Using Artificial Intelligence and Machine Learning Algorithms with Gene Expression Profiling to Predict Recurrence of Nonmuscle Invasive Urothelial Carcinoma of the Bladder at Initial Presentation

Georg Bartsch, Jr.,* Anirban P. Mitra,* Sheetal A. Mitra, Arpit A. Almal, Kenneth E. Steven, Donald G. Skinner, David W. Fry, Peter F. Lenehan, William P. Worzel and Richard J. Cote†

From the University of Southern California (GB, APM, SAM, DGS), Los Angeles, California, Goethe University Frankfurt (GB), Frankfurt, Germany, Everist Genomics (AAA, DWF, PFL, WPW), Ann Arbor, Michigan, University of Copenhagen (KES), Copenhagen, Denmark, and University of Miami (RJC), Miami, Florida

Purpose: Due to the high recurrence risk of nonmuscle invasive urothelial carcinoma it is crucial to distinguish patients at high risk from those with indolent disease. In this study we used a machine learning algorithm to identify the genes in patients with nonmuscle invasive urothelial carcinoma at initial presentation that were most predictive of recurrence. We used the genes in a molecular signature to predict recurrence risk within 5 years after transurethral resection of bladder tumor.

Materials and Methods: Whole genome profiling was performed on 112 frozen nonmuscle invasive urothelial carcinoma specimens obtained at first presentation on Human WG-6 BeadChips (Illumina®). A genetic programming algorithm was applied to evolve classifier mathematical models for outcome prediction. Cross-validation based resampling and gene use frequencies were used to identify the most prognostic genes, which were combined into rules used in a voting algorithm to predict the sample target class. Key genes were validated by quantitative polymerase chain reaction.

Results: The classifier set included 21 genes that predicted recurrence. Quantitative polymerase chain reaction was done for these genes in a subset of 100 patients. A 5-gene combined rule incorporating a voting algorithm yielded 77% sensitivity and 85% specificity to predict recurrence in the training set, and 69% and 62%, respectively, in the test set. A singular 3-gene rule was constructed that predicted recurrence with 80% sensitivity and 90% specificity in the training set, and 71% and 67%, respectively, in the test set.

Conclusions: Using primary nonmuscle invasive urothelial carcinoma from initial occurrences genetic programming identified transcripts in reproducible fashion, which were predictive of recurrence. These findings could potentially impact nonmuscle invasive urothelial carcinoma management.

Key Words: urinary bladder neoplasms; neoplasm recurrence, local; genome; algorithms; software

Abbreviations and Acronyms

C_t = cycle threshold
EORTC = European Organisation for Research and Treatment of Cancer
GP = genetic programming
NMIUC = nonmuscle invasive urothelial carcinoma
PCR = polymerase chain reaction
qPCR = quantitative PCR
TURBT = transurethral bladder tumor resection

Accepted for publication September 21, 2015.
No direct or indirect commercial incentive associated with publishing this article.
The corresponding author certifies that, when applicable, a statement(s) has been included in the manuscript documenting institutional review board, ethics committee or ethical review board study approval; principles of Helsinki Declaration were followed in lieu of formal ethics committee approval; institutional animal care and use committee approval; all human subjects provided written informed consent with guarantees of confidentiality; IRB approved protocol number; animal approved project number.
Supported by National Cancer Institute Grants CA132499, CA86871 and CA71921.
* Equal study contribution.
† Correspondence: Department of Pathology, University of Miami Miller School of Medicine, 1120 Northwest 14th St., Suite 1416, Miami, Florida 33136 (telephone: 305-243-1111; FAX: 305-243-1115; e-mail: RCote@med.miami.edu).

THE 2 major parameters for risk stratification of NMIUC of the bladder include tumor histological grading and pathological staging.[1] These parameters may help assess risk but they do not predict outcomes in individual

0022-5347/16/1952-0001/0
THE JOURNAL OF UROLOGY®

http://dx.doi.org/10.1016/j.juro.2015.09.090
Vol. 195, 1-6, February 2016
Printed in U.S.A.

patients.[2] Molecular alterations in cancerous tissues precede visually identifiable morphological changes and are responsible for biological behavior, prognosis and response to therapy.[3]

Many different molecular pathways are involved in tumor development, including cell cycle regulation, cell death, cell growth, gene regulation, signal transduction, angiogenesis and invasion.[4] Of note low grade NMIUCs are associated with *HRAS* and *FGFR3* mutations, homozygous deletions of $P16^{INK4a}$ and loss of heterozygosity of chromosome 9p.[5,6] Alterations in *p53* and retinoblastoma pathways have been reported in high grade NMIUC and carcinoma in situ.[7] However, to our knowledge no single molecular marker has been identified as significant enough for clinical implementation to determine NMIUC recurrence. Genetic marker panels may have higher power to predict recurrence.[8] Our prior efforts using pathway based approaches have led to the identification of marker panels that can predict prognosis in NMIUC and invasive disease.[9,10]

However, it has become increasingly clear that unbiased interrogation of the entire genome using high throughput technologies can identify novel prognostic markers in such scenarios.[11] In this study we used whole genome expression analysis with artificial intelligence based machine learning algorithms to identify markers predictive of NMIUC recurrence.

MATERIAL AND METHODS

Patient Selection

NMIUC specimens were obtained by cold cup biopsy taken under white light cystoscopy during TURBT at initial diagnosis in patients who presented to Herlev Hospital, University of Copenhagen, Denmark, between June 1993 and October 2004. Specimens were snap frozen and stored at −80C. Tissue sections and imaging were re-reviewed and TNM staging was standardized according to AJCC (American Joint Committee on Cancer) recommendations.[12]

Patients were followed by cystoscopy and cytology every 3 months in year 1 and every 6 to 12 months for at least 5 years. Recurrence was defined as 1 or more tumor recurrences without stage progression after initial presentation. Patients who experienced stage progression without initial recurrence after first presentation were not included when identifying genes associated with recurrence. All patients provided informed consent for use of tissues. The study was approved by the University of Copenhagen ethics committee and the University of Southern California institutional review board.

Whole Genome Profiling

Adequate quality RNA was extracted for whole genome profiling (ie whole mRNA expression profiling) using the RNeasy® kit. Whole genome expression profiling was performed using the HumanWG-6v3 BeadChip. In addition, 9 RNA control samples and 3 duplicate patient samples were run for quality control purposes. Data are expressed as gene intensity values and no-calls were assigned a value of −1 for analysis purposes.

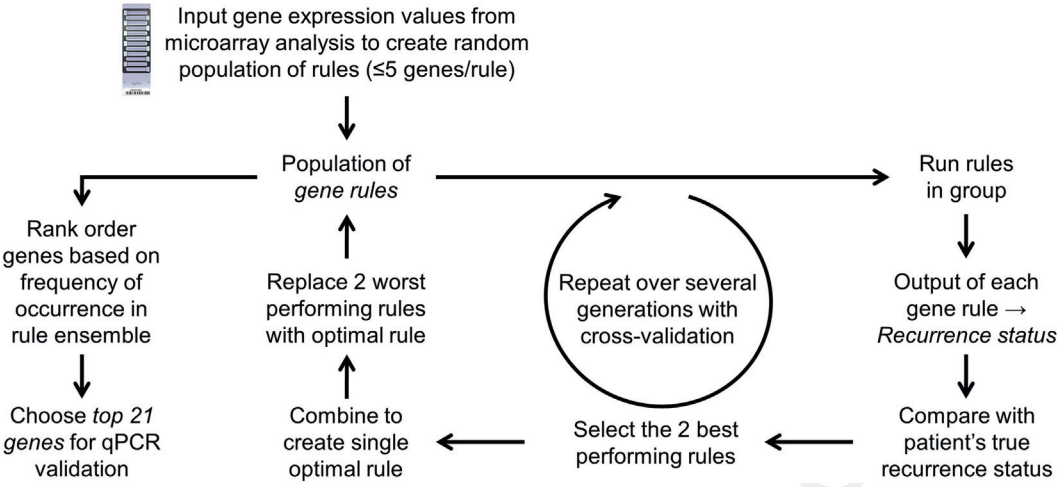
Key Genes

Selection of Genes Associated with Recurrence. A machine learning technique referred to as GP was used to select a minimal set of predictive markers associated with recurrence.[13–15] GP uses an approach that generates a unique collection of solutions to biodiagnostic problems by creating classifier programs using a subset of the available input. A genetic pool of candidate classification programs was created by randomly choosing inputs, and arithmetic and Boolean operators that work with these inputs (see figure). Successive generations of programs were then evolved through further selection and recombination. The accuracy of a program generated by the GP system to correctly classify the samples according to prespecified classes was determined by calculating the ROC AUC, thereby providing a measure of fitness.[16] Evolution was driven to a maximum by improving accuracy to yield rules with high sensitivity and specificity. [F1]

The complexity of the rules generated by this method was restricted to prevent overfitting the training data.[13] The programs were evolved over many generations to yield more accurate programs. The algorithm was terminated when a perfect classifier was generated within the limits of program size and complexity or after completion of the preset number of 75 cycles. Table 1 lists the inputs, operators and GP parameters used in the analysis. [T1]

To decrease the dimensionality of data gene frequency was used to identify markers that provided the most value to develop an accurate classifier. Cross-validation based resampling was applied to estimate the ability of the classifier to generalize to unseen samples, giving an approximation of its robustness.[17] Classifier sets (ensembles) were composed from the best rules from each fold. The ensemble was then polled with each rule in the ensemble voting on whether a sample belonged to a target class, in this case whether a patient experienced recurrence. If the majority of the rules agreed that the sample was in the target class, the ensemble flagged the sample as such. Aggregate performance of these ensembles on the test folds was considered the predictor of the classification error. The selected ensemble was the one with the smallest test error.

Validation by qPCR. To evaluate technical reproducibility and applicability to low throughput clinical platforms select genes were validated by qPCR. RNA from patient samples was reverse transcribed to cDNA and amplified using the Ovation® Pico WTA system. For qPCR assays the appropriate mRNA reference sequence accession number was identified for each gene. Primers and probes were designed for the consensus sequence and amplicon sizes were kept to a minimum with most being fewer than 100 bases. Gene expression was assayed using custom



Machine learning algorithm used approach that generates unique collection of solutions to biodiagnostic problems by creating classifier programs using subset of available input. Genetic pool of candidate classification programs was created by randomly choosing inputs, arithmetic and Boolean operators that work with these inputs.

TaqMan® low density array 384-well micro fluidic cards. The supplementary Appendix (http://jurology.com) lists Applied Biosystems® assay numbers for the genes of interest and for reference (ie housekeeping) genes.

cDNA (100 ng per 48 wells) was applied to the cards and all assays were performed in duplicate using the 7900HT Fast Real-Time PCR System (Applied Biosystems). Output data are expressed as the number of PCR cycles needed to reach a constant threshold set at 0.2 on the amplification curve, ie C_t. Replicates were inspected for congruence and a correlation coefficient was generated for each replicate set of genes. Replicates were averaged and expression data was normalized by subtracting the C_t for each rule gene (RG) C_t, ie genes of interest associated with recurrence, from the average of the 3 reference genes (Ave. 3Ref C_t) as described previously.[18] Since C_t values are expressed as logarithmic numbers to the base 2, the data were linearized by taking the antilog and the result was scaled by a factor of 10 to eliminate artifacts related with values approaching zero. The final form of the data normalized in this manner was described using the equation, normalized value $= 2^{(\text{Ave. 3Ref } C_t - \text{RG } C_t)} \times 10$.

To ensure high quality data key genes or reference genes with C_t values greater than 35 were not included in analysis. Also, samples were excluded if the linear regression coefficient of a scatterplot for the 2 replicates was less than 0.9.[19]

Table 1. *Genetic programing simulation parameters*

Elitism	True
Inputs	PCR generated normalized gene expression values
Operators	=, >, <, ≥, ≤, and, not, or, ?, nand, xor, nor
Fitness function	AUC, sensitivity, specificity
Population size	3,000
Crossover rate	0.9
Mutation rate	0.1
Generation limit	75

RESULTS

Study Cohort General Parameters

The study cohort comprised 112 patients a median of 67 years old who had NMIUC, of whom 87 (77.7%) were male. A total of 86 patients (76.8%) had Ta tumors and 26 (23.2%) had T1 tumors. Grade was G2 in 88 patients (78.6%) and G3 in 24 (21.4%). Median time to tumor recurrence was 8.8 months (IQR 4.8–18.2). Median followup was 7.6 years, during which recurrence developed in 88 patients (78.6%). Table 2 lists further clinical details on these patients. Application of EORTC risk tables to patients in the study cohort revealed good specificity (91.67%) but poor sensitivity (9.09%) to predict recurrence based on clinicopathological parameters alone (table 3).[20]

Table 2. *Clinical characteristics, outcomes and comparisons of study cohort and validation subset*

	Study Cohort		Validation Subset		p Value	
No. pts	112		83			
Median age (IQR)	67	(61–75)	67	(60–76)	0.95	(Mann-Whitney test)
No. male (%)	87	(77.7)	65	(78.3)		
No. female (%)	25	(22.3)	18	(21.7)	1.00	(Fisher exact test)
Tumor stage (%):					1.00	(Fisher exact test)
Ta	86	(76.8)	63	(75.9)		
T1	26	(23.2)	20	(24.1)		
Tumor grade (%):					0.73	(Fisher exact test)
G2	88	(78.6)	63	(75.9)		
G3	24	(21.4)	20	(24.1)		
No. tumors (%):					1.00	(Fisher exact test)
Single	88	(78.6)	66	(79.5)		
Multiple	24	(21.4)	17	(20.5)		
Outcome (%):					0.36	(Fisher exact test)
No recurrence	24	(21.4)	13	(15.7)		
Recurrence	88	(78.6)	70	(84.3)		

Table 3. *Performance of EORTC risk tables to predict patient outcomes*

Clinical Recurrence Probability*	No. Observed Clinical Outcome: No Recurrence	Recurrence	% Sensitivity	% Specificity
Study cohort:			9.09	91.67
Low	22	80		
Intermediate	2	8		
Validation subset:			8.57	84.62
Low	11	64		
Intermediate	2	6		

*Based on EORTC risk table recurrence score (low—1 to 4 and intermediate—5 to 9).

Whole Genome Expression Profiling Analysis

The expression of 37,846 RNA transcripts was measured per sample. Pairwise analysis of RNA controls and duplicate samples indicated good reproducibility and consistency between BeadChips and replicates (average Pearson correlation coefficient 0.91). GP analyses identified 21 key genes based on frequency of use in a collection of rules that predicted recurrence (supplementary Appendix, http://jurology.com/).

Real-Time qPCR Validation

Sufficient RNA was available from 100 patients for qPCR analysis of the identified 21 genes. Quality control identified 83 patients for further validation. Demographic and clinical parameters of patients in the validation subset did not differ significantly from those in the entire study cohort (table 2). These samples included 70 patients (84.3%) with recurrence. As noted in the study cohort application of the EORTC risk tables to patients in the validation subset also showed good specificity (84.62%) but poor sensitivity (8.57%) to predict recurrence (table 3).[20]

The samples were used to regenerate a set of rules to predict recurrence using the prognostic markers identified by whole genome expression profiling. The goal was to identify a majority of recurrent samples while keeping the number of misclassified samples as low as possible (ie maximizing sensitivity while minimizing loss of specificity). Fourfold cross-validation based resampling of the 83 samples was done since this provided a large enough training set while leaving a reasonable test set in each fold. This analysis resulted in selection of an optimal solution from a pool of rules with differing characteristics in terms of constitution, structure and accuracy.

Using normalized qPCR data an optimal 3-gene rule was developed, that is IF (*TMEM205* × (*NFKBIA* x *KRT17*)) ≥ 19637.3878 THEN RECURRENCE. This rule had 80.4% sensitivity and 90% specificity in the training set, and 70.6% sensitivity and 66.7% specificity in the test set (table 4).

In addition, a combined rule that used a total of 5 genes was produced with each rule independently developed in a different subset of the data. 1) IF ($RPS6 + (NFKBIA \times KRT17)) \geq 1195.5845$ THEN RECURRENCE. 2) IF ((*TMEM205* + (*KRT17* /1.2689386)) ≥ 59.5102 THEN RECURRENCE. 3) IF ((*NFKBIA* x (*TMEM205* + *GLTP*)) ≥ 1826.5199 THEN RECURRENCE. 4) IF (*TMEM205* × (*NFKBIA* x *KRT17*)) ≥ 19637.3878 THEN RECURRENCE.

Taken together in a voting algorithm if 3 or more of the rules considered a sample to be a recurrence, it was flagged as a recurrence. Otherwise it was considered a nonrecurrence. When blindly compared to actual outcomes, the combined rule had 77.1% sensitivity and 84.6% specificity in the training set, and 68.6% sensitivity and 61.5% specificity in the test set (table 4).

DISCUSSION

We previously reported the potential use of GP and artificial intelligence to predict nodal positivity for urothelial cancer in 60 patients with Ta to T4 disease.[10] The generated classifier rules using 70 genes demonstrated 81% accuracy in the validation set compared to pathological nodal status.

In the current study this technology was transferred to whole genome analyses in a unique population of patients with NMIUC that was obtained at initial presentation. We investigated the potential use of GP to predict tumor recurrence in NMIUC, investigating 37,846 RNA transcripts. With GP we were able to establish a 21 key gene panel from which a single 3-gene rule and a combined 5-gene rule were validated using qPCR. Gene profiling data obtained in this study were reproducible and consistent as determined by internal controls. The data were also amenable to analysis by GP algorithms to find a set of 21 key genes that strongly associated with our designated clinical outcomes of time to recurrence of bladder cancer after initial TURBT within 5 years.

While some of the 21 key genes identified have been shown to influence the NMIUC outcome, to our knowledge several others have not been associated with NMIUC to date. Nevertheless, these genes are

Table 4

Rule (set)	% Sensitivity	% Specificity
Combined:		
Training	77.1	84.6
Test	68.6	61.5
Singular:		
Training	80.4	90.0
Test	70.6	66.7

highly predictive of NMIUC recurrence when applied as the specified mathematical rule.

In a different whole genome analysis project using a 25k cDNA array and a 35k oligonucleotide array in 144 patients with bladder cancer Lindgren et al distinguished 2 molecular subtypes of Ta and T1 tumors.[21] A gene signature of 100 genes was defined that classified cases of NMIUC as low grade (G1/G2) and high grade (G3) as well as nonmuscle and muscle invasive tumors with high precision and sensitivity. These signatures were validated in 2 independent data sets. Another study identified a 38-gene panel that was validated by qPCR in another cohort subclassified between T1a and T1b stages.[22] The cluster classified 92.9% of invasive stages and 66.3% of superficial stages. Unlike prior reports the stringent requirement of a unique patient population with NMIUC with cold cup biopsies at initial presentation and concomitant recurrence data a minimum of 5 years after initial diagnosis to mirror our existing cohort precluded us from performing thorough external validation. However, the sustained and reproducible performance of the gene based rules identified in this study and the low sensitivity of clinicopathological models suggest that GP based approaches may be more promising to identify NMIUC associated prognostic markers.

Other groups investigated high throughput analysis with RNA sequencing strategies for nonmuscle invasive and invasive bladder cancer.[23–26] These studies confirmed mutations in genes previously identified as being altered in urothelial carcinoma. Furthermore, additional pathways have been identified that are implicated in NMIUC. Specifically Guo et al identified frequent alterations in *STAG2* and *ESPL1*, which are 2 genes involved in the sister chromatid cohesion and segregation process.[25] They also described recurrent fusion involving *FGF3* and *TACC3*. Overall 32% of tumors harbored genetic alterations affecting the sister chromatid cohesion and segregation process, making this process a potential target for therapeutic interventions.

In multicenter fashion TCGARN (The Cancer Genome Atlas Research Network) reported 32 mutations in genes regulating cell cycle, chromatin and kinase signaling pathways, and 9 genes without a history of significant mutation in carcinogenesis.[27] Besides recurrent in-frame activating *FGFR3-TACC3* fusions, the integration of several viruses and potential therapeutic targets was detected in 69% of tumors.

Besides our approach towards machine learning algorithms Catto et al applied artificial intelligence to microarray data to identify a gene signature for bladder cancer progression.[28] Neurofuzzy modeling and artificial neural networks were applied to a 2,800-gene data set of 66 patients. Six of the 11 progression associated genes that were identified were validated by immunohistochemistry. Rosser et al also identified bladder cancer associated expression signatures in exfoliated urothelial cells.[29] Hierarchical clustering and supervised learning algorithms were applied to investigate the samples based on the tumor burden. A panel of 319 genes was associated with the presence of bladder cancer. Supervised machine learning and cross validation were used to build a 14-gene molecular signature that could classify patients with or without cancer with 76% overall accuracy.

While there are several tools to analyze high throughput analysis in an adequate manner, to our knowledge our approach represents the initial attempt to integrate mathematical systems and machine learning algorithms to determining prognosis in NMIUC.

This study has some limitations. While the identified genes and rules provide novel prognostic and biological insights into NMIUC recurrence, it would be ideal to validate these findings in an external data set. However, given the uniqueness of the patient cohort (cold cup frozen tumor specimens of NMIUCs obtained at first presentation with a long followup), obtaining an external cohort with similar characteristics is extremely challenging. An effort was also made to identify genes associated with eventual stage progression. However, the number of patients with stage progression after experiencing at least 1 recurrence was limited in the validation subset, that is 14 patients (16.9%). This did not allow for the identification of robust signatures to predict progression without the potential for false discovery.

CONCLUSIONS

The GP technology enables the identification of key genes from whole genome analysis by evolution. It can construct a simple and comprehensive prognostic classifier for NMIUC. Several novel prognostic genes were identified, which can serve as a discovery tool for further investigations into pathways that are perturbed in NMIUC. The results of this study indicate that these data contain adequate information to predict recurrence within 5 years after TURBT. This is significant as it demonstrates the feasibility of creating a gene signature using machine learning algorithms that predicts recurrent NMIUC. With further validation such approaches may have the potential to yield gene signatures that can be applied to clinical practice.

ACKNOWLEDGMENTS

The late Dr. John P. Stein contributed to this study.

REFERENCES

1. Mitra AP, Datar RH and Cote RJ: Molecular staging of bladder cancer. BJU Int 2005; **96:** 7.
2. Mitra AP, Datar RH and Cote RJ: Molecular pathways in invasive bladder cancer: new insights into mechanisms, progression, and target identification. J Clin Oncol 2006; **24:** 5552.
3. Mitra AP, Lin H, Datar RH et al: Molecular biology of bladder cancer: prognostic and clinical implications. Clin Genitourin Cancer 2006; **5:** 67.
4. Mitra AP and Cote RJ: Molecular pathogenesis and diagnostics of bladder cancer. Annu Rev Pathol 2009; **4:** 251.
5. Pasin E, Josephson DY, Mitra AP et al: Superficial bladder cancer: an update on etiology, molecular development, classification, and natural history. Rev Urol 2008; **10:** 31.
6. Mitra AP and Cote RJ: Molecular screening for bladder cancer: progress and potential. Nat Rev Urol 2010; **7:** 11.
7. Mitra AP, Birkhahn M and Cote RJ: p53 and retinoblastoma pathways in bladder cancer. World J Urol 2007; **25:** 563.
8. Mitra AP, Bartsch CC and Cote RJ: Strategies for molecular expression profiling in bladder cancer. Cancer Metastasis Rev 2009; **28:** 317.
9. Mitra AP, Pagliarulo V, Yang D et al: Generation of a concise gene panel for outcome prediction in urinary bladder cancer. J Clin Oncol 2009; **27:** 3929.
10. Birkhahn M, Mitra AP, Williams AJ et al: Predicting recurrence and progression of noninvasive papillary bladder cancer at initial presentation based on quantitative gene expression profiles. Eur Urol 2010; **57:** 12.
11. Bartsch G, Mitra AP and Cote RJ: Expression profiling for bladder cancer: strategies to uncover prognostic factors. Expert Rev Anticancer Ther 2010; **10:** 1945.
12. Urinary bladder. In: AJCC Cancer Staging Manual, 7th ed. Edited by SB Edge, DR Byrd, CC Compton et al. New York: Springer 2010; pp 497—505.
13. Mitra AP, Almal AA, George B et al: The use of genetic programming in the analysis of quantitative gene expression profiles for identification of nodal status in bladder cancer. BMC Cancer 2006; **6:** 159.
14. Langdon WB and Buxton BF: Genetic programming for mining DNA chip data from cancer patients. Genetic Programming and Evolvable Machines 2004; **5:** 251.
15. Yu J, Almal AA, Dhanasekaran SM et al: Feature selection and molecular classification of cancer using genetic programming. Neoplasia 2007; **9:** 292.
16. Zou KH, Hall WJ and Shapiro DE: Smooth nonparametric receiver operating characteristic (ROC) curves for continuous diagnostic tests. Stat Med 1997; **16:** 2143.
17. Schaffer C: Selecting a classification method by cross-validation. Machine Learning 1993; **13:** 135.
18. Lenehan PF, Boardman LA, Riegert-Johnson D et al: Generation and external validation of a tumor-derived 5-gene prognostic signature for recurrence of lymph node-negative, invasive colorectal carcinoma. Cancer 2012; **118:** 5234.
19. Guthrie JL, Seah C, Brown S et al: Use of Bordetella pertussis BP3385 to establish a cutoff value for an IS*481*-targeted real-time PCR assay. J Clin Microbiol 2008; **46:** 3798.
20. Sylvester RJ, van der Meijden AP, Oosterlinck W et al: Predicting recurrence and progression in individual patients with stage Ta T1 bladder cancer using EORTC risk tables: a combined analysis of 2596 patients from seven EORTC trials. Eur Urol 2006; **49:** 466.
21. Lindgren D, Frigyesi A, Gudjonsson S 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.
22. Descotes F, Dessen P, Bringuier PP et al: Microarray gene expression profiling and analysis of bladder cancer supports the sub-classification of T1 tumours into T1a and T1b stages. BJU Int 2014; **113:** 333.
23. Sharron LX, Hu L, Sandy K et al: Differentiating progressive from nonprogressive T1 bladder cancer by gene expression profiling: applying RNA-sequencing analysis on archived specimens. Urol Oncol 2014; **32:** 327.
24. Morrison CD, Liu P, Woloszynska-Read A et al: Whole-genome sequencing identifies genomic heterogeneity at a nucleotide and chromosomal level in bladder cancer. Proc Natl Acad Sci U S A 2014; **111:** E672.
25. Guo G, Sun X, Chen C et al: Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation. Nat Genet 2013; **45:** 1459.
26. Mitra AP, Lam LL, Ghadessi M et al: Discovery and validation of novel expression signature for postcystectomy recurrence in high-risk bladder cancer. J Natl Cancer Inst 2014; **106:** dju290.
27. The Cancer Genome Atlas Research Network: Comprehensive molecular characterization of urothelial bladder carcinoma. Nature 2014; **507:** 315.
28. Catto JW, Abbod MF, Wild PJ et al: The application of artificial intelligence to microarray data: identification of a novel gene signature to identify bladder cancer progression. Eur Urol 2010; **57:** 398.
29. Rosser CJ, Liu L, Sun Y et al: Bladder cancer-associated gene expression signatures identified by profiling of exfoliated urothelia. Cancer Epidemiol Biomarkers Prev 2009; **18:** 444.