Published in final edited form as:

Urol Oncol. 2014 April ; 32(3): 327–336. doi:10.1016/j.urolonc.2013.06.014.

Differentiating progressive from nonprogressive T1 bladder cancer by gene expression profiling: Applying RNA-sequencing analysis on archived specimens

Xuanhui Sharron Lin, Ph.D.^a, Lan Hu, Ph.D.^b, Sandy Kirley, M.S.^a, Mick Correll, B.A., B.S.^b, John Quackenbush, Ph.D.^b, Chin-Lee Wu, M.D., Ph.D.^{a,c,**}, and William Scott McDougal, M.D.^{a,*}

^a Department of Urology, Massachusetts General Hospital and Harvard Medical School, Boston, MA

^b Department of Biostatistics and Computational Biology and Center for Cancer Computational Biology, Dana-Farber Cancer Institute, Boston, MA

^c Department of Pathology, Massachusetts General Hospital and Harvard Medical School, Boston, MA

Abstract

Objectives—To identify gene signatures in transitional cell carcinoma that can differentiate high-grade T1 nonprogressive (T1NP) bladder cancer (BCa) from those T1 progressive (T1P) tumors that progress to muscularis propria–invasive T2 tumors.

Materials and methods—We performed a high-throughput RNA sequencing (RNA-Seq) on formalin-fixed and paraffin-embedded BCa specimens with clinical pathologic characteristics best representing the general clinical development of the disease. For the T1NP group, only patients with long-term follow-up (6–17 y) and periodic examinations (average of 4 resections and 9 cytology tests) were selected. For the T1P group, only patients in whom a complete resection was performed after a minimum of 8 months after the initial T1 diagnosis were selected, therefore eliminating the possibility of underdiagnosis. Only samples in which muscularis propria was present and uninvolved were included, further assuring a correct diagnosis. The RNA-Seq reads were mapped to the human genome build NCBI 36 (hg18) using TopHat with no mismatch. After alignment to the transcriptome and expression quantification, a linear statistical model was built using Limma between T1NP and T1P samples to identify differentially expressed genes.

Results—Overall, 5,561 genes were mapped to all samples and used for RNA-Seq analysis to identify a gene signature that was significantly and differentially expressed between patients with T1NP BCa and patients with T1P BCa. Signature-based stratification indicated the gene signature correlated notably with the time of T1 development to T2 tumor, suggesting that the molecular signature might be used as an independent predictor for the pace of high-grade T1 BCa progression.

Conclusions—This is the first demonstration that RNA-Seq can be applied as a powerful tool to study BCa using formalin-fixed and paraffin-embedded specimens. We identified a gene signature that can distinguish patients diagnosed with high-grade T1 BCas that remain as non-muscle invasive tumors from those patients with cancers progressing to muscle-invasive tumors. Our findings will make future large-scale clinical cohort studies and clinical trial-based studies possible and help the development of prognostic tools for accurate prediction of T1 BCa progression that may considerably influence the clinical decision-making process, treatment regimen, and patient survival.

Keywords

T1 bladder cancer; RNA-Seq; Tumor progression; Molecular signature

1. Introduction

Management of high-grade T1 (lamina propria invasive) bladder cancer (BCa) poses a challenging dilemma to clinicians and patients. Most cases of patients with BCa present initially as non-muscle invasive transitional cell carcinoma (TCC). There are a significant number of high-grade T1 lesions that have the potential to progress to muscle-invasive BCa with increased risk for developing metastatic disease [1,2]. For high-grade T1 TCC, approximately 80% of treated tumors recur and 5% to 15% of recurrent tumors progress to invasive disease. Once BCa becomes metastatic, the cancer-related survival is approximately 5.4% at 5 years [3]. Currently, no definitive risk criteria exist to distinguish patients with T1 nonprogressive (T1NP) BCa who may suffer multiple recurrences of the disease without developing muscle-invasive tumors from those patients with the T1 progressive (T1P) BCa whose cancer first presented as T1 disease but eventually developed into muscle-invasive and metastatic disease. Therefore, establishing new prognostic criteria to distinguish high-grade T1NP from T1P would meet a great clinical need.

Previously, BCa prognostic factors have been reported and specific markers and molecular pathways have been linked with BCa tumor stage and progression [4–6]. Thus far, no standard guidelines have been adopted into clinical practice. Most of these studies used samples pooled from distinct groups of noninvasive BCa (Ta and T1) tumors and compared them with muscularis propria-invasive BCa (T2, T3, and T4) tumors. Specimens with long-term clinical T1 BCa follow-up were lacking. This makes distinguishing patients with T1NP BCa from patients with T1P particularly difficult, if not impossible from these studies.

Gene expression profiling has been used to develop prognostic signatures in a wide range of diseases [7,8], but its application has been limited, to an extent, by the fact that gene expression technologies work best with fresh frozen tissue [9–11]. By far, the vast majority of human disease tissue samples and those with the best outcome data are archived formalin-fixed paraffin-embedded (FFPE) surgical specimens. Important clinical information and disease outcome are often collected years after the initial specimen collection. However, FFPE tissue processing is known to cause fragmentation and chemical modification of RNA, presenting challenges for gene expression profiling. The analysis of BCa specimens is

potentially more difficult as they are typically obtained by transurethral resection (TUR) and are associated with cautery artifacts that may further degrade nucleic acids.

High-throughput RNA sequencing (RNA-Seq) is a recently developed gene expression profiling method that has several advantages over other expression profiling technologies, including higher sensitivities, low background noise, and the ability to detect splicing variants and somatic mutations, resulting in precise measurements of levels of transcripts and their isoforms [12,13]. Using RNA-Seq, expression profiling genes involved in various cancers were identified and showed potential for biomarker-driven clinical trials and targeted cancer therapy [14–16]. To this date, the RNA-Seq technology using FFPE samples on BCa analysis has not been published.

In this pilot study, our goal was to test whether RNA-Seq analysis with archival FFPE BCa specimens could be used to identify a genomic signature capable of differentiating high-grade BCa of T1NP from those T1 diseases that eventually progress to muscle-invasive tumors. We used FFPE samples of patients with BCa with long-term follow-up to determine the natural history of the disease. RNA-Seq analysis performed on these FFPE samples identified a gene expression signature associated with disease progression. Our results demonstrated the applicability of using RNA-Seq to study bladder tumors obtained by TUR and stored long term as FFPE specimens. The gene signature identified by this study may lead to a promising diagnostic tool for high-risk patients with tumors that rapidly progress to muscle-invasive BCa.

2. Materials and methods

2.1. Patients and tumor specimens

Our research was conducted with the approval by the hospital Institutional Review Board at Massachusetts General Hospital. All the samples were obtained from patients treated at Massachusetts General Hospital between 1994 and 2011. BCa tumor FFPE specimens from 3 patients with T1NP and 4 patients with T1P were analyzed for each experiment condition. An expert urologic pathologist (C.L.W) carefully evaluated specimens of papillary BCa, and the tumor tissues were identified, marked, and highlighted. Two cores with only papillary bladder tumors and without any stromal or muscular tissues were collected from highlighted areas using a biopsy punch with a plunger (1.5 mm in diameter) in a RNA-free environment. Cancer grade was assigned using the 1973 World Health Organization pathology (WHO) criteria for BCa, and cancer stage was assigned according to the seventh edition of American Joint Committee on Cancer TNM system [17,18].

2.2. RNA extraction, rRNA removal, and sequencing library construction

Total RNA was extracted using hot phenol with additional purification using the RNeasy mini kit (Qiagen, Germantown, MD) following the manufacturer's instructions. RNA integrity was assessed using an Agilent bioanalyzer (Agilent, Santa Clara, CA), and the RNA integrity number was calculated for each sample and the average RNA integrity number for all samples was 2.6, ranging from 2.3 to 3.3. A cDNA library was constructed for each sample using Illumina's mRNA-Seq Sample Prep Kit (Illumina San Diego, CA).

Briefly, for each sample, 100 ng of total RNA was used to generate a sequencing library and the RNA was directly subjected to fragmentation without the mRNA purification step. The resulting sample libraries were subjected to double-specific nuclease (DSN) treatment using the Trimmer-Direct cDNA Normalization kit (Evrogen, Moscow, Russia).

2.3. RNA-Seq read mapping and annotation

The cDNA library of each sample was loaded to a single lane of Illumina flow cell and the libraries were sequenced on Illumina Genome Analyzer II. Image deconvolution and calculation of quality value were performed using the Boat module (Firecrest v1.1.4.0 and Bustard v1.1.4.0 programs) of Illumina pipeline V1.4. Sequence base calls were assigned using Illumina CASAVA software. The sequence reads were 36 bases long and each lane produced on an average 30 million of 36-mer raw sequence reads. Reads were mapped to the human genome build NCBI 36 (hg18) using TopHat [19] with no mismatch. The mapped reads were assembled and annotated using Cufflinks software tools [20].

2.4. Transcript quantification and gene expression consolidation

Transcript abundances were quantified in fragments per kilobase of exon per million fragments mapped by Cufflinks by taking into account both the gene length and the mapped reads for each sample and normalized accordingly [21]. Owing to the fragmented nature of mRNA in FFPE samples, we focused on the abundance measurement at the gene level. When multiple transcript abundance measurements were reported for a gene, the maximum value was chosen to represent the expression level of that gene.

2.5. Differentially expressed genes identification in RNA-Seq data

After alignment to the transcriptome and expression quantification, a linear statistical model was built using limma [21,22] between 3 T1NP samples and 4 T1P samples to identify differentially expressed genes with T1NP samples as the reference. The analyses were accomplished using R and Bioconductor packages [23]. To validate our gene signature, Illumina cDNA-mediated Annealing, Selection, extension, and Ligation Cancer Panel (DASL) Assay was performed independently with the same specimens and genes were identified to show significantly differential expressions between patients with T1NP and patients with T1P using both the RNA-Seq and the DASL platforms.

2.6. Functional enrichment and network analysis

Network enrichment for the significantly differentially expressed genes was analyzed using Ingenuity Pathway Analysis (IPA) software. The network interaction of the focused genes in the network is based on their connectivity in Ingenuity Knowledge Base.

3. Results

3.1. Patient cohort characteristics

To ensure the high specificity of the gene signature obtained, we selected tissue samples from patients with the clinicopathologic characteristics that best represented the general clinical development of nonprogressive from progressive disease. We selected 7 patients in

this study based on several considerations. First, all patients had well-established lamina propria–invasive TCC confirmed by an expert urologic pathologist. Second, every patient's disease history and his/her pathology report was carefully evaluated to ascertain BCa progression. Third, only patients with extended follow-up and sufficient tumor cells in the paraffin blocks were included in this study. All specimens had muscularis propria present that was uninvolved. Patients had a short-interval follow-up biopsy after initial diagnosis, which demonstrated no tumor. As shown in Table 1, for the T1NP group, these 3 patients were confirmed as non-progressive disease by subsequent biopsies and with significantly extended follow-up times, ranging from 6 to 17 years (average 9 y). For patients with T1P, we excluded any cases with short interval between T1 and T2 diagnosis, eliminating the possibility of T1 BCa being understaged. The 4 patients with T1P had an average time to development of muscularis propria–invasive cancer of 2.43 years, ranging from 0.8 to 4.5 years and with an average follow-up time of 4.75 years, ranging from 1 to 8 years. All these data coincided with the general clinical observation of T1 BCa progression.

3.2. RNA-Seq analysis of T1NP and T1P tumors

We performed RNA-Seq using the Illumina GAII platform. The study design and the workflow for the RNA-Seq are illustrated in Fig. 1 and are described in detail in Materials and Methods. After transcript quantification, 11,092 genes were detected in at least 1 of the 7 samples, with 6,143 genes with multiple transcripts and 4,929 genes with one transcript. An unbiased analysis of the expression data revealed that 5,561 genes were found to be expressed in all samples and it is this final set that was used for further analysis. The characteristics of the RNA-Seq data are summarized in Table 2. On average, approximately 21 million sequencing reads were generated covering 47 million exon bases. Not all the genes could be mapped to the transcribed database, likely owing to genetic variations and repetitive elements and tandem repeats [24]. The average number of genes mapped was 8,149, which is substantially smaller than what has been reported when using RNA derived from fresh or frozen samples [24].

3.3. Identification of differentially expressed genes in T1NP compared with T1P samples

We used Limma [22] to construct a linear model and identified 181 significantly differentially expressed genes (Table 3; $P < 0.05$) between the nonprogressive and progressive patients whose original diagnosis was high-grade T1 BCa. Among them, we found 101 up-regulated and 80 down-regulated genes in T1P relative to T1NP. The results were validated with the DASL Cancer Panel analysis using the same samples. We then performed average-linkage hierarchical clustering using a Pearson correlation coefficient distance metric using the gene signature identified (Fig. 2A). The green bar indicates 3 T1NP samples and the red bar for 4 T1P samples. Rows are median-centered $\log_2$ (fragments per kilobase of exon per million fragments mapped) for each gene with relative high levels of expression in red and low levels in blue. On examining the hierarchical clustering dendrogram, one can see a correlation between gene expression levels and the time for disease progression for the 4 T1P samples. As shown in Fig. 2B, it took 0.8, 1.4, 3, and 4.5 years for patient BLK25, BLK20, BLK34, and BLK21, respectively, for the disease to progress from T1 to muscle-invasive BCa. This suggests that the biomarker we identified may also help to predict the pace of progression from high-grade T1 BCa to muscle-invasive

BCa, although a much larger sample size would be necessary to validate this observation. However, if validated, this result could have a direct effect on clinical decision-making. To further understand the biology underlying the patterns of differential expression, we used IPA software, to search for overrepresentation of biological pathways and annotated gene functional classes among the genes found to be significantly differentially expression between patients with T1P and T1NP and the top enriched network associated with T1 BCa progression was shown in Fig. 3.

4. Discussion

Genomics has created an unprecedented opportunity to survey expression patterns across the genome and to use the resulting data to develop diagnostic and prognostic bio-markers. However, doing this requires the availability of well-annotated clinical samples with extensive clinical data so that patterns of gene expression can be linked to outcome or other relevant end points.

Although there are many archival pathologic samples reserved as FFPE tissues, these have proven difficult to analyze using most of the available genomic technologies. This is largely owing to the fact that the process of creating FFPE samples is known to introduce chemical modification and cross-linking between DNA, RNA, and proteins in these samples. In BCa, the situation has proven particularly difficult as specimens are typically obtained through TUR and the cautery effect associated with the procedure may further degrade nucleic acids. To overcome these limitations, we used a variation on RNA-Seq technology that relies on short nucleic acid fragments coupled with DSN normalization. DSN removes ribosomal RNA and other abundant double-stranded DNA and DNA-RNA hybrid complexes, allowing creation of RNA-Seq libraries from most highly degraded samples [25,26]. Using this approach, we identified a gene signature that significantly differentially expressed between the nonprogressive and progressive T1 BCa (Table 3). Our data demonstrated that FFPE specimens obtained by TUR could be used in RNASeq, genome-wide expression analysis.

In our study, we went to great lengths to select samples that had an appropriate diagnosis and were of high quality. It is known that up to 25% of T1 G3 tumors are incorrectly staged on initial surgical resection (TUR of bladder [TURB]). Studies without systematically repeated TURB could have been hampered by the lack of patients with correctly documented T1 who have tumors that subsequently progress at a later date. It is therefore critical that a re-resection be performed, particularly if muscularis propria is not present in the specimen. In our study, a complete resection was performed for each and all patients with a minimum of at least 8 months after the initial T1 diagnosis. Only samples in which muscularis propria was present and uninvolved were included, further assuring a correct diagnosis. For each patient included in our study, there were longitudinal data, including TURB and cytology tests (between 14 and 24, ranging from 2–40), which followed the original diagnosis (data not shown). For the non-progressive group, one of the most critical issues is the length of the follow-up. In our study, patients with T1NP were followed up to 17 years with periodic examinations (average being more than 4 resections and 9 cytology tests), providing strong evidence to support our assignment of nonprogressive status. For the progressive group, the important issue was that a second resection should be performed

following an initial T1 diagnosis without evidence of T2 disease. In our study, a sample was required to have a second resection to be accepted as T1 progression, therefore eliminating the possibility of underdiagnosis. Only with careful exclusion of patients not fitting these criteria can one be certain of a true T1 diagnosis and not a misdiagnosed T2 BCa.

Previous studies of superficial papillary BCa have associated specific genomic alterations with the disease, including mutations in and dysregulation of FGFR3, PI3K, KRAS, HRAS, TP53, P16, TSC1, and PTEN and loss of chromosome 9, 9p, or 9q as well as loss of RB1 [9,10,27–29]. Recently, mutations in genes responsible for chromatin remodeling were identified [30], and high-throughput assays including RNA-Seq performed on fresh tissue samples have shown promising results, demonstrating molecular assessment can be used for the development of mutation and pathway-based trials of targeted therapy for cancer patients [16]. Ultimately, a list of genes provides little insight into the biology underlying the patterns of differential expression. Consequently, we used IPA software to search for overrepresentation of biological pathways and annotated gene functional classes among the genes found to be significant. Among the highest ranking were cancer-related pathways associated with cell death, cellular growth and proliferation, and cell cycle (Fig. 3). Future studies are needed to understand the involvement of these molecules in BCa progression. In this study, we selected the portion of BCa tissue that was primarily epithelial cells and limited any stromal contaminant. Even so, it is conceivable that a small amount of stromal contamination could have occurred, but in view of the small volume, it is unlikely that it influenced the results.

In this exploratory study, we were able to identify a gene signature that was able to distinguish patients with T1NP from patients with T1P. Although encouraging, this study has some limitations including a small number of samples and a potential undersampling of tumor suppressor genes that may have been lost in the T1P patient group. Regardless, the fact that we are able to find an apparently robust signature that seems to track with progression is encouraging and suggests a more thorough prospective study is in order. We hope our study would have an effect on T1 BCa clinical management, other bladder diseases, and different types of cancers, and ultimately, play a role in facilitating the use of biomarker and gene signatures in the clinical settings.

5. Conclusions

This study identifies a gene signature that can distinguish patients diagnosed with high-grade T1 BCa that remain as non-muscle invasive tumors from those patients with cancers progressing to muscle-invasive tumors. This is the first demonstration that RNA-Seq can be applied as a tool to study BCa using FFPE specimens. Our findings could influence clinical T1 BCa treatment regiment and thereby effect patient survival.

Acknowledgments

We thank Renee Rubio and Fieda Abderazzaq, Department of Biostatistics and Computational Biology and Center for Cancer Computational Biology, Dana-Farber Cancer Institute for excellent technical support.

References

1. Cookson MS, Hen HW, Zhang ZF, et al. The natural history of high risk superficial bladder cancer: 15 year outcome. *J Urol*. 1997; 158:62–7. [PubMed: 9186324]
2. William SG, Stein JP. Molecular pathways in bladder cancer. *Urol Res*. 2004; 32:373–85. [PubMed: 15551095]
3. (<http://seer.cancer.gov/statfacts/html/urinb.html#survival>)
4. Sanchez-Carbayo M, Socci ND, Lozano J, Saint F, Cordon-Cardo C. Defining molecular profiles of poor outcome in patients with invasive bladder cancer using oligonucleotide microarrays. *J Clin Oncol*. 2006; 24:778–89. [PubMed: 16432078]
5. Schiffer E, Vlahou A, Petrolekas A, Stravodimos K, Tauber R, Geschwend JE, et al. Prediction of muscle-invasive bladder cancer using urinary proteomics. *Clin Cancer Res*. 2009; 15:4935–43. [PubMed: 19602546]
6. Wiklund ED, Bramsen JB, Hulf T, Dyrskjot L, Ramanathan R, Hansen TB, et al. Coordinated epigenetic repression of the miR-200 family and miR-205 in invasive bladder cancer. *Int J Cancer*. 2011; 128:1327–34. [PubMed: 20473948]
7. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*. 1999; 286:531–7. [PubMed: 10521349]
8. Stuart RO, Wachsman W, Berry CC, et al. In silico dissection of cell-type-associated patterns of gene expression in prostate cancer. *Proc Natl Acad Sci USA*. 2004; 101:615–20. [PubMed: 14722351]
9. Blaveri E, Brewer JL, Roydasgupta R, et al. Bladder cancer stage and outcome by array-based comparative genomic hybridization. *Clin Cancer Res*. 2005; 11:7012–22. [PubMed: 16203795]
10. Blaveri E, Simko JP, Korkola JE, et al. Bladder cancer outcome and subtype classification by gene expression. *Clin Cancer Res*. 2005; 11:4044–55. [PubMed: 15930339]
11. Dyrskjot L, Kruhoffer M, Thykjaer T, et al. Gene expression in the urinary bladder: a common carcinoma in situ gene expression signature exists disregarding histopathological classification. *Cancer Res*. 2004; 64:4040–8. [PubMed: 15173019]
12. Mardis ER. The impact of next-generation sequencing technology on genetics. *Trends Genet*. 2008; 24:133–41. [PubMed: 18262675]
13. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009; 10:57–63. [PubMed: 19015660]
14. Huang Q, Lin B, Liu H, Ma X, Mo F, Yu W, et al. RNA-Seq analyses generate comprehensive transcriptomic landscape and reveal complex transcript patterns in hepatocellular carcinoma. *PLoS One*. 2011; 6:e26168. [PubMed: 22043308]
15. Berger MF, Levin JZ, Vijayendran K, Sivachenko A, Adiconis X, et al. Integrative analysis of the melanoma transcriptome. *Genome Res*. 2010; 20:413–27. [PubMed: 20179022]
16. Roychowdhury S, Iyer MK, Robinson DR, Lonigro RJ, Wu YM, Cao X, et al. Personalized oncology through integrative high-throughput sequencing: a pilot study. *Sci Transl Med*. 2011; 3:1–10.
17. Mostofi, FK.; Sobin, LH.; Torloni, H. Histological Typing of Urinary Bladder Tumours. World Organization; Geneva: International histology classification of tumours, No. 10.; p. 1973
18. Edge, SB.; Byrd, DR.; Compton, CC.; Fritz, AG.; Greene, FL.; Trotti, A., editors. American Joint Committee on Cancer (AJCC) cancer staging manual. 7th ed.. Springer-Verlag; New York, NY: p. 2010
19. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009; 25:1105–11. [PubMed: 19289445]
20. Trapnell C, Williams BA, Pertea G, Mortazavi AM, 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; 28:511–5. [PubMed: 20436464]
21. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*. 2008; 5:621–8. [PubMed: 18516045]

22. Smyth, GK. Limma: linear models for microarray data. *Bioinformatics and Computational Biology Solutions using R and Bioconductor*. Gentleman, R.; Carey, VJ.; Huber, W.; Irizarry, RA.; Dudoit, S., editors. Springer; New York: 2005. p. 397-420.
23. (Available at: <http://www.bioconductor.org/>).
24. Li B, Ruotti V, Stewart RM, Thomson JA, Dewey CN. RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics*. 2010; 26:493–500. [PubMed: 20022975]
25. Toung JM, Morley M, Li M, Cheung VG. RNA-sequence analysis of human B-cells. *Genome Res*. 2011; 21(6):991–8. [PubMed: 21536721]
26. Rupp GM, Locker J. Purification and analysis of RNA from paraffin-embedded tissues. *Biotechniques*. 1988; 6:56–60. [PubMed: 2483655]
27. Shujun L, Khrebtukova I, Perou C, Schroth GP. Complete RNA-seq analysis of cancer transcriptomes from FFPE samples. *Genome Biol*. 2010; 11(Suppl. 1):P35.
28. Lindgren D, Frigyesi A, Gudjonsson S, Sjödal G, Hallden C, Chebil G, et al. Combined gene expression and genomic profiling define two intrinsic molecular subtypes of urothelial carcinoma and gene signatures for molecular grading and outcome. *Cancer Res*. 2010; 70:3463–72. [PubMed: 20406976]
29. McConkey DJ, Lee S, Choi W, Tran M, Majewski T, Lee S, et al. Molecular genetics of bladder cancer: emerging mechanisms of tumor initiation and progression. *Urol Oncol*. 2010; 28:429–40. [PubMed: 20610280]
30. Gui Y, Guo G, Huang Y, Hu X, Tang A, Gao S, et al. Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nat Genet*. 2011; 43:875–8. [PubMed: 21822268]

Fig. 1.
Study design and RNA-Seq analysis workflow.

Fig. 2.

Heatmap and sample clustering analysis of the significantly differentially expressed genes between 3 T1NP and 4 T1P samples. Hierarchical cluster analysis of RNA-Seq data profiling bladder tissue obtained from 3 patients with T1NP of known BCa with nonprogressive status and 4 patients with T1P BCa with known muscle-invasive progressive disease, first diagnosed as nonprogressive (T1) and then progressed to T2 progressive BC. There were 181 genes significantly differentially expressed between 3 T1NP and 4 T1P samples with $P < 0.05$. (A) Each column represents a sample and each row a gene. Red represents a higher level of gene expression and blue represents a lower level gene expression, relative to the median across all samples for each gene. The green bar above the heatmap represents T1NP samples and red bar T1P samples. (B) the hierarchical clustering dendrogram from data obtained from the 4 T1P samples shown in (A) was plotted regarding the time for T1 progression to muscle-invasive tumor. For patients with BLK21, BLK20, BLK24, and BLK34, it took 4.5, 1.4, 0.8, and 3 years for their disease to progress from T1 to T2 (muscle invasive) BCa. y-Axis indicates the time (years) for the progression from T1 to muscle-invasive BCa.

Fig. 3.

The top enriched network associated with BCa progression. Map was created by Ingenuity Pathway Analysis (IPA) software. Red icons depict genes up-regulated and green icons depict gene down-regulated. Lines report interactions between proteins.

Table 1

Clinical pathologic characteristics of study cohort

	T1 nonprogressive (T1NP)	T1 progressive (T1P)
Samples	<i>n</i> = 3	<i>n</i> = 4
Median age, y (range)	69 (56–79)	73 (62–85)
Male:female	2:1	3:1
Clinical tumor stage	T1	T1
T1	3	4
Grade 1–2	0	0
Grade 3	3	4
Average follow-up years (range)	9 (6–17)	4.75 (1–8)
Average progression years (range)	0	2.43 (0.8–4.5)

Table 2

RNA-Seq data summary. The read coverage of the 3 T1NP and 4 T1P, patients' samples were shown.

Sample ID	Block age (y)	BCa group	Number of reads	Number of detected genes	Number of exon bases covered
BLK004	11	NP	21,879,070	8,790	42,922,505
BLK007	16	NP	26,555,690	9,092	47,994,767
BLK016	10	NP	33,927,499	7,567	43,266,436
BLK020	5	P	10,888,159	7,358	50,383,266
BLK021	15	P	25,313,779	8,935	48,868,444
BLK025	6	P	18,278,586	8,789	35,467,184
BLK034	6	P	12,497,222	6,513	61,692,969

Table 3

Annotation of significantly differentially expressed genes between patients with T1P BCa and patients with T1NP BCa

Gene symbol	log FC(P/NP)	t	P value
IGFBP5	5.28378713	6.394460723	0.000213397
LSP1	4.171875655	5.476182971	0.000596924
STIM1	3.420212531	4.654638098	0.001649362
APOL4	-3.221334132	-4.639699078	0.001681636
CCPG1	2.172448093	4.391558528	0.002331371
ANTXR2	2.254543518	4.232024004	0.002890112
C10orf76	-3.818675576	-4.214632711	0.002959271
ABCA5	-2.381645938	-4.029101417	0.003819003
TBC1D4	2.729309403	3.930572519	0.004381905
OPTN	2.1810596	3.889946098	0.004639408
CYP4Z2P	-3.076981378	-3.849316111	0.004913234
VEGF	-2.408409244	-3.748738239	0.00566838
MDK	-2.212910642	-3.745344698	0.005695932
CYP4B1	-3.622571589	-3.717678918	0.005925963
AGPS	2.694819961	3.687937586	0.006184361
NLRP1	3.727425324	3.64172336	0.006610026
NFIA	1.838046144	3.604198172	0.006978631
CDC42BPG	-2.284311666	-3.549514259	0.007555489
SH3D19	1.911486029	3.548235041	0.007569576
KALRN	-3.728133666	-3.544078123	0.007615546
ITGA2	2.800984034	3.511152117	0.007990317
PLA2G2F	-4.156008011	-3.508019408	0.008026982
ALDH16A1	-2.862696914	-3.505490848	0.008056706
LTBP1	2.656543111	3.478779402	0.008377941
DHRS7	2.279694661	3.462326805	0.008582544
5S_rRNA	5.028224923	3.436200666	0.008918409
CASP8	2.615453795	3.415985067	0.009187841
PDLIM5	1.708273735	3.39620038	0.009459889
TRAF4	-2.350257868	-3.381925006	0.00966147
RAB31	2.357119785	3.370061392	0.009832451
ANXA7	2.192966094	3.324721288	0.010515958
CAMK2N1	-2.520187721	-3.293483171	0.011015952
PICK1	-2.627547988	-3.274092328	0.011338883
AUH	1.566006828	3.258177995	0.01161137
ACVR1	2.242422498	3.245297375	0.011836961
HLA-K	2.547641918	3.2147809	0.012390042
RNF115	-1.854308544	-3.17808107	0.013091395
RNF145	1.921960747	3.165520893	0.013340906

Gene symbol	log FC(P/NP)	t	P value
AMD1	1.872288678	3.157376316	0.013505363
F3	3.925845101	3.152687846	0.013600998
SLC19A2	-2.408176319	-3.14173746	0.013827139
MAN2A1	1.853521362	3.141501987	0.013832045
SNX14	2.178844054	3.10806374	0.014547586
LY75	1.982609555	3.089435701	0.014962925
TCIRG1	1.519819127	3.078512561	0.015212227
TOR1B	1.645119041	3.057636916	0.01570084
PTGR1	-2.445310476	-3.044717278	0.016011429
CRISPLD2	2.737330915	3.035132503	0.016245991
BHLHE41	-2.305189141	-3.028038491	0.016421907
SIPA1L2	1.830611683	3.025830923	0.016477054
CEP164	3.297527769	3.008038912	0.016928636
PAPSS1	1.73903819	2.975601371	0.017785457
GNPDA1	1.762492611	2.933743377	0.018958336
LONP1	-1.56359523	-2.933095568	0.018977107
KIAA0182	-1.622904521	-2.913291303	0.019560385
LIG1	-2.313584528	-2.906708453	0.019758362
CD44	4.002931343	2.898246336	0.020015917
PLEKHG6	-1.982202313	-2.880303583	0.020573615
FAM46A	2.044259415	2.880296689	0.020573832
PPFIBP1	1.789928588	2.879589157	0.020596151
RP11-345P4.4	-1.757416661	-2.875835802	0.020714972
POLE	-2.623991911	-2.851535329	0.021501641
PON2	2.032836251	2.826353505	0.022349529
HMG3	1.72280555	2.825036704	0.022394807
COL7A1	3.970018254	2.813795234	0.022785223
C19orf33	1.440515121	2.792784855	0.023533875
POFUT2	-1.487815721	-2.787720939	0.023718083
AKAP1	-1.803718104	-2.786099847	0.023777367
SLC35D1	1.466010414	2.777427063	0.02409714
ASCC2	-2.033023461	-2.773047563	0.024260298
CCND1	3.259008872	2.772126521	0.024294756
BTN3A3	1.72513362	2.762907815	0.024642439
TAF10	1.657211243	2.759334651	0.024778575
SIPA1L3	-1.669399938	-2.745252431	0.025322692
ANXA1	4.206731688	2.745215675	0.025324128
TINF2	1.890822482	2.743158705	0.025404634
CSPG2	3.90783955	2.73594918	0.025688889
KIAA0415	-2.105622992	-2.730688823	0.025898357
ITGB7	-1.908655609	-2.725256153	0.02611653
NUMB	1.394564947	2.722490684	0.026228315

Gene symbol	log FC(P/NP)	t	P value
GPC1	2.378547736	2.703836999	0.026995286
C20orf194	2.388977829	2.678685101	0.02806606
GGA1	-1.922870984	-2.672079037	0.028354455
PAK1	1.547808712	2.663626083	0.028727912
SLC12A6	1.791246724	2.660450487	0.028869511
DDB2	-1.762351611	-2.659926863	0.028892928
SAMD9L	2.379492442	2.657272638	0.029011927
UNC13B	-1.938490147	-2.654444283	0.029139285
RNF141	1.824007956	2.64865981	0.02940154
PSEN1	1.614933918	2.644278862	0.029601769
MAPKAPK3	2.450534752	2.640518671	0.029774739
TMEM201	-1.59423922	-2.619510743	0.030760314
MORC2	-1.628752784	-2.618769431	0.030795695
NEDD9	-1.961975214	-2.618627696	0.030802465
GATA2	-2.511628364	-2.618350233	0.030815721
FBXL5	1.345326223	2.607768178	0.031325674
FARP2	1.725692393	2.60722749	0.03135196
AARS2	-1.977567987	-2.60127093	0.03164304
FAM73B	-2.597450137	-2.59632296	0.031886929
AKR1C3	-1.563447818	-2.595660351	0.031919735
DOT1L	-1.418412546	-2.590124605	0.03219516
DHX8	-1.793769773	-2.585869975	0.03240849
C9orf84	2.742453953	2.581254025	0.032641569
DSCR1	1.822394025	2.570407591	0.033195996
RECQL5	-1.830666864	-2.563814258	0.033537701
REEP3	1.698774983	2.562182046	0.033622843
MAP3K7IP2	1.596373857	2.559860853	0.033744304
SS18L1	-1.39546114	-2.558763081	0.033801903
SKAP2	1.821616802	2.549410379	0.034296703
FKBP3	1.642847336	2.546576449	0.03444808
SLC25A24	1.433158668	2.544613401	0.034553337
PDE4DIP	-2.144480334	-2.541263209	0.034733726
KLHL22	-1.580374914	-2.539279442	0.034840992
DST	2.17427338	2.536334286	0.035000864
AHCTF1	1.574229907	2.532632073	0.035202889
CEACAM1	-1.726409623	-2.531454922	0.035267373
ENGASE	-2.276362919	-2.525395678	0.035601197
DCBLD1	1.875591159	2.521535273	0.035815552
RPL36AL	1.54280083	2.511514843	0.036378087
PLSCR3	1.420903374	2.51142349	0.036383256
PPP2R5C	1.307050755	2.510440903	0.036438906
GRB7	-1.720623113	-2.508716287	0.03653679

Gene symbol	log FC(P/NP)	t	P value
UGCGL2	-2.647291621	-2.50680118	0.036645799
CARD14	-1.409730182	-2.502937857	0.036866705
USP36	-1.252316221	-2.491910656	0.037504699
C22orf9	1.721359342	2.486442274	0.037825219
SH3BGRL	2.009799067	2.48497519	0.03791168
TNFAIP2	-1.661063451	-2.484647383	0.037931027
MAP3K7	1.282293995	2.479406568	0.038241689
OTUD7B	-1.896353997	-2.476276125	0.038428481
PDSS2	-2.338248486	-2.476129154	0.038437273
MGST2	1.462079037	2.472563455	0.038651211
NUPL2	1.628957352	2.467877477	0.038934198
DPYSL3	1.91576424	2.464112227	0.0391631
ZBTB4	1.21710635	2.463520042	0.039199225
ACOT11	-2.070513114	-2.462763559	0.039245421
HEG1	2.584471663	2.460518133	0.039382866
UGT1A8	2.849211631	2.458255547	0.039521854
PAN2	-2.293607666	-2.455013779	0.039721857
RAD51L3	-1.29829818	-2.45280216	0.039858891
CYLD	1.374783471	2.448055195	0.040154632
CRAT	-2.755364275	-2.448014399	0.040157183
ACACA	-1.275498781	-2.445917734	0.040288521
HDAC5	-1.451246253	-2.441489292	0.040567351
CHAF1A	-1.874461565	-2.427956095	0.041431578
PNMA1	1.646612223	2.427569109	0.041456562
UBE2D3	1.170890055	2.426386474	0.041533008
ZNF704	-2.638607084	-2.424684105	0.041643299
FAM117B	1.969329664	2.421711392	0.041836601
ACOX3	-1.370497072	-2.421331019	0.041861401
SETD7	1.668513412	2.415543953	0.042240531
C4orf34	1.345863238	2.411436977	0.042511691
IL6ST	1.491818909	2.41008791	0.042601144
PTEN	1.251488632	2.408825886	0.042684997
FAM179B	1.536096482	2.405452694	0.04290994
CBLC	-1.352667824	-2.404041938	0.04300437
HOOK2	-1.902609985	-2.403003647	0.043074003
CGN	-2.341892273	-2.401814619	0.043153884
EZH2	-1.845958401	-2.390117617	0.043947702
XPNPEP3	-1.735687018	-2.388371373	0.044067465
LSG1	-1.353533592	-2.382805061	0.044451413
CCDC109A	1.486079427	2.382529921	0.044470479
GPT2	-2.301677625	-2.378141626	0.044775671
CACNA1D	-3.552311409	-2.371912571	0.045212499

Gene symbol	log FC(P/NP)	t	P value
AFMID	-1.738027331	-2.364302478	0.045751988
ZNF330	1.623902688	2.359784763	0.046075305
DGKZ	-1.828659304	-2.357378201	0.046248469
CSAD	-2.268177751	-2.354431508	0.046461388
ESD	1.479256457	2.351836341	0.04664972
DOPEY2	-1.502561646	-2.34891318	0.046862771
GPRC5A	-1.80242518	-2.342122595	0.047361468
C6orf134	-1.73080824	-2.340904003	0.047451521
MARCH5	1.421978347	2.337675959	0.047690903
CD164	1.263919165	2.337024623	0.04773935
LFNG	1.815659694	2.333772259	0.047982005
BIRC2	2.272732375	2.321656288	0.048896876
AEBP1	3.476393655	2.319486911	0.049062517
CPSF3	1.433022399	2.319243225	0.049081158
COPB2	-1.174259775	-2.318070989	0.049170931
RPS23	1.364880755	2.31757051	0.049209308
NARF	-2.194601067	-2.307580216	0.04998168

Annotation of significantly differentially expressed genes between the nonprogressive and progressive groups of patients with T1 BCa.

The 181 most significantly differentially expressed genes ($P < 0.05$) that can distinguish high-grade T1NP tumors with nonprogressive recurrence from those T1P muscle-invasion tumors are listed. The genes were sorted based on the P value and their expression levels were indicated by the log FC (P/NP) values, genes that overexpressed were with positive log FC (P/NP) values and genes that underexpressed were with negative log FC (P/NP) values.