Disease map-based biomarker selection and pre-validation for bladder cancer diagnostic

Marine De Paoli[1], Paul Perco[2], Irmgard Mühlberger[2], Arno Lukas[2], Hardev Pandha[3], Richard Morgan[3], Gui Jie Feng[3], Marquette Christophe[1*]

[1]AXO Science SAS, 34 Rue du Mail, 69004 Lyon, France

[2]emergentec biodevelopment GmbH, Gersthofer Strasse 29-31, 11080 Vienna, Austria

[3]Oncology, Faculty of Health and Medical Sciences, Leggett Building, University of Surrey, Surrey GU2 7WG, UK

[*] Corresponding author: AXO Science SAS, 34 Rue du Mail, 69004 Lyon, France. Tel.: +334 72 43 13 69; Fax: +334 57 74 77 69; e-mail: christophe.marquette@axoscience.com

Keywords: biomarkers, bladder cancer, invasiveness, recurrence

Abstract

Context: Urinary biomarkers are promising as simple alternatives to cystoscopy for the diagnosis of *de novo* and recurrent bladder cancer.

Objective: To identify a highly sensitive and specific biomarker candidate set with potential clinical utility in bladder cancer.

Materials and methods: Urinary biomarkers concentrations were determined by ELISA. The performance of individual markers and marker combinations was assessed using ROC analysis.

Results: A 5-biomarker panel (IL8, MMP9, VEGFA, PTGS2 and EN2) was defined from the candidate set.

Discussion and conclusion: This panel showed a better overall performance than the best individual marker. Further validation studies are needed to evaluate its clinical utility in bladder cancer.

Introduction

Bladder cancer is the ninth most common cancer worldwide based on incidence (Ferlay et al., 2013). About 75-85 % of patients have non-muscle invasive bladder cancer (NMIBC) at first diagnosis (Babjuk et al., 2011). Although these tumors are associated with good prognosis, there is a 50 to 70% recurrence rate with a probability of progression to muscle invasiveness in 10 to 30% (Clark et al., 2013; Hall et al., 2007; Sylvester et al., 2006). Consequently, lifelong surveillance is essential for early recurrence detection.

Both diagnosis and surveillance of bladder cancer rely on cystoscopy and urine cytology methods (Babjuk et al., 2011). Cystoscopy is highly sensitive and while considered the 'gold standard', it has a significant false-negative rate due to operator-dependent variability or to the difficulty of detecting *in situ* lesions. Furthermore, it is a costly and invasive procedure which is unpleasant for patients, and carries an additional 10% risk of developing urinary tract infection (Budman et al., 2008; Shariat et al., 2008). Urine cytology has a higher specificity, ranging from 85 to 100%, but lacks sensitivity (13-75%), especially when it comes to the detection of low-grade tumors (van Rhijn et al., 2005).

In this context and taking into account the limitations mentioned above, alternative non-invasive methods such as measurements of urinary markers appear more and more appropriate for the diagnosis and surveillance of bladder cancer.

Urine-based biomarker discovery was enabled by the advent of high-throughput Omics technologies (Ramachandran et al., 2008; Urquidi et al., 2012). Many markers were identified for their potential utility in the detection and monitoring of bladder cancer such as NMP22, BTA, BLCA-4, CYFRA21-1, survivin, hyaluronic acid and hyaluronidase (Konety et al., 2000; Shariat et al., 2009, 2008; Van Tilborg et al.,

2009; Vrooman and Witjes, 2008). Urinary tests have been developed for some of them, including bladder tumor antigen (BTA) and nuclear matrix protein 22 (NMP22). Two BTA assays, BTA TRAK® and BTA stat® (Polymedco, USA) are approved by the FDA for monitoring of bladder cancer in conjunction with cystoscopy. The Alere NMP22® BladderChek® Test is FDA-approved for the diagnosis and monitoring of bladder cancer patients in conjunction with standard diagnostic procedures. Compared to urine cytology which has a median sensitivity of 35% and a median specificity of 94% (van Rhijn et al., 2005), these tests exhibit better sensitivity (24-89% for BTA TRAK®, 57-79% for BTA stat® and 49.5-65% for NMP22) but lower specificity (52-93% for BTA TRAK®, 48-95% for BTA stat® and 40-87.3 % for NMP22) as they are influenced by benign urological conditions such as inflammation, urinary lithiasis or benign prostatic hyperplasia. (Vrooman and Witjes, 2008).

The ideal marker should have both high sensitivity and high specificity in order to replace or decrease the need for cystoscopy and guide the surveillance scheme. Lack of specificity as outlined above for BTA and NMP2 is common to all described urinary markers and the best approach to address this specificity problem of individual markers may be to combine several biomarkers in a panel detectable by a multiplex assay.

In this report, the selection and pre-validation of bladder cancer biomarkers are described. Basis for selection of appropriate marker candidates was the construction of a molecular disease model for bladder cancer and the identification of bladder cancer disease relevant molecular processes taking into account background Omics and literature datasets. Marker selection focused on covering relevant molecular

disease processes by selecting at least one marker for every detected biological process also considering available biomarker evidence from scientific literature.

The biomarker candidate set was then evaluated in a technical feasibility study using urinary samples from both bladder cancer patients and healthy donors. The evaluation led to a selection of markers with potential clinical utility to be included in a multiplex assay.

Materials and methods

Specimen and data collection

Urine samples were collected from patients with cystoscopic and histological evidence of bladder cancer and controls. The latter were age and sex matched individuals who had no ongoing or previous cancer, were non-smokers, were not on medication, no urinary symptoms or history of prior bladder disorders apart from occasional urinary tract infection. First pass urines were collected according to a standard operating procedure and were spun at 150*g* for 10 minutes. The supernatant was aliquoted into 1 ml samples and frozen and stored at -80°C. All individuals gave informed consent for sample donation and the collection was approved by local ethical committee (ref 13/LO/0739). For patients with cancer, the histological report was prepared by an uro-pathologist additionally containing information to tumor stage and grade for each patient.

Bladder cancer model and selection of biomarker candidate set

We first created a set of bladder cancer disease genes consisting (i) of molecular features derived via an automatic literature mining approach for bladder cancer associated genes and (ii) of molecules reported as being deregulated in non-muscle-invasive bladder cancer (NMIBC) Omics studies. The literature derived set was generated via gene2pubmed mappings for publications retrieved with the PubMed query *("Urinary Bladder Neoplasms/diagnosis" [majr]) OR ("Urinary Bladder Neoplasms/genetics" [majr]) OR ("Urinary Bladder Neoplasms/pathology" [majr]) OR ("Urinary Bladder Neoplasms/physiopathology" [majr]) OR ("Urinary Bladder Neoplasms/urine" [majr])* in March 2013.

Deregulated molecular features were furthermore extracted from high-throughput Omics studies, namely genome-wide association studies (GWAS), transcriptomics mRNA and miRNA studies, proteomics, as well as metabolomics studies, in the focus of disease diagnosis of bladder cancer fulfilling the following requirements:

(i) A study on human samples

(ii) Sample material had to be from untreated patients

(iii) Focus of the study had to be disease diagnosis

(iv) The analysis approach of the study had to be unbiased

(v) Tumor category: superficial bladder cancer (NMIBC), early stage (pTis, pTa, pT1); early/intermediate grade (1 or 2)

(vi) Tumor type: transitional cell carcinoma

(vii) Origin: mucosa

(viii) For proteomics and metabolomics, studies conducted in urine were included

Studies passing this manual curation step were forwarded to molecular feature extraction thus complementing the list of features available after the automatic literature mining approach. Metabolites were mapped to their enzymes via information in the Human Metabolome Database [http://www.hmdb.ca/] and miRNAs were mapped to their target genes via information in miRTarBase [http://mirtarbase.mbc.nctu.edu.tw/]. Data sources used for molecular features as well as mapping for building the molecular model are listed in Appendix B.

The molecular model was constructed for NMIBC using the bladder cancer molecular features set following the procedure as outlined in Heinzel et al. (Heinzel et al., 2014) with the omicsNET protein dependency network as underlying biological network (Fechete et al., 2013).

Each process unit of the resulting network was evaluated regarding evidence of association to bladder cancer by comparing the distribution of paper counts for each process unit member to a reference distribution of all human genes having at least one disease association as derived from literature mining.

All process units were further evaluated regarding the capability of individual process unit members to separate healthy controls [n=48] from diseased samples [n=28] in the publicly available transcriptomics dataset from Sanchez-Carbayo (Sanchez-Carbayo et al., 2006). Distribution of single process unit member AUC values were tested against the reference set of all mRNA features on the array of the transcriptomics study in order to identify those process units showing good performance in separating cases from controls.

Association to bladder cancer was evaluated for each member of the resulting molecular model based on gene2pubmed mappings of scientific articles retrieved with the following PubMed query: *Urinary Bladder Neoplasms[majr].*

In addition a list of marker candidates was generated based on a literature search using the following query followed by the extraction of genes via gene2pubmed: *"Urinary Bladder Neoplasms/urine"[majr] AND "Biological Markers"[mh].*

A third number was derived for each molecule in the molecular model on the number of publications reporting the respective molecule as biomarker for bladder cancer in urine using the following PubMed query: *"Urinary Bladder Neoplasms/urine"[majr] AND "Biological Markers"[mh].*

Enzyme-linked immunosorbent assays (ELISA) for urinary biomarkers

Commercially available ELISA kits were used to measure urinary levels of the different markers. Appendix C gives the 20 kits sources and references.

The assays were conducted according to the manufacturer's instructions using an EVO75 robotic platform (Tecan, France).

Engrailed-2 (EN2) protein levels were also determined by ELISA (Morgan et al., 2013). A monoclonal mouse anti-EN2 antibody, APS1, was generated (Antibody Production Services Ltd., UK) using the synthetically produced EN2 C-terminal 100 amino acids (Biosynthesis Inc., USA). APS1 was used to detect EN2 coated on a 96-well plate (Immuno Clear Standard Modules, Cat # 434797 Thermo Scientific, USA).

APS1 was then detected using an anti-mouse IgG (γ chain specific) - peroxidase conjugate (Calbiochem, Germany).

For each assay, a calibration curve in assay diluent and a calibration curve in normal human urine (UTAK Laboratories, USA) were prepared. EN2 calibration curves were generated from dilution series of one of the bladder cancer urinary samples. For all the other markers, calibration curves were prepared using protein standards provided in the ELISA kits. Curve fitting was accomplished using four-parameter logistic regression.

Due to the high degree of variability of voided urine (Thomas et al., 2010; Urquidi et al., 2012), the concentrations of all markers were normalized to urinary creatinine (measured with ADVIA 1800 Chemistry System, Siemens) and expressed as a ratio to urinary creatinine values.

Data analysis

A receiver operating characteristic (ROC) analysis was performed to assess the performance of individual markers using the PanelomiX software (http://www.panelomix.net/) (Robin et al., 2013). PanelomiX also allowed ROC analysis of panels by testing various combinations of markers based on the iterative combination of biomarkers and thresholds (ICBT) method.

Results

Bladder cancer model and selection of biomarker candidate set

The bladder cancer disease gene set held 1384 unique protein coding genes with 587 genes extracted from scientific literature complemented by 897 molecules

extracted from the Omics datasets on NMIBC. The constructed molecular model of NMIBC held 520 proteins in 30 process units ranging in size from 3 to 162. Five process units were significantly enriched in proteins associated with bladder cancer as compared to the reference set of genes being associated with at least one disease term via gene2pubmed which were considered as relevant for biomarker selection. AUC values in separating cases from controls in the transcriptomics dataset from Sanchez-Carbayo were significantly higher for members of four additional process units which were also considered for biomarker selection.

A schematic representation of the molecular model is given in the Figure 1. Based on the number of associations to bladder cancer the most promising molecules were selected for relevant molecular process units. This marker set was complemented by four molecules with high evidence in the scientific literature but not being part of the molecular model resulting in a set of 20 marker candidates as given in Table 1.

Following the identification of this biomarker candidate set, the evaluation of markers was performed in a two-step selection process: a first evaluation for measurability and detectability in urine samples followed by a second evaluation for confirmation of the markers' selectivity for bladder cancer.

First evaluation of candidate biomarkers: measurability and detectability in urine

The biomarker candidate set consisting of 20 markers was evaluated in urine samples of bladder cancer patients and healthy donors.

Measurability in urine was first ascertained as the performance of many ELISAs may be significantly affected by the urinary environment due to changes in pH or urea/creatinine levels. Columns 2 and 3 of Table 2 present the measurability of each marker (standard solution from kit) in kit manufacturer buffer and standardized urine, respectively. Four out of the 20 ELISA tests weren't able to detect their specific target in urine (CD44, AP1M2, CDH1 and ERBB2). On the contrary, MYC was detected in urine but not in manufacturer buffer.

Detectability in urine samples from bladder cancer patients was then evaluated using the kits able to detect standard solution in urine. Out of the 16 remaining tests, 10 were able to detect markers in patient urine (MMP9, EN2, VEGFA, EGFR, IL8, DCN, MYC, UPK3A, PTGS2 and FGFR3).

Second evaluation of candidate biomarkers: markers' selectivity for bladder cancer

The 10 selected markers underwent a second evaluation step using a set of 32 samples (16 bladder cancer patients and 16 healthy donors). Clinical and pathological characteristics of the 32 subjects are presented in Appendix A.

Figure 2 presents the marker concentrations in samples determined with commercial ELISA kits using the four-parameter logistic regression and normalized to urinary creatinine. For each marker, the cut-off value was calculated as the mean concentration of all healthy donor samples + 3SD (99.7% of the healthy population is below this value). Eight of the ten markers showed at least one positive patient sample concentration, i.e., above the cut-off value. Three markers (VEGFA, DCN and UPK3A) showed a cut-off value of zero as no protein was detected in the healthy donor samples. Patient samples showing detectable marker signals were

considered positive. Four other markers (MMP9, IL8, MYC and EN2) showed one healthy donor marker concentration above the cut-off value, thus generating false positives. Finally, PTGS2 was the only marker for which none of the samples from both populations had a concentration above the cut-off value.

Marker performance was evaluated using ROC analysis with performance values of the individual markers given in Table 3. MMP9 was the most accurate marker with an AUC of 0.742 (95% CI: 0.560-0.923). Using the optimal threshold (determined by PanelomiX ICBT method), MMP9 revealed a sensitivity of 86.7% (95% CI: 66.7-100.0%) and specificity of 68.8% (95% CI: 43.8-87.5%). The markers VEGFA and DCN were only detected in patient samples and thus provided high specificity values of 100% but low sensitivity values of 20.0% (95% CI: 0.0-40.0%) and 13.3% (95% CI: 0.0-33.3%) for VEGFA and DCN respectively. UPK3A could not be included in the ROC analysis as the protein was detected only in one sample.

ROC analysis was also performed for multiple markers and PanelomiX threshold-based algorithm was used to test all combinations of biomarkers (excluding UPK3A). The best performing marker panel achieved an AUC of 0.871 (95% CI: 0.723-1.0) and of 0.81 (95% CI: 0.640-0.981) after cross-validation including the following six markers: MMP9, VEGFA, IL8, PTGS2, FBLN3, and EN2 (Table 4).

Excluding FBLN3 from the model resulted in only marginally reduced overall performance but drastically increased specificity at the expense of reduced sensitivity. As specificity is the parameter of major interest in clinical practice the following marker panel was defined as the one to be multiplexed on a chip (Table 5 and Figure 3).

Discussion

Patient-specific biomarker profiles could be of predictive and prognostic utility and could contribute to evidence-based bladder cancer patient management. As no single biomarker can achieve the desired accuracy, a multi-marker test would logically be more likely to be successful. In this regard, we have identified a urinary biomarker candidate set and evaluated it in bladder cancer urine samples, leading to the selection of a panel of five markers: IL8, MMP9, VEGFA, PTGS2 and EN2.
This panel showed a better overall performance than the best individual marker (MMP9), achieving an AUC of 0.865 (0.727-1.0) whereas MMP9 reached an AUC of 0.742 (95% CI: 0.560-0.923). Even though the panel sensitivity is slightly reduced compared to MMP9 (80.0% and 86.7, respectively), its specificity, the most relevant parameter in this context, is considerably increased (93.8% vs. 68.8%).
It is also important to note that due to a relatively small sample size (32 patients), confidence intervals are considerably large.

Non-muscle-invasive bladder tumors have a tendency to recur with a risk of progression to muscle-invasive disease, highlighting the importance of monitoring patients to allow for early recurrence or progression detection (21). We analyzed the potential of our five-marker panel for predicting invasiveness. The role of the various markers with regards to invasiveness was thus evaluated based on the bladder cancer model and on a literature search about the mechanistic link to invasiveness. Each marker measured in the study was searched in the scientific literature in combination with the search term invasiveness or invasive. Resulting publications were manually screened and paragraphs discussing the respective marker in the

context of invasiveness were extracted and collected. In addition, links to other markers from the lists in the publications on invasiveness were extracted and consolidated.

A schematic pathway diagram was constructed based on the information extracted from publications (Figure 4). This diagram emphasizes the roles of MMP9, VEGFA, EGFR, IL8 and PTGS2 in bladder cancer invasiveness. Indeed, it was found that matrix metalloproteinases, including MMP9, promote tumor invasion and alter microenvironment. They may be thus associated with the development of invasive bladder cancer (Kader et al., 2007). Moreover, the mechanism of MMP9 activation involves EGF/EGFR signaling activities (Pei et al., 2014) as well as IL8 expression (Inoue et al., 2000). Angiogenic activity is also enhanced by the expression of VEGF and its receptors VEGFR1/VEGFR2, promoting tumor proliferation and invasion (Kopparapu et al., 2013; Nakanishi et al., 2009). Additionally, PTGS2 has a role in bladder cancer development and invasion and is associated with angiogenesis as well (Gee et al., 2008; Margulis et al., 2007).

Four of the markers included in the final panel are thus linked to invasiveness: IL8, MMP9, VEGFA and PTGS2. The remaining marker, EN2, has no reported link to invasiveness so far in the literature. Nevertheless, its inclusion in the final panel is justified by its contribution to overall performance of the marker panel and previous reports on this marker as bladder cancer marker (Morgan et al., 2013). EGFR on the other hand was not included in the panel as it did not increase overall performance although mechanistically linked to invasiveness.

Conclusions

Marker candidates were selected on the basis of a molecular disease model for bladder cancer. A 5-biomarker panel (IL8, MMP9, VEGFA, PTGS2 and EN2) was then defined from the candidate set by evaluating the measurability and detectability as well as the selectivity for bladder cancer of the candidate markers in urine samples.

Although this preliminary study only investigated the diagnostic aspect of bladder cancer (in terms of detectability), it was followed by an analysis of the selected panel showing its potential for the prediction of invasiveness.

A scoring system, developed by the European Organization for Research and Treatment of Cancer (EORTC), is already available to predict the risks of both recurrence and progression in individual superficial bladder cancer patients (Sylvester et al., 2006). However, it is based only on clinical and pathological factors such as the number of tumors, tumor size and T category.

Biomarker profiles, such as defined by our selected panel, could have a great impact on clinical management of bladder cancer patients. Indeed, the surveillance scheme as well as the treatment strategy could be modified according to the risk of invasiveness and/or recurrence assessed at the time of diagnosis and based on the urinary biomarker levels of the defined panel.

Finally, it is important to note that the present study includes a small number of samples (32 patients). Thus, there is a need for further clinical validation with a greater sample size. A multiplex assay shall then be now developed to detect this panel in high-throughput to consolidate this first validation study.

Declaration of interest

This work has been supported in part by the European Commission Program DIPROMON - HEALTH-F5-2012-306157-2: Development of an integrated protein- and cell-based device for non-invasive diagnostics in the urogenital tract.

The authors report no declarations of interest.

Appendices

Appendix A: Patient characteristics of the study cohort

	Healthy donors (%) N = 16	Bladder cancer patients (%) N = 16
Age, y		
Mean [median]	71.1 [71]	73.3 [73]
Range	59 - 82	65 -83
Sex		
Male	16 (100)	14 (87.5)
Female	0 (0)	2 (12.5)
Clinical stage*		
Ta	N/A	10 (66.7)
T1	N/A	4 (26.7)
T2	N/A	1 (6.6)
Tumor grade*		
G1	N/A	1 (6.7)
G2	N/A	11 (73.3)
G3	N/A	3 (20)

*One subject with metastatic bladder cancer, not included in clinical stage and tumor grade
N/A: not applicable

Appendix B: Data sources of the molecular model

Author-Year	PubMed ID	Omics type	Study title
Garcia-Closas-2011	21824976	genomics	A genome-wide association study of bladder cancer identifies a new susceptibility locus within SLC14A1, a urea transporter gene on chromosome 18q12.3
Rafnar-2011	21750109	genomics	European genome-wide association study identifies SLC14A1 as a new urinary bladder cancer susceptibility gene
Rothman-2010	20972438	genomics	A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci
Wu-2009	19648920	genomics	Genetic variation in the prostate stem cell antigen gene PSCA confers susceptibility to urinary bladder cancer
Golka-2011	21380501	genomics	Genetic variants in urinary bladder cancer: collective power of the "wimp SNPs"
Mengual-2009	19539325	mRNA transcriptomics	DNA microarray expression profiling of bladder cancer allows identification of noninvasive diagnostic markers
Sanchez-Carbayo-2006	16432078	mRNA transcriptomics	Defining molecular profiles of poor outcome in patients with invasive bladder cancer using oligonucleotide microarrays
Dyrskjot-2004	15173019	mRNA transcriptomics	Gene expression in the urinary bladder: a common carcinoma in situ gene expression signature exists disregarding histopathological classification
Catto-2009	19843843	miRNA transcriptomics	Distinct microRNA alterations characterize high- and low-grade bladder cancer
Linden-2012	22065568	proteomics	Proteomic analysis of urinary biomarker candidates for nonmuscle invasive bladder cancer
Niu-2009	18818109	proteomics	Differences in shotgun protein expression profile between superficial bladder transitional cell carcinoma and normal urothelium
Jobu-2012	22466574	metabolomics	Metabolomics study on the biochemical profiles of odor elements in urine of human with bladder cancer
Putluri-2011	21990318	metabolomics	Metabolomic profiling reveals potential markers and bioprocesses altered in bladder cancer progression
Pasikanti-2010	20337499	metabolomics	Noninvasive urinary metabonomic diagnosis of human bladder cancer.

Appendix C: ELISA kits references

Marker	Full name	Company	Cat No
MMP9	matrix metalloproteinase 9	Abcam	ab100610
EN2	engrailed-2	Cusabio	CSB-EL007660HU
VEGFA	vascular endothelial growth factor	Cusabio	CSB-E11718h
VIM	vimentin	Abnova	KA3127
CD44	CD44 molecule (Indian blood group)	Abnova	KA0119
IL8	interleukin 8	Abcam	ab46032
EGFR	epidermal growth factor receptor	Sigma-Aldrich	RAB0160-1KT
RARB	retinoic acid receptor, beta	USCN Life Science	SED951Hu
PTGS2	prostaglandin-endoperoxide synthase 2	Uscn Life Science	SEA699Hu
DCN	decorin	Abcam	ab99998
MYC	V-Myc myelocytomatosis viral oncogene homolog	BlueGene Biotech	E01C0774
AP1M2	adaptor-related protein complex 1, mu 2 subunit	Cusabio	CSB-EL001865HU
FGFR3	fibroblast growth factor receptor 3	Cusabio	CSB-EL008646HU
TMOD1	tropomodulin 1	Cusabio	CSB-EL023909HU
UPK3A	uroplakin 3A	Cusabio	CSB-EL025657HU
BIRC5	baculoviral IAP repeat containing 5 (survivin)	Abnova	KA0441
CDH1	cadherin-1	Abnova	KA0433
FBLN3	fibulin 3	USCN Life Science	SEF422Hu
p53	cellular tumor antigen p53	Sigma-Aldrich	RAB0500-1KT
ERBB2	receptor tyrosine-protein kinase erbB-2	Sigma-Aldrich	RAB0173-1KT

References

Babjuk M, Oosterlinck W, Sylvester R, Kaasinen E, Böhle A, Palou-Redorta J, Rouprêt M. (2011). EAU guidelines on non-muscle-invasive urothelial carcinoma of the bladder, the 2011 update. *Eur Urol* 59:997–1008.

Bettum IJ, Vasiliauskaite K, Nygaard V, Clancy T, Pettersen SJ, Tenstad E, Maelandsmo GM, Prasmickaite L. (2014). Metastasis-associated protein S100A4 induces a network of inflammatory cytokines that activate stromal cells to acquire pro-tumorigenic properties. *Cancer Lett* 344:28–39.

Budman LI, Kassouf W, Steinberg JR. (2008). Biomarkers for detection and surveillance of bladder cancer. *Can Urol Assoc J* 2:212–221.

Clark PE, Agarwal N, Biagioli MC, Eisenberger MA, Greenberg RE, Herr HW, Inman BA, Kuban DA, Kuzel TM, Lele SM, Michalski J, Pagliaro LC, Pal SK, Patterson A, Plimack ER, Pohar KS, Porter MP, Richie JP, Sexton WJ, Shipley WU, Small EJ, Spiess PE, Trump DL, Wile G, Wilson TG, Dwyer M, Ho M. (2013). Bladder Cancer. *J Natl Compr Canc Netw* 11:446–475.

Donmez G, Sullu Y, Baris S, Yildiz L, Aydin O, Karagoz F, Kandemir B. (2009). Vascular endothelial growth factor (VEGF), matrix metalloproteinase-9 (MMP-9), and thrombospondin-1 (TSP-1) expression in urothelial carcinomas. *Pathol Res Pract* 205:854–857.

Fechete R, Heinzel A, Söllner J, Perco P, Lukas A, Mayer B. (2013). Using information content for expanding human protein coding gene interaction *Networks J Compu. Sci Syst Biol* 06.

Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin D, Forman D, Bray F. (2013). GLOBOCAN 2012 v1.0, Cancer Incidence and

Mortality Worldwide: IARC CancerBase No. 11 [Online].Available at: http://globocan.iarc.fr/ Accessed on 12 November 2014.

Gee J, Lee I-L, Grossman HB, Sabichi AL. (2008). Forced COX-2 expression induces PGE(2) and invasion in immortalized urothelial cells. *Urol Oncol* 26: 641–645.

Hall MC, Chang SS, Dalbagni G, Pruthi RS, Seigne JD, Skinner EC, Wolf JS, Schellhammer PF. (2007). Guideline for the management of nonmuscle invasive bladder cancer (stages Ta, T1, and Tis): 2007 update. *J Urol* 178: 2314–2330.

Heinzel A, Mühlberger I, Fechete R, Mayer B, Perco P. (2014). Functional molecular units for guiding biomarker panel design. *Methods Mol Biol* 1159:109-33.

Inoue K, Slaton JW, Kim SJ, Perrotte P, Eve BY, Bar-Eli M, Radinsky R, Dinney CP. (2000). Interleukin 8 expression regulates tumorigenicity and metastasis in human bladder cancer. *Cancer Res* 60:2290–2299.

Kader AK, Liu J, Shao L, Dinney CP, Lin J, Wang Y, Gu J, Grossman HB, Wu X. (2007). Matrix metalloproteinase polymorphisms are associated with bladder cancer invasiveness. *Clin Cancer Res* 13:2614–2620.

Konety BR, Nguyen T-ST, Dhir R, Day RS, Becich MJ, Stadler WM, Getzenberg RH. (2000). Detection of bladder cancer using a novel nuclear matrix protein, BLCA-4. *Clin Cancer Res* 6:2618–2625.

Kopparapu PK, Boorjian SA, Robinson BD, Downes M, Gudas LJ, Mongan NP, Persson JL. (2013). Expression of VEGF and its receptors VEGFR1/VEGFR2 is associated with invasiveness of bladder cancer. *Anticancer Res* 33:2381–2390.

Margulis V, Shariat SF, Ashfaq R, Thompson M, Sagalowsky AI, Hsieh J-T, Lotan Y. (2007). Expression of cyclooxygenase-2 in normal urothelium, and superficial and advanced transitional cell carcinoma of bladder. *J Urol* 177:163–1168.

Morgan R, Bryan RT, Javed S, Launchbury F, Zeegers MP, Cheng KK, James ND, Wallace DMA, Hurst CD, Ward DG, Knowles MA, Pandha H. (2013). Expression of Engrailed-2 (EN2) protein in bladder cancer and its potential utility as a urinary diagnostic biomarker. *Eur J Cancer* 49:2214–2222.

Nakanishi R, Oka N, Nakatsuji H, Koizumi T, Sakaki M, Takahashi M, Fukumori T, Kanayama H. (2009). Effect of vascular endothelial growth factor and its receptor inhibitor on proliferation and invasion in bladder cancer. *Urol Int* 83: 98–106.

Pei J, Lou Y, Zhong R, Han B. (2014). MMP9 activation triggered by epidermal growth factor induced FoxO1 nuclear exclusion in non-small cell lung cancer. *Tumour Biol* 35:6673–6678.

Ramachandran N, Srivastava S, LaBaer J. (2008). Applications of protein microarrays for biomarker discovery. *Proteomics Clin Appl* 2:1444–1459.

Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M. (2013). PanelomiX: A threshold-based algorithm to create panels of biomarkers. Transl. *Proteomics* 1:57–64.

Sanchez-Carbayo M, Socci ND, Lozano J, Saint F, Cordon-Cardo C. (2006). Defining molecular profiles of poor outcome in patients with invasive bladder cancer using oligonucleotide microarrays. *J Clin Oncol* 24:778–789.

Shariat SF, Karakiewicz PI, Godoy G, Karam JA, Ashfaq R, Fradet Y, Isbarn H, Montorsi F, Jeldres C, Bastian PJ, Nielsen ME, Müller SC, Sagalowsky AI, Lotan Y. (2009). Survivin as a prognostic marker for urothelial carcinoma of

the bladder: a multicenter external validation study. *Clin Cancer Res* 15:7012–7019.

Shariat SF, Karam JA, Lotan Y, Karakiewizc PI. (2008). Critical evaluation of urinary markers for bladder cancer detection and monitoring. *Rev Urol* 10:120–135.

Sylvester RJ, van der Meijden APM, Oosterlinck W, Witjes JA, Bouffioux C, Denis L, Newling DWW, Kurth K. (2006). 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* 49: 466–465; discussion 475–477.

Thomas CE, Sexton W, Benson K, Sutphen R, Koomen J. (2010). Urine collection and processing for protein biomarker discovery and quantification. *Cancer Epidemiol Biomark Prev* 19:953–959.

Urquidi V, Kim J, Chang M, Dai Y, Rosser CJ, Goodison S. (2012). CCL18 in a multiplex urine-based assay for the detection of bladder cancer. PLoS ONE 7: e37797.

Van Rhijn BWG, van der Poel HG, van der Kwast TH. (2005). Urine markers for bladder cancer surveillance: a systematic review. *Eur Urol* 47:736–748.

Van Tilborg AA, Bangma CH, Zwarthoff EC. (2009). Bladder cancer biomarkers and their role in surveillance and screening. *Int J Urol* 16:23–30.

Vrooman OPJ, Witjes JA. (2008). Urinary markers in bladder cancer. *Eur Urol* 53: 909–916.

Table 1. Marker list

Process unit	Symbol	AUC	# BC papers	# BC marker papers	# BC marker papers (urine)
4	PTGS2	0,908	8	4	1
4	RARB	0,958	3	2	1
5	TP53	0,866	70	12	0
5	IL8	0,692	6	2	0
5	VEGFA	0,876	10	4	0
3	MYC	0,763	10	2	0
1	EGFR*	0,684	10	5	1
1	ERBB2*	0,981	16	15	0
1	FGFR3*	0,990	46	7	0
7	CD44	0,929	7	6	1
7	MMP9	0,539	5	2	1
23	EFEMP1	0.993	0	0	0
12	VIM	0.896	2	2	0
12	TMOD1	1.000	0	0	0
25	DCN	0.999	0	0	0

29	AP1M2	0.997	0	0	0
-	BIRC5	0,918	11	7	2
-	CDH1	0,984	19	9	0
-	UPK3A*	0,842	3	1	0
-	EN2	0.531	1	1	1

Process unit: process unit number of the molecular NMIBC model

Symbol: Official Gene Symbol

AUC: Single AUC value in the transcriptomics dataset by Sanchez-Carbayo (red if upregulated in diseased samples; green if downregulated in diseased samples)

BC papers: number of bladder cancer papers addressing the respective gene

BC marker papers: number of BC papers addressing the respective gene as marker

BC marker papers: number of BC papers addressing the respective gene as marker in urine

* indicates if the protein is a receptor

Table 2. Marker evaluation in urine samples

Marker	Calibration curve in assay diluent	Calibration curve in urine	LOD (urine)†	Marker detection in patient samples
MMP9	+	+	0.222 ng/mL	+
EN2*	+	+	1.25 ng/mL	+
VEGFA	+	+	0.0625 ng/mL	+
VIM	+	+	28.125 ng/mL	-
CD44	+	-	N/A	-
EGFR	+	+	0.1 ng/mL	+
IL8	+	+	0.1 ng/mL	+
RARB	+	+	0.625 ng/mL	-
PTGS2	+	+	2.5 ng/mL	+
DCN	+	+	25.9 pg/mL	+
MYC	-	+	1 ng/mL	+
AP1M2	+	-	NA	-
FGFR3	+	+	>250 pg/mL	-
TMOD1	+	+	1.25 ng/mL	-
UPK3A	+	+	100 pg/mL	+
BIRC5	+	+	200 pg/mL	-
CDH1	+	-	N/A	-
FBLN3	+	+	>3.12 ng/mL	+
p53	+	+	2.5 ng/mL	-
ERBB2	+	-	N/A	-

+ : presence of calibration curve or marker detection in samples; **-** :absence of calibration curve or no marker detection in samples; **N/A:** not applicable; **LOD**: limit of detection

*EN2 was detected using direct coating of the urine sample.

†The limits of detection were determined visually on the calibration curves.

Table 3. Individual biomarker performance sorted by AUC

Marker	**% AUC (95% CI)**	**% Sp (95% CI)**	**% Se (95% CI)**
MMP9	74.2 (56.0-92.3)	68.8 (43.8-87.5)	86.7 (66.7-100.0)
EN2	66.0 (47.8-84.3)	87.5 (68.8-100.0)	53.3 (26.7-80.0)
IL8	65.4 (48.9-81.9)	81.2 (62.5-100.0)	46.7 (20.0-73.3)
VEGFA	60.0 (49.5-70.5)	100.0 (100.0-100.0)	20.0 (0.0-40.0)
PTGS2	57.7 (38.7-76.7)	68.8 (43.8-87.5)	53.3 (26.7-80.0)
DCN	56.7 (47.8-65.6)	100.0 (100.0-100.0)	13.3 (0.0-33.3)
EGFR	54.0 (38.5-69.4)	87.5 (68.8-100.0)	26.7 (6.7-46.7)
FBLN3	47.5 (26.1-68.9)	56.2 (31.2-81.2)	60.0 (33.3-86.7)
MYC	47.1 (36.5-57.7)	93.8 (81.2-100.0)	6.7 (0.0-20.0)

AUC: area under the curve; CI: confidence interval; Sp: specificity; Se: sensitivity

No performance characteristics could be obtained for UPK3A since the protein was detected in one sample only and thus only one nonzero value was available for the analysis.

Table 4. Best panel performance

Panel		**% AUC (95% CI)**	**% Sp (95% CI)**	**% Se (95% CI)**
MMP9 VEGFA IL8 PTGS2 FBLN3 EN2	Training	87.1 (72.3-100.0)	81.2 (62.5-100.0)	93.3 (80.0-100.0)
	Cross-validation	81.0 (64.0-98.1)	68.8 (43.8-93.8)	93.3 (80.0-100.0)

AUC: area under the curve; CI: confidence interval; Sp: specificity; Se: sensitivity

Cross-validation procedure for panel verification: 10-fold cross-validation repeated 5 times

Table 5. Performance of final biomarker panel

Panel		**% AUC (95% CI)**	**% Sp (95% CI)**	**% Se (95% CI)**
MMP9 VEGFA IL8 PTGS2 EN2	Training	86.5 (72.7-100.0)	93.8 (81.2-100.0)	80.0 (60.0-100.0)
	Cross-validation	81.9 (66.3-97.5)	87.5 (68.8-100.0)	73.3 (53.3-93.3)

AUC: area under the curve; CI: confidence interval; Sp: specificity; Se: sensitivity

Cross-validation procedure for panel verification: 10-fold cross-validation repeated 5 times

Figure captions

Figure 1. Schematic representation of the constructed molecular model of bladder cancer. Each node represents a molecular process unit whereas the size corresponds to number of proteins within this unit.

Figure 2. Marker concentrations in healthy donor and bladder cancer patient samples, normalized to urinary creatinine.
Concentrations were determined with commercial ELISA kits with the exception of EN2. Cut-off values were calculated as the mean concentration of all healthy donor samples + 3SD.
a.u.: arbitrary unit

Figure 3. Receiver operating characteristics (ROC) curves of the final panel (black) and of MMP9 (grey) on training set (left) and test set (right)

Figure 4. Schematic pathway for bladder cancer invasiveness development (Kader et al., 2007; Pei et al., 2014; Donmez et al., 2009; Bettum et al., 2014).
BC: bladder cancer