decreased while relative usage increased.

Table 2. The top 10 enriched GO terms of differentially expressed genes.

Ontology	ID	Functional Term	Gene Count	Adjusted p-value
CC	GO:0031012	extracellular matrix	218	8.59E-38
CC	GO:0062023	collagen-containing extracellular matrix	194	5.26E-37
BP	GO:0003012	muscle system process	192	1.19E-25
BP	GO:0006936	muscle contraction	159	1.21E-25
BP	GO:0030198	extracellular matrix organization	147	5.62E-21
BP	GO:0043062	extracellular structure organization	163	5.62E-21
MF	GO:0005201	extracellular matrix structural constituent	85	1.16E-18
BP	GO:0007517	muscle organ development	157	3.25E-16
BP	GO:0002694	regulation of leukocyte activation	181	6.04E-16
CC	GO:0009897	external side of plasma membrane	93	8.64E-16

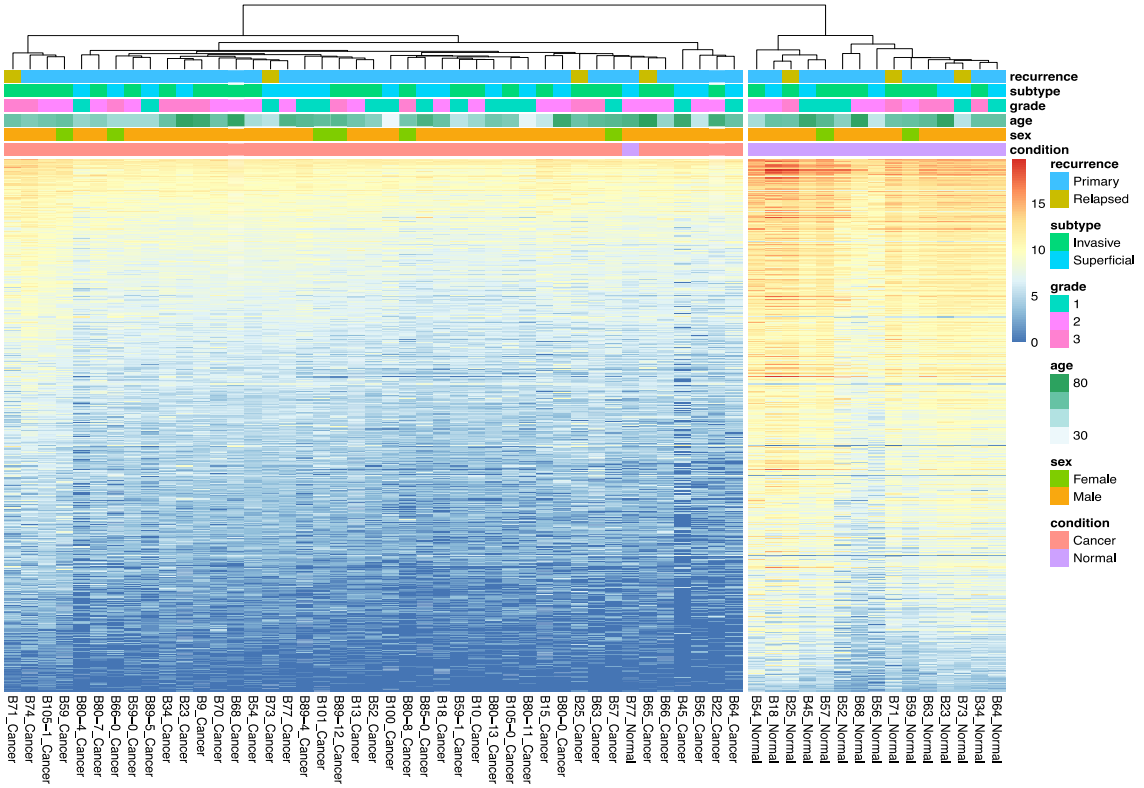

Figure 4.The hierarchical clustering result of the top 500 DE genes. Columns stood for samples while rows indicated genes. The vertical white band separated samples into two subgroups. The degree of gene expression corresponded to the transition from blue to red. Independent variables were shown as well.

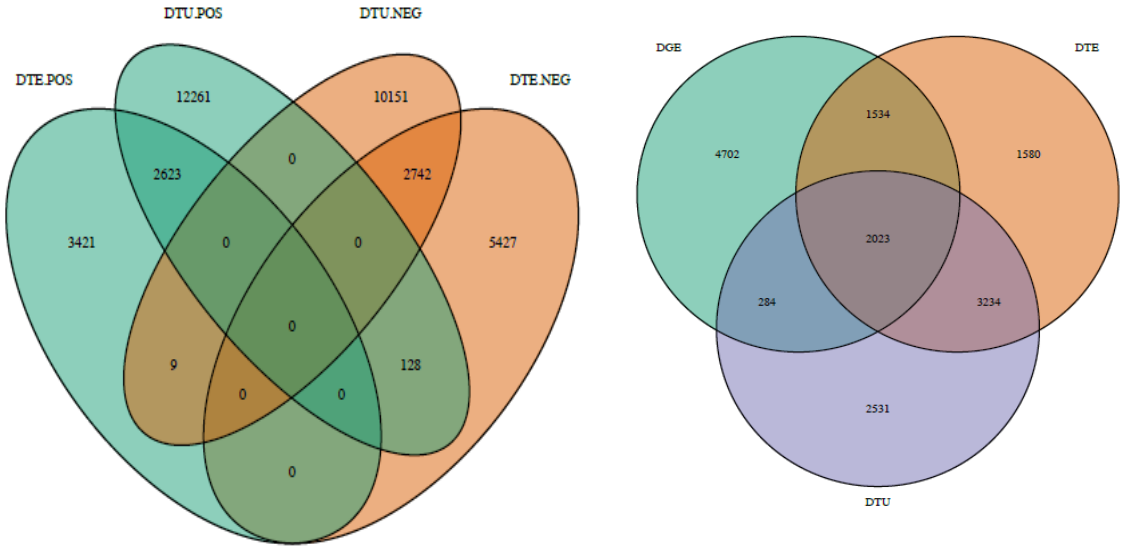

Figure 5. (a) the left Venn diagram exhibits the transcript overlaps among DTE.POS, DTE.NEG, DTU.POS and DTU.NEG and (b) the right diagram exposes gene intersections from DGE, DTE, and DTU analysis. POS and NEG stand for over-expressed and down-expressed features, respectively.

Table 3. The top 10 differentially expressed transcripts.

Transcript	Gene	Mean Expression	log2FoldChange	Adjusted p-value
ENST00000588553	AC005786.3	11.53	27.66	2.37E-14
ENST00000618836	UBE2D3	13.20	16.16	1.64E-05
ENST00000557860	ACTC1	550.95	-12.24	9.05E-19
ENST00000492726	DES	466.95	-11.39	1.21E-13
ENST00000290378	ACTC1	831.71	-11.13	7.32E-27
ENST00000611814	PI16	614.92	-10.87	1.55E-70
ENST00000373960	DES	11908.26	-10.85	4.78E-48
ENST00000560765	FGF7	287.42	-10.78	1.37E-15
ENST00000461273	RP11-274B21.5	11.39	10.54	1.63E-02
ENST00000347557	SMTN	159.59	-10.47	2.57E-19

Table 4. The top 10 differentially used transcripts

Transcript	Gene	Mean Expression	log2FoldChange	Adjusted p-value
ENST00000543780	IGJ	3632.15	-60.74	1.25E-22
ENST00000305046	ADH1B	28.92	-49.90	0.00E+00
ENST00000355722	TRPM8	10.63	-48.37	1.66E-06
ENST00000577017	MAPT	11.82	-47.00	3.37E-06
ENST00000510545	CLDND1	54.47	44.93	2.80E-13
ENST00000466266	PRUNE2	30.03	-39.34	6.97E-20
ENST00000355426	EFEMP1	426.46	-36.24	2.52E-119
ENST00000472859	SGK1	32.38	-35.94	1.15E-02
ENST00000457773	PLCD4	36.87	-33.52	0.00E+00
ENST00000618157	HLA-DRB4	13.78	-27.55	1.26E-07

Table 5. The top 10 enriched GO terms of genes that produced differentially expressed transcripts

Ontology	ID	Functional Term	Gene Count	Adjusted p-value
CC	GO:0062023	collagen-containing extracellular matrix	244	1.91E-22
CC	GO:0005925	focal adhesion	241	8.23E-22
CC	GO:0005924	cell-substrate adherens junction	241	1.25E-21
CC	GO:0030055	cell-substrate junction	243	1.25E-21
CC	GO:0031012	extracellular matrix	273	1.25E-21
CC	GO:0005912	adherens junction	286	1.25E-21
BP	GO:0030198	extracellular matrix organization	209	2.06E-19
BP	GO:0043062	extracellular structure organization	228	7.88E-17
CC	GO:0015629	actin cytoskeleton	238	1.30E-14
MF	GO:0005201	extracellular matrix structural constituent	112	2.68E-14

Table 6. The top 10 enriched GO terms of genes that produced differentially used transcripts

Ontology	ID	Functional Term	Gene Count	Adjusted p-value
CC	GO:0005925	focal adhesion	306	2.19E-33
CC	GO:0005924	cell-substrate adherens junction	306	4.56E-33
CC	GO:0030055	cell-substrate junction	309	4.56E-33
CC	GO:0005912	adherens junction	364	1.31E-32
CC	GO:0005813	centrosome	320	7.28E-19
CC	GO:0031252	cell leading edge	268	1.29E-18
BP	GO:0000226	microtubule cytoskeleton organization	314	3.02E-17
BP	GO:0043087	regulation of GTPase activity	285	3.02E-17
CC	GO:0005819	spindle	229	2.09E-16
BP	GO:0006914	autophagy	315	1.51E-15

We then look at genes that generated significant DE and DU transcripts. In total, 8371 genes produced all the significant DE transcripts compared to 8072 genes that generated such DU transcripts. A Venn diagram in Figure 5b exhibits the overlaps among genes from distinct analysis methods (i.e., DGE, DTE, and DTU). From Figure 5b, 3841 (45%) DE genes generated either significant DE transcripts or significant DU transcripts or both. We also explored the transcript-level changes in the 37 significantly mutated genes. We found that 19 of 37 (51.4%) genes produced DE transcripts, and 25 of such genes (67.6%) had DU transcripts (Table 1).
We also did GO enrichment analysis for the genes that generated DE transcripts and DU transcripts. GO enrichment analysis on DTE genes found that collagen-containing extracellular matrix (GO:0062023), focal adhesion (GO:0005925), cell-substrate adherens junction (GO:0005924), cell-substrate junction (GO:0030055), extracellular matrix (GO:0031012), adherens junction (GO:0005912), extracellular matrix organization (GO:0030198), extracellular structure organization (GO:0043062), actin cytoskeleton (GO:0015629), extracellular matrix structural constituent (GO:0005201) were the top 10 significantly enriched terms (Table 5). In contrast, the top 10 terms found from DTU genes were focal adhesion (GO:0005925), cell-substrate adherens junction (GO:0005924), cell-substrate junction (GO:0030055), adherens junction (GO:0005912), centrosome (GO:0005813), cell leading edge (GO:0031252), microtubule cytoskeleton organization (GO:0000226), regulation of GTPase activity (GO:0043087), spindle (GO:0005819), autophagy (GO:0006914) (Table 6).

4. DISCUSSION

This study of RNA-seq of human bladder reveals some crucial genes and transcripts as well as functional characteristics related to bladder cancer development. From the differential analysis of expression of gene and transcript and usage of the transcript, we identified potential biomarkers that may help in bladder cancer diagnosis, treatment, and prognoses.

Human bladder cancer is a type of disease that full of complex genetic causes. DGE analysis discovered 8543 differentially expressed genes that enriched in plenty of GO terms in distinct biological processes and molecular functions, while DTE analysis discovered 14350 transcripts originated from 8371 genes, and DTU analysis revealed 27914 transcripts from 8072 genes, all of which may contribute to the development of BC.

Although DGE analysis discovered a large range of genes related to BC, DTE analysis and DTU analysis provided a new dimension to explore cancer RNA-seq data. There were 4814 and 5765 novel genes found by DTE analysis and DTU analysis, respectively. Furthermore, changes in the expression of some DE genes were attributable to its transcripts. We identified ACTC1 as the

most down-regulated gene in bladder cancer, which was previously recognized as a commonly down-regulated gene in BC [26]. Gene ACTC1 had expressed three transcripts, i.e., ENST00000290378, ENST00000557860, and ENST00000560563, in the dataset. DTE analysis found that such three transcripts were significantly decreased (log2FC < -8.2, adjusted p-value < 7E-8). DTU analysis revealed that only transcript ENST00000560563 displayed a significant change in relative usage (log2FC = 1.3, adjusted p-value < 1E-2). A similar situation happened on two of the most suppressed genes, DES and PI16, where their transcripts' expression decreased in tumor and each had only one significant DU transcript. Another two genes, ASB5 and PCP4, however, showed a more complicated pattern. Although expression degrees at the gene level and the transcript level both inhibited in BC, some of their isoforms' usage presented an opposite direction of change. For example, isoform ENST00000510578 from gene ASB5 gained an increase in its relative usage level.

Differential analysis verified that some significantly mutated BC-related genes also experienced variations in expression (Table 1). DGE analysis found that 10 of 37 significantly mutated genes were differentially expressed. Moreover, extra genes were detected to contain either DE transcripts or DU transcripts (19 genes and 25 genes, respectively). It suggests that there may be an underlying link from gene mutation to gene expression and transcriptional composition that contributed to the development of bladder cancer.

This work was based on the analysis of RNA-seq data and revealed potential biomarkers associated with bladder cancer. Further expansion of the study may be an experimental validation to fortify and narrow the findings so that promising therapies can be derived.

5. CONCLUSION

In conclusion, we conducted both gene-level and transcript-level differential analyses on 42 bladder cancer samples, including differential expression analysis and differential usage analysis. Transcript level analysis results revealed details contributing to the significant changes in gene expression level. Furthermore, we discovered additional genes that didn't detect by gene-level analysis and may relate to the development of bladder cancer. We also did GO enrichment analysis based on the differential analysis results and disclosed candidate pathways that potentially associated with bladder cancer. Despite the analytical study we completed, experimental validation is expected to fortify our findings.

REFERENCE

[1] Bray, Freddie, et al. "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries." CA: a cancer journal for clinicians 68.6 (2018): 394-424.

[2] Kaufman, Donald S., William U. Shipley, and Adam S. Feldman. "Bladder cancer." The Lancet 374.9685 (2009): 239-249.

[3] Sylvester, Richard J., Willem Oosterlinck, and ADRIAN PM van der MEIJDEN. "A single immediate postoperative instillation of chemotherapy decreases the risk of recurrence in patients with stage Ta T1 bladder cancer: a meta-analysis of published results of randomized clinical trials." The Journal of urology 171.6 Part 1 (2004): 2186-2190.

[4] Mant, Jonathan, et al. "Systematic review and individual patient data meta-analysis of diagnosis of heart failure, with modelling of implications of different diagnostic strategies in primary care." NIHR Health Technology Assessment programme: Executive Summaries. NIHR Journals Library, 2009.

[5] Shariat, Shahrokh F., et al. "Outcomes of radical cystectomy for transitional cell carcinoma of the bladder: a contemporary series from the Bladder Cancer Research Consortium." The Journal of urology 176.6 (2006): 2414-2422.

[6] McAlpine, Kristen, et al. "Radiotherapy with radical cystectomy for bladder cancer: A systematic review and meta-analysis." Canadian Urological Association Journal 12.10 (2018): 351.

[7] Kim, Jaegil, et al. "Invasive bladder cancer: genomic insights and therapeutic promise." Clinical Cancer Research 21.20 (2015): 4514-4524.

[8] Inamura, Kentaro. "Bladder cancer: new insights into its molecular pathology." Cancers 10.4 (2018): 100.

[9] Abbosh, Philip H., and Elizabeth R. Plimack. "Molecular and clinical insights into the role and significance of mutated DNA repair genes in bladder cancer." Bladder Cancer 4.1 (2018): 9-18.

[10] Jung, Ichabod, and Edward Messing. "Molecular mechanisms and pathways in bladder cancer development and progression." Cancer control 7.4 (2000): 325-334.

[11] Pelucchi, Claudio, et al. "Mechanisms of disease: the epidemiology of bladder cancer." Nature Reviews Urology 3.6 (2006): 327.

[12] McConkey, David J., et al. "Molecular genetics of bladder cancer: Emerging mechanisms of tumor initiation and progression." Urologic Oncology: Seminars and Original Investigations. Vol. 28. No. 4. Elsevier, 2010.

[13] Ahmad, Imran, Owen J. Sansom, and Hing Y. Leung. "Exploring molecular genetics of bladder cancer: lessons learned from mouse models." Disease models & mechanisms 5.3 (2012): 323-332.

[14] Fantini, Damiano, et al. "A Carcinogen-induced mouse model recapitulates the molecular alterations of human muscle invasive bladder cancer." Oncogene 37.14 (2018): 1911.

[15] Hussain, Syed A., et al. "Gene expression profiling in bladder cancer identifies potential therapeutic targets." International journal of oncology 50.4 (2017): 1147-1159.

[16] Sanchez-Carbayo, Marta, Paola Capodieci, and Carlos Cordon-Cardo. "Tumor suppressor role of KiSS-1 in bladder cancer: loss of KiSS-1 expression is associated with bladder cancer progression and clinical outcome." The American journal of pathology 162.2 (2003): 609-617.

[17] Soneson, Charlotte, Michael I. Love, and Mark D. Robinson. "Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences." F1000Research 4 (2015).

[18] Guo, Guangwu, et al. "Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation." Nature genetics 45.12 (2013): 1459.

[19] Liao, Yang, Gordon K. Smyth, and Wei Shi. "The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote." Nucleic acids research 41.10 (2013): e108-e108.

[20] Patro, Rob, et al. "Salmon provides fast and bias-aware quantification of transcript expression." Nature methods 14.4 (2017): 417.

[21] Dobin, Alexander, et al. "STAR: ultrafast universal RNA-seq aligner." Bioinformatics 29.1 (2013): 15-21.

[22] Love, Michael I., Wolfgang Huber, and Simon Anders. "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." Genome biology 15.12 (2014): 550.

[23] Anders, Simon, Alejandro Reyes, and Wolfgang Huber. "Detecting differential usage of exons from RNA-seq data." Genome research 22.10 (2012): 2008-2017.

[24] Reyes, Alejandro, et al. "Drift and conservation of differential exon usage across tissues in primate species." Proceedings of the National Academy of Sciences 110.38 (2013): 15377-15382.

[25] Yu, Guangchuang, et al. "clusterProfiler: an R package for comparing biological themes among gene clusters." Omics: a journal of integrative biology 16.5 (2012): 284-287.

[26] Zaravinos, Apostolos, et al. "Identification of common differentially expressed genes in urinary bladder cancer." PloS one 6.4 (2011): e18135.