Optimizing treatment and follow-up of patients with bladder cancer

Risk stratification and biomarkers



Tim Muilwijk, 2022

KU Leuven
Biomedical Sciences Group
Faculty of Medicine
Department of Development and Regeneration

KU LEUVEN
DOCTORAL SCHOOL
BIOMEDICAL SCIENCES

OPTIMIZING TREATMENT AND FOLLOW-UP OF PATIENTS WITH BLADDER CANCER

RISK STRATIFICATION AND BIOMARKERS

Tim Muilwijk

Promotor:	Steven Joniau
Copromotors:	Thomas Gevaert
	Maarten Albersen
Jury members:	Aleyde Van Eynde
	Peter de Witte
	Karel Decaestecker
	Morgan Rouprêt
	Bart Schrier
Chair:	Frank Van der Aa
Secretary:	Aleyde Van Eynde

Dissertation presented in partial fulfilment of the requirements for the degree of Doctor of Biomedical Sciences

June 21, 2022

™ 2022 - Optimizing treatment and follow-up of patients with bladder cancer: risk stratification and biomarkers

© 2022 Tim Muilwijk

Uitgegeven in eigen beheer, Tim Muilwijk, Antwerpen.

Cover image design and layout: Scheepens (www.scheepens.nl)

Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotokopie, microfilm, elektronisch of op welke andere wijze ook zonder voorafgaandelijke schriftelijke toestemming van de uitgever.

All rights reserved. No part of the publication may be reproduced in any form by print, photoprint, microfilm, electronic or any other means without written permission from the publisher.

I know that I know nothing.
Ik weet dat ik niets weet.

Sokrates

469-399 BC

Table of Contents

List of Abbreviations

Chapter 1: General Introduction p. 11

1.1 Non-muscle-invasive bladder cancer
1.2 Muscle-invasive bladder cancer
1.3 Molecular subtypes
1.4 Tumor immune microenvironment
1.5 Conclusion

Chapter 2: Hypotheses and aims p. 21

Chapter 3: Quality control indicators for transurethral resection of bladder tumor: results from an embedded Belgian multicenter prospective registry p. 25

3.1 Abstract
3.2 Introduction

3.3 Materials and methods
3.4 Results
- 3.4.1 Complete resection status (CR)
- 3.4.2 Presence of detrusor muscle (DM)
- 3.4.3 Single instillation of Mitomycin C (MMC)
- 3.4.4 Re-TURBT
- 3.4.4 Re-TURBT
- 3.4.5 Start of BCG induction (BCG)
- 3.4.6 TURBT ≤6 weeks after primary diagnosis (T≤6)

3.5 Discussion
3.6 Conclusion
3.7 Acknowledgments
3.8 Authors Contributions
3.9 Conflict of Interest
3.10 Supplementary Figures
3.11 Supplementary Tables
3.12 References

Chapter 4: The Cancer of the Bladder Risk Assessment (COBRA) score for Estimating Cancer-Specific Survival after Radical Cystectomy: External Validation in a Large Bi-institutional Cohort p. 53

4.1 Abstract
4.2 Introduction
4.3 Patients and methods
4.3.1 Patient identification
4.3.2 Lymphadenectomy
4.3.3 Statistical analysis
4.4 Results
4.4.1 Clinical and pathological variables
4.4.2 Survival analyses
4.4.3 Survival analyses – COBRA score
4.4.4 Survival analyses – COBRA risk categories
4.4.5 Comparison of COBRA score, COBRA risk categories and AJCC staging system
4.5 Discussion
4.6 Conclusion
4.7 Acknowledgments
4.8 Conflict of Interest
4.9 Supplementary Figures
4.10 Supplementary Tables
4.11 References

Chapter 5: CK5 and CK20 inversely correlate with tumor grading in Ta non-muscle invasive bladder cancer p. 83

5.1 Abstract
5.2 Introduction
5.3 Materials and methods
5.3.1 Patients
5.3.2 Immunohistochemistry
5.3.3 Real-time quantitative polymerase chain reaction (qPCR)
5.3.4 Statistical analysis
5.4 Results
5.4.1 Clinical and pathological characteristics
5.4.2 IHC and CK5 expression patterns
5.4.3 Survival analysis

5.5 Discussion
5.6 Acknowledgments
5.7 Authors Contributions
5.8 Conflict of Interest
5.9 Supplementary Figures
5.10 Supplementary Tables
5.11 References

Chapter 6: Stromal marker fibroblast activation protein drives out come in T1 non-muscle invasive bladder cancer p. 113

6.1 Abstract
6.2 Introduction
6.3 Materials and methods
6.4 Results
6.5 Discussion
6.6 Acknowledgments
6.7 Authors Contributions
6.8 Conflict of Interest
6.9 Supplementary data
6.10 References

Chapter 7: Tumor immune microenvironment of high-risk non-muscle invasive bladder cancer: fibroblast activation protein and carcinoma in situ drive progression-free survival p. 129

7.1 Introduction
7.2 Materials and methods
7.2.1 Patients
7.2.2 Immunohistochemistry and quantitative analysis
7.2.3 Statistical analysis
7.3 Results
7.3.1 Patient characteristics
7.3.2 Differences in IHC markers
7.3.2.1 Progressors vs. non-progressors
7.3.2.2 Regions of interest: T1 vs. Ta vs. CIS
7.3.2.3 Stroma vs. tumor
7.3.3 Immune phenotypes
7.3.4 Progression-free survival correlated with clinicopathological and IHC variables
7.3.5 Subgroup analyses of the tumor immune microenvironment

7.4 Discussion
7.5 Acknowledgments
7.6 Supplementary data
7.7 References

Chapter 8: Discussion and Conclusion p. 159
8.1 Discussion
8.1.1 Quality control indicators correlate with clinical outcome
8.1.2 The COBRA score is outperformed by the AJCC staging system
8.1.3 Luminal and basal markers have a different role in Ta NMIBC compared to MIBC
8.1.4 FAP expression correlates strongly with PFS in T1 NMIBC
8.1.5 FAP remains the strongest driver for PFS
8.2 Concluding remarks and future perspectives

Chapter 9: Scientific Summary p. 169
9.1 English summary
9.2 Nederlandstalige samenvatting

Chapter 10: Final statements p. 175
10.1 Scientific acknowledgement and personal contribution
10.2 Conflict of interest

List of Abbreviations

AJCC	American Joint Committee on Cancer
AUC	Area under the curve
BC	Bladder cancer
BCG	Bacillus Calmette-Guerin
CAF	Cancer-associated fibroblast
CI	Confidence interval
CIS	Carcinoma in situ
COBRA	Cancer of the Bladder Risk Assessment
CHT	Chemotherapy
C-index	Concordance index
CPI	Immune checkpoint inhibitor
CR	Complete resection status
CSS	Cancer-specific survival
CSM	Cancer-specific mortality
CUETO	Club Urológico Español de Tratamiento Oncológico
DCA	Decision curve analysis
EAU	European Association of Urology
eCRFs	Electronic case report forms
EMR	Electronic medical record
EMT	Epithelial-mesenchymal transition
E-LND	Extended lymph node dissection
EORTC	European Organization for Research and Treatment of Cancer
FAP	Fibroblast activation protein-a
FDA	United States Food and Drug Administration
FFPE	Formalin-fixed paraffin-embedded
FGFR	Fibroblast growth factor receptor
GZMB	Granzyme B
HG	High grade
IHC	Immunohistochemistry
IQR	Interquartile range
LG	Low grade
LN	Lymph node
LN+	Lymph node-positive
LN-	Lymph node-negative
LND	Lymph node dissection
MIBC	Muscle-invasive bladder cancer
MMC	Mitomycin C

NA	Not available/not applicable
NAC	Neoadjuvant chemotherapy
NMIBC	Non-muscle-invasive bladder cancer
OS	Overall survival
pN+	Lymph node metastases
PFS	Progression-free survival
QCI	Quality control indicators
qPCR	Real-time quantitative polymerase chain reaction
RC	Radical cystectomy
RCT	Randomized controlled trials
RFS	Recurrence-free survival
RMA	Relative marker area
ROC	Receiver operating characteristic
ROI	Region of interest
SEER	Surveillance, Epidemiology, and End Results
SE-LND	Super-extended lymph node dissection
SI	Single instillation
S-LND	Standard lymph node dissection
TURBT	Transurethral resection of the bladder tumor
TGF-β	Transforming growth factor β
T≤6	Timing of TURBT ≤6 weeks after initial diagnosis
WHO	World Health Organization
ρ	Correlation
χ2	Chi-squared test
+ve	Positive
-ve	Negative

Chapter 1:
General Introduction



Chapter 1:

General Introduction

Bladder cancer (BC) is worldwide the seventh most prevalent cancer and has an estimated incidence of 573.278 patients per year, and an estimated mortality of 212.536 patients per year in 2020 (1). A total of 2.438 patients were diagnosed with BC in Belgium in 2019, and the Belgian cancer registry predicts that the number of new diagnoses per year will increase with 17% between 2014 and 2025 (2). Patients with BC present usually with macroscopic hematuria, and sometimes with lower urinary tract symptoms such as frequency or urgency. BC is diagnosed by cystoscopy at an outpatient clinic, followed by a transurethral resection of the bladder tumor (TURBT) under general or spinal anesthesia in the operating room. Using the resected tissue, pathologist stage BC using the TNM classification for the depth of invasion (**Figure 1**), and grade the tumor histologically using the WHO grading system (3–5). Additional staging is done by computed tomography urography to assess local invasiveness of the tumor, tumors in the upper urinary tract, invasion of lymph nodes, or metastases in other organs.

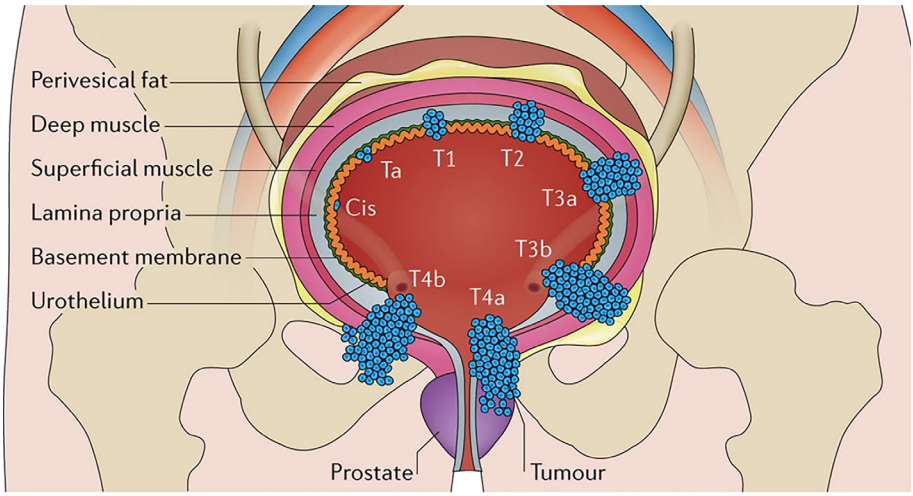


Figure 1: Tumor stages of bladder cancer. Adopted from Comperat et al. (2018) (6).

BC costed the European Union a total of €4.9 billion in 2012, which represents 5% of total health care costs of all cancers (7). Despite the high incidence and significant societal costs, research in BC is largely underfunded compared to other cancers that have similar incidence and societal costs (8). This lack of investment is one of the reasons why the age-standardized mortality of BC did

not improve significantly the last two decades (1). The European Association of Urology (EAU) Guidelines provide urologists recommendations on the clinical management of NMIBC, and MIBC (9,10). However, adherence to these guidelines is suboptimal, which can impact clinical outcome of patients with BC (11). Therefore, quality control indicators (QCI) can be used to evaluate guideline adherence, to increase guideline awareness and compliance, and most importantly to improve clinical outcome of patients (12).

1.1 Non-muscle-invasive bladder cancer

Patients with BC are diagnosed in 2/3rd of the cases with non-muscle-invasive bladder cancer (NMIBC; Ta, T1 and Tis) and in 1/3rd of the cases with muscle-invasive bladder cancer (MIBC; ≥T2) **(Figure 1)** (9). Patients with NMIBC are at risk for disease progression to MIBC which encompasses significant treatment-related and cancer-specific morbidity and mortality (13,14). Therefore, patients with NMIBC are stratified based on clinicopathological characteristics to predict their risk of progression, on which treatment and follow-up of patients are based (15,16). Using this risk stratification, patients can be grouped as: low-risk, intermediate-risk, high-risk, or very high-risk (9). All risk groups are primarily treated by a TURBT, which also serves as a diagnostic and staging procedure. Low-risk patients have a probability of progression to MIBC of less than 1% at 5 years follow-up, and are followed up with regular cystoscopies without need for adjuvant treatment (16). Contrarily, patients with intermediate-risk, high-risk, or very high-risk require adjuvant bladder instillations with chemotherapy or bacillus Calmette-Guerin (BCG), combined with intensive follow-up with more frequently cystoscopies. BCG has been used for more than 40 years in patients with NMIBC, and reduces recurrence of NMIBC and progression to MIBC (17–20). Patients in these higher risk groups are at significant risk for progression at 5 years follow-up: intermediate risk 3.6-4.9%, high-risk 9.6-11%, and very high-risk 40-44% (16). Unfortunately, this classification overestimates the risk of progression in the higher risk groups, which is illustrated by a recent validation of this classification in a cohort of BCG treated patients. The 5 year progression rate of high-risk NMIBC was calculated at 9.6%, but was actually 5.3%; whilst very-high-risk NMIBC was calculated at 40%, but was actually 14.9% both after receiving adequate BCG (21,22). Treatment options for patients with high-risk, or very high-risk NMIBC that recur after adequate BCG (BCG-unresponsive NMIBC) are limited, and most of these patients are treated by radical cystectomy (9). Multiple clinical trials are enrolling patients for treatment with novel agents such as

immune checkpoint inhibitors, oncolytic virotherapy, and antibody-drug conjugates. These trials will potentially provide bladder-sparing alternatives for patients with BCG-unresponsive NMIBC.

Patients with BCG-naive high-risk T1 NMIBC with ≥2 risk factors of either >1 lesion, size >3 cm, or carcinoma in situ (CIS), can be counseled for an early cystectomy, due to the fact cancer-specific survival (CSS) is significantly higher in patients treated with an early cystectomy compared to a deferred cystectomy in this subgroup (23). Radical cystectomy should thus be considered and discussed with patients as a primary treatment for very high-risk NMIBC (9). However, there is a proportion of these patients with high-risk NMIBC that can be successfully treated in a conservative way with BCG, thus sparing the bladder and maintaining quality of life. Current risk stratification lacks accuracy and fails to identify exactly these patients that are at highest risk for progression, and insufficiently predicts response to BCG. This is of importance, because patients with NMIBC that progress to MIBC have a lower CSS after cystectomy than patients with primary diagnosis of MIBC treated with cystectomy (24). Therefore, there is currently a clear unmet need to identify new predictive biomarkers which can be used in combination with current risk classification. This is of importance to identify which patients with high-risk NMIBC will respond to BCG, and which patients are at the highest risk for progression.

1.2 Muscle-invasive bladder cancer

As stated before, patients with MIBC are challenged with significant treatment-related and cancer-specific morbidity and mortality (13,14). Patients with localized MIBC are treated with a radical cystectomy with pelvic lymph node dissection combined with a urinary diversion, with or without neoadjuvant chemotherapy (10). Alternatively, tri-modality therapy can be considered in a very selected subgroup of patients (T2 disease-only plus absence of CIS) unwilling or unfit for surgery, which consists of a maximally safe TURBT, chemotherapy, and radiotherapy (10). CSS after radical cystectomy ranges between 58-67% at 10-years follow-up. The perioperative mortality after radical cystectomy is estimated at 3-5% within 90 days after surgery (14,25). Treatment is definitely necessary, since the 5-year CSS of untreated patients with MIBC is 14%, compared to 52% of treated patients (26). CSS of patients depends on the tumor stage at time of radical cystectomy, with a 10-year CSS of 78-89% in T1 disease compared to 23-31% for T4 disease (13,14).

Additionally, presence of lymph node metastases (pN+) found at pelvic lymph node dissection impacts the prognosis of patients significantly. The 10-year CSS of lymph node positive disease is only 14-25%, regardless of tumor stage (13,14). Cisplatin-based neoadjuvant chemotherapy before radical cystectomy provides an absolute overall survival (OS) benefit of 5–8% at 5-years follow-up (27,28). However, cisplatin-based chemotherapy is underused mainly due to patient frailty, and significant comorbidities (29). Additionally, 53% of patients with MIBC are deemed ineligible for cisplatin before radical cystectomy due to an impaired renal function (glomerular filtration rate <60 ml/min) (30,31).

Due to the significant cancer-specific mortality of MIBC, it is of interest to counsel patients more adequately on their prognosis, and to identify patients that are potentially in need of adjuvant therapy. Therefore, several nomograms been developed that predict CSS after radical cystectomy in MIBC (32–36). However, none of these nomograms is currently being used in clinical practice which can be attributed to their complexity, or to the inclusion of variables that can be considered irrelevant in some clinical scenarios. Additionally, most of these nomograms were not externally validated which impairs their clinical usefulness in other cohorts of patients. Identification of patients that are at highest risk for progression is becoming more relevant due promising data on adjuvant therapeutic options such as radiotherapy, chemotherapy, and immunotherapy. The efficacy of adjuvant radiotherapy has recently been illustrated in a phase 2 trial with a CSS 2-year of 62% in patients high-risk disease after radical cystectomy (definition: pT3-4, pN+, presence of lymphovascular invasion, <10 lymph nodes removed, or positive surgical margins) with acceptable toxicity compared to historical series (37). Immediate versus deferred (at time of recurrence) chemotherapy after radical cystectomy are also possibilities in patients with pT3-4 and/or pN+ disease. Immediate chemotherapy after radical cystectomy had a significant impact on progression-free survival (5-year PFS: 48% vs. 32%), but did not improve OS (38). However, in a post-hoc subgroup analysis, OS differed significantly for patients without lymph node involvement in favor of the immediate chemotherapy subgroup (5-year OS: 80% vs. 59%). If these findings are validated, a subgroup of patients could potentially benefit from direct adjuvant chemotherapy. Additionally, in unresectable locally advanced or metastatic disease, maintenance therapy with immune checkpoint inhibitor avelumab after chemotherapy improved OS significantly (1-year OS: 71% vs. 58%) (39). Therefore, patients with high-risk MIBC, based on risk stratification after radical cystectomy, could potentially be counseled in the future for either adjuvant

radiotherapy, chemotherapy, or immunotherapy, to improve their survival.

1.3 Molecular subtypes

The lack of accuracy of current risk stratification of BC, which is solely based on clinicopathological characteristics, may be attributed to the intrinsic heterogeneity of BC. Tumor heterogeneity occurs at multiple levels, and can be divided into intertumoral and intratumoral heterogeneity. Intertumoral heterogeneity refers to heterogeneity of the same type of tumors between patients, whilst intratumoral heterogeneity refers to heterogeneity of a tumor within a single patient (40). BC has one of the highest tumor mutational burden of all cancer types, and multiple studies have identified synchronous and metachronous inter- and intratumoral heterogeneity in BC (41–43). Synchronous heterogeneity means differences at the same time within or between a tumor within one patient, whilst metachronous heterogeneity means differences in time during the evolution of the disease. Genomic and transcriptomic analyses of BC have provided us additional information regarding this heterogeneity, and have led to the identification of different molecular subtypes in both NMIBC and MIBC, with luminal and basal subtypes as most distinct (44–46) (**Figure 2**).

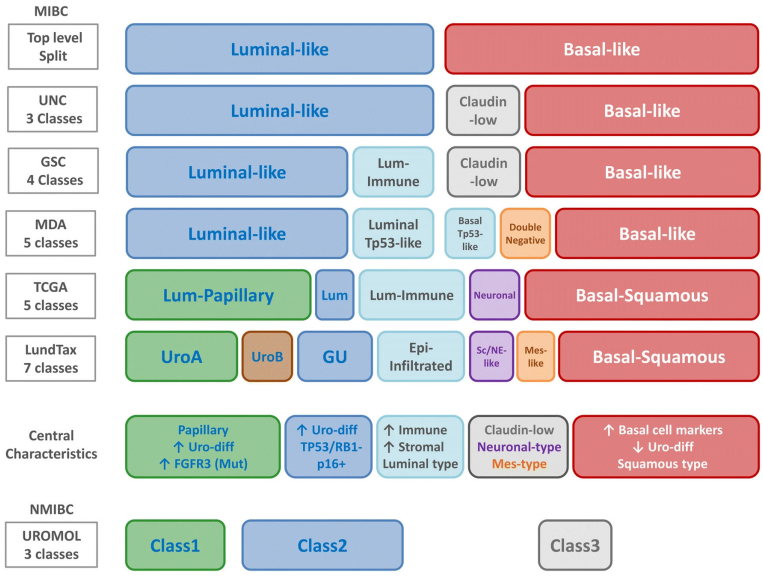


Figure 2: Different molecular subtype classifications of bladder cancer. Adopted from McConkey et al. (2019) (47).

Basal subtypes have squamous differentiation with epithelial-mesenchymal transition (EMT), and generally have more immune cell infiltration. Interestingly, basal and luminal subtypes differ in prognosis, and can potentially be used for treatment selection. For example, the presence of a basal subtype was predictive for a better response to neoadjuvant chemotherapy prior to radical cystectomy (48). Patients with a luminal subtype are characterized by fibroblast growth factor receptor (FGFR) mutations, fusions, or amplification. Treatment with FGFR TKIs is already approved in patients with FGFR alterations in patients with advanced disease who progress after chemotherapy (49). Therefore, multiple clinical trials are currently enrolling patients in earlier stages of the disease, for example with BCG-unresponsive NMIBC, for treatment with FGFR tyrosine kinase inhibitors. Treatment of NMIBC with FGFR tyrosine kinase inhibitors makes sense, as FGFR alterations are more frequently observed in patients with NMIBC compared to MIBC (44). Implementation of these molecular subtypes is not yet standard of care, as the predictive value of these molecular subtypes needs to be validated in prospective randomized studies. Additionally, transcriptomic analyses of tumors are still costly and time-consuming. Therefore, immunohistochemistry (IHC) panels, based on correlation with have transcriptomic data, have been suggested as an alternative to stratify patients in MIBC, such as GATA3 for luminal subtypes and CK5-6 for basal subtypes (50,51).

1.4 Tumor immune microenvironment

BCG has been used in high-risk NMIBC as immunotherapeutic agent for more than 40 years and its efficacy has been proven, although not all patients with high-risk NMIBC respond to this treatment. The efficacy of BCG can be attributed to a response of the tumor immune microenvironment which will recruit lymphocytes, neutrophils, macrophages, cytokines, and anti-tumoral cytotoxic responses (52). Our immune system has auto-regulatory checkpoints which are crucial for immune tolerance, for example to prevent auto-immune diseases. Tumors evade our immune system by upregulation of these immune checkpoints (e.g. PD-1, PD-L1, and CTLA4) thus promoting tumor immune evasion, which can be reversed by treatment with immune checkpoint inhibitors (CPIs). CPIs have recently provided us new immunotherapeutic options for patients with advanced disease that progressed after platinum-based chemotherapy, or were platinum-ineligible (53,54,39). Clinical trials with CPI have also shown promising results in patients with NMIBC in the BCG-unresponsive setting (55). The efficacy of BCG and CPIs in the treatment of BC

stresses the importance of the immune system, and has further emphasized the role of the tumor immune microenvironment and its associated anti-cancer immune response (56,57). The anti-cancer immune response requires infiltration of immune cells from the peritumoral stroma into the tumoral epithelium of the cancer cells (58). Based on these findings, immune phenotypes have been defined based on the presence of cytotoxic T cells in the tumor and/or stroma (**Figure 3**). The inflamed phenotype is characterized by the presence of both stromal and tumoral T cells, the excluded phenotype by the presence of stromal T-cells only, while the desert phenotype lacks stromal and tumoral T cells (60). The excluded phenotype is further characterized by the presence of stromal components such as cancer-associated fibroblasts (CAF) (60,61). Lack of response to ICP in advanced disease has been associated with a signature of transforming growth factor β (TGF-β) signaling in CAFs (60). Additionally, inflamed phenotypes are more prone to respond to immunotherapy whilst desert phenotypes are less likely to respond. Activation of the immune infiltrate also matters, as sole presence of cytotoxic T cells did not predict response to ICP in the neoadjuvant setting, but co-expression of cytotoxic T cells with Granzyme B (GZMB) did (62). GZMB is a a serine protease that plays a role in induction of target cell apoptosis, and can be used as a marker for activated cytotoxic T cells (56). Multiple additional immune cell populations and markers for activation within the tumor immune microenvironment have been identified as prognostic for clinical outcome of patients with BC, such as regulatory T-cells, tumor-associated macrophages and PD-L1 expression (63).

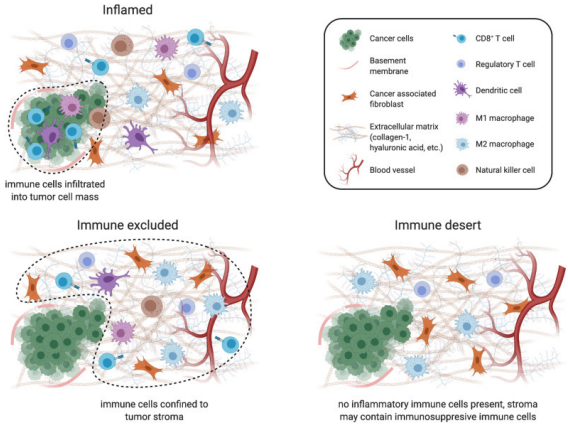


Figure 3: Immune phenotypes and composition of the tumor immune microenvironment. Adopted from Lee et al. (59).

1.5 Conclusion

BC is a heterogeneous disease in which risk stratification in both NMIBC and MIBC is important for counseling of patients, treatment decisions, and adequate follow-up. Over- and undertreatment of patients with high-risk NMIBC should be avoided through optimizing risk stratification. Furthermore, patients with MIBC are at high risk for cancer-specific mortality after radical cystectomy, and should also be counseled adequately on their prognosis, and perhaps for adjuvant treatment. Molecular subtypes, which have been discovered based on large transcriptomic studies, can potentially play a role as prognostic and predictive biomarkers once validated. Additionally, the tumor immune microenvironment provides another avenue which warrants further exploration to improve risk stratification of patients with BC.

Chapter 2:
Hypotheses and aims



Chapter 2:

Hypotheses and aims

This PhD thesis will focus on optimizing risk stratification of patients with BC, especially for those patients that are at high risk for treatment-related and disease-related morbidity and mortality. Additionally, we will assess whether treatment recommendations based on risk stratification are correctly implemented in daily clinical practice.

Based on the findings that guideline adherence is suboptimal in the clinical management of NMIBC, we sought to identify and assess QCIs in a large cohort of patients with NMIBC (11). In **Chapter 3**, we check the feasibility of extracting QCIs from our multicenter prospective TURBT registry using a stepwise approach. First, we assess the quality of reporting of each QCI. Second, we assess compliance to each QCI, which will be compared between centers, and assessed over time. Finally, we assess correlation between compliance to each QCI and recurrence-free survival (RFS).

As patients with MIBC are at high risk for cancer-specific mortality after radical cystectomy, we sought to compare methods of risk stratification for cancer-specific survival for these patients. In **Chapter 4**, we will externally validate the performance of the Cancer of the Bladder Risk Assessment (COBRA) score in a large bi-institutional cohort of patients, and compare its performance with the American Joint Committee on Cancer (AJCC) staging system. Additionally, we aim to improve and simplify the COBRA score for easier adoption in clinical practice.

In **Chapter 5**, we investigate the protein expression of GATA3 and CK5 as surrogate markers for luminal and basal molecular subtypes in a group of patients with mostly high-risk Ta NMIBC. This set-up provided us new insights on these markers in Ta NMIBC, as most studies used mixed cohorts of NMIBC and/or MIBC. We use a granular approach to assess these markers, which were correlated with clinical outcome of patients. Additionally, we correlate CK5 protein expression with RNA expression.

To further investigate the prognostic role of these surrogate markers, we focused on a homogeneous cohort of T1 NMIBC in **Chapter 6**. Patients with T1 NMIBC are considered high-risk NMIBC, and are therefore at the highest risk for disease progression. CAFs have been recognized as an important player in the tumor microenvironment, and their role in T1 NMIBC has not been thoroughly investigated yet. Therefore, we assess protein expression of fibroblast activation protein-α (FAP) as a marker of CAFs, and GATA3/CK5 as surrogate

markers for luminal/basal molecular subtypes in a retrospective cohort of patients with high-risk T1 NMIBC and long follow-up. We designed a pair-matched nested case-control study of patients with T1 NMIBC, who did/did not progress to muscle-invasive disease (≥T2). This approach allows for detection of differences between progressors and non-progressors, that could not be predicted based on current risk stratification. To fully capture tumor heterogeneity, we assess these markers using digital image analysis software separately for T1, Ta, and CIS regions, which often co-occur in the same patient sample.

As the tumor microenvironment consists of more elements than CAFs, we set out to further investigate the interplay and prognostic value of immune cells and CAFs in a cohort of high-risk NMIBC. In **Chapter 7**, we assess protein expression of multiple components of the tumor immune microenvironment: cytotoxic T cells (CD8), activated cytotoxic T cells (CD8-GZMB), regulatory T cells (FOXP3), M2 macrophages (CD163), PD-L1, and FAP. Using a similar approach as used in the previous chapter, we assess these markers using digital image analysis software separately for T1, Ta, and CIS regions. The anti-cancer immune response requires immune cell infiltration from the peritumoral stroma into the tumoral epithelium. Therefore, we assess expression of these immune cell markers separately for the peritumoral stroma and the tumoral epithelium. This way, it is possible to assess whether immune cell infiltration contributed to the prognosis of patients with high-risk NMIBC. This approach also enables us to define different immune phenotypes based on stromal and tumoral presence of cytotoxic T cells. We will also correlate these phenotypes with clinical outcome.

Chapter 3:

Quality control indicators for transurethral resection of bladder tumor: results from an embedded Belgian multicenter prospective registry

Chapter 3:

Quality control indicators for transurethral resection of bladder tumor: results from an embedded Belgian multicenter prospective registry

Tim Muilwijk[1], Murat Akand[1], Yannic Raskin[1], Cedric Jorissen[1], Kathy Vander Eeckt[2], Siska Van Bruwaene[3], Ben Van Cleynenbreugel[1], Steven Joniau[1], Frank Van Der Aa[1]

[1] Department of Urology, University Hospitals Leuven, Belgium

[2] Department of Urology, AZ Sint Blasius Dendermonde, Belgium

[3] Department of Urology, AZ Groeninge Kortrijk, Belgium

Submitted to European Urology Oncology, 2022 (IF 7.48)

Revised manuscript after receiving major revision

3.1 Abstract

Objective: Quality control indicators (QCIs) can objectively evaluate guideline adherence and benchmark quality among urologists and centers. We assessed six QCIs for non-muscle-invasive bladder cancer (NMIBC) using a prospective transurethral resection of the bladder tumor (TURBT) registry.

Materials and Methods: Clinical and pathological data of TURBT procedures were prospectively collected using electronic case report forms (eCRFs) embedded in the electronic medical record (EMR) in three Belgian centers from 2013 to 2017. We assessed six QCIs: complete resection status (CR), presence of detrusor muscle (DM), re-TURBT, single instillation of Mitomycin C (MMC), start of Bacillus Calmette-Guérin (BCG), and therapy in ≤6 weeks after diagnosis (T≤6W). We assessed quality of reporting on QCIs, assessed compliance to QCIs, compared compliance between centers and in time, and assessed correlation of compliance to QCIs with recurrence-free survival (RFS).

Results: A total of 1350 TURBTs were collected, of which 1151 TURBTs met inclusion criteria for a total of 907 unique patients. EAU risk categories after TURBT were: 271 low-, 464 intermediate-, and 416 high-risk. Quality of reporting on two QCIs was suboptimal: DM: 34.5%; and BCG: 50.8%. Compliance to QCIs was: CR: 97.0%; DM: 30.7%; MMC: 65.1%; re-TURBT: 33.0%; BCG: 39.4%; and T≤6W 87.7%. Compliance to all QCIs differed significantly between centers. Compliance to MMC and re-TURBT increased significantly over time, which could both be attributed to one center. Compliance to MMC and BCG correlated significantly with RFS.

Conclusion: We demonstrate the feasibility of assessing QCIs of TURBTs in a real-life setting using eCRFs integrated within the EMR. Compliance to QCIs differed significantly between centers, and scores on two QCIs increased over time. Two QCIs were inadequately reported, which will be improved in by implementation of an additional follow-up eCRF. Additionally, two QCIs correlated with clinical outcome of patients.

3.2 Introduction

Quality control indicators (QCIs) can be used to evaluate guideline adherence, to increase guideline awareness and compliance, and most importantly to improve clinical outcome (1–3). In our recent review on QCIs in non-muscle-

invasive bladder cancer (NMIBC), we concluded that most QCIs were based on a low level of evidence, and identified complete resection status (CR), re-TURBT (transurethral resection of the bladder tumor), presence of detrusor muscle (DM), and intravesical instillations as most valid QCIs (4). A correlation between compliance to QCIs and clinical outcome is however lacking, apart from a single instillation (SI) of chemotherapy (5).

We assessed the feasibility of extracting QCIs from our multicenter prospective TURBT registry embedded in our electronic medical record (EMR) within the Vlaams Ziekenhuisnetwerk KU Leuven. In our assessment of QCIs, we consistently used a stepwise approach. Firstly, we assessed the quality of reporting of each QCI. Secondly, we assessed compliance to each QCI, which was then compared between centers, and assessed over time. Finally, we assessed correlation between compliance to each QCI and recurrence-free survival (RFS).

3.3 Materials and methods

Ethical approval was obtained at the Ethical Commission of UZ/KU Leuven and the local ethical commissions of the other participating hospitals (internal study number S55725) in accordance with the ethical standards of the Declaration of Helsinki of 1964 and its later amendments. Written informed consent was obtained preoperatively from participants.

Clinical data of patients treated with TURBT were collected in a prospective registry in three Belgian hospitals. Data were collected using standardized electronic case report forms (eCRFs), which were incorporated within the EMR as reported earlier (6). A reader-friendly surgery report was generated upon completion (6). The collected data within the eCRFs were queried, while pathological and outcome data were retrieved retrospectively from the EMR. The following variables were extracted from eCRFs regarding patient history: date of birth, date of diagnosis, date of surgery, gender, center, recurrence status, muscle-invasive before current TURBT; and regarding the TURBT procedure: tumor number, tumor size, resection of scar, macroscopically CR, macroscopically muscle invasion, SI of mitomycin C (MMC), bladder perforation. The following variables were retrieved retrospectively: tumor stage, tumor grade, presence of DM, carcinoma in situ (CIS), variant histology, EORTC/CUETO recurrence score, European Association of Urology (EAU) risk categories, start of Bacillus Calmette-Guérin (BCG), and date of start of BCG, re-TURBT, and date of

re-TURBT, recurrence, and date of recurrence. Exclusion criteria were benign disease, and muscle-invasive disease before current TURBT. In all participating hospitals, the chemotherapy of choice for a SI was MMC.

Six QCIs were assessed: complete resection status (CR), presence of DM, a SI of MMC, re-TURBT, start of BCG, and TURBT ≤6 weeks after initial diagnosis (T≤6W) (4). The quality of reporting of QCIs was assessed for all patients, while compliance to QCIs was assessed for each QCIs specific subgroup of relevance as specified in **Table 1**. Variables were reported in absolute numbers and percentages, and compared using chi-squared test (categorical), ANOVA (numerical), or Jonckheere-Terpstra test (ordinal). RFS was defined as time from TURBT to first recurrence at cystoscopy. Kaplan-Meier plots and log-rank tests were used for survival analyses. Analyses were performed in R (version 4.0.0) using the 'survival, 'survminer', and 'lubridate' packages.

3.4 Results

Between June 2013 and March 2017, 1350 TURBT procedures were registered for 1043 individual patients (Supplementary **Figure 1**). After applying inclusion and exclusion criteria, 1151 TURBTs in 907 patients were retained (**Figure 1**). Clinicopathological characteristics are summarized in **Table 2**. Patients had the following EAU risk categories after TURBT: 271 low-, 464 intermediate-, and 416 high-risk NMIBC. The 3-year RFS was 61.0% (95% confidence interval [CI]: 56.9-65.4), with no significant difference in RFS between centers. **Table 1** illustrates each QCIs specific subgroup of relevance used for further analyses.

3.4.1 Complete resection status (CR)

CR was collected as a compulsory variable within the eCRF, thus reporting on CR was 100% (n=1151). The subgroup of patients to calculate compliance to CR were patients with: Ta disease, T1 disease, or in case of a re-TURBT for new or residual disease (n=706). Compliance to CR was 97.0% with a significant difference between centers (p=0.02; **Figure 2A; Supplementary Table 1**). There was no significant change in compliance to CR over time. The 3-year RFS in this subgroup was 60.1% (95%CI: 55.6-64.8) with no significant difference when stratifying for CR.

Table 1: Overview of selection criteria for each QCIs subgroup of patients of relevance. The following is reported: inclusion and exclusion criteria per subgroup; number of patients that were included per subgroup; and compliance to each QCI (yes vs. no vs. not reported).

QCI	Inclusion criteria	Exclusion criteria	n	Yes	No	NR
Complete resection status	• Ta • T1 *Type of TURBT:* • first TURBT • TURBT for recurrent disease • re-TURBT for new lesion • re-TURBT for incomplete resection (at prior TURBT)	• Tis • T2 *Type of TURBT:* • TURBT for ad random biopsy • re-TURBT of scar • re-TURBT for no detrusor muscle in specimen	706	685 (97.0%)	21 (3.0%)	0* (0%)
Presence of detrusor muscle	• Ta HG/G3 • T1 • T2 *Type of TURBT:* • first TURBT • TURBT for recurrent disease • re-TURBT for new lesion • re-TURBT for incomplete resection (at prior TURBT) • re-TURBT for no detrusor muscle in specimen	• Tis • Ta LG/G1-2 *Type of TURBT:* • TURBT for ad random biopsy • re-TURBT of scar	694	213 (30.7%)	249 (35.9%)	232 (33.4%)
Single instillation of Mitomycin C	• Low-risk tumors (Primary + n=1 + <3 cm) • Intermediate-risk tumors (Primary + n=1 + ≥3 cm) (Primary + n=2-7 + <3 cm) (≤1 rec/yr + n=1 + <3 cm) *Type of TURBT:* • first TURBT • TURBT for recurrent disease • re-TURBT for new lesion • re-TURBT for incomplete resection (at prior TURBT)	• Incomplete resection • Macroscopically muscle-invasive tumors • Bladder perforation • Bleeding *Type of TURBT:* • TURBT for ad random biopsy • re-TURBT of scar • re-TURBT for no detrusor muscle in specimen	172	112 (65.1%)	60 (34.9%)	5* (0.4%)
re-TURBT	• T1 • Incomplete resection • No detrusor muscle (except Ta LG/G1-2 or Tis)	• Macroscopically muscle-invasive tumors	303	100 (33.0%)	203 (67.0%)	0# (0%)
Start of BCG induction	• Intermediate-risk tumors • High-risk tumors	• Prior BCG instillations	477	188 (39.4%)	289 (60.6%)	0# (0%)
TURBT ≤6 weeks after diagnosis	• All patients with first TURBT • TURBT date within ≤6 weeks after diagnosis	• NA	595	522 (87.7%)	73 (12.3%)	0# (0%)

*Collected as a compulsory variable within the eCRF, therefore reporting was 100%, except for MMC.#Collected retrospectively from the electronic medical record, therefore reporting was 100%. Abbreviations: BCG: Bacillus Calmette-Guérin; NA: not applicable; NR: not reported; QCI: quality control indicator; TURBT: transurethral resection of the bladder tumor.

There were significant differences in tumor stage, and tumor grade between patients that had a complete vs. incomplete resection (**Supplementary Table 2**). Of these variables, we found that tumor stage (p=0.02) correlated significantly with RFS in this subgroup of patients. Additionally, tumor size (≤3 cm vs. >3 cm; p=0.004) correlated significantly with RFS, while tumor number and presence of CIS did not. Blue light cystoscopy was used in 172 TURBTs (15%) with no significant difference between centers (p=0.5).

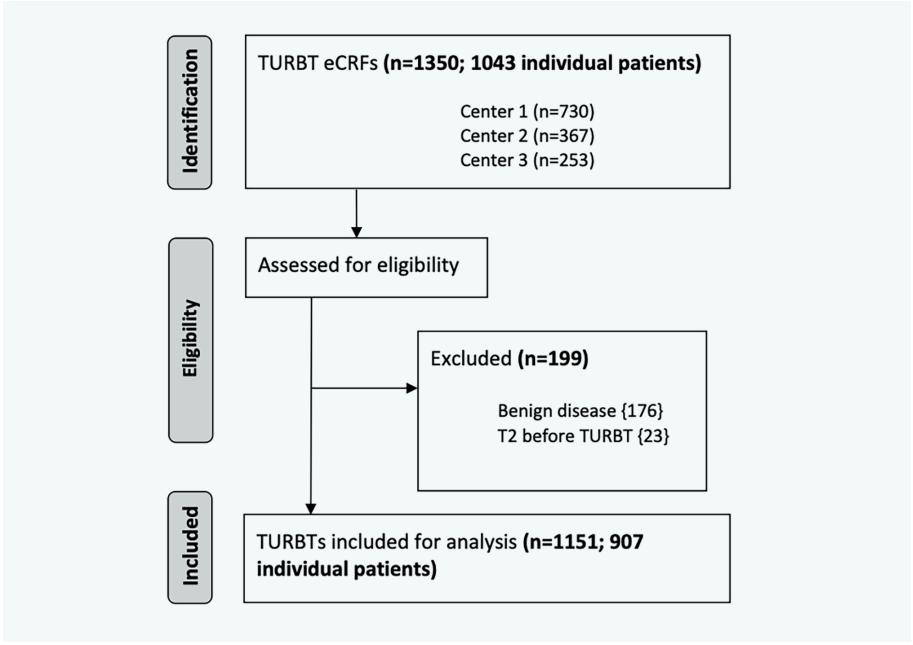


Figure 1: Flow diagram of the procedure for the selection of patients for our analyses.

3.4.2 Presence of detrusor muscle (DM)

DM was collected retrospectively from the EMR, and therefore depended on the quality of the pathologists reporting. DM was reported in 65.5% (n=754), and not reported in 34.5% (n=397). Lack of reporting on DM could be attributed for 94.4% (n=375) to one center, that did not routinely report the presence of DM. The subgroup of patients to calculate compliance to DM were all patients, except those with Tis or Ta LG/G1-2 (n=694). Compliance to DM was 30.7%, with a significant difference between centers ($p<0.0001$; **Figure 2B; Supplementary Table 1**). There was no significant change in compliance to this QCI over time. The 3-year RFS in this subgroup was 63.2% (95%CI: 58.8-68.0) with no significant differences when stratifying for DM. There were significant differences in tumor stage, tumor grade, CIS, EORTC/CUETO recurrence scores between patients stratified by DM (yes vs. no vs. not reported; **Supplementary Table 3**). Of these variables, we found that tumor stage ($p<0.0001$), and tumor grade ($p<0.0001$) correlated significantly with for RFS in this subgroup of patients.

Table 2: Clinical and pathological characteristics of patients per center, and all patients combined. Comparisons between centers were made with chi-squared test ($\chi 2$) or analysis of variance (ANOVA).

		All centers		Center 1		Center 2		Center 3		
		n	%	n	%	n	%	n	%	p-value ($\chi 2$/ ANOVA)
Procedures	Total	1151		572	49.7	345	30.0	234	20.3	
Gender										0.27
	Male	928	80.6	470	82.2	286	82.9	180	76.9	
	Female	223	19.4	108	18.9	64	18.6	54	23.1	
Age										0.27
	Mean (95 %CI)	71.9	71.2-72.5	71.3	70.4-72.3	72.5	71.4-73.5	72.2	70.8-73.5	
	Median (IQR)	73.0	66.0-79.0	73.0	65.0-80.0	73	66.0-79.0	73	66.0-79.0	
	<50	40	3.5	25	4.4	7	2.0	8	3.4	
	50 – 59	90	7.8	47	8.2	27	7.8	16	6.8	
	60 – 69	309	26.8	156	27.3	95	27.5	58	24.8	
	70 – 79	427	37.1	198	34.6	135	39.1	94	40.2	
	≥80	285	24.8	146	25.5	81	23.5	58	24.8	
Tumor stage										<0.0001
	T0	206	17.9	154	26.9	36	10.4	16	6.8	
	Ta	539	46.8	209	36.5	199	57.7	131	56.0	
	T1	171	14.9	106	18.5	33	9.6	38	16.2	
	Tis	77	6.7	41	7.2	25	7.2	11	4.7	
	≥T2	152	13.2	62	10.8	52	15.1	38	16.2	
Tumor grade										<0.0001
	G1	463	40.2	322	56.3	111	32.2	30	12.8	
	G2	208	18.1	32	5.6	92	26.7	84	35.9	
	G3	272	23.6	62	10.8	105	30.4	105	44.9	
	No tumor	186	16.2	136	23.8	35	10.1	15	6.4	
	NA	22	1.9	20	3.5	2	0.6	0	0.0	

Carcinoma in situ									0.65
Yes	100	8.7	47	8.2	34	9.9	19	8.1	
No	1050	91.2	530	92.7	311	90.1	215	91.9	
NR	1	0.1	1	0.2	0	0.0	0	0.0	
EORTC score* Median (IQR)									
Recurrence	4	3-6	4	2-5	4	2-6	5	3-7	<0.0001
CUETO score* Median (IQR)									
Recurrence	4	2-6	4	2-6	4	2-6	5	4-7	<0.0001
Prior recurrence rate									<0.0001
Primary	597	51.9	238	41.6	219	63.5	140	59.8	
≤1 per year	257	22.3	172	30.1	53	15.4	32	13.7	
>1 per year	195	16.9	90	15.7	55	15.9	50	21.4	
re-TURBTs	101	8.8	71	12.4	18	5.2	12	5.1	
NR	1	0.1	1	0.2	0	0.0	0	0.0	
Number of tumors									0.46
Single	652	56.6	315	55.1	203	58.8	134	57.3	
2-7	424	36.8	210	36.7	126	36.5	88	37.6	
≥8	70	6.1	42	7.3	16	4.6	12	5.1	
NR	5	0.4	5	0.9	0	0.0	0	0.0	
Largest tumor size									0.33
≤3cm	736	63.9	358	62.6	224	64.9	154	65.8	
>3cm	116	10.1	59	10.3	28	8.1	29	12.4	
NR	299	26.0	155	27.1	93	27.0	51	21.8	
Variant histology									<0.0001
Yes	34	3.0	12	2.1	9	2.6	13	5.6	
No	212	18.4	83	14.5	110	31.9	19	8.1	
NR	905	78.6	477	83.4	226	65.5	202	86.3	
Perforation									0.08
Yes	128	11.1	52	9.1	43	12.5	33	14.1	
No	1014	88.1	516	90.2	299	86.7	199	85.0	
NR	9	0.8	4	0.7	3	0.9	2	0.9	

*: calculated for Ta and T1 disease only

Abbreviations: n: number; χ2: Chi squared; ANOVA: analysis of variance; 95%CI: 95% confidence interval; IQR: Interquartile range; NR: not reported; EAU: European Association of Urology; TURBT: transurethral resection of bladder tumor; CUETO: Club Urológico Español de Tratamiento Oncológico.

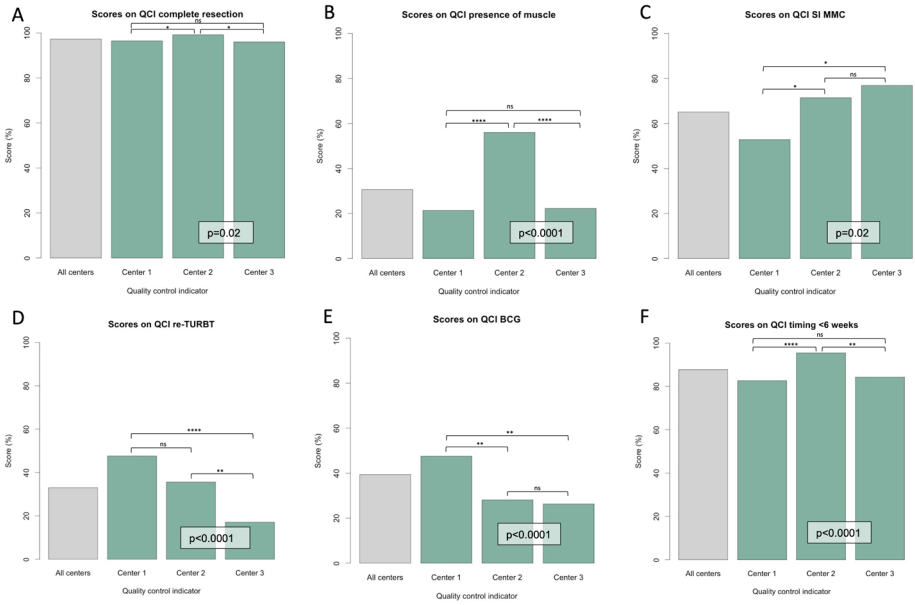


Figure 2: Compliance to each QCI visualized per center, and all centers combined. QCIs were compared between centers by Chi-squared test. Compliance to all QCIs differed significantly between centers. **A**: complete resection; **B**: presence of detrusor muscle; **C**: single instillation of Mitomycin C (SI MMC); **D**: re-TURBT; **E**: start of Bacillus Calmette-Guérin (BCG) induction; **F**: start of initial therapy ≤6 weeks after diagnosis (timing ≤6 weeks).

3.4.3 Single instillation of Mitomycin C (MMC)

MMC was collected as a compulsory variable within the eCRF. It was reported in 99.6% (n=1146), and not reported in 0.4% (n=5). The subgroup of patients to calculate compliance to MMC were patients with low- and intermediate-risk tumors that underwent a macroscopically complete resection without bladder perforation (n=172). Compliance to MMC was 65.1%, with a significant difference between centers (p=0.02; **Figure 2C; Supplementary Table 1**). There was a significant increase in compliance to this QCI over time (p=0.01; **Figure 3A**), which could be attributed to a significant increase in one center (p=0.01; **Figure 3B**). The 3-year RFS in this subgroup was 69.0% (95%CI: 60.9-78.3) with a significant difference when stratifying for MMC (p=0.04; **Figure 4A**). There were significant differences in tumor stage, tumor grade, CIS, EORTC/CUETO recurrence scores between patients stratified by MMC (yes vs. no; **Supplementary Table 4**). Of these variables, we found that tumor stage

(p<0.0001), and presence of CIS (p<0.0001) correlated significantly with RFS in this subgroup of patients. Additionally, tumor size differed significantly between patients who did and did not receive MMC (p=0.04), while tumor number did not differ significantly. The main reasons recorded in the eCRF for not administering a SI of MMC after TURBT were deep/extensive resection (35.0%), patient comorbidities (18.3%), previous intravesical therapy (15.0%), surgeons' decision (13.3%), and absence of clinical suspicion for malignancy (6.7%) (**Supplementary Table 5**).

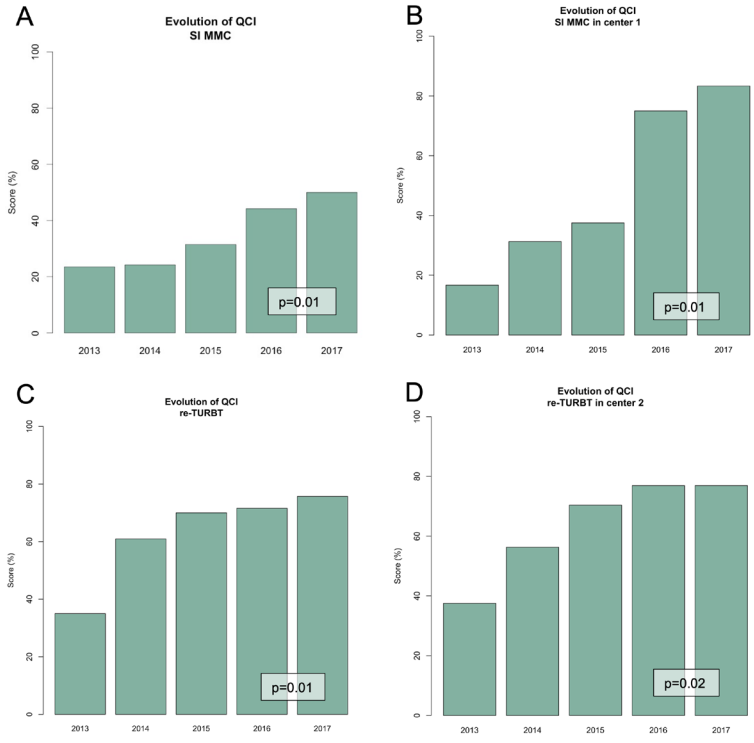


Figure 3: Evolution of compliance with QCI visualized per year between 2013 and 2017

3.4.4 Re-TURBT

Each performed TURBT and subsequent re-TURBT were collected within the eCRF, and thus reporting of re-TURBT was 100% (n=1151). The subgroup of patients to calculate compliance re-TURBT were patients with a macroscopically incomplete resection, with no DM (Ta LG/G1-2 excluded), or

with high-risk disease (at first resection, Tis excluded) (n=303). Compliance to re-TURBT was 33.0%, with a significant difference between centers (p<0.0001; **Figure 2D; Supplementary Table 1**). There was a significant increase in compliance to this QCI over time (p=0.01; **Figure 3C**), which could be attributed to a significant increase in one center (p=0.02; **Figure 3D**). The 3-year RFS in this subgroup was 59.3% (95%CI: 52.7-66.7) with no significant differences when stratifying for re-TURBT. There were significant differences in tumor stage, tumor grade between the patients stratified by re-TURBT (yes vs. no; **Supplementary Table 6**). However, none of the variables that differed correlated significantly with RFS in this subgroup of patients.

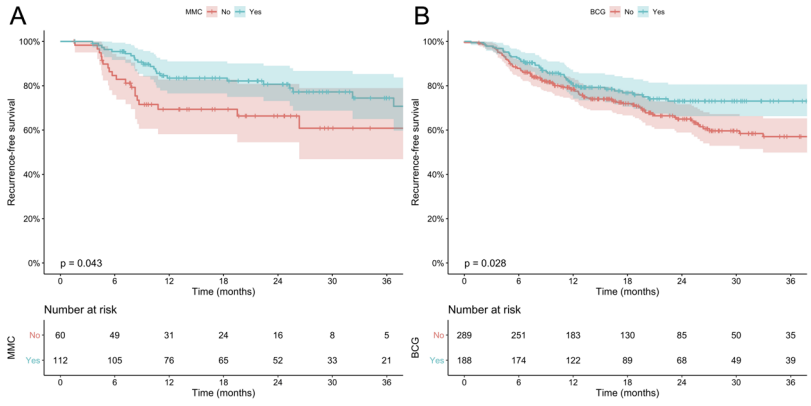


Figure 4: Recurrence-free survival (RFS) of patients stratified by compliance to QCI, for each QCIs specific subgroup of patients of relevance (as specified in Table 2). **A:** A single postoperative instillation of Mitomycine C; **B:** Start of Bacillus Calmette–Guérin (BCG) induction. Compliance to both QCIs correlated to a significant better RFS of patients.

3.4.5 Start of BCG induction (BCG)

BCG was collected retrospectively from the EMR, and was thus dependent on the quality of the urologists reporting. BCG was reported in 49.2% (n=566), and not reported in 50.8% (n=585). The subgroup of patients to calculate compliance to this QCI were patients with intermediate- or high-risk tumors and no prior BCG instillations (n=477). Compliance to BCG was 39.4%, with a significant difference between the centers (p<0.0001; **Figure 2E**; **Supplementary Table 1**). There was no significant change in compliance to this QCI over time. The 3-year RFS in this subgroup was 63.7% (95%CI: 58.5-69.3) with a significant difference when stratifying for BCG (p=0.03; **Figure 4A**). There

were significant differences in tumor stage, tumor grade, and EORTC recurrence score between patients stratified by BCG (yes vs. no; **Supplementary Table 7**). Of these variables, we found that tumor stage (p<0.0001), and tumor grade (p<0.0001) correlated significantly with RFS in this subgroup of patients. A total of 188 patients completed BCG induction (at least 5/6 induction courses), 91 patients completed 1 year of maintenance BCG (19%), and 17 patients completed 3 years of maintenance BCG (4%). Additionally, 31 of these patients (16%) stopped BCG therapy during maintenance due to intolerance or side effects. A total of 27 patients underwent cystectomy, of which 11 patients (41%) started with BCG, and 16 patients (59%) did not.

3.4.6 TURBT ≤6 weeks after primary diagnosis (T≤6W)

T≤6W was calculated using the date of diagnosis, and the date of TURBT and was thus reported in 100% (n=1151). The subgroup of patients to calculate compliance to T≤6W were patients with a primary diagnosis of BC (n=597). Compliance to T≤6W was 87.7%, with a significant difference between centers (p<0.0001; **Figure 2F; Supplementary Table 1**). There was no significant change in compliance to this QCI over time. The 3-year RFS in this subgroup was 68.7% (95%CI: 64.2-73.5) with no significant differences when stratifying for T≤6W. There was a significant difference in tumor stage between the patients who had and did not have initial TURBT within 6 weeks (**Supplementary Table 8**). Tumor stage (p<0.0001) correlated significantly with RFS in this subgroup of patients.

3.5 Discussion

We report on assessment of six QCIs from our prospective TURBT registry. First, we identified that quality of reporting on some QCIs was suboptimal. Secondly, compliance to all QCIs differed significantly between centers (**Figure 2**). Thirdly, compliance to two QCIs (MMC and re-TURBT) increased over time, which could be attributed to a significant increase in one center (**Figure 3**). Fourthly, we identified that compliance to two QCIs (MMC and BCG) correlated with the RFS in their specific subgroup of patients (**Figure 4**). Differences in reporting on QCIs can be explained by not filling in the eCRFs or by lack of collecting variables in the EMR, whilst differences in compliance to QCIs can be explained by differences in clinical practice. Lastly, there was no significant difference in RFS between centers despite significant differences in clinicopathological characteristics (**Table 2**). This conflicting finding could also be explained by differences in clinical practice.

The six assessed QCIs are accepted as standard of care in the EAU guidelines (7). However, adherence to these guidelines is reportedly low (8–14). This is illustrated by comparing compliance to three QCIs in a series of European centers with our registry, which are somewhat in line: MMC 43% vs. 65%, re-TURBT 43% vs. 33%, and BCG 65% vs. 39% (15). This demonstrates the need for strategies to improve guideline adherence. Additionally, without inferring causal relationship from observational data, the correlation of compliance to QCIs with RFS will potentially motivate urologists to improve compliance. Furthermore, urologists will subconsciously modify their behavior due to a Hawthorne effect by awareness of being observed in a TURBT registry (16,17). Therefore, participation in a TURBT registry could be seen as a strategy to improve patient care.

Interestingly, we found that compliance to two QCIs increased over time. Additionally, reporting of QCIs increased in time in all three centers, which can be explained by an increasing reporting trend after the eCRFs implementation (**Supplementary Figure 1**). We hypothesize that an eCRF contributes to an increase in compliance. Urologists are actively prompted to provide relevant clinical information, which ensures standardized reporting of variables (6). Postoperative instructions, for example a SI of MMC, are also automatically recorded, and suggested based on patients characteristics. The reporting rate on MMC of 99.6%, illustrates that urologists were actively reminded to administer MMC. Not surprisingly, the subgroup of patients of QCI MMC (yes vs. no) differed significantly in pathological factors that were not known at time of TURBT. Macroscopically judgment of tumor characteristics (e.g. larger tumor size) potentially influenced administration of MMC. Significantly larger tumors were indeed reported in patients that did not receive MMC.

Our study is not devoid of limitations. Firstly, pathological data, outcome data, and three QCIs were not automatically recorded, and retrospectively retrieved. We implemented an eCRF to collect this data during the follow-up of patients. Using this eCRF, it will for example be possible to differentiate why patients did not start with BCG induction, or why start of BCG was delayed. Secondly, this primary analysis required manual input for data collection, and calculation of reporting and compliance to QCIs. Nonetheless, we demonstrated the feasibility of using EMR-embedded data to generate QCIs as an example of lean development. We are working to automatize the reporting to ultimately create a direct feedback loop for participants, and to give periodic feedback to increase awareness. The Scottish QCI program has implemented such a

feedback loop, which provides real time data on QCIs at multidisciplinary meetings (5). Thirdly, compliance to CR was relatively high with 97%, which could be attributed to the patient selection of this subgroup, in which T2 and Tis were excluded. Patients with a higher tumor stage, and higher tumor grade were still at higher risk for an incomplete resection.

We demonstrate that a prospective multi-center TURBT registry allows for standardized data collection and assessment of reporting and compliance to QCIs for TURBTs and NMIBC. Currently, ten Belgian centers are already participating in this prospective TURBT registry, and we are recording >1000 TURBT procedures per year. Other initiatives also collect QCIs, such as the Scottish quality performance indicator program and the international RESECT study (5,18). Combining know-how regarding QCIs allows for optimalisation of data collection and comparisons of QCIs internationally. Implementation of QCIs will allow for critical reviewing of clinical practice and adherence to guidelines, and will, in the end, improve the outcome of patients with NMIBC.

3.6 Conclusion

We demonstrate the feasibility of assessing QCIs of TURBTs in a real-life setting using eCRFs integrated within the EMR. Compliance to QCIs differed significantly between centers, and scores on two QCIs increased over time. Two QCIs were inadequately reported, which will be improved in by implementation of an additional follow-up eCRF. Additionally, two QCIs correlated with clinical outcome of patients.

3.7 Acknowledgments

Steven Joniau is a senior clinical researcher of the Research Foundation of Flanders (FWO).

3.8 Authors Contributions

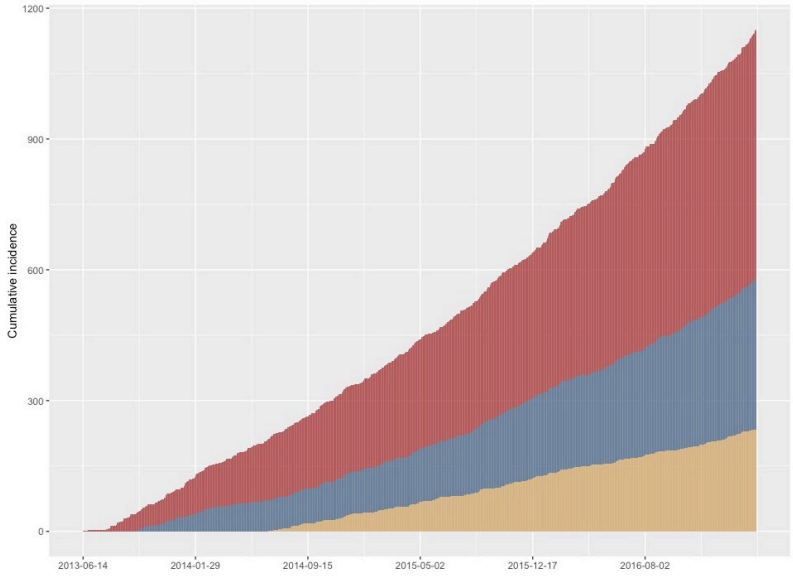
TM: Acquisition of data, analysis and interpretation of data, drafting of the manuscript, critical revision of the manuscript for important intellectual content, statistical analysis. MA: Acquisition of data, analysis and interpretation of data, drafting of the manuscript, critical revision of the manuscript for important intellectual content. YR: Acquisition of data. CY: Acquisition of data. KV: Acquisition of data, critical revision of the manuscript for important

intellectual content. SV: Acquisition of data, critical revision of the manuscript for important intellectual content. BV: Acquisition of data, critical revision of the manuscript for important intellectual content. SJ: Analysis and interpretation of data, critical revision of the manuscript for important intellectual content, supervision. FV: Conception and design, acquisition of data, analysis and interpretation of data, critical revision of the manuscript for important intellectual content, supervision.

3.9 Conflict of Interest

The authors declare that they have no conflict of interest related to this manuscript.

3.10 Supplementary Figures



Supplementary Figure 1: Cumulative recruitment plot of three centers between 2013 and 2017. Center 1 in red, center 2 in blue, and center 3 in yellow.

3.11 Supplementary Tables

Supplementary table 1: Compliance with QCIs of each center and the combined score between June 2013 and March 2017. A comparison between centers was performed using a Chi-squared test.

QCI	Center	Compliance (%)	Procedures (n)	p-value (χ2)
Complete resection status	1	95.8	296	0.02
	2	99.6	229	
	3	95.8	160	
	Total	97.0	706	
Presence of detrusor muscle	1	21.4	346	<0.0001
	2	56.0	182	
	3	22.3	166	
	Total	30.7	694	
SI of MMC	1	52.9	71	0.02
	2	71.4	63	
	3	76.9	39	
	Total	65.1	173	
re-TURBT	1	47.6	105	<0.0001
	2	35.6	87	
	3	17.1	111	
	Total	33.0	303	
Start of BCG	1	47.6	286	<0.0001
	2	28.1	96	
	3	26.3	95	
	Total	39.4	477	
Timing ≤6 weeks	1	82.6	236	<0.0001
	2	95.4	219	
	3	84.3	140	
	Total	87.7	595	

Abbreviations: QCI: quality control indicator; n: number; χ2: Chi squared; CR: complete resection status; re-TURBT: re-transurethral resection of the bladder tumor; MMC: Mitomycin C; BCG: Bacillus Calmette-Guérin.

Supplementary Table 2: Clinical and pathological characteristics of the subgroup of patients of interest to calculate compliance with QCI: complete resection status (n=706). Differences between continuous and categorical variables were assessed using T-test and Chi-squared test, respectively.

		Yes		No		
		n	%	n	%	p-value
Procedures	Total	685	97,0	21	3,0	
Tumor stage						<0,0001
	T0	NA	NA	NA	NA	
	Ta	529	77,2	5	23,8	
	T1	156	22,8	16	76,2	
	Tis	NA	NA	NA	NA	
	≥T2	NA	NA	NA	NA	
Tumor grade						0,02
	G1	374	54,6	10	47,6	
	G2	187	27,2	2	9,5	
	G3	122	17,8	9	42,9	
	No tumor	1	0,2	0	0,0	
	NA	1	0,2	0	0,0	
Carcinoma in situ						0,5
	Yes	17	2,5	0	0,0	
	No	668	97,5	21	100,0	
	NR	0	0,0	0	0,0	
EORTC score	Median (IQR)					
	Recurrence	4	3-7	4	2-7	0,8
CUETO score	Median (IQR)					
	Recurrence	5	2-6	5	3-6	0,8

Abbreviations: n: number; NR: not reported; IQR: Interquartile range; EORTC: European Organization for Research and Treatment of Cancer; CUETO: Club Urológico Español de Tratamiento Oncológico

Supplementary Table 3: Clinical and pathological characteristics of the subgroup of patients of interest to calculate compliance with QCI: presence of detrusor muscle (n=694). Differences between continuous and categorical variables were assessed using ANOVA and Chi-squared test, respectively.

		Yes		No		NR		
		n	%	n	%	n	%	p-value
Procedures	Total	213	30,7	249	35,9	232	33,4	
Tumor stage								<0,0001
	T0	59	27,7	31	12,4	116	50,0	
	Ta	63	29,6	169	67,9	2	0,9	
	T1	54	25,4	37	14,9	86	37,1	
	Tis	37	17,4	12	4,8	28	12,1	
	≥T2	NA	NA	NA	NA	NA	NA	
Tumor grade								<0,0001
	G1	1	0,5	1	0,4	111	47,8	
	G2	60	28,2	129	51,8	2	0,9	
	G3	93	43,7	89	35,7	3	1,3	
	No tumor	45	21,1	24	9,6	116	50,0	
	NA	14	6,6	6	2,4	1	0,4	
Carcinoma in situ								<0,0001
	Yes	47	22,1	15	6,0	32	13,8	
	No	166	77,9	234	94,0	199	85,8	
	NR	0	0,0	0	0,0	1	0,4	
EORTC score*	Median (IQR)							
	Recurrence	6	4-8	5	4-7,8	5	4-7	<0,002
CUETO score*	Median (IQR)							
	Recurrence	6	5-8	6	4-8	6	4-7	<0,0001

*: calculated for Ta and T1 disease only

Abbreviations: n: number; NR: not reported; IQR: Interquartile range; EORTC: European Organization for Research and Treatment of Cancer; CUETO: Club Urológico Español de Tratamiento Oncológico

Supplementary Table 4: Clinical and pathological characteristics of the subgroup of patients of interest to calculate compliance with QCI: SI of MMC (n=172). Differences between continuous and categorical variables were assessed using ANOVA and Chi-squared test, respectively.

		Yes		No		
		n	%	n	%	p-value
Procedures	Total	112	65,1	60	34,9	
Tumor stage						<0,0001
	T0	2	1,8	7	11,7	
	Ta	91	81,3	28	46,7	
	T1	11	9,8	13	21,7	
	Tis	2	1,8	6	10,0	
	≥T2	9	8,0	6	10,0	
Tumor grade						0,0006
	G1	68	60,7	25	41,7	
	G2	28	25,0	12	20,0	
	G3	14	12,5	16	26,7	
	No tumor	1	0,9	7	11,7	
	NA	1	0,9	0	0,0	
Carcinoma in situ						0,006
	Yes	2	1,8	7	11,7	
	No	110	98,2	53	88,3	
	NR	0	0,0	0	0,0	
EORTC score*	Median (IQR)					
	Recurrence	3	2-4	4	3-5	0,006
CUETO score*	Median (IQR)					
	Recurrence	3	2-5	5	2,75-6	0,04

*: calculated for Ta and T1 disease only

Abbreviations: n: number; NR: not reported; IQR: Interquartile range; EORTC: European Organization for Research and Treatment of Cancer; CUETO: Club Urológico Español de Tratamiento Oncológico

Supplementary Table 5: Reasons for not administering a SI of MMC

Reason	Procedures (n)	%
Deep/extensive resection	21	35
Comorbidities	11	18
Previous intravesical therapy	9	15
Preference surgeon	8	13
Not suspect for UC	4	7
Practical considerations	2	3
Prostatic urethral lesion	2	3
Small lesion	2	3
Resection of ostium	1	2
Total	60	

Abbreviations: SI: single instillation; MMC: Mitomycin C; UC: urothelial cancer.

Supplementary Table 6: Clinical and pathological characteristics of the subgroup of patients of interest to calculate compliance with QCI: re-TURBT (n=303). Differences between continuous and categorical variables were assessed using ANOVA and Chi-squared test, respectively.

		Yes		No		
		n	%	n	%	p-value
Procedures	Total	100	33,0	203	67,0	
Tumor stage						<0,0001
	T0	0	0,0	3	1,5	
	Ta	36	36,0	136	67,0	
	T1	62	62,0	61	30,0	
	Tis	2	2,0	3	1,5	
	≥T2	NA	NA	NA	NA	
Tumor grade						<0,0001
	G1	35	35,0	28	13,8	
	G2	31	31,0	107	52,7	
	G3	34	34,0	64	31,5	
	No tumor	0	0,0	4	2,0	
	NA	0	0,0	0	0,0	
Carcinoma in situ						0,049
	Yes	8	8,0	6	3,0	
	No	92	92,0	197	97,0	
	NR	0	0,0	0	0,0	
EORTC score*	Median (IQR)					
	Recurrence	4	4-6	4	2-6,75	0,5
CUETO score*	Median (IQR)					
	Recurrence	4	3-6	5	3-7	0,2

*: calculated for Ta and T1 disease only

Abbreviations: n: number; NR: not reported; IQR: Interquartile range; EORTC: European Organization for Research and Treatment of Cancer; CUETO: Club Urológico Español de Tratamiento Oncológico

Supplementary Table 7: Clinical and pathological characteristics of the subgroup of patients of interest to calculate compliance with QCI: start of BCG (n=477). Differences between continuous and categorical variables were assessed using ANOVA and Chi-squared test, respectively.

		Yes		No		
		n	%	n	%	p-value
Procedures	Total	188	39,4	289	60,6	
Tumor stage						<0,0001
	T0	114	60,6	57	19,7	
	Ta	36	19,1	166	57,4	
	T1	17	9,0	39	13,5	
	Tis	21	11,2	27	9,3	
	≥T2	NA	NA	NA	NA	
Tumor grade						<0,0001
	G1	28	14,9	126	43,6	
	G2	15	8,0	54	18,7	
	G3	31	16,5	52	18,0	
	No tumor	101	53,7	54	18,7	
	NA	13	6,9	3	1,0	
Carcinoma in situ						0,5
	Yes	24	12,8	31	10,7	
	No	164	87,2	258	89,3	
	NR	0	0,0	0	0,0	
EORTC score*	Median (IQR)					
	Recurrence	7	6-8	7	4-8	0,01
CUETO score*	Median (IQR)					
	Recurrence	6	6-8,5	6,5	6-8	0,06

*: calculated for Ta and T1 disease only

Abbreviations: n: number; NR: not reported; IQR: Interquartile range; EORTC: European Organization for Research and Treatment of Cancer; CUETO: Club Urológico Español de Tratamiento Oncológico

Supplementary Table 8: Clinical and pathological characteristics of the subgroup of patients of interest to calculate compliance with QCI: timing ≤6 weeks (n=597). Differences between continuous and categorical variables were assessed using ANOVA and Chi-squared test, respectively.

		≤6 weeks		>6 weeks		
		n	%	n	%	p-value
Procedures	Total	522	87,7	73	12,3	
Tumor stage						0,02
	T0	0	0,0	0	0,0	
	Ta	284	54,4	50	68,5	
	T1	102	19,5	11	15,1	
	Tis	26	5,0	6	8,2	
	≥T2	110	21,1	6	8,2	
Tumor grade						0,4
	G1	250	47,9	39	53,4	
	G2	114	21,8	19	26,0	
	G3	155	29,7	15	20,5	
	No tumor	2	0,4	0	0,0	
	NA	1	0,2	0	0,0	
Carcinoma in situ						0,5
	Yes	38	7,3	7	9,6	
	No	484	92,7	66	90,4	
	NR	0	0,0	0	0,0	
EORTC score*	Median (IQR)					
	Recurrence	3	1-4	3	1-4	0,8
CUETO score*	Median (IQR)					
	Recurrence	4	2-5	3	2-5	0,9

*: calculated for Ta and T1 disease only

Abbreviations: n: number; NR: not reported; IQR: Interquartile range; EORTC: European Organization for Research and Treatment of Cancer; CUETO: Club Urológico Español de Tratamiento Oncológico

3.12 References

1. De Vos M, Graafmans W, Kooistra M, Meijboom B, Van Der Voort P, Westert G. Using quality indicators to improve hospital care: a review of the literature. Int J Qual Health Care. 2009 Apr 1;21(2):119–29.
2. Khare SR, Aprikian A, Black P, Blais N, Booth C, Brimo F, et al. Quality indicators in the management of bladder cancer: A modified Delphi study. Urol Oncol Semin Orig Investig. 2017;
3. Leow JJ, Catto JWF, Efstathiou JA, Gore JL, Hussein AA, Shariat SF, et al. Quality Indicators for Bladder Cancer Services: A Collaborative Review. Vol. 0, European Urology. Elsevier B.V.; 2019.
4. Akand M, Muilwijk T, Raskin Y, De Vrieze M, Joniau S, Van Der Aa F. Quality Control Indicators for Transurethral Resection of Non–Muscle-Invasive Bladder Cancer. Clin Genitourin Cancer. 2019 Apr 19;
5. Mariappan P, Johnston A, Padovani L, Clark E, Trail M, Hamid S, et al. Enhanced Quality and Effectiveness of Transurethral Resection of Bladder Tumour in Non-muscle-invasive Bladder Cancer: A Multicentre Real-world Experience from Scotland's Quality Performance Indicators Programme. Eur Urol. 2020 Oct;78(4):520–30.
6. Akand M, Muilwijk T, Cornelissen J, Van Bruwaene S, Vander Eeckt K, Baekelandt F, et al. Development of a Prospective Data Registry System for Non-muscle-Invasive Bladder Cancer Patients Incorporated in the Electronic Patient File System. Front Oncol. 2019 Dec 11;9.
7. Babjuk M, Burger M, Capoun O, Cohen D, Compérat EM, Dominguez Escrig JL, et al. European Association of Urology Guidelines on Non-muscle-invasive Bladder Cancer (Ta, T1, and Carcinoma in Situ). Eur Urol. 2022 Jan;81(1):75–94.
8. Gontero P, Oderda M, Altieri V, Bartoletti R, Cai T, Colombo R, et al. Are referral centers for non-muscle-invasive bladder cancer compliant to EAU guidelines? A report from the vesical antiblastic therapy Italian study. Urol Int. 2011 Feb;86(1):19–24.
9. Reek C, Rink M, Bloch M, Hansen J, Chun FK, Schneider A, et al. Versorgungsstandard bei neu diagnostiziertem Blasenkarzinom: Eine prospektive Erhebung in Norddeutschland. Urol - Ausg A. 2013 Jul;52(7):986–90.
10. Siddins MT, Wong VV, Fitzgerald JT, Bamberg LJ. Challenges in non-muscle invasive bladder cancer: Lessons from a regional review. ANZ J Surg. 2011 Dec;81(12):889–94.

11. Madeb R, Golijanin D, Noyes K, Fisher S, Stephenson JJ, Long SR, et al. Treatment of nonmuscle invading bladder cancer: Do physicians in the United States practice evidence based medicine? The use and economic implications of intravesical chemotherapy after transurethral resection of bladder tumors. Cancer. 2009 Jun 15;115(12):2660–70.
12. Gotto GT, Shea-Budgell MA, Dean Ruether J. Low compliance with guidelines for re-staging in high-grade T1 bladder cancer and the potential impact on patient outcomes in the province of Alberta. J Can Urol Assoc. 2016 Feb 1;10(1-2February):33–8.
13. Huang GJ, Hamilton AS, Lo M, Stein JP, Penson DF. Predictors of Intravesical Therapy for Nonmuscle Invasive Bladder Cancer: Results From the Surveillance, Epidemiology and End Results Program 2003 Patterns of Care Project. J Urol. 2008 Aug;180(2):520–4.
14. Hendricksen K, Aziz A, Bes P, Chun FK-H, Dobruch J, Kluth LA, et al. Discrepancy Between European Association of Urology Guidelines and Daily Practice in the Management of Non-muscle-invasive Bladder Cancer: Results of a European Survey. European Urology Focus. 2017;
15. Van Rhijn BWG, Burger M. Bladder cancer: Low adherence to guidelines in non-muscle-invasive disease. Vol. 13, Nature Reviews Urology. Nature Publishing Group; 2016. p. 570–1.
16. Roethlisberger F, Dickson W. Management and the Worker. Cambridge, Mass: Harvard University Press; 1939.
17. Leonard K, Masatu MC. Outpatient process quality evaluation and the Hawthorne Effect. Soc Sci Med. 2006 Nov 1;63(9):2330–40.
18. RESECT – BURST Urology [Internet]. [cited 2020 Apr 29]. Available from: https://bursturology.com/resect/

Chapter 4:

The Cancer of the Bladder Risk Assessment (COBRA) score for Estimating Cancer-Specific Survival after Radical Cystectomy: External Validation in a Large Bi-institutional Cohort

Chapter 4:

The Cancer of the Bladder Risk Assessment (COBRA) score for Estimating Cancer-Specific Survival after Radical Cystectomy: External Validation in a Large Bi-institutional Cohort

Tim Muilwijk[1*], Murat Akand[1,2*], Francesco Soria[3], Andrea Giordano[3], Uros Milenkovic[1], Lisa Moris[1], Thomas Gevaert[4,5], Hendrik Van Poppel[1], Maarten Albersen[1], Paolo Gontero[3], Steven Joniau[1]

[1] Department of Urology, University Hospitals Leuven, Leuven, Belgium;

[2] Department of Urology, Selçuk University, Konya, Turkey;

[3] Department of Urology, Molinette Hospital, University of Turin, Turin, Italy;

[4] Organ Systems, KU Leuven, Leuven, Belgium;

[5] Department of Pathology, AZ Klina, Brasschaat, Belgium

*Shared first authorship

BJU International, July 8th, 2020 (IF 5.59)

4.1 Abstract

Objective: To perform an external validation of the Cancer of the Bladder Risk Assessment (COBRA) score for estimating cancer-specific survival (CSS) after radical cystectomy (RC) in a large bi-institutional cohort of patients.

Patients and methods: Patients treated with RC and lymph node dissection (LND) between May 1996 and July 2017 were retrieved from the RC databases of Leuven and Turin. Collected variables were age at RC, tumor stage, lymph node (LN) density, neoadjuvant chemotherapy, the extent of LND, and nodal stage. The primary outcome was CSS visualized using Kaplan-Meier plots. Cox proportional hazard models were used to assess the impact of variables on CSS. We performed a pairwise comparison between the COBRA score levels using a log-rank test corrected by Bonferroni, and developed a simplified COBRA score with three risk categories. To compare models, we assessed concordance indices (C-indices), receiver operating characteristic (ROC) curves with area under the curve (AUC), calibration plots, and decision curve analysis (DCA). Finally, we compared both COBRA and simplified COBRA models with the established AJCC model.

Results: A total of 812 patients were included. All COBRA score variables had a significant impact on CSS in a Cox proportional hazard model. However, pairwise comparison of the COBRA subscores could not differentiate significantly between all COBRA score levels. Based on these findings, we developed a simplified COBRA score by introducing three categories within the following COBRA score ranges: low- [0-1] vs. intermediate- [2-4] vs. high-risk [5-7]. A pairwise comparison could discriminate significantly between all COBRA risk categories. When finally comparing COBRA and simplified COBRA models with the AJCC model, AJCC performed better than both. C-indices, AUCs, calibration plots and DCA for AJCC were all better compared with the original and simplified COBRA models.

Conclusion: We performed an external validation of the COBRA score in a large bi-institutional cohort. We observed that several risk groups had overlapping CSS, demonstrating suboptimal performance of the COBRA score. Therefore, we constructed a simplified model with three COBRA score risk categories. This model resulted in demarcated risk groups with non-overlapping CSS and good predictive accuracy. However, both COBRA score models were outperformed by the AJCC staging system. Therefore, we conclude that the AJCC staging system should remain the current standard for stratifying patients after RC for CSS.

4.2 Introduction

Radical cystectomy (RC) with pelvic lymph node dissection (LND) with or without neoadjuvant chemotherapy is the standard of care for patients with muscle-invasive bladder cancer (MIBC) (1,2). It has been demonstrated that both tumor stage and nodal stage have a prognostic value for survival after RC with LND (3–5). Furthermore, lymph node (LN) density has been identified as a better prognostic marker than nodal staging, and is calculated by dividing the total number of positive resected LNs by the total number of resected LNs (6–8). The Cancer of the Bladder Risk Assessment (COBRA) score was developed to predict cancer-specific mortality after surgery and was based on the Surveillance, Epidemiology, and End Results (SEER) cancer database (9). The COBRA score is calculated using three variables: age at RC, tumor stage, and LN density. The COBRA score ranges from 0 to 7, and each point increase is associated with a 1.61-fold increase (95% confidence interval [CI]: 1.56-1.65) in the risk of cancer-specific mortality (CSM) (9). The COBRA score can be used to counsel patients more adequately on their cancer-specific prognosis after RC.

Although several nomograms predicting survival after RC have been developed, they are not widely adopted due to their complexity and a high number of included variables (6,7,10). To simplify the prediction of cancer-specific survival (CSS) after RC, the COBRA score was developed based on a series of patients (n=14.828) from the SEER database in the United States (9). One of the differences with the other nomograms is the inclusion of nodal density as a parameter instead of the nodal stage based on the American Joint Committee on Cancer (AJCC) lymph node status, which has low prognostic value because it does not account for the positive nodal burden of a patient (8). Although neoadjuvant chemotherapy (NAC) is standard of care in most countries for patients with MIBC, this variable was not included in the development of the COBRA score because the SEER database lacks data regarding NAC usage. The exclusion of NAC as a variable could impact the accuracy of the COBRA score.

Recently, a Korean group performed the first external validation of the COBRA score in an Asian patient cohort (n=855) and concluded that the COBRA score did not provide completely accurate risk stratification but provided better prognostic information than the AJCC staging system (11). We aimed to provide the first external validation of the COBRA score in a European cohort of patients. Furthermore, we sought to improve the COBRA score by adding the

usage of NAC as an extra variable to the model, and by simplifying the model which could increase its clinical use and implementation in clinical practice.

4.3 Patients and methods

4.3.1 Patient identification

The data of patients who underwent RC and LND, between May 1996 and July 2017, were retrospectively retrieved from the institutional RC databases of Leuven and Turin. Patients without LND at the time of RC, patients with missing data regarding variables of interest and outcome, or patients with clinically node-positive disease and/or metastatic disease at the time of RC were excluded from our analysis.

4.3.2 Lymphadenectomy

In both centers the templates for LND were used as defined by the European Association of Urology (EAU) guidelines: standard LND (S-LND): up to the common iliac arteries, extended LND (E-LND): up to the crossing of the ureters with the common iliac vessels or the aortic bifurcation, with or without the presacral LNs, and super-extended LND (SE-LND): up to the inferior mesenteric artery (12).

4.3.3 Statistical analysis

The COBRA score comprises three variables, which were retrieved from the institutional databases: age at RC, tumor stage, and LN density (**Table 1**). Other collected variables were sex, administration of NAC, the extent of LND, and nodal stage. The primary outcome was CSS that was defined as the time from surgery to the time of cancer-related death. Differences between the SEER cohort and Leuven/Turin cohort were not assessed as continuous variables due to the categorical reporting of variables in the original COBRA paper. However, we used a Chi-squared ($\chi 2$) tests to compare the distribution in categories of variables between the SEER cohort and Leuven/Turin cohort.

The CSS of the patient cohort was plotted using Kaplan-Meier stratified by the total COBRA score. A Cox proportional hazard model was used to assess the impact of the COBRA score and other variables. Using the log-rank test, we performed a pairwise comparison between the COBRA score levels, which was

corrected by Bonferroni for multiple testing. Based on the results of pairwise testing, we developed a simplified COBRA score by introducing three risk categories within specific COBRA score ranges. CSS was stratified by the simplified COBRA score for CSS, and a pairwise comparison by log-rank was performed.

Variable	COBRA Points
Age at RC, y	
<80	0
≥80	1
Tumor stage	
≤T1	0
T2	1
T3, T4	3
Lymph node density	
0	0
>0-0.33	1
≥0.33-0.5	2
>0.5-1	3

Table 1: The Cancer of the Bladder Risk Assessment Score Abbreviations: COBRA, Cancer of the Bladder Risk Assessment Score; RC: radical cystectomy

To compare the COBRA score with the simplified COBRA score and the AJCC staging system, we assessed the concordance indices (C-indices) of both Cox proportional hazard models and plotted receiver operating characteristic (ROC) curves, and compared the area under the curve (AUC). We evaluated the impact of the use of NAC as an extra variable on the C-indices and AUCs of both models. We performed a calibration plot and calculated the time-dependent Brier score and the time-dependent AUC. A decision curve analysis (DCA) was performed to compare the performance of the COBRA score compared with the simplified COBRA score (13).

Analyses were performed in R (version 3.6.0) using the 'dplyr' package for data manipulation, using the 'survival' package for survival analyses, using the 'ROCR' package for visualizing the performance of models, and using 'ggplot2', 'ggpubr', 'survminer', 'pROC' and 'riskRegression' packages for drafting of graphs.

4.4 Results

4.4.1 Clinical and pathological variables

We identified 914 patients who underwent RC for bladder cancer (BC) in Leuven and Turin between May 1996 and July 2017 (**Figure 1**). Of these, we excluded 101 patients who had no LND and 1 patient who lacked survival data. A total of 812 patients were included for final analyses. Characteristics of the original development cohort from the SEER database and from the Leuven/Turin validation cohort are visualized in **Table 2**. Characteristics were stratified by cohort (SEER vs. Leuven/Turin) and by pathological LN status at RC (lymph node-positive [LN+] vs. lymph node-negative [LN-]).

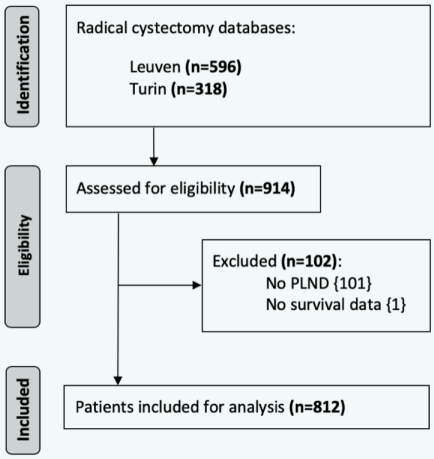


Figure 1: Flow diagram of patients' selection process.

The median follow-up was 29 months (Interquartile range [IQR]: 11-45 months) in the Leuven/Turin cohort and 29 months (IQR: 12-71 months) in the SEER cohort. The median total LN count was 15 (IQR: 9-22) in the Leuven/Turin cohort and 10 (IQR: 5-18) in the SEER cohort and differed significantly in both LN+ and LN- patients ($\chi 2$: LN+ and LN-: $p<0.001$). The median number of positive LNs, in patients with node-positive disease-only, was 2 (IQR: 1-7) in the Leuven/Turin cohort and 2 (IQR: 1-4) in the SEER cohort ($\chi 2$: $p<0.001$). Median LN densities were, however, comparable between both cohorts: 0.21

(IQR: 0.08-0.50) in the Leuven/Turin cohort and 0.22 (IQR: 0.1-0.5) in the SEER cohort ($\chi 2$: LN+: p=0.2). The distribution of the tumor stage in patients in the Leuven/Turin and SEER cohort were pT1:
23% vs. 13%, pT2: 14% vs. 40%, pT3: 24% vs. 30%, and pT4: 39% vs. 17%. Tumor stage was more advanced in patients in the Leuven/Turin cohort compared to the SEER cohort, in both LN+ and LN- patients ($\chi 2$: LN+ and LN-: p<0.001). Age at RC was evenly distributed between the Leuven/Turin cohort and the SEER cohort ($\chi 2$: LN-: p=0.2; LN+: p=0.6) with a median age of 69.0 years (IQR: 61.8-75.1) in our LN- cohort, and 69.0 years (IQR: 62.3-75.8) in our LN+ cohort (**Table 2**).

	SEER [n=14.828]				Leuven/Turin [n=812]			
	LN-		LN+		LN-		LN+	
	No.	*%*	*No.*	*%*	*No.*	*%*	*No.*	*%*
Gender								
Male	8713	78	2699	74	499	79	148	82
Female	2466	22	950	26	132	21	33	18
Total	11179	100	3649	100	631	100	181	100
Age at RC, y								
Median [IQR]	NR	NR	NR	NR	69.0	[61.8-75.1]	69.0	[62.3-75.8]
≤60	2473	22	894	24	132	21	37	20
60-69	3537	32	1175	32	207	33	64	35
70-79	3845	34	1152	32	233	37	59	33
≥80	1324	12	428	12	59	9	21	12
Total	11179	100	3649	100	631	100	181	100
Tumor stage								
pT0	NR	NR	NR	NR	117	19	5	3
pTa, pTis, pT1	1801	16	68	2	149	24	12	7
pT2	5022	46	783	22	85	13	11	6
pT3	2783	25	1590	45	125	20	42	23
pT4	1387	13	1123	32	155	25	111	61
Total	10993	100	3564	100	631	100	181	100
Total LNs								
Median [IQR]	NR*	NR*	NR*	NR*	14	[8-21]	10	[10-25]
0-5	3042	27	847	23	99	16	20	11
6-10	2660	24	872	24	110	17	27	15
11-15	1960	18	648	18	133	21	39	22
16-20	1228	11	481	13	120	19	24	13
≥21	2289	20	801	22	169	27	71	39
Total	11179	100	3649	100	631	100	181	100
Positive LNs								
Median [IQR]	0	0	2	[1-4]	0	0	2	[1-7]
0	11179	100	0	0	631	100	0	0
1	0	0	1497	41	0	0	63	35
2-4	0	0	1448	40	0	0	57	31
≥5	0	0	704	19	0	0	61	34
Total	11179	100	3649	100	631	100	181	100
LN density								
Median [IQR]	0	0	0.22	[0.1-0.5]	0	0	0.21	[0.08-0.50]
≤0.125	11179	100	1184	32	631	100	70	39
>0.125-0.25	0	0	847	23	0	0	35	19
>0.25-0.50	0	0	881	24	0	0	36	20
>0.50	0	0	737	20	0	0	40	22
Total	11179	100	3649	100	631	100	181	100
Neoadjuvant chemotherapy								
No	NR	NR	NR	NR	535	85	153	85
Yes	NR	NR	NR	NR	96	15	28	15
Total	NR	NR	NR	NR	631	100	181	100
LND template								
Limited/Standard	NR	NR	NR	NR	367	58	94	52
Extended	NR	NR	NR	NR	229	36	79	44
Super-extended	NR	NR	NR	NR	35	6	8	4
Total	NR	NR	NR	NR	631	100	181	100
Total	11179	100	3649	100	631	100	181	100

Table 2. Characteristics of the patients from the SEER and Leuven/Turin cohort.*Combined LN- and LN+: median 10 [IQR: 5-18]; Abbreviations: SEER: Surveillance, Epidemiology, and End Results; RC: radical cystectomy; y: years; LN-: lymph node-negative at final pathology; LN+: lymph node-positive at finale pathology; LN(s): lymph node(s); LND: lymph node dissection; NR: not reported; NA: not applicable.

4.4.2 Survival analyses

Out of the 812 patients, a total of 256 patients (32%) experienced cancer-specific death during follow-up. Median follow-up for all patients was 29 months (IQR: 11-45 months) and was 44 months (IQR: 16-55 months) for patients that did not die of BC. On Kaplan-Meier analysis, in which patients were stratified based on their tumor stage (<pT2 vs. pT2 vs. >pT2) and nodal positivity (pN0 vs. pN+), we found that patients with pN+ disease had significantly worse CSS (**Supplementary Figure 1**). On separate log-rank comparisons, we found that both tumor stage and nodal density stratified the patients significantly for CSS, which was not the case for age at RC (<80 or ≥80 y) and administration of NAC. Pairwise comparison between each of the tumor stages could not stratify significantly between the following tumor stages: pT0 vs. pT1; pT0 vs. pT2; pT1 vs. pT2; and pT1 vs. pT3 (**Supplementary Table 1**). A pairwise comparison between each of the different LN density scores (0, >0-0.33, >0.33-0.5, and >0.5-1) only demonstrated significant differences between a score of 0 and any of the higher scores, but could not differentiate significantly between any of the higher scores (>0-0.33, >0.33-0.5, and >0.5-1).

4.4.3 Survival analyses – COBRA score

Figure 2 demonstrates the CSS of the Leuven/Turin cohort stratified by the COBRA score. A Cox proportional hazard model that incorporated the components of the COBRA score [LN density (continuous), age at RC (continuous), and tumor stage (factor)] as independent variables showed a significant impact of all variables on CSS **(Supplementary Table 2**). On Cox proportional hazard model with the COBRA score as a variable, we found that the hazard ratio (HR) did not increase between COBRA score levels 2 vs. 3, and between COBRA score levels 5 vs. 6 (**Supplementary Table 3**). On pairwise comparison by log-rank, the COBRA score could not differentiate significantly between all of the COBRA score levels (**Supplementary Table 4**).

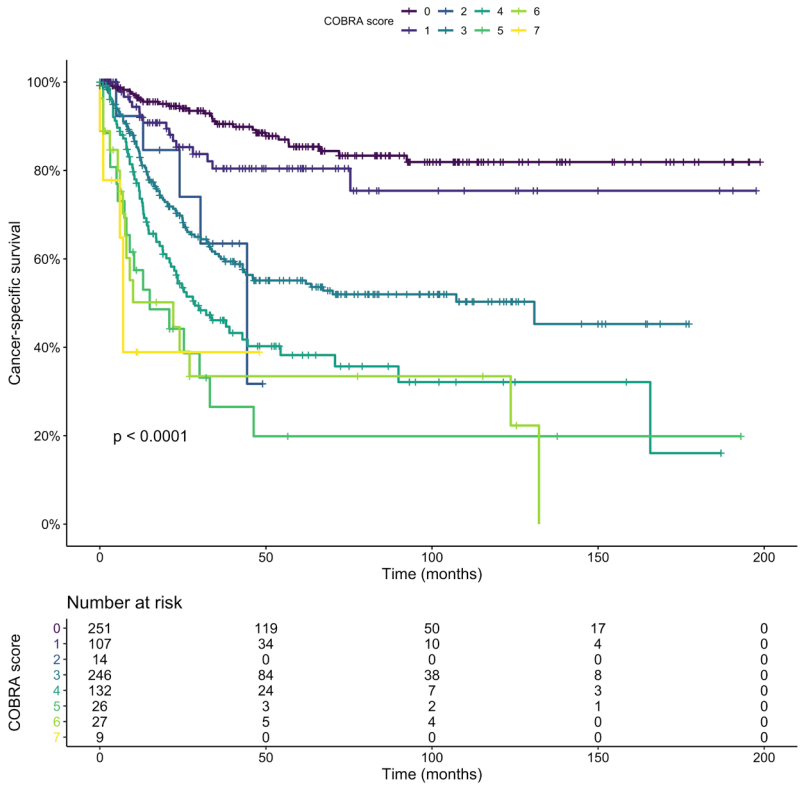


Figure 2: Estimated cancer-specific survival (CSS) stratified by COBRA score.

4.4.4 Survival analyses – COBRA risk categories

Based on the results of the pairwise testing, we developed a simplified COBRA score by introducing three risk categories within the following COBRA score ranges: low-risk [0-1] vs. intermediate-risk [2-4] vs. high-risk [5-7]. **Figure 3** demonstrates the CSS stratified by these three novel risk categories. The 5-year CSS of patients with low-, intermediate-, or high-risk were 83.9%, 49.5%, and 26.8%, respectively. The median CSS for intermediate- and high-risk patients were 54.4 months (95%CI: 37.7-NR), and 15.0 months (95%CI: 8.0-30.0), respectively. The median CSS for low-risk patients was not reached. A pairwise comparison could discriminate significantly between all COBRA risk categories (**Supplementary Table 5**).

4.4.5 Comparison of COBRA score, COBRA risk categories and AJCC staging system

The C-indices of the COBRA score and COBRA risk categories were 0.69 and 0.71, respectively. When subgroup analyses based on the extent of LND (S-LND, E-LND, and SE-LND) were performed, the performance of the COBRA score increased with C-indices of 0.70, 0.74, and 0.75, respectively. ROC curves for both models showed comparable AUC: 72.9% for the COBRA score and 71.0% for the COBRA risk categories (**Supplementary Figures 2 and 3**).

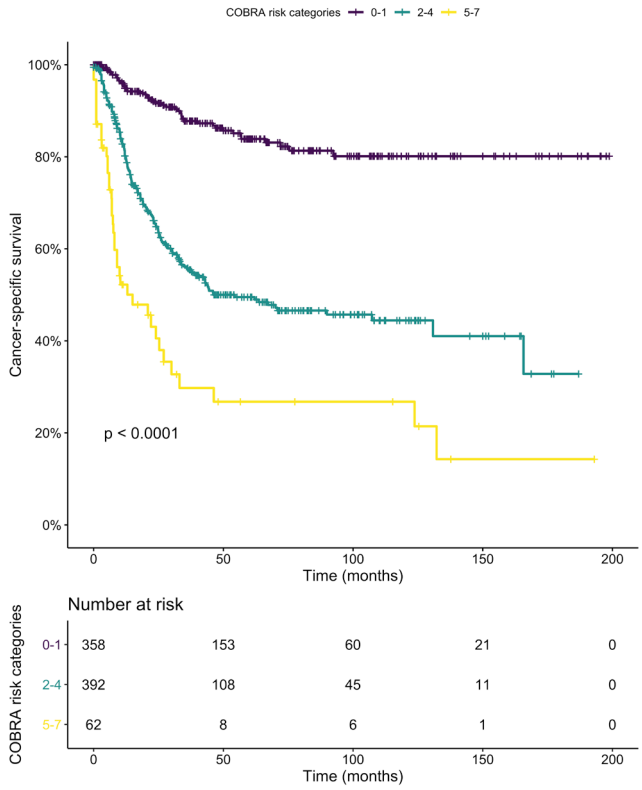


Figure 3: Kaplan-Meier curves for cancer-specific survival (CSS) stratified by COBRA risk categories.

We evaluated the impact of the use of NAC as an extra variable in both models, which caused an increase of the C-index (0.69 to 0.72) and an increase in the AUC (72.9% to 73.3%) for original COBRA score; while it caused a decrease in the C-index (0.71 to 0.69) and an increase in the AUC (71.0% to 71.7%) for the COBRA risk categories. Additionally, we also checked the performance of the AJCC staging system: pairwise comparison by log-rank could not differentiate significantly between stages 0, I, and II, C-index of the AJCC staging system was 0.73, and AUC was 76.8% (14). C-index and AUC were higher compared to the COBRA score and compared to the COBRA risk categories (**Supplementary Figures 4 and 5**).

A calibration plot at a 5-years follow-up showed comparable results for both the COBRA score and COBRA risk categories on time-dependent AUC and time-dependent Brier score, although the calibration plot of the COBRA score risk categories followed the observed-predicted plot more accurately (**Figure 4**). The DCA shows comparable net benefits of the COBRA score compared with the COBRA score risk categories with a more favorable profile of the original COBRA score between the 0.4 and 0.6 threshold probability range (**Figure 5**). These findings were also valid when comparing the COBRA score and the COBRA risk categories with the AJCC staging system, with a more favorable profile of the AJCC staging system on DCAs (**Supplementary Figures 6A-B-C-D**).

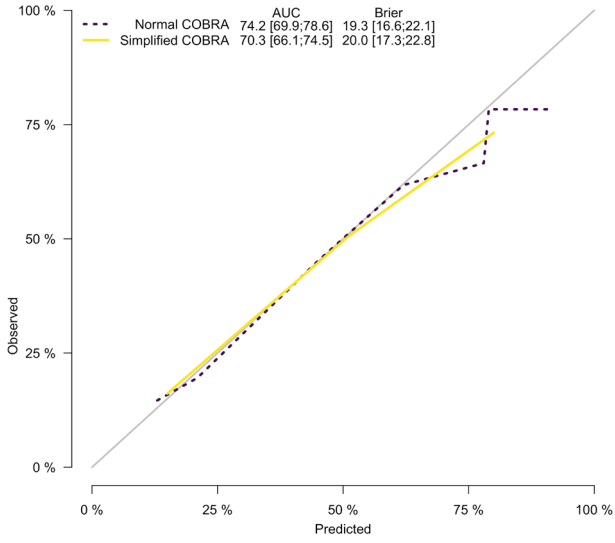


Figure 4: Calibration plot at 5-years for COBRA score and COBRA risk categories.

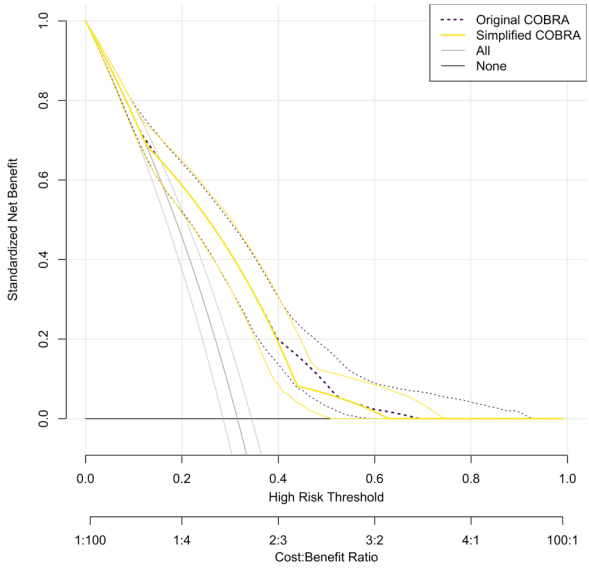


Figure 5: Decision curve analysis for estimating cancer-specific survival (CSS) which compares the COBRA score (dotted, purple) and COBRA risk categories (straight, yellow). The standardized net benefit (NB) is visualized on the y-axis. High-risk threshold probabilities are visualized on the x-axis which can be directly translated to the harm-benefit ratio (Cost:Benefit Ratio). The grey line holds the assumption that all patients will die. The horizontal black line (with Standardized Net Benefit = 0.0) holds the assumption that no patients will die.

4.5 Discussion

We report the first validation of the COBRA score in a European cohort of patients based on cystectomy databases of two tertiary referral centers. First, our study demonstrates that the COBRA score withholds its prognostic value in a large cohort of European patients from two high-volume centers. Second, we suggest a simplified COBRA score based on three risk categories with comparable performance on discrimination (C-indices), calibration plots, and DCA. Third, we found that the implementation of NAC as an extra variable in the original model increased the prognostic accuracy of the COBRA score marginally, but not in the simplified model. Lastly, both COBRA score systems were outperformed by the AJCC staging system, which had a higher C-index and better results on DCAs.

In the original COBRA paper, the authors found a CSM of 31% that is comparable with our CSM of 32%. Follow-up was also comparable in the Leuven/Turin cohort and the SEER cohort. We found in our cohort that all variables of the original COBRA score had a significant impact on CSS on multivariable cox regression analysis. There was a significant difference in the total number of removed LNs between the SEER and Leuven/Turin cohort, with a higher number of LNs removed in our series. A higher number of total removed LNs could theoretically improve the prognostic value of the AJCC staging system, due to a more accurate nodal staging. Indeed, the AJCC staging system outperformed both the original COBRA scoring and the COBRA risk categories in our series, with a C-index of 0.73, and AUC of 76.8%. LN density has also shown its prognostic value on CSS in patients undergoing RC with LND in our patient cohort (HR: 4.8; 95%CI: 2.7-8.5) (15–17). We found a comparable distribution of LN densities in our patient cohort as was reported in the SEER cohort. Interestingly, the discriminative ability of the COBRA score increased with the extension of the LND template (C-indices: L-LND 0.70, E-LND 0.74, SE-LND 0.75). This increase in discriminative ability illustrates that the prediction of CSS using the COBRA score is also highly dependent on the extent of LND, and shows that a more extended template yields a higher prognostic value in our cohort. Whether a more extensive LND yields also a better oncological outcome is currently not clear. Randomized controlled trials (RCTs) in other types of cancers, such as esophageal, gastric, and pancreatic head cancer, could not demonstrate a survival benefit of extending the template of the LND (18–21). The LEA trial was the first prospective randomized phase III trial comparing the oncological outcome between S-LND and SE-LND, and could not demonstrate a survival benefit of extending the LND template (22,23). The SWOG-1011 trial (NCT01224665) is awaited which could provide us the final answer in this matter. Based on the results of the current study and that of the LEA trial, we hypothesize that a more extensive LND template will result in a higher prognostic value of the LN density, and will provide a better adjuvant therapy choice, but whether it yields an improved oncological outcome remains to be seen.

Several nomograms have been proposed in the past, but are currently not implemented in clinical practice (**Supplementary Table 7**) (6,7,24–27). Most of these nomograms use >3 variables, or even up to 10 variables, and are difficult to calculate without using their specific chart of the nomogram or calculator(10). Furthermore, some of the included variables are not relevant in most hospitals: e.g., adjuvant radiotherapy is not standard-of-care in most

countries(7). The two other nomograms with 3 variables apart from the COBRA score were based on relatively small cohorts and were never validated internally or externally. The reasons for lack of implementation of these models can be the multitude of included parameters and the difficulty to calculate the specific risk prediction of a patient. The major strength of the COBRA score itself, is that it is easy-to-use. The necessary clinical parameters to calculate the COBRA score can be easily gathered from patients' files, and the original COBRA score or COBRA risk categories will allow quick determination of the risk category of a patient.

Our study is not devoid of limitations. First, the retrospective design of both RC databases leads to an inherent bias. However, although retrospective in nature, both databases contain complete data regarding the parameters of interest for calculating the COBRA score. Second, we cannot assure that the templates of LND were exactly those as defined by the EAU Guidelines, as differences might have occurred between surgeons and in time, and LNDs might not have been performed equally vigorously in every case by every surgeon. Furthermore, the performance of the COBRA score in our cohort could have been influenced by the small number of patients in some COBRA score categories, as is visible in the risk table of **Figure 2**. There was also a significant difference in tumor stage distribution between the SEER cohort and Leuven/Turin cohort, which could explain differences in performance of the COBRA score and could have influenced the development of the COBRA risk categories (**Table 2**). Lastly, we developed a simplified COBRA score based on risk categories, but we did not perform a validation of this model. This is of interest for future research in which combinations of institutional, national, or international cystectomy databases can be used.

The strengths of our study are the experience of both centers in the treatment of BC and the high number of included patients. The included variables to calculate the COBRA score – age at RC, tumor stage, and LN density – are objective and subject to a low risk of bias, and are easy to retrieve from retrospective data. Furthermore, we suggested an even more simplified model of the COBRA score by introducing three risk categories. Easy-to-use risk categories have shown their clinical benefit, for example, in the prognosis and treatment of patients with low-, intermediate- or high-risk non–muscle-invasive bladder cancer (NMIBC) (28). However, the AJCC staging system outperformed both the original COBRA score and the COBRA score with risk categories, which could be explained by the higher total number of removed

LNs in our series. Contrarily, in the external validation of the COBRA score by the Korean group, the COBRA score provided better prognostic information than the AJCC staging system (11).

4.6 Conclusion

We performed an external validation of the COBRA score in a large bi-institutional cohort. We observed that several risk groups had overlapping CSS, demonstrating suboptimal performance of the COBRA score. Therefore, we constructed a simplified model with three COBRA score risk categories. This model resulted in demarcated risk groups with non-overlapping CSS and good predictive accuracy. However, both COBRA score models were outperformed by the AJCC staging system. Therefore, we conclude that the AJCC staging system should remain the current standard for stratifying patients after RC for CSS.

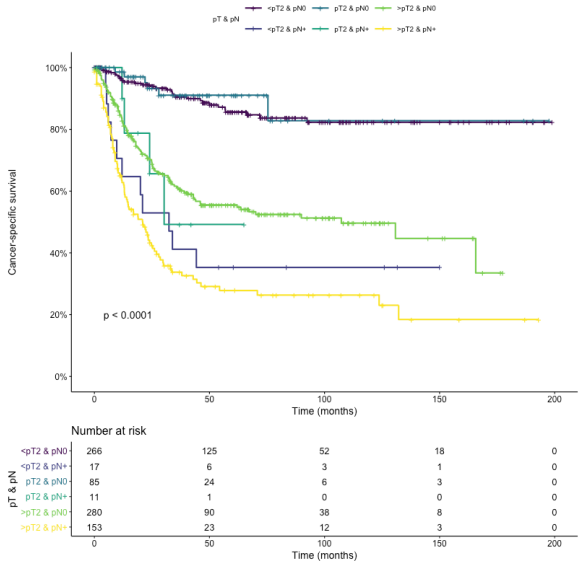
4.7 Acknowledgments

Steven Joniau is a senior clinical researcher of the Research Foundation of Flanders (FWO).

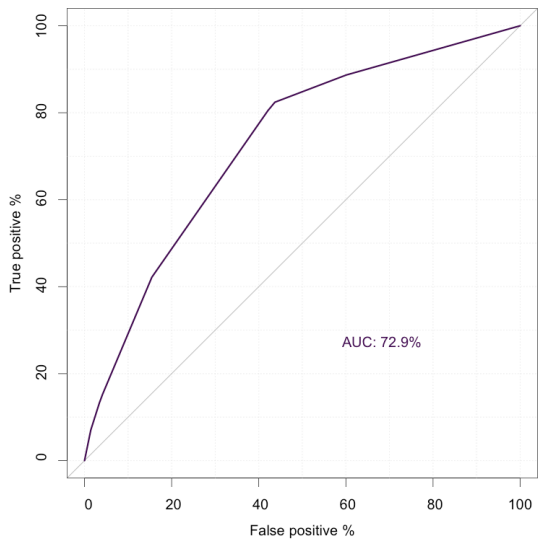
4.8 Conflict of Interest

The authors declare that they have no conflict of interest related to this manuscript.

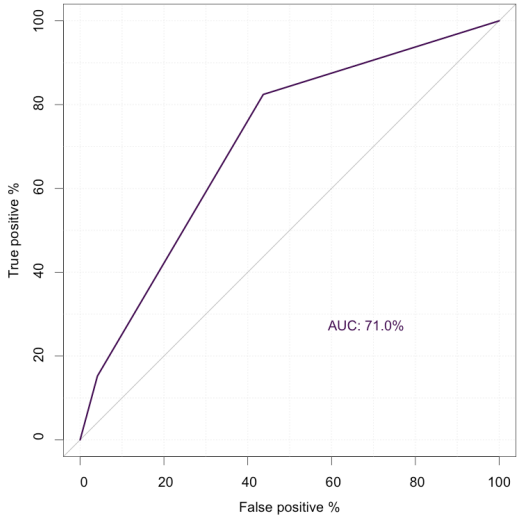
4.9 Supplementary Figures



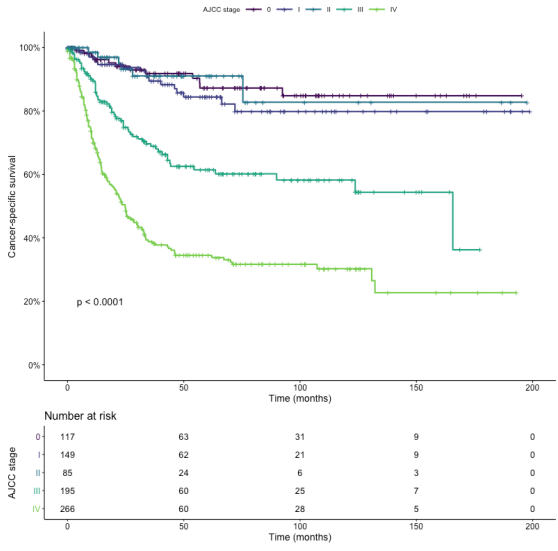
Supplementary Figure 1: Estimated cancer-specific survival (CSS) stratified by tumor stage (<pT2 vs. pT2 vs. >pT2) and nodal positivity (pN0 vs. pN+). Patients with pN+ disease had significant worse CSS.



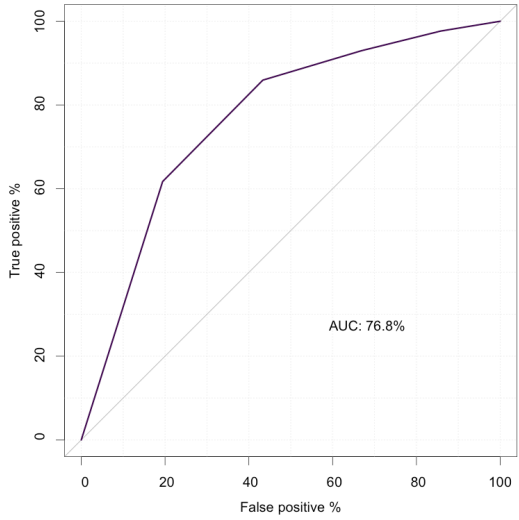
Supplementary Figure 2: ROC curve with AUC for COBRA score.



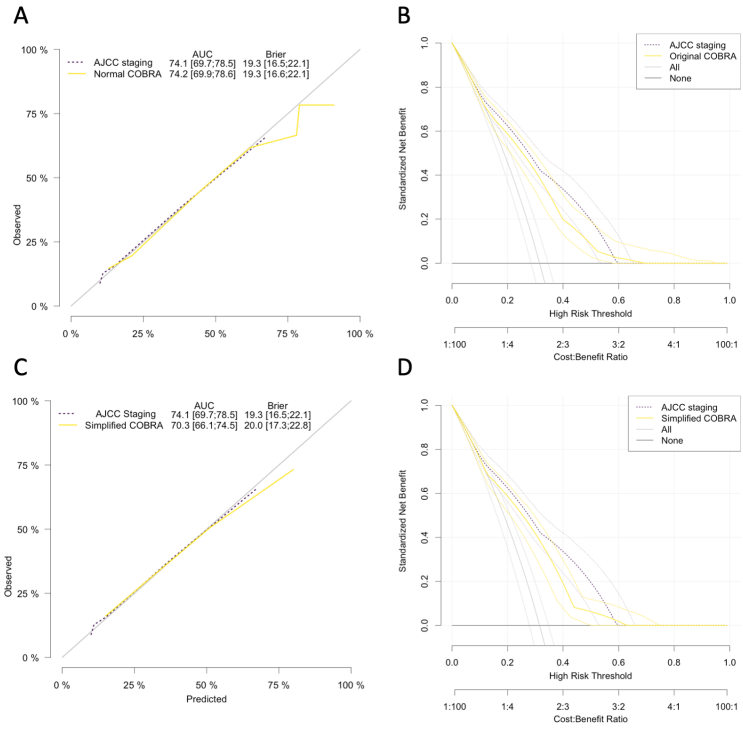
Supplementary Figure 3: ROC curve with AUC for COBRA risk categories.



Supplementary Figure 4: Kaplan-Meier curves for cancer-specific survival (CSS) stratified by the AJCC staging system.



Supplementary Figure 5: ROC curve with AUC for AJCC staging system.



Supplementary Figure 6: A: Calibration plot at 5-years for COBRA score and AJCC staging system. B: Decision curve analysis for estimating cancer-specific survival (CSS) which compares the AJCC staging system (dotted, purple) and the original COBRA score (straight, yellow). The standardized net benefit (NB) is visualized on the y-axis. High-risk threshold probabilities are visualized on the x-axis which can be directly translated to the harm-benefit ratio (Cost:Benefit Ratio). The grey line holds the assumption that all patients will die. The horizontal black line (with Standardized Net Benefit = 0.0) holds the assumption that no patients will die. C: Calibration plot at 5-years for COBRA risk categories and AJCC staging system. D: Decision curve analysis for estimating cancer-specific survival (CSS) which compares the AJCC staging system (dotted, purple) and COBRA risk categories (straight, yellow). The standardized net benefit (NB) is visualized on the y-axis. High-risk threshold probabilities are visualized on the x-axis which can be directly translated to the harm-benefit ratio (Cost:Benefit Ratio). The grey line holds the assumption that all patients will die. The horizontal black line (with Standardized Net Benefit = 0.0) holds the assumption that no patients will die.

4.10 Supplementary Tables

Supplementary Table 1: Pairwise comparison of tumor stage.

Tumor stage	pT0	pT1	pT2	pT3
pT1	0.98	-	-	-
pT2	1.00	1.00	-	-
pT3	<0.001	0.080	0.048	-
pT3	<0.001	<0.001	<0.001	<0.001

Abbreviations: pT: pathological tumor stage.

Supplementary Table 2: Cox proportional hazards model of the separate COBRA score variables. All variables that are incorporated into the COBRA score have a significant effect on cancer-specific survival.

	No.	HR	Lower 95%CI	Upper 95%CI	p-value
LN density	812	4.8	2.7	8.5	<0.001
Age at RC	812	1.02	1.00	1.03	0.02
pT0	110	Ref	-	-	-
pT1	117	1.98	0.98	3.98	0.06
pT2	69	1.24	0.51	3.04	0.64
pT3	115	2.66	1.37	5.16	0.004
pT4	228	6.79	3.73	12.35	<0.001

Abbreviations: LN: lymph node; RC: radical cystectomy; pT: pathological tumor stage; HR: hazard ratio; CI: confidence interval.

Supplementary Table 3: Cox proportional hazards model of COBRA score. The COBRA score can stratify correctly with significant differences compared to the reference level, although there is not a subsequent increase in HR between each COBRA score level.

COBRA score	No.	HR	Lower 95%CI	Upper 95%CI	p-value
Continuous	812	1.53	1.42	1.64	<0.001
0	251	Ref	-	-	-
1	107	1.60	0.86	2.95	0.13
2	14	4.18	1.61	10.81	0.003
3	246	4.17	2.76	6.32	<0.001
4	132	6.76	4.37	10.44	<0.001
5	26	10.84	6.00	19.56	<0.001
6	27	10.60	5.75	19.55	<0.001
7	9	16.98	6.54	44.12	<0.001

Abbreviations: HR: hazard ratio; CI: confidence interval.

Supplementary Table 4: Pairwise comparison of the COBRA score. Log-rank test corrected by Bonferroni stratifies significantly between lower COBRA scores (0 and 1) and the higher COBRA scores. COBRA score 2 shows no significant difference compared to the other COBRA scores. COBRA score 3 shows a significant difference with COBRA score 5, 6, and 7, but not with COBRA score 2 and 4. There are no significant differences in survival between COBRA scores 5, 6, and 7.

COBRA score	0	1	2	3	4	5	6
1	1	-	-	-	-	-	-
2	0.005	1	-	-	-	-	-
3	<0.001	0.006	1	-	-	-	-
4	<0.001	<0.001	1	0.07	-	-	-
5	<0.001	<0.001	1	0.004	1	-	-
6	<0.001	<0.001	1	0.01	1	1	-
7	<0.001	<0.001	1	0.04	0.98	1	1

Supplementary Table 5: Pairwise comparison between the three COBRA score risk categories [0-1] vs. [2-4] vs. [5-7]. Log-rank test corrected by Bonferroni stratifies significantly between all three risk categories.

COBRA score	0-1	2-4
2-4	<0.001	-
5-7	<0.001	<0.001

Supplementary Table 6: Published nomograms stratifying patients after RC for CSS or RFS.

Supplementary Table 6: Published nomograms stratifying patients after RC for CSS or RFS.

	Williams Urology 2017	Welty* Cancer 2017	Todenhöfer BJUI 2012	Gakis BJUI 2011	Shariat Clin C. Res 2006	el-Mekresh J Urol 2009	IBCN J Clin Onc. 2006
C-index	0.66	0.71	0.75	0.79	0.79	0.74	0.75
Outcome	CSS	CSS	CSS	CSS	CSS	RFS	RFS
Original cohort	SEER [n=5.325]	SEER [n=14.828]	Single center [n=258]	Single center [n=246]	Multi-center (n=3) [n=731]	Single center [n=1.133]	Multi-center (n=12) [n=9.064]
Validation cohort	TCR [n=1257]	NCDB [n=19.362]	NR	NR	Bootstrapping	Single center [n=357]	Cross-validation by leaving out centers
Number of variables	10	3	3	3	5	4	6
Variables	RC (yes vs. no)	pT stage (≤pT1 vs. pT2 vs. pT3-4)	pT stage (<pT3a vs. ≥pT3a)	pT stage (<pT3a vs. ≥pT3a)	pT stage (≤pT1 vs. pT2 vs. pT3 vs. pT4)	pT stage (≤pT1 vs. pT2 vs. pT3a vs. pT3b vs. pT4)	pT stage (pT0 vs. pTis vs. pTa vs. pT1 vs. pT2 vs. pT3 vs. pT4)
	LND (yes vs. no)	LN density (0 vs. >0-0.33 vs. ≥0.33-0.5 vs. ≥0.5-1)	Positive margins (yes vs. no)	Positive margins (yes vs. no)	pN stage (pN0 vs. pN1 vs. pN2 vs. pN3)	pN stage (pN0 vs. pN1 vs. pN2 vs. pN3)	pN stage (pNx vs. pN0 vs. pN+)
	NAC (yes vs. no)	Age, y (<80 vs. ≥80)	PLT count (≤450 vs. >450 × 10^9/L)	CRP (≤0.5 vs. >0.5 mg/dL)	LVI (yes vs. no)	LVI (yes vs. no)	Grade (low vs. high. vs. unknown)
	Age at RC, y (65-69 vs. 70-74 vs. 75-79 vs. ≥80)				NAC (yes vs. no)	Type of UD (orthotopic vs. continent cutaneous vs. rectal diversion vs. ileal loop)	Histology (AC vs. SCC vs. TCC)
	Marital status (single vs. married vs. other)				Adjuvant radiotherapy (yes vs. no)		Age at RC, y (continuous scale)
	Sex (male vs. female)						Sex (male vs. female)
	Stage (II vs. III vs. IV)						Time from diagnosis to RC, d (continuous scale)
	Hydronephrosis (yes vs. no)						
	Grade (low vs. high vs. unknown)						
	CCI (0 vs. 1 vs. 2 vs. ≥3)						

*Original COBRA publication (9). Abbreviations: [n]: number of patients; (n): number of centers; RC: radical cystectomy; CSS: cancer-specific survival; RFS: recurrence-free survival; C-index: concordance-index; LND: lymph node dissection; NAC: neoadjuvant chemotherapy; CCI: Charlson comorbidity index; pT stage: pathological tumor stage; pN stage: pathological nodal stage; CRP: C-reactive protein; PLT: platelet; LVI: lymphovascular invasion; RC: radical cystectomy; UD: urinary diversion; AC: adenocarcinoma; SCC: squamous cell carcinoma; TCC: transitional cell carcinoma

4.11 References

1. Alfred Witjes J, Lebret T, Compérat EM, Cowan NC, De Santis M, Bruins HM, et al. Updated 2016 EAU Guidelines on Muscle-invasive and Metastatic Bladder Cancer. European Urology. 2017 Mar;71(3):462–75.
2. Chang SS, Bochner BH, Chou R, Dreicer R, Kamat AM, Lerner SP, et al. Treatment of Non-Metastatic Muscle-Invasive Bladder Cancer: AUA/ASCO/ASTRO/SUO Guideline. The Journal of Urology. 2017 Sep 1;198(3):552–9.
3. Stein JP, Lieskovsky G, Cote R, Groshen S, Feng A-C, Boyd S, et al. Radical Cystectomy in the Treatment of Invasive Bladder Cancer: Long-Term Results in 1,054 Patients. Journal of Clinical Oncology. 2001 Feb;19(3):666–75.
4. Rink M, Lee DJ, Kent M, Xylinas E, Fritsche H-M, Babjuk M, et al. Predictors of cancer-specific mortality after disease recurrence following radical cystectomy. BJU International. 2013 Mar 1;111(3b):E30–6.
5. Hautmann RE, de Petriconi RC, Pfeiffer C, Volkmer BG. Radical Cystectomy for Urothelial Carcinoma of the Bladder Without Neoadjuvant or Adjuvant Therapy: Long-Term Results in 1100 Patients. European Urology. 2012 May;61(5):1039–47.
6. El-Mekresh M, Akl A, Mosbah A, Abdel-Latif M, Abol-Enein H, Ghoneim MA. Prediction of Survival After Radical Cystectomy for Invasive Bladder Carcinoma: Risk Group Stratification, Nomograms or Artificial Neural Networks? Journal of Urology. 2009 Aug;182(2):466–72.
7. Shariat SF, Karakiewicz PI, Palapattu GS, Amiel GE, Lotan Y, Rogers CG, et al. Nomograms Provide Improved Accuracy for Predicting Survival after Radical Cystectomy. Clinical Cancer Research. 2006 Nov 15;12(22):6663–76.
8. Bruins HM, Dorin RP, Rubino B, Miranda G, Cai J, Daneshmand S, et al. Critical Evaluation of the American Joint Committee on Cancer TNM Nodal Staging System in Patients with Lymph Node–Positive Disease after Radical Cystectomy. European Urology. 2012 Oct;62(4):671–6.
9. Welty CJ, Sanford TH, Wright JL, Carroll PR, Cooperberg MR, Meng M V, et al. The Cancer of the Bladder Risk Assessment (COBRA) score: Estimating mortality after radical cystectomy. Cancer. 2017;123(23):4574–82.

10. Williams SB, Huo J, Chu Y, Baillargeon JG, Daskivich T, Kuo Y-F, et al. Cancer and All-cause Mortality in Bladder Cancer Patients Undergoing Radical Cystectomy: Development and Validation of a Nomogram for Treatment Decision-making. Urology. 2017 Dec 1;110:76–83.
11. Kim HS, Kwak C, Kim HH, Ku JH. The Cancer of the Bladder Risk Assessment (COBRA) score for predicting cancer-specific survival after radical cystectomy for urothelial carcinoma of the bladder: External validation in a cohort of Korean patients. Urologic Oncology: Seminars and Original Investigations. 2019 Mar 29;
12. Bruins HM, Veskimae E, Hernandez V, Imamura M, Neuberger MM, Dahm P, et al. The impact of the extent of lymphadenectomy on oncologic outcomes in patients undergoing radical cystectomy for bladder cancer: A systematic review. Vol. 66, European Urology. 2014. p. 1065–77.
13. Vickers AJ, Elkin EB. Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. Medical Decision Making. 2006 Nov 5;26(6):565–74.
14. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, et al. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. CA: A Cancer Journal for Clinicians. 2017 Mar 1;67(2):93–9.
15. Lee D, Yoo S, You D, Hong B, Cho YM, Hong JH, et al. Lymph node density vs. the American Joint Committee on Cancer TNM nodal staging system in node-positive bladder cancer in patients undergoing extended or super-extended pelvic lymphadenectomy. Urologic Oncology: Seminars and Original Investigations. 2017 Apr 1;35(4):151.e1-151.e7.
16. Ku JH, Kang M, Kim HS, Jeong CW, Kwak C, Kim HH. Lymph node density as a prognostic variable in node-positive bladder cancer: a meta-analysis. BMC Cancer. 2015 Dec 2;15(1):447.
17. Wright JL, Lin DW, Porter MP. The association between extent of lymphadenectomy and survival among patients with lymph node metastases undergoing radical cystectomy. Cancer. 2008 Jun 1;112(11):2401–8.

18. Farnell MB, Pearson RK, Sarr MG, DiMagno EP, Burgart LJ, Dahl TR, et al. A prospective randomized trial comparing standard pancreatoduodenectomy with pancreatoduodenectomy with extended lymphadenectomy in resectable pancreatic head adenocarcinoma. Surgery. 2005 Oct;138(4):618–30.
19. Lagergren J, Mattsson F, Zylstra J, Chang F, Gossage J, Mason R, et al. Extent of lymphadenectomy and prognosis after esophageal cancer surgery. JAMA Surgery. 2016 Jan 1;151(1):32–9.
20. Omloo JMT, Lagarde SM, Hulscher JBF, Reitsma JB, Fockens P, Van Dekken H, et al. Extended transthoracic resection compared with limited transhiatal resection for adenocarcinoma of the mid/distal esophagus: Five-year survival of a randomized clinical trial. Annals of Surgery. 2007 Dec;246(6):992–1000.
21. Sasako M, Sano T, Yamamoto S, Kurokawa Y, Nashimoto A, Kurita A, et al. D2 Lymphadenectomy Alone or with Para-aortic Nodal Dissection for Gastric Cancer. New England Journal of Medicine. 2008 Jul 31;359(5):453–62.
22. Gschwend JE, Heck MM, Lehmann J, Rübben H, Albers P, Wolff JM, et al. Extended Versus Limited Lymph Node Dissection in Bladder Cancer Patients Undergoing Radical Cystectomy: Survival Results from a Prospective, Randomized Trial. European Urology. 2018 Oct 15;0(0).
23. Muilwijk T, Akand M, Gevaert T, Joniau S. No survival difference between super extended and standard lymph node dissection at radical cystectomy: what can we learn from the first prospective randomized phase III trial? Translational Andrology and Urology. 2019 Mar;8(S1):S112–5.
24. International Bladder Cancer Nomogram Consortium, Bochner BH, Kattan MW, Vora KC. Postoperative Nomogram Predicting Risk of Recurrence After Radical Cystectomy for Bladder Cancer. Journal of Clinical Oncology. 2006 Aug 20;24(24):3967–72.
25. Gakis G, Todenhöfer T, Renninger M, Schilling D, Sievert KD, Schwentner C, et al. Development of a new outcome prediction model in carcinoma invading the bladder based on preoperative serum C-reactive protein and standard pathological risk factors: The TNR-C score. BJU International. 2011 Dec;108(11):1800–5.
26. Todenhöfer T, Renninger M, Schwentner C, Stenzl A, Gakis G. A new prognostic model for cancer-specific survival after radical cystectomy including pretreatment thrombocytosis and standard pathological risk factors. BJU international. 2012 Dec;110(11 Pt B):E533-40.

27. Williams SB, Huo J, Chu Y, Baillargeon JG, Daskivich T, Kuo Y-F, et al. Cancer and All-cause Mortality in Bladder Cancer Patients Undergoing Radical Cystectomy: Development and Validation of a Nomogram for Treatment Decision-making. Urology. 2017 Dec;110:76–83.

28. Babjuk M, Burger M, Compérat EM, Gontero P, Mostafid AH, Palou J, et al. European Association of Urology Guidelines on Non-muscle-invasive Bladder Cancer (TaT1 and Carcinoma In Situ) - 2019 Update. European urology. 2019 Aug 20;

Chapter 5:

CK5 and CK20 inversely correlate with tumor grading in Ta non-muscle invasive bladder cancer



Chapter 5:

CK5 and CK20 inversely correlate with tumor grading in Ta non-muscle invasive bladder cancer

Tim Muilwijk[1,2*], Murat Akand[1,2*], Frank Van der Aa[1], Vincent De Coninck[3], Marc Claessens[3], Robert Hente[3], Markus Eckstein[4], Yves Allory[5], Louis Libbrecht[6,7], Steven Joniau[1#], Thomas Gevaert[2,8#]

[1] University Hospitals Leuven, Department of Urology, Leuven, Belgium

[2] KU Leuven, Organ Systems, Leuven, Belgium

[3] AZ Klina, Department of Urology, Brasschaat, Belgium

[4] University Hospital Erlangen, Department of Pathology, Erlangen, Germany

[5] Curie Institute, Department of Pathology, Paris, France

[6] AZ Groeninge, Kortrijk, Department of Pathology, Belgium

[7] University Hospital Saint-Luc, Department of Pathology, Brussels, Belgium

[8] AZ Klina, Department of Pathology, Brasschaat, Belgium

*Shared first authorship

#Shared senior authorship

Journal of Cellular and Molecular Medicine, June 29th, 2021 (IF 5.31)

5.1 Abstract

CK5 is a marker of basal molecular subtypes of muscle-invasive bladder cancer (MIBC), which correlate with worse overall survival compared to luminal subtypes. Our observations have not confirmed CK5 as a marker of high grade (HG) disease in Ta non-muscle invasive bladder cancer (NMIBC). Therefore, to understand the basal-luminal immunohistochemistry profile in Ta NMIBC, we performed immunohistochemistry for CK5, P40, P63 (basal), GATA3, and CK20 (luminal) and studied the correlation with HG and clinical outcome in 109 patients with Ta NMIBC. HG and low grade (LG) disease were scored in each patient. Four different CK5 patterns were evaluated: absent (41.3%), normal (72.5%), rising (84.4%), and full-thickness (23.9%). The median percentage of GATA3 was 100%. HG disease and CK5 expression and rising CK5 pattern had a significant inverse correlation, whereas HG disease and CK20 expression had a significant positive correlation. We also found a significant inverse correlation between CK5 expression and CK20 expression. Quantitative PCR confirmed that the presence of CK5 correlated with upregulation of CK5 RNA. None of the markers could differentiate patients with regard to clinical outcome. Our results suggest a role for CK5 and CK20 in differentiating between LG and HG disease in Ta NMIBC.

5.2 Introduction

Bladder cancer (BC) is the seventh most common cancer worldwide in men, with a yearly incidence of approximately 549,000 cases (1). Three out of four patients present with non-muscle-invasive BC (NMIBC), which can be treated non-invasively by local resection, additional intravesical instillations, and stringent follow-up (2). However, NMIBC progresses to muscle-invasive BC (MIBC) in 31%-78% patients depending on clinical and pathological risk factors (3). Furthermore, 28% of patients with NMIBC, when treated with intravesical BCG instillations, become BCG-refractory, which significantly impacts their prognosis (4). Treatment of BCG-unresponsive NMIBC and MIBC is aggressive with either surgical resection of the bladder (cystectomy) or trimodal therapy (maximal transurethral resection, chemotherapy, and radiotherapy).

Transcriptional analysis of MIBC has revealed several molecular subtypes, with luminal and basal as the most distinct (5,6). Interestingly, patients with tumors of the basal subtype have worse overall survival (OS) compared to patients

with the luminal subtype. In MIBC, CK5 has been suggested as a surrogate marker for the basal subtype, whereas CK20 and GATA3 have been linked to the luminal subtype (5,7,8). Based on these findings, immunohistochemistry (IHC) using CK5 and GATA3 has been used to stratify patients in MIBC between the luminal and basal molecular subtypes (9). Basal MIBC is recognized by its CK5+ and GATA3- profile, whereas luminal MIBC has a CK5- and GATA3+ profile (9). Other commonly used basal markers are P63 and its isoform P40 (ΔNp63). P63 and P40 may act as either an oncogene or a tumor suppressor gene (10,11). In MIBC, P40 correlates with high grade (HG) disease and higher European Organisation for Research and Treatment of Cancer (EORTC) risk score (12). Furthermore, patients with loss of P40 have a higher risk of NMIBC relapse and progression (12).

In contrast to the findings in MIBC, transcriptional analysis of NMIBC Ta and T1 disease identified a basal subtype with improved progression-free survival (PFS) compared to the luminal subtype (13). Interestingly, this basal NMIBC subtype has increased RNA expression of CK5, whereas luminal NMIBC subtypes have higher RNA expression of CK20, a marker of apical umbrella cells (13). The luminal marker GATA3 was expressed at even higher levels in the basal subtype in NMIBC than the other subtypes (13). Interestingly, subclassification of T1 NMIBC using the RNA expression of CK5 and CK20 can be used for the successful stratification of patients, demonstrating worse recurrence-free survival (RFS) and PFS in patients with high CK20 and low CK5 expression (14).

Similarly, our routine pathological observations did not confirm the potential of CK5-IHC as a marker of HG Ta NMIBC. Furthermore, we observed significant inter-tumor and intra-tumor heterogeneity of CK5 in Ta NMIBC. Given the contradictory findings reported in the RNA and IHC studies between NMIBC and MIBC, we studied the IHC of the basal and luminal markers in a large group of patients with Ta NMIBC for the first time (15). To fully capture the heterogeneity of CK5 expression at the IHC level, we applied a novel approach using different CK5 expression patterns and correlated them with the percentages of HG and low grade (LG) disease within one patient. Lastly, we aimed to identify correlations between these CK5 IHC patterns and CK5 RNA expression, tumor grading (%), the other basal/luminal IHC markers, and RFS.

5.3 Materials and methods

5.3.1 Patients

Patients with Ta NMIBC diagnosed after transurethral resection of the bladder tumor (TURBT) between May 2016 and February 2019 were included in the study cohort. Ethical approval was obtained from the local ethics committee (Ethics Committee AZ Klina no. OG 146; approval no. 086/200/019). Patients with invasive disease (≥T1), clinical node-positive disease, or metastatic disease at the time of TURBT were excluded from the analysis. Histopathology was reassessed in all cases by an expert uro-pathologist (T.G.) blinded to clinical outcome. The following parameters were recorded: tumor stage, tumor grade, presence of carcinoma in situ (CIS), and the presence or absence of detrusor smooth muscle. As heterogeneity of tumor grading within one patient is possible, we scored the tumor grade traditionally as either LG or HG according to the WHO 2004 classification, but also scored the percentage of LG and HG disease within each patient. This scoring is a novel method of assessing HG disease burden that can be used to assess the heterogeneity of tumor grading within a patient. To promote reproducibility, grades were scored in increments of 10%.

A retrospective chart review was performed for the following clinico-pathological data: age, gender, number of lesions, lesion size, primary vs. recurrent NMIBC, and prior intravesical treatment. The following survival outcomes were collected: RFS, PFS, cancer-specific survival (CSS), and OS. All survival endpoints were defined as the time from date of TURBT; for RFS the endpoint was the date of first recurrence of disease at cystoscopy, for PFS the date of first diagnosis of ≥T2 disease, for CSS the date of death due to BC, and for OS the date of death due to any cause.

5.3.2 Immunohistochemistry

Depending on the extent of the TURBT, 1 or >1 formalin-fixed paraffin-embedded (FFPE) blocks were available per patient. In the case of the availability of >1 FFPE block, we selected a representative FFPE block after the presence of a relevant tumor was assessed on all hematoxylin and eosin-stained slides. Sections (4 μm) were cut and mounted on poly-L-lysine-coated glass slides. IHC was performed using the automated Benchmark Ultra IHC system (Roche Diagnostics, Belgium). The automated procedure consisted of

blocking endogenous peroxidase activity using 0.3% H2O2 in methanol, with heat-induced epitope retrieval in Tris-EDTA buffer (pH 7.8) at 95°C for 44 min (standard CC1). Slides were incubated with primary antibodies for 15 min, a peroxidase-labelled polymer for 30 min, and then a substrate-chromogen (mixed DAB Refine) for 10 min. Nuclear counterstaining was performed with hematoxylin.

Antibody clones were selected for their epitope-selectivity (see **Table 1** for details); all have been extensively validated for clinical diagnostic practice (www.nordiqc.com). Prior to enrollment in the study, antibodies were validated on control tissue (tissue with known endogenous expression) for staining specificity and reliability. The primary antibodies used were anti-CK5, anti-GATA3, anti-P40, anti-P63, and anti-CK20 **(Table 1**).

Immunogen	Clone	Manufacturer/ Code	Host	Titer	Control
Cytokeratin 5 (CK5) (Immunizing human cytokeratin 5 protein)	D5/16B4	Roche Diagnostics, Belgium 790-4554	Mouse	Ready to use	Tonsil, Prostate
Cytokeratin 20 (CK20) (C-terminus of the human cytokeratin 20 protein)	SP33	Roche Diagnostics, Belgium 790-4431	Mouse	Ready to use	Colon
GATA3 (Specific immunogen not provided)	L50/823	Roche Diagnostics, Belgium 760-4897	Mouse	Ready to use	Breast, Urothelium
P40 (amino acid sequence from amino acid 5 to 17) (PVLEPGDKPRKAT)	BC28	Roche Diagnostics, Belgium 790-4950	Mouse	Ready to use	Prostate (basal cells)
P63 (N-terminus of the human P63 protein)	4A4	Roche Diagnostics, Belgium 790-4509	Mouse	Ready to use	Prostate (basal cells)

Table 1: Properties of the antibody clones used

CK5 is a marker of undifferentiated basal cells in the urothelium and has been correlated with the basal molecular subtype in RNA expression studies in both MIBC and NMIBC. Based on the clinical observation by our uro-pathologists (T.G. and L.L.) that CK5 expression can be heterogeneous within each patient, we developed a scoring system for CK5 expression using different patterns to fully capture the heterogeneity of CK5 IHC within each patient.

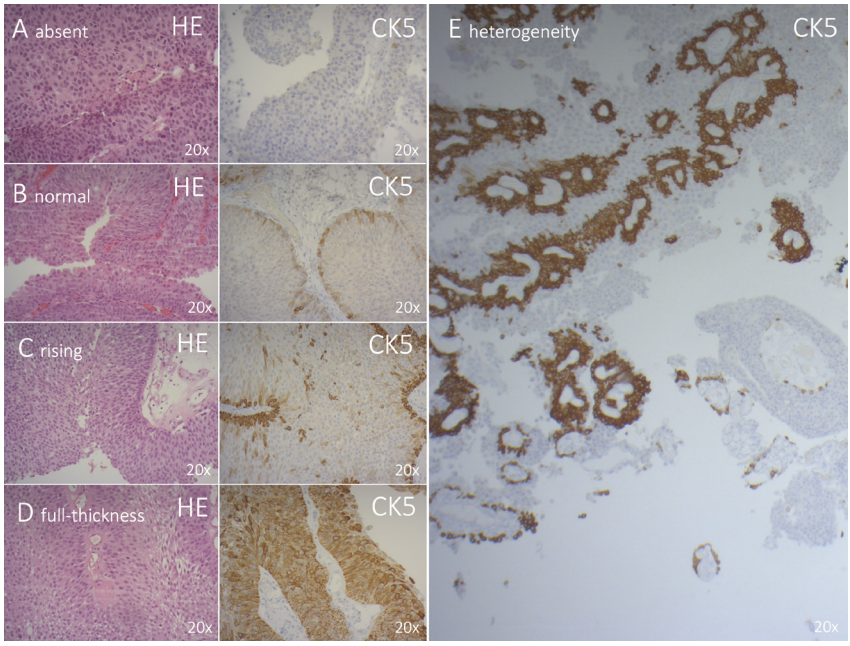


Figure 1: Patterns of CK5 IHC expression: A: absent, B: normal, C: rising, and D: full-thickness. For each pattern: on the left side hematoxylin and eosin stain, and on the right side CK5 stain; on 20x magnification. E: Illustration of heterogeneity of CK5 IHC expression patterns present.

The CK5 expression pattern in IHC could be either 'absent' (i.e., no CK5 expression present), or 'normal', 'rising', or 'full-thickness' when CK5 expression was present. These four CK5 expression patterns are illustrated in **Figure 1**. The normal pattern is CK5 in the basal layer, as is expected from a basal marker, whereas the rising pattern comprises CK5 in the basal layer with increasing CK5 expression towards the luminal side. Comparable expression patterns have been described in both NMIBC and MIBC, which strengthens our methodological approach.16 To capture heterogeneity, we scored the presence of each pattern as the percentage of each pattern per patient using a visual estimation (**Figure 2**). In addition, we scored the presence of the other basal

(P40, P63) and luminal (GATA3, CK20) markers per CK5 pattern present in each patient (**Supplementary Figure 1**). Finally, an additional assessment was made of the apical expression pattern of CK20; in normal urothelium, CK20 is present in umbrella cells only. Based on observational findings, we discriminated three patterns (apical only, apical negative, apical attenuated), and the percentages of these patterns were scored per CK5 pattern (**Supplementary Figure 2**).

Our workflow of assessing heterogeneity in tumor grading, CK5 IHC patterns, and other IHC markers is illustrated in **Figure 2**. IHC staining was evaluated and scored by a specialized uro-pathologist (T.G.).

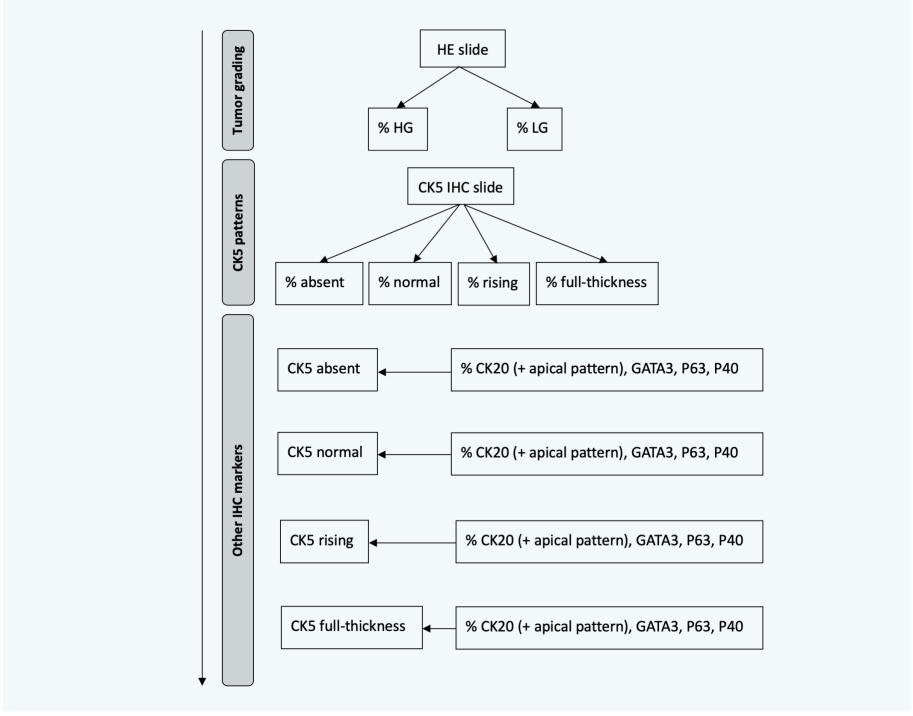


Figure 2: Work flow to assess intra-tumor heterogeneity by scoring tumor grading, CK5 IHC expression patterns, and other IHC markers. Firstly, tumor grading was assessed on an HE slide and scored by % of presence of HG and % of presence of LG disease. Secondly, CK5 patterns were assessed on a CK5 IHC slide by scoring each % of presence of each pattern per patient. Lastly, for each CK5 pattern we assessed IHC expression of the other luminal and basal markers by scoring % of presence of each IHC marker per CK5 expression pattern.

5.3.3 Real-time quantitative polymerase chain reaction (qPCR)

To correlate the results of CK5 IHC with CK5 RNA expression, we used the sample function in R to randomly select 10 samples with the presence of any CK5 expression based on IHC (combination of 'normal', 'rising', or 'full-thickness') and 10 samples that had >50% absence of CK5 expression. As all patient samples had some CK5 expression based on IHC, it was impossible to select patients without CK5 expression. Total RNA was extracted from 80-µm tissue slides using the RecoverAll Total Nucleic Acid Isolation Kit (Life Technologies Corporation, Carlsbad CA, USA) according to the manufacturer's instructions. The RNA integrity number (RIN; 28S/18S) and RNA concentration were assessed using the Experion RNA StdSens (Bio-Rad, Hercules CA, USA) according to the manufacturer's instructions. Only samples with a concentration >50 ng/µL were selected for further downstream experiments.

First-strand cDNA was synthesized from 500 ng of RNA by reverse transcription using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Thermo-Fisher) according to the manufacturer's protocol. Real-time qPCR was performed using the TaqMan® PCR kit (Thermo-Fisher). TaqMan assay primers and probes (Assays-on-Demand Gene Expression Products) were obtained (Applied Biosystems, Thermo-Fisher). The reaction mixture (25 µL) containing 1 µL of cDNA template, 1 µL of each primer and probe mix, and TaqMan Universal PCR master mix (Applied Biosystems, Thermo-Fisher) was amplified as follows: denaturation at 95°C for 10 min and 40 cycles of 95°C for 10 s and 60°C for 20 s. Direct detection of PCR products was monitored by measuring the fluorescence produced by the result of TaqMan probe hydrolysis after every cycle. Data were analyzed using the 2–ΔΔCt method for relative quantification compared to GAPDH and S18 as internal controls.

5.3.4 Statistical analysis

Analyses were performed in R (version 3.6.0) using the 'tidyverse,' 'lubridate,' 'survival,' 'survminer,' 'pheatmap,' 'Hmisc,' and 'corrplot' packages. Summary statistics are presented for continuous variables as median with interquartile range (IQR) and for categorical variables as frequencies and proportions. Spearman correlation coefficients were used to assess the relationship between continuous variables of interest: percentage of CK5 expression patterns, percentage of LG vs. HG, and expression of IHC markers. Spearman correlation was used due to the variables of interest being skewed and having

outliers. The normality of variables was assessed using the Lilliefors normality test. A heatmap was created using complete-linkage clustering, which is a method of agglomerative hierarchical clustering. Kaplan-Meier plots were used to visualize the survival data. Log-rank tests and Cox proportional hazard models were used to assess correlations between RFS and categorical or numerical variables, respectively.

5.4 Results

5.4.1 Clinical and pathological characteristics

A total of 109 Ta tumors met our inclusion criteria and were included in the analysis. Relevant clinical and pathological characteristics are summarized in **Table 2**. The median age at the time of TURBT was 76 years (IQR 67-82). Median follow-up was 19 months (IQR 15-24). HG disease was present in 82% (89/109) of patients, which means that 82% would be considered as having high-risk NMIBC following the EAU guidelines risk stratification.2 The median percentage of HG and LG disease was 50% (IQR 20-70) and 50% (IQR 30-80), respectively. CIS was present in 1 patient (1%). A total of 48 patients (44%) had recurrent disease, 6 (13%) of whom had prior intravesical BCG. Adjuvant intravesical chemotherapy instillations were administered in 42 patients (39%). Adjuvant intravesical BCG instillations were administered in 5 patients (5%), 3 (3%) of whom remained recurrence-free at follow-up and 2 (2%) became BCG-unresponsive and progressed to MIBC. A total of 35 patients (32%) had disease recurrence at follow-up, 4 (4%) of whom progressed to MIBC. These four patients had recurrent disease at inclusion, and three of them had BCG-refractory T1 HG disease at TURBT.

5.4.2 IHC and CK5 expression patterns

All Ta NMIBC samples presented CK5 expression by IHC, but with variable intra-tumoral heterogeneity. We correlated the percentage of any CK5 IHC, regardless of the pattern, with the percentage of HG disease and identified a moderate inverse correlation (ρ) of -0.31 (p=0.001). We also correlated each CK5 IHC pattern with the percentage of HG disease and identified an inverse correlation between HG disease and the percentage of rising pattern (ρ=-0.28; p=0.003).

Using hierarchical clustering for the percentages of HG disease, CK5 patterns, and CK20, we identified five distinct clusters (**Figure 3**): CK5 absent/CK20+, CK5 normal, CK5 normal/CK20+, CK5 normal/rising, and CK5 rising/full-thickness. Interestingly, patients with a high percentage of HG disease mostly had an absence of CK5 or normal CK5 expression, whereas patients with a low percentage of HG disease had mostly rising and full-thickness CK5 expression (**Figure 3**). The full-thickness CK5 expression pattern almost did not co-occur with the other CK5 expression patterns, as is visible in the first cluster in **Figure 3**. More detailed data regarding the Spearman correlation between the percentage of HG and CK5 patterns are given in **Supplementary Table 1**.

Next, we looked for correlations between the CK5 expression patterns and the other IHC markers (CK20, GATA3, P40, P63; **Table 3**). The luminal marker GATA3 and basal markers P40 and P63 were diffuse and uniformly expressed in Ta NMIBC (only umbrella cells were consistently P40- and P63-), whereas the expression of luminal marker CK20 was highly variable (median 50%; IQR 5-75; **Table 4**). No significant correlations were found between the expression of any of these markers by IHC. As GATA3, P40, and P63 presented uniform expression in Ta NMIBC, no difference in correlation was observed between their expression patterns and the different CK5 patterns. For CK20, we found a moderate inverse correlation (ρ=-0.28; p=0.01) with the percentage of any CK5 expression (normal, rising, and full-thickness combined) and the percentage of normal CK5 (ρ=0.25; p=0.02; **Figure 3**). Furthermore, we found a positive correlation between the percentage of CK20 and percentage of HG disease (ρ=0.21; p=0.04). The findings regarding CK5, CK20, and tumor grading are further illustrated in **Supplementary Figure 3**.

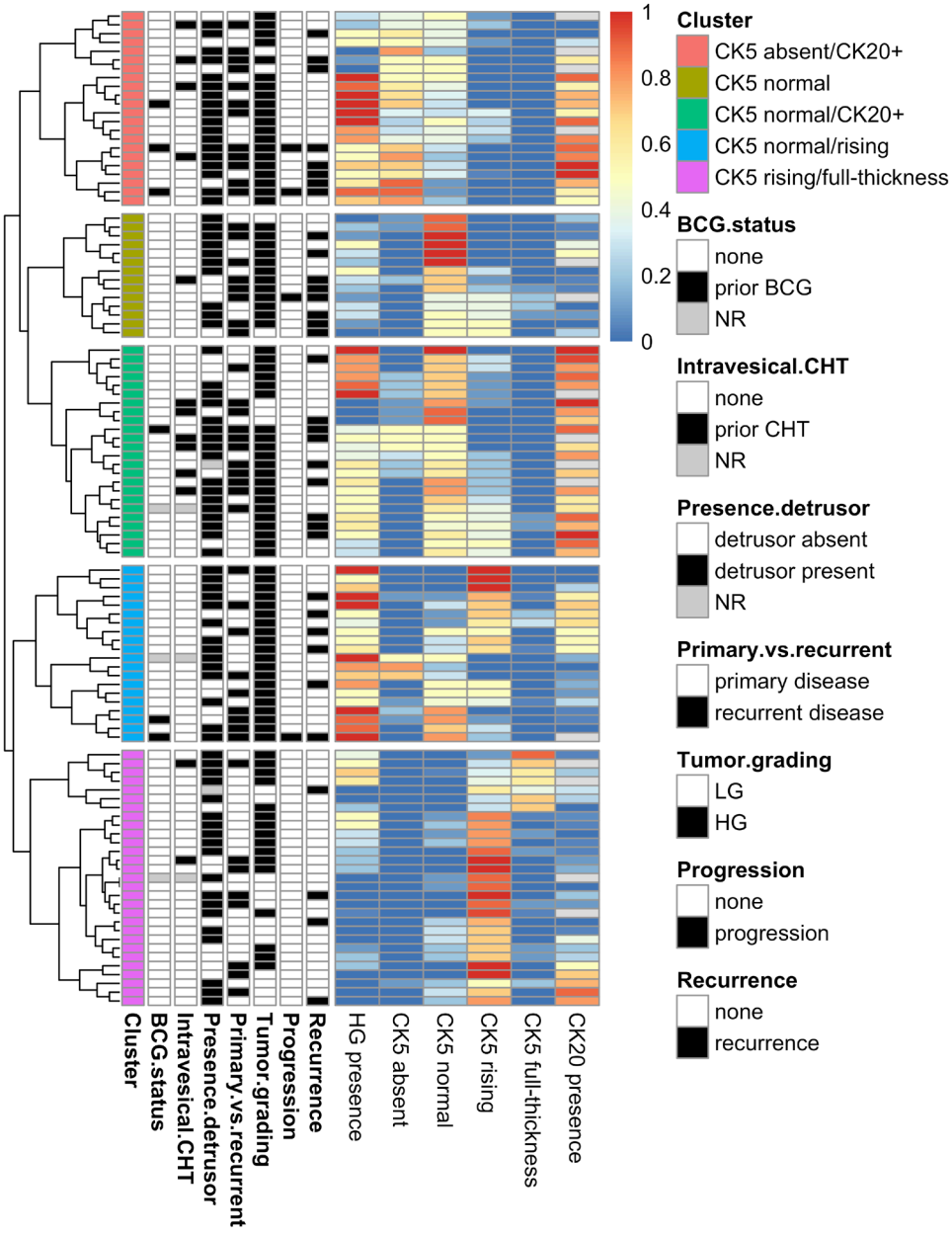


Figure 3: Heatmap using hierarchical clustering for % of presence of HG disease, % of presence of CK5 patterns, and % of presence of CK20, we identified 5 distinct clusters. Clusters from top to below are: CK5 absent/CK20+, CK5 normal, CK5 normal/CK20+, CK5 normal/rising , and CK5 rising/full-thickness.

Table 2: Clinical and pathological characteristics of all patients

	No.	*% / [IQR]*
Total patients	109	100
Unique patients	95	NA
Median follow-up (months)	19	[15-24]
Age at time of TURBT		
Median	76	[67-82]
≤65	18	17
>65	91	83
Gender		
Male	89	82
Female	20	18
Tumor stage		
Ta	109	100
Tumor grade (WHO 2004)		
HG presence in patients (n)	89	82
Median % of tumor HG (%)	50	[20-70]
LG presence in patients (n)	96	88
Median % of tumor LG (%)	50	[30-80]
Carcinoma in situ		
Present	1	1
Absent	108	99
EAU risk categories		
High	90	83
Intermediate	10	9
Low	8	7
NR	1	1
Detrusor in resection specimen		
Present	78	71
Absent	29	27
NR	2	2
Number of lesions		
≥8	24	22
2-7	22	20
1	61	56
NR	2	2
Size of lesions (cm)		
≥3	21	19
<3	86	79
NR	2	2
Single postoperative instillation of CHT		
Yes	75	69
No	32	29
NR	2	2

Chemotherapy maintenance		
Yes	42	38
Median number of courses	7	[6-9]
No	63	58
NR	4	4
BCG induction		
Yes	5	5
Median number of courses	6	[6-6]
Followed by BCG maintenance	3	3
Median number of courses	3	[3-6]
No	99	91
NR	2	2
Primary vs. recurrent disease		
Primary	59	54
Recurrent	48	44
Median time to recurrence (months)	14	[8-40]
≤1 recurrence/year	20	42
>1 recurrence/year	28	58
NR	2	2
Prior highest tumor stage *(recurrent disease-only; n=48)*		
Ta	36	75
T1	9	19
Tis	1	2
Tx	1	2
NR	1	2
Prior tumor grade *(recurrent disease-only; n=48)*		
G1	28	58
G2	3	6
G3	12	25
LG	2	4
NR	3	6
Prior intravesical instillations *(recurrent disease-only; n=48)*		
BCG	6	13
Chemotherapy	13	27
None	26	54
NR	3	6
Outcome		
Recurrence of NMIBC	35	32
Progression to MIBC	4	4
Death	4	4
Cancer-specific death	1	1

Abbreviations: IQR: interquartile range; TURBT: transurethral resection of bladder tumor; WHO: World Health Organization; HG: high grade; LG: low grade; NR: not reported; CHT: chemotherapy; BCG: Bacillus Calmette-Guerin.

We also assessed the differences in IHC and CK5 expression patterns between patients with primary (n=59) and recurrent disease (n=48; **Supplementary Table 2**). Any CK5 expression differed significantly (p=0.03), as well as CK5 full-thickness expression (p=0.02). We performed a similar analysis based on EAU risk categories for patients with low (n=8), intermediate (n=10), or high-risk disease (n=90; **Supplementary Table 3**). We found no significant differences between EAU risk categories in IHC and CK5 expression patterns. The CK5 full-thickness pattern was present in 8% of cases (n=4) with recurrent disease. The other IHC markers (GATA3, P40, P63, and CK20) and CK5 patterns did not differ significantly between patients with primary and recurrent disease.

We examined the correlation between the relative CK5 RNA expression levels and CK5 IHC patterns and found positive correlations between CK5 RNA expression and the presence of any CK5 in IHC (ρ=0.79; p=0.0008), and between CK5 RNA expression and the presence of the rising CK5 pattern (ρ=0.74; p=0.002; **Supplementary Figure 4**).

5.4.3 Survival analysis

RFS at 1 and 2 years (**Figure 4A**) was 80% (95% CI 72-88) and 57% (95% CI 46-71), respectively, with a median follow-up of 19 months (IQR 15-24). Follow-up data were missing for two patients. The median RFS was not reached (95% CI 23.1-NR). Only four patients progressed to MIBC disease, with PFS at 1 and 2 years of 98% (95% CI 96-100) and 95% (95% CI 90-100), respectively. CSS at 1 and 2 years was 99% (95% CI 97-100), and OS at 1 and 2 -years was 96% (95% CI 93-100).

Table 3: Median percentage of presence of expression of immunohistochemical markers (GATA3, P40, P63, and CK20) per CK5 pattern (absent, normal, rising, and full-thickness). Percentages of CK20 patterns (umbrella cells positive, umbrella cells negative, and accentuated) within the present CK20 as is illustrated in Supplementary Figure 2.

CK5 pattern	*median %*	*IQR*	*missing*	*completion %*
Absent n=45 (41%)				
GATA3	100	100-100	2	96
P40	90	80-90	24	47
P63	90	90-95	38	16
CK20	78	25-90	7	84
umbrella +ve	0	0-0		
umbrella -ve	80	50-100		
accentuated	5	0-45		
Normal n=92 (84%)				
GATA3	100	100-100	5	95
P40	90	88-95	52	44
P63	90	90-95	80	13
CK20	50	10-80	12	87
umbrella +ve	0	0-10		
umbrella -ve	80	50-100		
accentuated	3	0-40		
Rising n=79 (72%)				
GATA3	100	100-100	4	95
P40	90	90-95	44	44
P63	90	90-95	68	14
CK20	25	5-75	11	86
umbrella +ve	0	0-20		
umbrella -ve	70	43-100		
accentuated	0	0-43		
Full-thickness n=26 (24%)				
GATA3	100	100-100	3	89
P40	93	90-95	16	39
P63	90	90-93	22	15
CK20	15	5-29	4	85
umbrella +ve	10	0-30		
umbrella -ve	60	30-98		
accentuated	5	0-44		

Abbreviations: n: number; IQR: interquartile range; NA: not available; +ve: positive; -ve: negative.

Table 4: Presence and median percentage of presence of immunohistochemical markers, and numbers regarding completion of percentage.

IHC marker	*presence (%)*	*presence (n)*	*median %*	*IQR*	*missing*	*completion %*
CK5	100%	109	100	70-100	0	100
GATA3	100%	103	100	100-100	6	95
P40	100%	47	90	90-95	62	43
P63	88%	15	90	90-95	92	16
CK20	87%	82	50	5-75	15	86
umbrella +ve			0	0-18		
umbrella -ve			77	40-100		
accentuated			2	0-40		

Abbreviations: IHC: immunohistochemistry; IQR: interquartile range.

We found no significant effect of any of the CK5 expression patterns or CK5 RNA expression on RFS in univariable or multivariable Cox proportional hazard analyses (**Supplementary Table 4**). All clinical, pathological, and IHC expression data were checked to identify any prognostic variable for RFS, PFS, CSS, and OS in log-rank (for categorical variables) and Cox proportional hazard analyses (for numerical variables). The five distinct clusters of our heatmap (**Figure 3**) were not prognostic for RFS (**Supplementary Figure 5**), PFS, CSS, or OS. The only significant prognostic variable was the postoperative administration of a single instillation of intravesical chemotherapy, which increased the RFS of patients (hazard ratio: 0.43; 95% CI 0.22-0.86; p=0.01; C-index: 0.63; **Figure 4B**). We identified no additional prognostic variables when these analyses were performed separately for patients with primary (n=59) or recurrent disease (n=48).

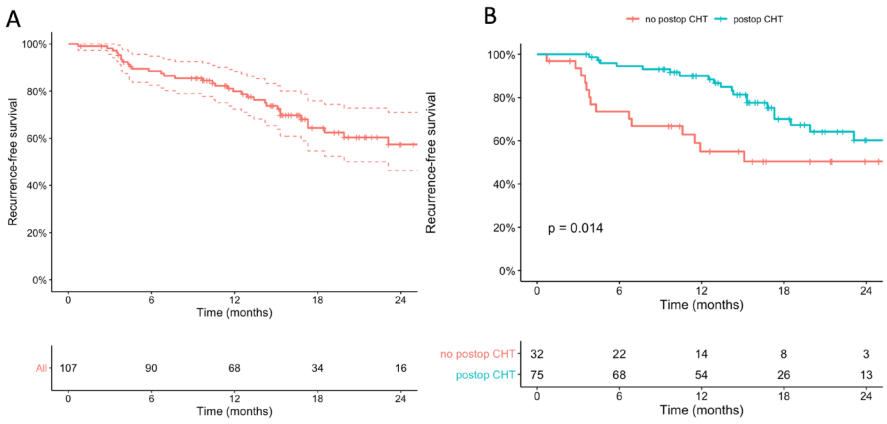


Figure 4: A: Kaplan-Meier plot of recurrence-free survival (RFS) of all patients with follow-up data (n=107). RFS at 1-year and 2-year was 80% (95%CI: 72-88), and 57% (95%CI: 46-71) respectively. Median RFS was not reached (95%CI: 23.1-NR). B: Kaplan-Meier plot of RFS of all patients with follow-up data stratified by administration of a single postoperative instillation of intravesical chemotherapy (n=107). RFS in the not administered group at 1-year and 2-year was 55% (95%CI: 39-77), and 50% (95%CI: 35-73); and in the administered group 90% (95%CI: 83-97), and 60% (95%CI: 47-77), respectively. Median RFS was not reached in both groups.

5.5 Discussion

Here, we describe the expression patterns of luminal (GATA3, CK20) and basal (CK5, P40, P63) IHC markers in a group of patients with Ta NMIBC. To fully assess the heterogeneity of Ta NMIBC, we used a granular methodology in which we scored the tumor grade as the percentage of HG and LG disease in each patient and addressed four specific patterns of CK5 expression in IHC. The study set-up is unique because we used a homogenous Ta NMIBC cohort, compared to the often-used mixed Ta/T1 NMIBC cohorts, and we report four different types of CK5 expression patterns for the first time.

Our findings are novel and interesting for several reasons. First, expression of the basal IHC marker CK5 was present in all patients, with remarkable heterogeneity in the CK5 expression patterns within individual patients. Second, we observed a significant inverse correlation between the percentage of HG disease and expression of CK5. In addition, we confirmed that CK5

protein expression in IHC correlated with CK5 RNA expression. Third, we observed a significant positive correlation between the percentage of HG disease and expression of the luminal IHC marker CK20. We also observed that the luminal marker GATA3 was omnipresent in Ta disease and did not correlate with any specific tumor feature. Fourth, CK5 inversely correlated with CK20. Finally, we found no significant impact of any of the IHC variables on RFS.

In recent years, research on the molecule profile of BC has expanded quickly, with various genomic studies being published. Most of these studies have focused on MIBC, and comprehensive classification schemes have been proposed by different groups (6–8,17–22). Recently, a consensus paper proposed a set of six molecular classes, with the basal/squamous (Ba/Sq) group as the most prominent subgroup (35%) (6). Although less molecular data is available for NMIBC, three molecular subgroups have been defined by several groups. Hedegaard et al. reported three major tumor classes with basal- and luminal-like characteristics and different clinical outcomes in Ta and T1 NMIBC (13,22). To the best of our knowledge, profiling studies with a unique focus on Ta NMIBC have not been published previously.

Several groups have tried to characterize these molecular subgroups by IHC, and GATA-3 and CK5 have emerged as the most frequently used luminal and basal marker, respectively (9,23). Our data suggest different biology underlying CK5/CK20 expression in Ta NMIBC and MIBC (6). Loss of CK5 and gain of CK20 at the protein level in Ta NMIBC can be hypothesized to mark an evolution to a more aggressive Ta tumor type, whereas the opposite has been reported in MIBC (6,24). In MIBC, CK5 expression is a marker of basal differentiation and CK20 of luminal differentiation, which is not the case in Ta NMIBC. The diffuse expression of the basal cell markers P40 and P63 in Ta NMIBC, regardless of the CK5 pattern, supports this thesis. As mentioned above, the role of CK5 in T1 NMBIC is not well-elucidated in the literature but, based on current knowledge, we can hypothesize a dichotomous role. In some T1 subgroups, low CK5 could correlate with high-risk biology, whereas in other T1 subgroups there may be a correlation between high CK5 and high-risk biology (22). This may reflect the existence of two different pathways of CK5 expression in T1 NMIBC. Differences in the molecular biology of NMIBC and MIBC have been reported by others, demonstrating similar gene expression subtypes between NMIBC and MIBC but opposite prognostic results (13). Sjödahl et al. reported inconsistencies between mRNA expression and tumor-cell phenotypes in MIBC but not NMIBC (15). We observed a concordance between CK5 protein and mRNA expression in

Ta NMIBC, which is in line with other reports (15,23). Most of the molecular classification studies on BC have used gene expression methods, but the discordant correlations between mRNA and protein levels in MIBC versus NMIBC show that gene expression data should always be validated at the protein level. As suggested by others, a bi-nominal classification system consisting of both tumor-cell phenotype and gene expression clusters would be more appropriate (15).

Studies on the prognostic value of CK5 expression in NMIBC are rather diffuse. Our study could not link any of the CK5 patterns to recurrence, which is in line with the findings of Lobo et al (23). Patschan et al. identified a subgroup of T1 NMIBC patients with high expression of CK5 in IHC and poor PFS, but also reported a subgroup with little or no detectable CK5 and poor PFS (22). In contrast to these findings, Breyer et al. found that the combination of high KRT20 and low KRT5 expression is a significant predictor of worse RFS and PFS in patients with T1 NMIBC (14). Two separate studies on non-muscle invasive upper urinary tract carcinoma (UTUC) found that CK5 negativity is an independent prognostic factor for shorter PFS and CSS (25,26). For both PFS and OS, it was not possible to stratify patients in the present study due to the low number of events, possibly due to the lower risk of progression of Ta disease compared to T1 disease and by the limited follow-up of 19 months.

This study has several limitations. First, clinical data were retrieved retrospectively, which could have caused a bias in data collection. Second, our cohort consisted of a mix of patients with primary and recurrent disease with HG disease present in 82% of patients. Third, follow-up was limited to a median of 19 months, and we had relatively few patients progress to MIBC with subsequent cancer-specific death. Finally, we validated CK5 expression with qPCR in only 14 of the 20 patients, based on a sufficient RNA concentration for further downstream analyses, and we did not perform a standard curve of RNA dilutions due to the limited availability of RNA.

The inverse correlation between CK5 and HG disease, the positive correlation between CK20 and HG disease, the inverse correlation between CK5 and CK20, and the absent correlation between CK5 and GATA3 in our Ta NMIBC cohort support the presence of different molecular biology between NMIBC and MIBC. Our study set-up is unique in that we used a homogenous Ta NMIBC cohort compared to the often-used Ta/T1 NMIBC cohorts. The present findings suggest different applicability and relevance of the GATA3/CK5/CK20 panel for the

classification and grading of NMIBC compared to MIBC. The added value of these markers to clinico-pathological grading has to be proven in larger prospectively designed trials with longer term data on clinical outcomes. If validated, CK5 and CK20 may serve as additional markers for differentiating between HG and LG Ta NMIBC disease.

5.6 Acknowledgments

Steven Joniau is a senior clinical researcher of the Research Foundation of Flanders (FWO).

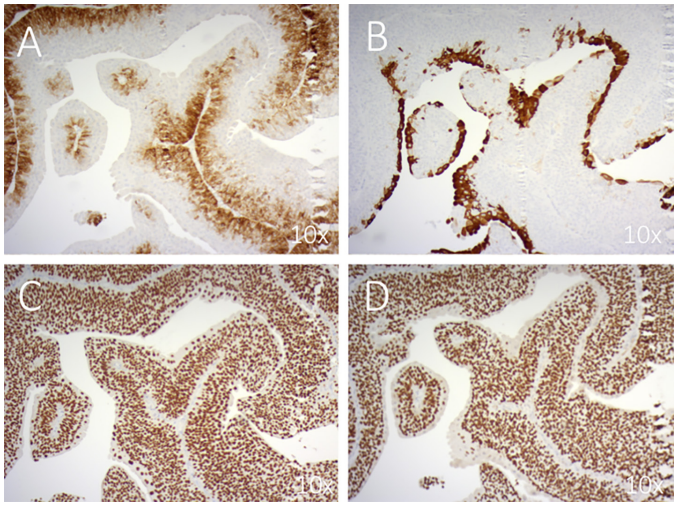
5.7 Authors Contributions

LL and TG conceived the study design. LL and TG scored H&E and IHC. TM performed qPCR and the data analysis and generated the figures. TM, VD, MC, and RH performed clinical data collection. TM, MA, FV, LL, SJ, and TG were involved in further data interpretation. All authors were involved in writing the paper and gave final approval of the submitted and published versions.

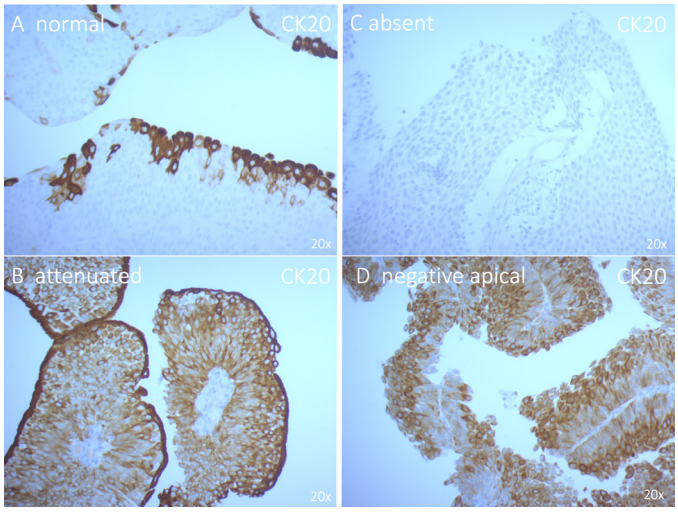
5.8 Conflict of Interest

The authors declare that they have no conflict of interest related to this manuscript.

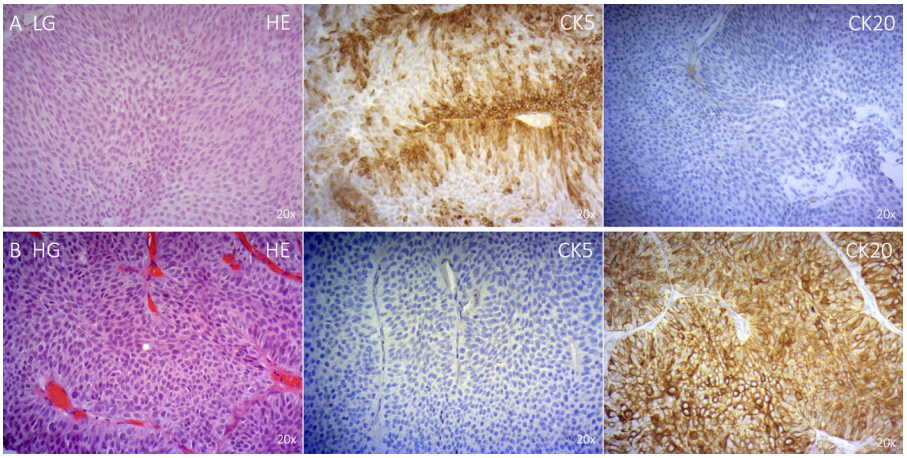
5.9 Supplementary Figures



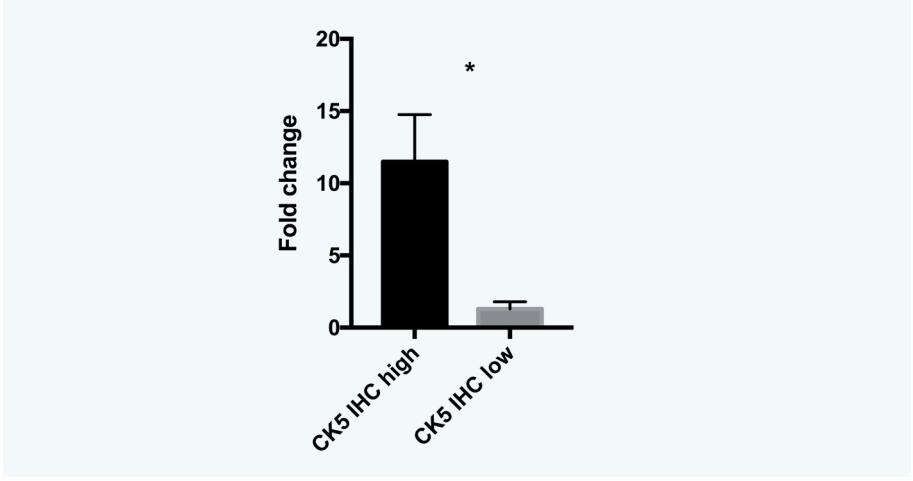
Supplementary Figure 1: Patterns of IHC expression of basal and luminal markers: **A**: CK5, **B**: CK20, **C**: GATA3, and **D**: P40; on 10x magnification. Expression pattern of P63 was similar to that of P40. Similar tissue areas are shown.



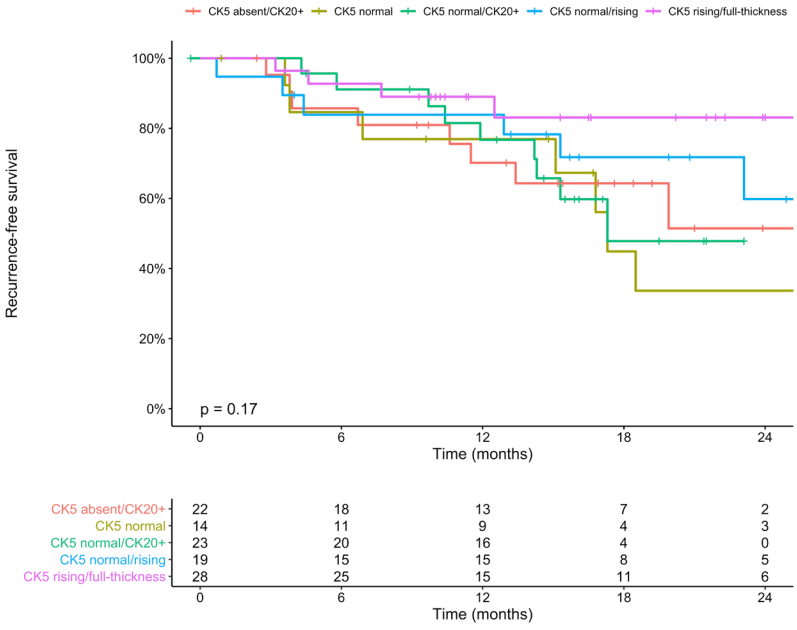
Supplementary Figure 2: Patterns of IHC expression of CK20: **A**: apical only, **B**: apical attenuated, **C**: absent, and **D**: apical negative; on 20x magnification. Patterns A and B both express apical CK20 staining, whilst patterns C and D lack apical CK20 staining.



Supplementary Figure 3: Illustrative figure showing inverse correlation between tumor grading on HE, % of CK5 on IHC, % of CK20 on IHC in Ta NMIBC.



Supplementary Figure 4: There was a significant difference in RNA expression of CK5 between patients that had 100% of CK5 IHC expression (combinations of CK5 IHC patterns: 'normal', 'rising', or 'full-thickness') and patients that had <50% presence of CK5 IHC expression (unpaired T-test: p=0.02*).Data were visualized using the 2–ΔΔCt method for relative quantifications compared to internal control with GAPDH and S18.



Supplementary Figure 5: Kaplan-Meier plot of recurrence-free survival (RFS) of all patients by clusters identified using hierarchical clustering: CK5 absent/CK20+, CK5 normal, CK5 normal/CK20+, CK5 normal/rising , and CK5 rising/full-thickness.: CK5 absent/CK20+, CK5 normal, CK5 normal/CK20+, CK5 normal/rising , and CK5 rising/full-thickness.

5.10 Supplementary Tables

Supplementary Table 1: Spearman correlation index (ρ) between % of HG disease and % of CK5 and CK5 immunohistochemical expression patterns.

Correlation with % HG disease	ρ	p-value
Absent CK5	0.31	0.001
Normal CK5 pattern	0.17	0.07
Rising CK5 pattern	-0.28	0.003
Full-thickness CK5 pattern	-0.18	0.06
Any CK5 (combined)	-0.31	0.001

Abbreviations: HG: high grade; ρ: Spearman correlation index; p-value: probability value.

Supplementary Table 2: Comparison between primary vs. recurrent disease for % of tumor grading, CK5 patterns, and other IHC markers.

	Primary disease (n=59)		**Recurrent disease** (n=48)		
	median %	*IQR*	*median %*	*IQR*	*p-value*
HG disease	50	25-70	50	10-70	0.7
Any CK5	100	85-100	95	50-100	0.03
CK5 absent	0	0-15	5	0-50	0.03
CK5 normal	35	10-50	50	28-70	0.1
CK5 rising	40	10-70	20	0-50	0.06
CK5 full-thickness	0	0-10	0	0-0	0.02
GATA3	100	100-100	100	100-100	0.07
P40	90	90-95	90	90-95	0.2
P63	90	90-90	90	85-95	0.3
CK20	40	5-75	55	7-75	0.9

Abbreviations: IHC: immunohistochemistry; IQR: interquartile range; p-value: probability value.

Supplementary Table 3: Comparison between European Association of Urology (EAU) risk categories for CK5 patterns, and other IHC markers.

EAU risk category	**High** (n=90)		**Intermediate** (n=10)		**Low** (n=8)		
	median %	*IQR*	*median %*	*IQR*	*median %*	*IQR*	*p-value*
Any CK5	100	60-100	100	90-100	100	98-100	0.2
CK5 absent	0	0-40	0	0-10	0	0-2	0.2
CK5 normal	43	20-60	25	0-73	23	18-30	0.4
CK5 rising	28	0-50	55	3-85	70	45-76	0.06
CK5 full-thickness	0	0-0	0	0-0	0	0-5	0.7
GATA3	100	100-100	100	100-100	100	100-100	0.7
P40	90	90-95	95	95-95	93	91-94	0.6
P63	90	90-95	48	24-71	NA	NA-NA	0.1
CK20	50	5-75	70	25-80	22	4-49	0.4

Abbreviations: EAU: European Association of Urology; IHC: immunohistochemistry; IQR: interquartile range; p-value: probability value.

Supplementary Table 4: Multivariable cox proportional hazards model of CK5 immunohistochemical expression patterns for recurrence-free survival (p = 0.4; C-index = 0.55)

	HR	95%CI	p-value
Normal CK5 pattern	1.50	0.32-7.15	0.61
Rising CK5 pattern	0.75	0.20-2.80	0.67
Full-thickness CK5 pattern	0.28	0.01-5.42	0.40

Abbreviations: HR: hazard ratio; 95%CI: 95% confidence interval; p-value: probability value.

5.11 References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394–424.
2. Babjuk M, Burger M, Compérat EM, Gontero P, Mostafid AH, Palou J, et al. European Association of Urology Guidelines on Non-muscle-invasive Bladder Cancer (TaT1 and Carcinoma In Situ) - 2019 Update. Eur Urol. 2019 Aug 20;
3. Sylvester RJ, van der Meijden APM, Oosterlinck W, Witjes JA, Bouffioux C, Denis L, 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 Mar;49(3):466–77.
4. Denzinger S, Fritsche H-M, Otto W, Blana A, Wieland W-F, Burger M. Early Versus Deferred Cystectomy for Initial High-Risk pT1G3 Urothelial Carcinoma of the Bladder: Do Risk Factors Define Feasibility of Bladder-Sparing Approach? Eur Urol. 2008 Jan 1;53(1):146–52.
5. Robertson AG, Kim J. Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer. Cell. 2017;
6. Kamoun A, de Reyniès A, Allory Y, Sjödahl G, Robertson AG, Seiler R, et al. A Consensus Molecular Classification of Muscle-invasive Bladder Cancer. Eur Urol. 2020 Apr 1;77(4):420–33.
7. Choi W, Porten S, Kim S, Willis D, Plimack ERRR, Hoffman-Censits J, et al. Identification of Distinct Basal and Luminal Subtypes of Muscle-Invasive Bladder Cancer with Different Sensitivities to Frontline Chemotherapy. Cancer Cell. 2014 Feb;25(2):152–65.

8. Sjödahl G, Lövgren K, Lauss M, Patschan O, Gudjonsson S, Chebil G, et al. Toward a Molecular Pathologic Classification of Urothelial Carcinoma. Am J Pathol. 2013 Sep;183(3):681–91.
9. Dadhania V, Zhang M, Zhang L, Bondaruk J, Majewski T, Siefker-Radtke A, et al. Meta-Analysis of the Luminal and Basal Subtypes of Bladder Cancer and the Identification of Signature Immunohistochemical Markers for Clinical Use. EBioMedicine. 2016 Oct;12:105–17.
10. Higgins JPT, Kaygusuz G, Wang L, Montgomery K, Mason V, Zhu SX, et al. Placental S100 (S100P) and GATA3: Markers for Transitional Epithelium and Urothelial Carcinoma Discovered by Complementary DNA Microarray. Am J Surg Pathol. 2007 May;31(5):673–80.
11. Chen Y, Peng Y, Fan S, Li Y, Xiao ZX, Li C. A double dealing tale of p63: an oncogene or a tumor suppressor. Vol. 75, Cellular and Molecular Life Sciences. Birkhauser Verlag AG; 2018. p. 965–73.
12. Papadimitriou MA, Avgeris M, Levis PK, Tokas T, Stravodimos K, Scorilas A. ΔNp63 transcript loss in bladder cancer constitutes an independent molecular predictor of TaT1 patients post-treatment relapse and progression. J Cancer Res Clin Oncol. 2019 Dec 1;145(12):3075–87.
13. Hedegaard J, Lamy P, Nordentoft I, Algaba F, Høyer S, Ulhøi BP, et al. Comprehensive Transcriptional Analysis of Early-Stage Urothelial Carcinoma. Cancer Cell. 2016 Jul 11;30(1):27–42.
14. Breyer J, Wirtz RM, Otto W, Erben P, Kriegmair MC, Stoehr R, et al. In stage pT1 non-muscle-invasive bladder cancer (NMIBC), high KRT20 and low KRT5 mRNA expression identify the luminal subtype and predict recurrence and survival. Virchows Arch. 2017 Mar 1;470(3):267–74.
15. Sjödahl G, Eriksson P, Liedberg F, Höglund M. Molecular classification of urothelial carcinoma: global mRNA classification versus tumour-cell phenotype classification. J Pathol. 2017 May 1;242(1):113–25.
16. Bernardo C, Eriksson P, Marzouka N al dain, Liedberg F, Sjödahl G, Höglund M. Molecular pathology of the luminal class of urothelial tumors. J Pathol. 2019 Nov 1;249(3):308–18.
17. Choi W, Ochoa A, McConkey DJ, Aine M, Höglund M, Kim WY, et al. Genetic Alterations in the Molecular Subtypes of Bladder Cancer: Illustration in the Cancer Genome Atlas Dataset. Eur Urol. 2017 Sep 1;72(3):354–65.
18. Aine M, Eriksson P, Liedberg F, Sjödahl G, Höglund M. Biological determinants of bladder cancer gene expression subtypes. Sci Rep. 2015 Jun 8;5:10957.

19. McConkey DJ, Choi W, Dinney CPN. New insights into subtypes of invasive bladder cancer: Considerations of the clinician. Eur Urol. 2014 Oct;66(4):609–10.
20. Eich ML, Dyrskjøt L, Netto GJ. Toward personalized management in bladder cancer: the promise of novel molecular taxonomy. Virchows Arch. 2017 Aug 1;471(2):271–80.
21. Lerner SP, McConkey DJ, Hoadley KA, Chan KS, Kim WY, Radvanyi F, et al. Bladder Cancer Molecular Taxonomy: Summary from a Consensus Meeting. Bladder Cancer Amst Neth. 2016;2(1):37–47.
22. Patschan O, Sjödahl G, Chebil G, Lövgren K, Lauss M, Gudjonsson S, et al. A molecular pathologic framework for risk stratification of stage T1 urothelial carcinoma. Eur Urol. 2015 Nov 1;68(5):824–32.
23. Lobo J, Monteiro-Reis S, Guimarães-Teixeira C, Lopes P, Carneiro I, Jerónimo C, et al. Practicability of clinical application of bladder cancer molecular classification and additional value of epithelial-to-mesenchymal transition: prognostic value of vimentin expression. J Transl Med [Internet]. 2020 Aug 5 [cited 2021 Mar 16];18. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7405371/
24. Choi W, Porten S, Kim S, Willis D, Plimack ERRR, Hoffman-Censits J, et al. Identification of Distinct Basal and Luminal Subtypes of Muscle-Invasive Bladder Cancer with Different Sensitivities to Frontline Chemotherapy. Cancer Cell. 2014 Feb;25(2):152–65.
25. Jung M, Kim B, Moon KC. Immunohistochemistry of cytokeratin (CK) 5/6, CD44 and CK20 as prognostic biomarkers of non-muscle-invasive papillary upper tract urothelial carcinoma. Histopathology. 2019 Feb 1;74(3):483–93.
26. Sikic D, Keck B, Wach S, Taubert H, Wullich B, Goebell PJ, et al. Immunohistochemiocal subtyping using CK20 and CK5 can identify urothelial carcinomas of the upper urinary tract with a poor prognosis. PLoS ONE [Internet]. 2017 Jun 20;12(6). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5478149/

Chapter 6:

Stromal marker fibroblast activation protein drives outcome in T1 non-muscle invasive bladder cancer



Chapter 6:

Stromal marker fibroblast activation protein drives outcome in T1 non-muscle invasive bladder cancer

Tim Muilwijk[1,2], Murat Akand[1,2], Sofie Daelemans[3,4], Koen Marien[3], Yannick Waumans[3], Mark Kockx[3], Loïc Baekelandt[1], Thomas Van den Broeck[1], Frank Van der Aa[1], Thomas Gevaert[1,2,5*], Steven Joniau[1*]

*Shared senior authorship

[1] University Hospitals Leuven, Department of Urology, Leuven, Belgium

[2] KU Leuven, Organ Systems, Leuven, Belgium

[3] CellCarta, Pathology – Histology, Imaging and Quantification, Antwerp, Belgium

[4] Medical Biochemistry, Faculty of Pharmaceutical, Biomedical and Veterinary Sciences, University of Antwerp, Belgium

[5] AZ Klina, Department of Pathology, Brasschaat, Belgium

PLoS One, September 15th, 2021 (IF 3.24)

6.1 Abstract

Fibroblast activation protein-α (FAP) is a transmembrane peptidase and a surrogate marker for cancer-associated fibroblasts (CAFs). FAP has been linked to worse prognosis and therapy resistance in several cancers. We hypothesised that FAP might have a prognostic 3biomarker potential to stratify patients with high-grade (HG) T1 non-muscle-invasive bladder cancer (NMIBC). We selected 30 patients with HG T1 NMIBC that progressed to ≥T2 disease which were pair-matched based on CUETO progression score variables with 90 patients that did not progress. After revision a final cohort of 86 patients was retained. Slides were stained for FAP, the luminal marker GATA3 and the basal marker CK5. All HG T1 tumour regions of interest (ROIs) within each patient were annotated, analysed and scored using image analysis software. FAP expression in HG T1 ROIs was significantly higher in progressors vs. non-progressors and was prognostic for recurrence-free survival, progression-free survival, cancer-specific survival, and overall survival. FAP expression in HG T1 ROIs remained strongly prognostic for these outcomes in a bivariable model corrected for adequate BCG per FDA definition. Expression of GATA3 and CK5 did not differ between progressors vs. non-progressors, and were not prognostic for these outcomes. FAP might serve as an easily applicable prognostic biomarker to risk-stratify patients with HG T1 NMIBC if these results are prospectively validated in a larger series.

6.2 Introduction

Due to a significant heterogeneity in clinical behaviour of stage T1 non-muscle-invasive bladder cancer (NMIBC), there is an unmet need for prognostic biomarkers to aid in patient counseling and treatment decisions(1). Fibroblast activation protein-α (FAP) is a transmembrane peptidase expressed on stromal cells and acts as a surrogate marker for cancer-associated fibroblasts (CAFs) which play a major role in the tumour microenvironment (2,3). FAP overexpression in CAFs has been linked to resistance to therapy in several solid cancers, such as colorectal, pancreatic, gastric and ovarian cancer, highlighting the importance of stromal components in determining outcome (2–5). Furthermore, FAP overexpression in colorectal, pancreatic and ovarian cancer is associated with an increased risk of lymph node metastasis and worse prognosis (2,6). We hypothesised that FAP expression might have a prognostic biomarker potential in patients with T1 NMIBC. To that end, we designed a retrospective study with high grade (HG) T1 NMIBC and pair-matched patients,

who did experience progression to ≥T2 disease, with patients who did not progress. Given the emerging data on the correlation between molecular subtypes and clinical outcome in bladder cancer (BC), we also included CK5 and GATA3 as surrogate immunohistochemical (IHC) markers for basal and luminal differentiation, respectively (7,8).

6.3 Materials and methods

From our institutional database of HG T1 NMIBC, we retrospectively selected 30 patients who progressed to ≥T2. Those were pair-matched using the Club Urologico Español de Tratamiento Oncologico (CUETO) progression score variables with 90 patients who never progressed despite having even longer follow-up (9). All patients completed at least 5/6 doses of their initial BCG induction course. Haematoxylin slides and the corresponding formalin-fixed paraffin-embedded blocks from the transurethral resection of bladder tumour of the first diagnosis of HG T1 of the selected patients were retrieval from our institutional pathology archives. All procedures were performed between March 1993 and April 2015. After single pathologist revision of all slides and tissue quality assessment, we retained a final cohort of 86 patients (64 non-progressors and 22 progressors) (**S1 Figure**). There were no significant differences in clinical and pathological variables between progressors and non-progressors, except for the presence of carcinoma in situ (CIS) (p=0.02) (**Table 1**). Slides were stained for FAP and CK5-GATA3 (double stain), using validated protocols on the Ventana Discovery Ultra and Ventana Benchmark XT platform (**S2 Table**). Stained slides were scanned and all HG T1 tumour regions within each patient were digitally annotated on a serial H&E-stained slide by an expert uro-pathologist (TG). Multiple tumour stages (T1, Ta or CIS) and/or regions of interest (ROIs) were often present within one patient, but only HG T1 ROIs were used for our analyses. Digital annotations of ROIs were transferred to the image analysis software Visiopharm, and stains were scored via a quantitative method per ROI. CK5 and GATA3 were analysed in the tumour area of HG T1 ROIs, and were measured as relative marker area (RMA) (formula: IHC-positive tumour area/total tumour area). FAP was analysed in the stromal area of HG T1 ROIs, and was also measured as RMA (formula: IHC-positive stromal area/total stromal area; **Figure 1**). Ethical approval was obtained at the Ethics Committee Research UZ/KU Leuven (internal study number S59191) in accordance with the ethical standards of the Declaration of Helsinki of 1964 and its later amendments, with a waiver of informed consent due to the retrospective nature and the secondary use of tissue.

Table 1: Clinical and pathological characteristics of included patients

	All patients n=86		Non-progressors n=64		Progressors n=22		*P-value**
	No.	*%*	*No.*	*%*	*No.*	*%*	
Gender							
Male	69	80	49	77	20	91	0.1
Female	17	20	15	23	2	9	
Age							
<60 years	13	15	12	19	1	5	0.2
60-70 years	29	34	19	30	10	45	
>70 years	44	51	33	52	11	50	
Smoking history							
Active	12	14	8	13	4	18	0.8
Never	36	42	27	42	9	41	
Stopped	38	44	29	45	9	41	
T1 status[&]							
Primary T1	71	83	54	84	17	77	0.4
Secondary T1	15	17	10	16	5	23	
Prior intravesical therapy							
BCG	1	1	0	0	1	5	0.2
Chemotherapy	4	5	3	5	1	5	
Size of lesions (cm)							
<3	31	36	22	34	9	41	0.6
≥3	55	64	42	66	13	59	
Number of lesions (n)							
≤3	39	45	28	44	11	50	0.6
>3	47	55	36	56	11	50	
Presence of CIS	25	29	14	22	11	50	0.02
T1 substage[#]							
extensive invasion	51	59	39	61	12	55	0.6
microinvasion	35	41	25	39	10	45	
Presence of detrusor	73	85	52	81	21	95	0.1
Re-TURBT performed							
Yes – before BCG ind.	24	28	16	25	8	36	0.6
Yes – after BCG ind.	53	62	41	64	12	55	
No	9	10	7	11	2	9	
Detrusor at re-TURBT if absent at T1 TURBT (n=13)							
Present detrusor	4		3		1		0.2
Absent detrusor	6		6		0		
No re-TURBT	3		3		0		
BCG							
Induction (≥5/6)	86	100	64	100	22	100	
Maintenance (≥2/3 1st)[%]	23	27	20	31	3	14	0.1

Abbreviations: BC: bladder cancer; CIS: carcinoma in situ; ind.: induction; HG: high grade; LG: low grade; NMIBC: non-muscle-invasive bladder cancer; ROIs: regions of interest; TURBT: transurethral resection of bladder tumour.

*Comparison between progressors vs. non-progressors by chi-squared test. &Primary T1: no prior NMIBC; Secondary T1: prior Ta and/or Tis NMIBC; #Lamina propria invasion depth cut-off 0.5mm. %Adequate BCG as defined by the FDA.

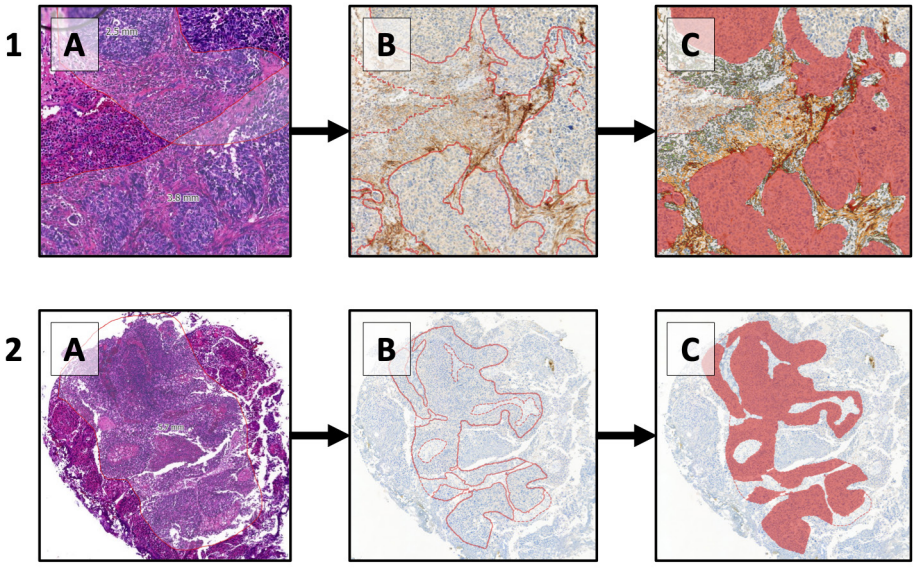


Figure 1: Patient 1 and patient 2: Illustrative figure of the analytical workflow. Patient 1 shows high FAP expression and patient 2 shows absent FAP expression. **A**: T1 tumour regions of interest (ROI) within each patient were digitally annotated on a serial H&E-stained slide by an expert uro-pathologist (TG) within Pathomation digital pathology software. **B**: Digital annotations of ROIs were transferred from the H&E slides to the FAP-IHC slides with the image analysis software Visiopharm, and areas with tumour and stroma were separated. **C**: FAP expression was measured in the stromal areas only as relative marker area (RMA) with integration of low, medium and high expression levels (formula, calculated using Visiopharm: [RMA low FAP] + [RMA medium FAP] + [RMA high FAP]) (red coloured areas are omitted tumour areas).

6.4 Results

The 5-year progression-free survival (PFS) of the total population was 73% (95% Confidence Interval [CI]: 64-84). Median time to progression was not reached, and median follow-up of non-progressors was 7.8 years (Interquartile range [IQR]: 4.7-11.7). FAP RMA was significantly higher in HG T1 ROIs of progressors vs. non-progressors ($p<0.05$). There were no significant differences between progressors and non-progressors in CK5 and GATA3 RMAs in T1 ROIs. There were no significant differences in FAP, GATA3 and CK5 RMAs between patients with a primary diagnosis of HG T1 with no prior NMIBC (primary T1; n=71), and a secondary diagnosis of T1 with prior Ta and/or Tis NMIBC (secondary T1; n=15). We constructed a bivariable Cox proportional hazards model with FAP RMA in HG T1 ROIs, and adequate BCG (per Food and Drug Administration [FDA] definition: at least 5/6 doses of induction plus 2/3 doses of maintenance, or at least 5/6 doses of induction plus at least 2/6 doses of a second induction course; yes vs. no) for recurrence-free survival (RFS), PFS, cancer-specific survival (CSS), and overall survival (OS). CUETO progression score was not included in the model as our primary cohort was pair-matched using this score, and CUETO progression scores remained balanced between groups following single pathologist revision of all slides and tissue quality assessment ($p=0.8$) (9). There was no significant difference in RFS ($p=0.08$) or PFS ($p=0.8$) when stratifying for presence of CIS (yes vs. no). Additionally, there was no correlation between FAP RMA in HG T1 ROIs and CUETO progression score ($\rho=-0.02$; $p=0.8$), and CUETO progression score was also not prognostic for PFS ($p=0.5$), and RFS ($p=0.2$). Interestingly, FAP RMA in HG T1 ROIs was prognostic for RFS, PFS, CSS, and OS in these bivariable models (for PFS: **S3 Figure**). CK5 and GATA3 were not prognostic for these endpoints in comparable bivariable models replacing FAP with either CK5 or GATA3. Adequate BCG, as defined by the FDA, only impacted RFS, but not PFS, CSS, and OS, which stresses the fact that FAP is indeed an independent prognostic variable for all endpoints. To visualize the impact of FAP RMA in HG T1 ROIs, we dichotomized FAP as a categorical variable and compared low vs. high using log-rank test with the median FAP RMA as cut-off (**Figure 2**).

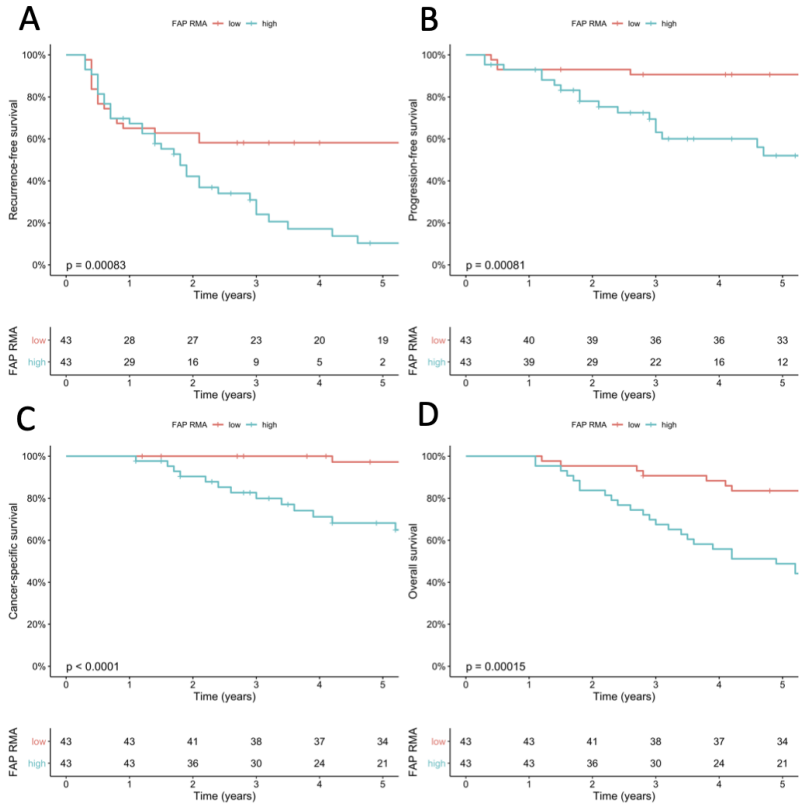


Figure 2: Kaplan-Meier plots of clinical outcome stratified by FAP expression as relative marker area (RMA) in HG T1 regions of interest (low [≤median] vs. high [>median]). FAP expression stratified all patients (n=86) significantly for each of the investigated clinical outcomes. **A**: Recurrence-free survival; **B**: Progression-free survival; **C**: Cancer-specific survival; **D**: Overall survival.

Additionally, we visualized PFS stratified by FAP RMA in T1 ROIs (low vs. high) and adequate BCG (yes vs. no), which suggested a potential predictive role of FAP **(Figure 3)**. However, no significant differences in pairwise comparison were found in PFS in FAP low patients between adequate vs. inadequate BCG (p=0.5), and in FAP high patients between adequate vs. inadequate BCG (p=1). We did find a significant difference in RFS in FAP low patients between adequate vs. inadequate BCG (p=0.01) on separate analyses for RFS and PFS for FAP high and FAP low patients (**Figure 4**). These findings suggest a potential predictive role with regard to RFS of FAP expression for adequate BCG treatment per FDA definition.

6.5 Discussion

This is the first time that a stromal factor is shown to be prognostic for progression of T1 NMIBC. Additionally, luminal and basal markers were not prognostic for clinical outcome in T1 NMIBC. The stromal marker FAP is an emerging biomarker and has been linked to impaired prognosis and therapy resistance in several cancers (2–4).

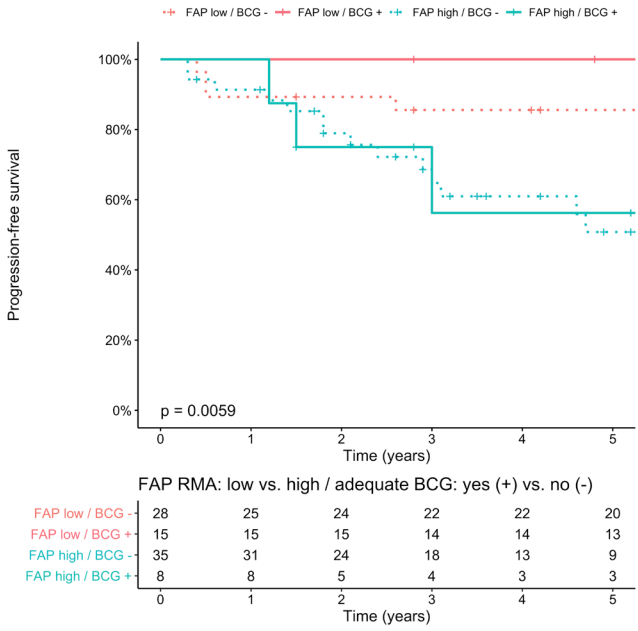


Figure 3: Kaplan-Meier plot of progression-free survival (PFS) stratified by FAP expression as relative marker area (RMA) (low [≤median] vs. high [>median]) in T1 regions of interest, and adequate BCG (as defined by FDA; yes vs. no). Four groups are visualized: FAP low with inadequate BCG (n=28), FAP low with adequate BCG (n=15), FAP high with inadequate BCG (n=34), and FAP high with adequate BCG (n=8). All patients received al least By performing a pairwise comparison between groups using log-rank corrected by Bonferroni for multiple testing, we identified significant differences in PFS between: 1. FAP low with adequate BCG vs. FAP high with inadequate BCG (p=0.02), and 2. FAP low with adequate BCG vs. FAP high with adequate BCG (p=0.04). However, no significant differences were found between 1. FAP low with adequate BCG vs. FAP low with inadequate BCG (p=0.5), and 2. FAP high with adequate BCG vs. FAP high with inadequate BCG (p=1).

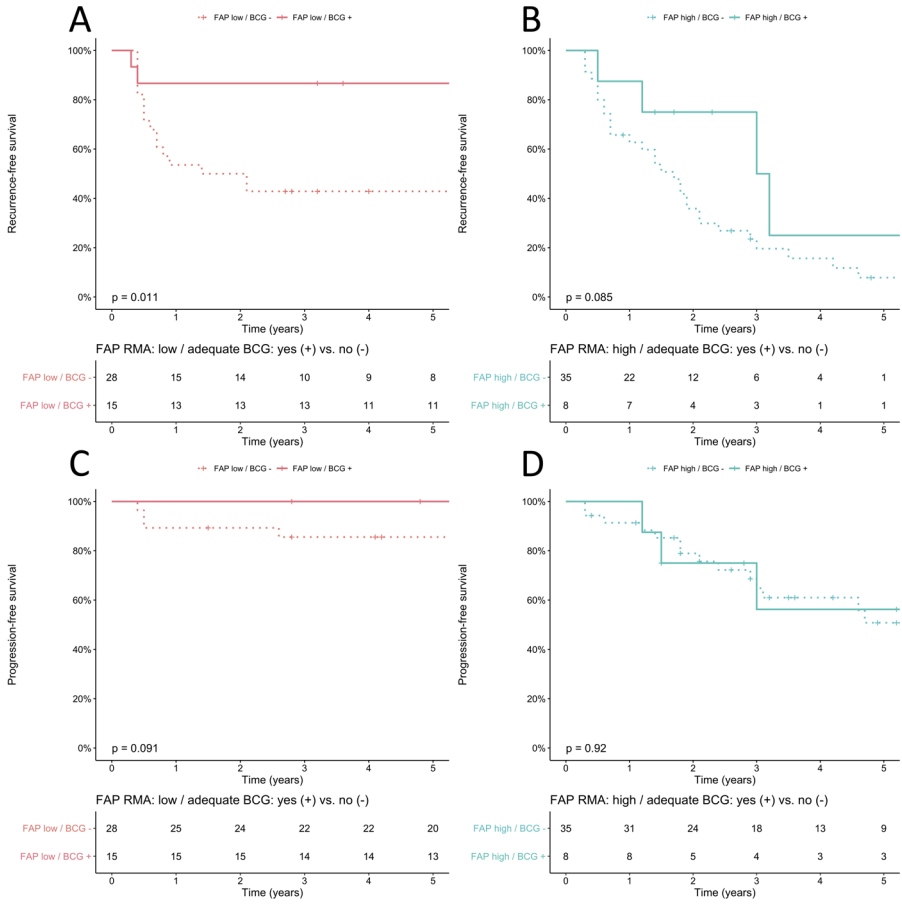


Figure 4: Kaplan-Meier plot of recurrence-free survival (RFS), or progression-free survival (PFS) stratified by adequate BCG (as defined by FDA; yes vs. no). Four groups are visualized: **A**: RFS for FAP low stratified by adequate BCG (n=43), **B**: PFS for FAP high stratified by adequate BCG (n=43), **C**: PFS for FAP low stratified by adequate BCG (n=43), **D**: PFS for FAP high stratified by adequate BCG (n=43). There was a significant difference in RFS in FAP low patients stratified by adequate BCG (yes vs. no; p=0.01).

Moreover, FAP-specific PET represents a promising imaging modality for radiotherapy planning (10), and various approaches to clinically target FAP are currently being studied in other cancers (4). At present, the possible role of FAP in BC is largely understudied. The single arm, phase 2 ABACUS-trial with neo-adjuvant atezolizumab in patients with MIBC found that FAP expression remained high in relapsing tumours whereas a decrease was seen in

responders, suggesting that FAP is linked to resistance to therapy (11). Another study on a mixed MIBC/NMIBC population identified a FAP-dominant cluster prognostic for decreased OS (12), and an additional study on MIBC found that presence of FAP in CAFs implies worse CSS (13). Our data, focusing exclusively on a HG T1 population, are in line with these findings, strongly suggesting a prognostic role for FAP expression. Clinical response to PD-L1 inhibitors in patients with metastatic UC is associated with presence of CD8+ immune cells in the tumour epithelium, whilst lack of response is associated with presence of CD8+ immune cells in the fibroblast-rich peritumoural stroma, excluded from tumour cells (14). Future research regarding FAP in T1 disease should focus on this interaction, for example by assessing tumor infiltrating lymphocytes or CD8+ cells in the tumour area and the peritumoural stroma, and their correlation with FAP (15). Strengths of this study are the pair-matched design, analysis of an exclusive T1 NMBIC cohort, and the detailed dissection and inclusion of all HG T1 ROIs within each single patient. FAP expression in HG T1 ROIs remained strongly prognostic for RFS, PFS, CSS, and OS in a bivariable model corrected for adequate BCG. Limitations are its retrospective design, a relatively small sample size, and absence of detrusor muscle at re-TURBT in a limited number of patients (n=9). FAP might serve as an easily applicable prognostic biomarker to risk-stratify patients with HG T1 NMIBC if these results are prospectively validated in a larger series.

6.6 Acknowledgments

Steven Joniau is a senior clinical researcher of the Research Foundation of Flanders (FWO).

6.7 Authors Contributions

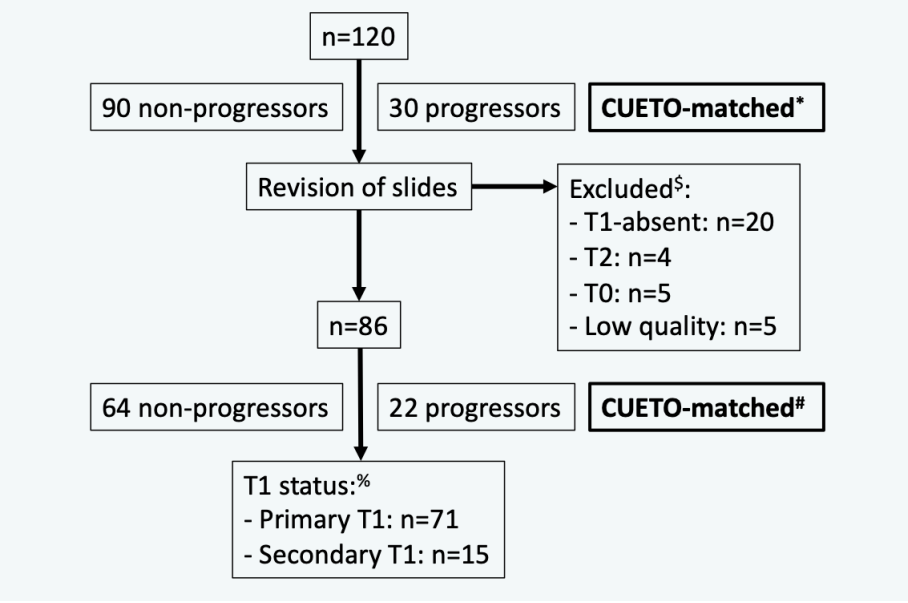
TM: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization; MA: Conceptualization, Methodology, Resources, Writing - Review & Editing; SD: Methodology, Software; KM: Methodology, Software; Yannick Waumans: Methodology, Software; MK: Methodology, Software; LB: Writing - Review & Editing; TV: Writing - Review & Editing; FV: Resources, Writing - Review & Editing, Supervision; TG: Conceptualization, Methodology, Investigation, Supervision; SJ: Conceptualization, Methodology, Resources, Supervision, Project administration, Funding acquisition

6.8 Conflict of Interest

CellCarta provided support in the form of salaries for authors SD, KM, YW, and MK, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the author statement.

6.9 Supplementary data

S1 Figure:

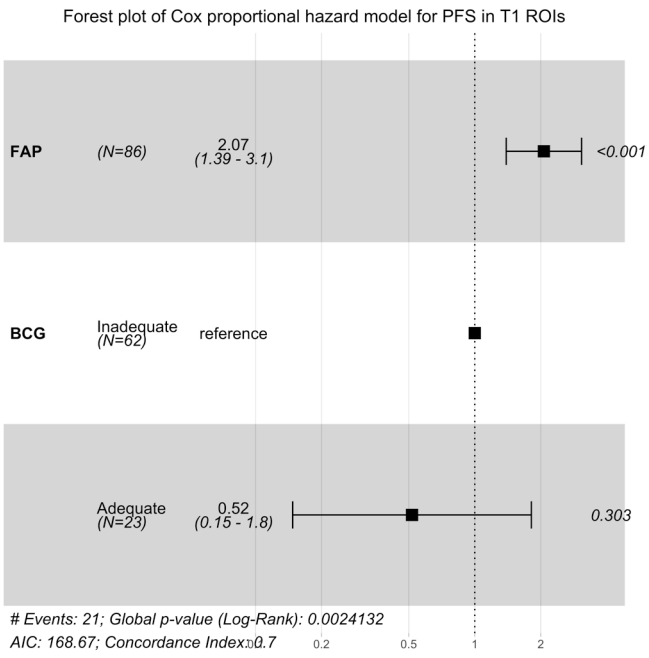


S1 Figure: Selection flowchart of included patients diagnosed with high-grade T1 NMIBC (n=120). *T1 patients were pair-matched in 1:3 for CUETO progression score variables. $Patients were excluded due to: T1-absent: absence of T1 on the serial slides following the initially evaluated clinical H&E slide; T2: presence of T2 disease at revision; T0: absence of relevant tumour at revision; Low quality: FAP stain was of low quality with no more left-over slides. #After revision of slides, patients were still matched for CUETO progression score variables (p=0.8). %Primary T1: no prior NMIBC; Secondary T1: prior Ta and/or Tis NMIBC.

S2 Table: Properties of the antibody clones used

Immunogen	Clone	Manufacturer/Code	Host
Fibroblast activation protein-α (FAP)	SP325	Spring Bioscience M6250	Rabbit
Cytokeratin 5 (CK5)	SP27	Ventana 760-4935	Rabbit
GATA binding protein 3 (GATA3)	L50/823	Ventana 760-4935	Mouse

S3 Figure:



S3 Figure: Forest plot of bivariable Cox proportional hazards model for progression-free survival (PFS) incorporating FAP expression as relative marker area (RMA) in T1 regions of interest (per factor 10; continuous variable), and adequate BCG (as defined by FDA; yes vs. no).

6.10 References

1. Bessa A, Maclennan S, Enting D, Bryan R, Josephs D, Hughes S, et al. Consensus in Bladder Cancer Research Priorities Between Patients and Healthcare Professionals Using a Four-stage Modified Delphi Method. European Urology. 2019 Aug;76(2):258–9.
2. Liu F, Qi L, Liu B, Liu J, Zhang H, Che D, et al. Fibroblast activation protein overexpression and clinical implications in solid tumors: a meta-analysis. PLoS One. 2015;10(3):e0116683.
3. Wynn TA, Ramalingam TR. Mechanisms of fibrosis: therapeutic translation for fibrotic disease. Nat Med. 2012 Jul 6;18(7):1028–40.
4. Fitzgerald AA, Weiner LM. The role of fibroblast activation protein in health and malignancy. Cancer Metastasis Rev. 2020 Sep;39(3):783–803.
5. Mhawech-Fauceglia P, Yan L, Sharifian M, Ren X, Liu S, Kim G, et al. Stromal Expression of Fibroblast Activation Protein Alpha (FAP) Predicts Platinum Resistance and Shorter Recurrence in patients with Epithelial Ovarian Cancer. Cancer Microenviron. 2015 Apr;8(1):23–31.
6. Zhang M, Xu L, Wang X, Sun B, Ding J. Expression levels of seprase/FAPα and DPPIV/CD26 in epithelial ovarian carcinoma. Oncol Lett. 2015 Jul;10(1):34–42.
7. Kamoun A, de Reyniès A, Allory Y, Sjödahl G, Robertson AG, Seiler R, et al. A Consensus Molecular Classification of Muscle-invasive Bladder Cancer. European Urology. 2020 Apr 1;77(4):420–33.
8. Dadhania V, Zhang M, Zhang L, Bondaruk J, Majewski T, Siefker-Radtke A, et al. Meta-Analysis of the Luminal and Basal Subtypes of Bladder Cancer and the Identification of Signature Immunohistochemical Markers for Clinical Use. EBioMedicine. 2016 Oct;12:105–17.
9. Fernandez-Gomez J, Madero R, Solsona E, Unda M, Martinez-Piñeiro L, Gonzalez M, et al. Predicting nonmuscle invasive bladder cancer recurrence and progression in patients treated with bacillus Calmette-Guerin: the CUETO scoring model. The Journal of urology. 2009 Nov;182(5):2195–203.
10. Windisch P, Zwahlen DR, Koerber SA, Giesel FL, Debus J, Haberkorn U, et al. Clinical Results of Fibroblast Activation Protein (FAP) Specific PET and Implications for Radiotherapy Planning: Systematic Review. Cancers (Basel). 2020 Sep 15;12(9).

11. Powles T, Kockx M, Rodriguez-Vida A, Duran I, Crabb SJ, Van Der Heijden MS, et al. Clinical efficacy and biomarker analysis of neoadjuvant atezolizumab in operable urothelial carcinoma in the ABACUS trial. Nature Medicine. 2019;25(11):1706–14.
12. Mezheyeuski A, Segersten U, Leiss LW, Malmström P-U, Hatina J, Östman A, et al. Fibroblasts in urothelial bladder cancer define stroma phenotypes that are associated with clinical outcome. Sci Rep. 2020 Jan 14;10(1):281.
13. Calvete J, Larrinaga G, Errarte P, Martín AM, Dotor A, Esquinas C, et al. The coexpression of fibroblast activation protein (FAP) and basal-type markers (CK 5/6 and CD44) predicts prognosis in high-grade invasive urothelial carcinoma of the bladder. Hum Pathol. 2019 Sep;91:61–8.
14. Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, et al. TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. Nature. 2018 Feb 14;554(7693):544.
15. Hendry S, Salgado R, Gevaert T, Russell PA, John T, Thapa B, et al. Assessing tumor infiltrating lymphocytes in solid tumors: a practical review for pathologists and proposal for a standardized method from the International Immuno-Oncology Biomarkers Working Group. Adv Anat Pathol. 2017 Nov;24(6):311–35.

Chapter 7:

Tumor immune microenvironment of high-risk non-muscle invasive bladder cancer: fibroblast activation protein and carcinoma in situ drive progression-free survival

Chapter 7:

Tumor immune microenvironment of high-risk non-muscle invasive bladder cancer: fibroblast activation protein and carcinoma in situ drive progression-free survival

Tim Muilwijk[1,2], Murat Akand[1,2], Sofie Daelemans[3,4], Koen Marien[3], Yannick Waumans[3], Mark Kockx[3], Loïc Baekelandt[1], Thomas Van den Broeck[1], Ben Van Cleynenbreugel[1], Frank Van der Aa[1], Thomas Gevaert[1,2,5*], Steven Joniau[1*]

*Shared senior authorship

[1] University Hospitals Leuven, Department of Urology, Leuven, Belgium;

[2] KU Leuven, Organ Systems, Leuven, Belgium;

[3] CellCarta, Pathology – Histology, Imaging and Quantification, Antwerp, Belgium;

[4] Medical Biochemistry, Faculty of Pharmaceutical, Biomedical and Veterinary Sciences, University of Antwerp, Belgium;

[5] AZ Klina, Department of Pathology, Brasschaat, Belgium

7.1 Introduction

Patients with high-risk non-muscle invasive bladder cancer (NMIBC) are at risk for disease progression despite treatment with bacillus Calmette-Guérin (BCG) (1). Progression to muscle-invasive disease implies significant treatment-related and cancer-specific morbidity and mortality, thus it is paramount importance to tailor treatment to each patient's risk of disease progression (2,3). Currently, there are no readily applicable prognostic biomarkers for progression that are used in clinical practice, apart from clinical and pathological features (4,5). In earlier work, we identified the stromal marker fibroblast activation protein-α (FAP) as a strong prognostic biomarker for clinical outcome in T1 NMIBC (6). In addition, FAP has been identified as a poor prognostic marker in muscle-invasive disease (≥T2), and other disease entities (7). However, the interplay between FAP, which is a marker for cancer-associated fibroblasts (CAFs), and the tumor immune microenvironment remains to be elucidated further in high-risk NMIBC.

The tumor immune microenvironment plays an important role in the anti-cancer immune response. It is important that immune cells infiltrate from the peritumoral stroma into the cancer epithelium in order to have an anti-cancer effect (8). Therefore, we designed a pair-matched nested case-control study of patients with high-risk NMIBC treated with BCG, who progressed or did not progress to muscle-invasive disease (≥T2). Matching was done based on clinical and pathological features, so progressors and non-progressors could not be distinguished based on current risk stratification (4). Using IHC markers, we aimed to identify prognostic components or combinations of components of the tumor immune microenvironment which correlate with progression-free survival (PFS) in this matched cohort of patients with high-risk NMIBC.

7.2 Materials and methods

7.2.1 Patients

From our institutional database of HG T1 NMIBC we retrospectively selected 30 patients who progressed to ≥T2. Those were matched 1:3 pairwise for CUETO progression score variables with 90 patients who never progressed to ≥T2 disease, despite an even longer follow-up(4). The CUETO progression score comprises of: age (≤70 years vs. >70 years); recurrent status (primary vs. recurrent); number of tumors (n≤3 vs. n>3); tumor stage (Ta vs. T1); presence of

carcinoma in situ (CIS; yes vs. no); and tumor grade (G3 in all of our cases). There were no significant differences in clinical and pathological variables between progressors and non-progressors at the time of pair-matching. Clinical and pathological variables were collected by patient chart review (**Table 1**). Haematoxylin slides and their respective formalin-fixed paraffin-embedded blocks were retrieved from the transurethral resection of bladder tumor at the time of diagnosis of HG T1 disease. All TURBTs were performed between March 1993 and December 2010. The following survival outcomes were collected: recurrence-free survival (RFS), progression-free survival (PFS), cancer-specific survival (CSS), and overall survival (OS). All survival endpoints were defined as the time from date of first diagnosis of HG T1 at TURBT; for RFS until the date of first recurrence of disease at cystoscopy, for PFS until the date of first diagnosis of ≥T2 disease, for CSS until the date of death due to BC, and for OS until the date of death due to any cause.

7.2.2 Immunohistochemistry and quantitative analysis

To accurately define the tumor immune microenvironment, it is important to distinguish between tumoral and stromal expression of immune cell markers. Therefore, we assessed the presence of immune cells in tumor and peritumoral stroma in each patient separately. In addition, using this approach, we defined immune phenotypes (inflamed, excluded, and desert) for each patient based on the presence of cytotoxic T cells in the tumor and peritumoral stroma (8). Because high-risk NMIBC is a heterogeneous disease, multiple tumor stages can often be present simultaneously in a single patient. Therefore, we also differentiated for the presence of immune cells and other markers of the tumor microenvironment in T1, Ta, or CIS regions within each patient.

A panel of immunohistochemical (IHC) markers was used to define the tumor immune microenvironment. A CD8-PanCK double stain was used to distinguish between tumoral and peritumoral stromal regions by the PanCK stain, and CD8 to identify cytotoxic T cells. A GZMB-CD8-FOXP3 triple stain allowed for the assessment of activated cytotoxic T cells in case of co-localization of CD8 with granzyme B (GZMB), a serine protease involved in the induction of apoptosis of target cells (9). FOXP3 is a marker for regulatory T cells that play a tumor suppressor role by recruiting cytotoxic T cells, which can additionally be assessed in the same slide. CD163 is a marker for M2 macrophages, which are considered tumor-associated macrophages and play a role in producing anti-inflammatory cytokines such as TGF-β (10). Upregulation of PD-L1 in immune

and tumor cells aids tumors in evading the immune system. FAP is surrogate marker for CAFs as mentioned earlier. CK5 and GATA3 can be used as surrogate IHC markers for basal and luminal differentiation, respectively (11,12).

After single pathologist revision of all slides and assessment of tissue quality, we maintained a final cohort of 111 patients (83 non-progressors and 28 progressors) (**Figure 1**). Serial slides were stained for PanCK-CD8 (double stain), GZMB-CD8-FOXP3 (triple stain), CD163, PD-L1 (clones SP142 and SP263), FAP and CK5-GATA3 (double stain), using validated protocols on the Ventana Discovery Ultra and Ventana Benchmark XT platforms. Stained slides were scanned and all T1, Ta, and CIS tumor regions within each patient were digitally annotated on a serial H&E-stained slide by an expert uro-pathologist (TG) within the Pathomation digital pathology software. Multiple tumor stages (T1, Ta or CIS) or regions of interest (ROIs) were often present within one patient. Digital annotations of ROIs were transferred to the image analysis software Visiopharm and applied on the PanCK-CD8 slide (double stain). A quality check was performed excluding areas of necrosis and artefacts (e.g. by cauterisation). Tumor and stromal regions within each of the annotated ROIs were differentiated with the Pan-CK stain, using an semi-automated application within Visiopharm. The remaining serial stains were then aligned with the PanCK-CD8 stain using the Tissuealign software within Visiopharm, a virtual multiplexing technique. A new quality check was performed for each of the aligned stains. Using the ROIs, it was now possible to differentiate between T1, Ta or CIS ROIs and stroma and tumor ROIs within these T1, Ta or CIS ROIs for each staining. All stains were scored by a quantitative method per ROI as previously described (6). Two methods of quantification were used: relative marker area (RMA; formula: IHC-positive area/total area), and density (formula: number per area; n/mm2). Density was calculated within Visiopharm for immune cell populations separately within stroma and tumor: CD8, CD8-GZMB+ (using automated application for recognition of co-localization of CD8 and GZMB), and FOXP3. RMA was calculated within Visiopharm for CD163 (stroma and tumor), FAP (stroma alone), CK5 (tumor alone), and GATA3 (tumor alone). PD-L1 SP142 (positive tumor-infiltrating immune cells; cut-off ≥5%) and SP263 (combined positive score; cut-off ≥10) were scored manually according to the Ventana protocol for T1, Ta and CIS ROIs by an expert uro-pathologist (TG). Based on stromal and tumoral CD8 expression in all ROIs (T1, Ta and CIS) we defined three different immune phenotypes(8): inflamed with high tumoral CD8 expression, excluded with low tumoral and high stromal CD8 expression, and desert with low tumoral and low stromal CD8 expression. Using

the RMA of CD8 expression, we were able to differentiate between these phenotypes by choosing a cut-off for both tumoral and stromal CD8 expression. We set the thresholds for tumoral CD8 at 10/mm2, and for stromal CD8 at 1000/mm2, and changed thresholds in steps of 10/mm2 as previously described(13). Using this approach, we identified the most prognostic cut-offs for PFS, and chose the following definitions: inflamed with tumoral CD8 >150/mm2; excluded with tumoral CD8 ≤150/mm2 and stromal CD8 >100/mm2; and desert with tumoral CD8 ≤150/mm2 and stromal CD8 ≤100/mm2. In addition, we defined 'T1 immune phenotypes' based on CD8 presence in T1 ROIs-only.

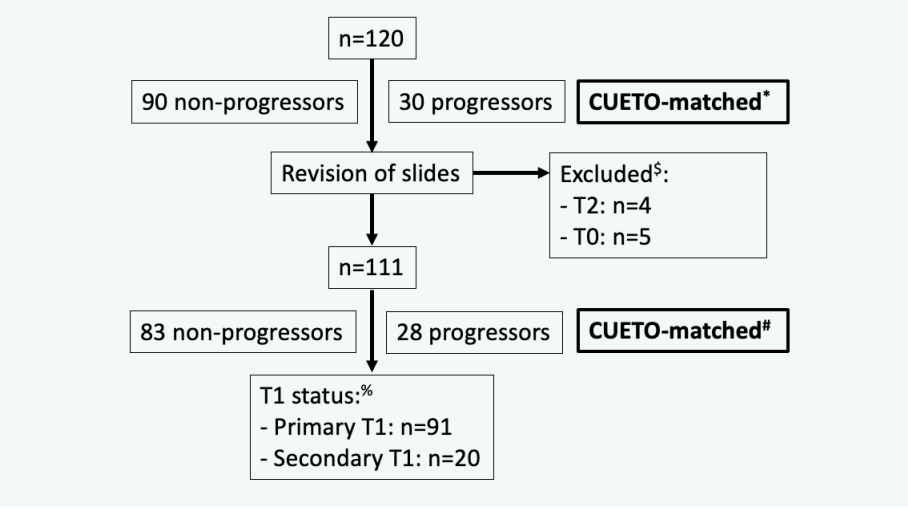


Figure 1: Selection flowchart of included patients diagnosed with high-grade (HG) T1 NMIBC (n=120). *HG T1 patients were pair-matched in 1:3 for CUETO progression score variables. $Patients were excluded due to: T2: presence of T2 disease at revision; T0: absence of relevant tumor at revision. #After revision of slides, patients were still matched for CUETO progression score variables (p=0.8). %Primary T1: no prior NMIBC; Secondary T1: prior Ta and/or Tis NMIBC.

7.2.3 Statistical analysis

Analyses were performed in R (version 4.0.0) using the 'tidyverse', 'lubridate', 'survival', 'survminer', 'my.stepwise', 'partykit', and 'pheatmap' packages. Summary statistics are presented for continuous variables as median with interquartile range (IQR), and for categorical variables as frequencies and proportions. Assumption of normality was assessed using the Shapiro-Wilk test, and equality of variance was assessed using the Levene test. Differences in

continuous variables between groups were assessed using the T-test. Analysis of variance (ANOVA) was assessed between continuous variables, and comparisons between >2 groups were performed using the Tukey's test. Kaplan-Meier plots were used to visualize the survival data. Log-rank tests and Cox proportional hazard models were used to assess correlations between survival outcome and categorical or continuous variables, respectively. Variables were checked for relevance to be included in a multivariable Cox proportional hazards model using stepwise variable selection, with iterations between forward and backward steps. Proportional hazards assumptions were tested using scaled Schoenfeld residuals. We constructed a non-parametric survival tree based on conditional inference which recursively performs univariate splits by log-rank tests. For continuous variables, the optimal cut-off is chosen by the model using the maximally selected log-rank statistic. RMAs and densities were analysed as continuous variables, and dichotomized as categorical variables (low vs. high) with the median as cut-off.

7.3 Results

7.3.1 Patient characteristics (Table 1)

A total of 111 patients were enrolled after slide revision, of which 28 patients progressed to T2 disease and 83 patients never progressed. A total of 91 patients had a primary diagnosis of HG T1 with no prior NMIBC (primary T1), and a total of 20 patients had a secondary diagnosis of T1 with prior Ta and/or CIS NMIBC (secondary T1). Patients with primary and secondary T1 were balanced between progressors and non-progressors (p=0.6). T1 substage, which was defined as lamina propria invasion depth with a cut-off of 0.5mm (extensive invasion vs. microinvasion), did not differ between progressors and non-progressors (p=0.4). Presence of lymphovascular invasion was visualized for 2 progressors and not visualized in any non-progressor (p=0.01). There was a significant difference in the presence of CIS (p=0.003) between progressors and non-progressors after single pathologist revision (**Table 1**), but CUETO progression scores remained balanced between groups (p=0.8). Other pathological and clinical variables of interest were balanced between groups (**Table 1**). After single pathologist revision, the distribution of the presence of ROIs was: T1 in 92 patients (83%), Ta in 96 patients (86%), and CIS in 31 patients (28%). The 5-year RFS was 41% (95% Confidence Interval [CI]: 32-52), with a median RFS of 3.0 years (95% CI: 2.1-5.8; **Figure 2**). The 5-year PFS was 73% (95% CI: 65-83), while the median time to progression was not

reached. The median follow-up of non-progressors was 7.8 years (Interquartile range [IQR]: 4.3-10.3). The 5-year was CSS 83% (95% CI: 76-91), while the median CSS was not reached. The 5-year OS was 65% (95% CI: 57-75), with a median OS of 8.4 years (95% CI: 6.4-11.6).

Table 1: Clinicopathological characteristics of included patients

	All patients n=111		**Non-progressors** n=83		**Progressors** n=28		*P-value**
	No.	*%*	*No.*	*%*	*No.*	*%*	
Gender							
Male	90	81	66	80	24	86	0.5
Female	21	19	17	20	4	14	
Age							
<60 years	20	18	18	22	2	7	0.2
60-70 years	33	30	22	27	11	39	
>70 years	58	52	43	51	15	54	
Smoking history							
Active	20	18	13	16	7	25	0.3
Never	39	35	28	34	11	40	
Stopped	52	47	42	50	10	35	
T1 status[&]							
Primary T1	91	82	69	83	22	79	0.6
Secondary T1	20	18	14	17	6	21	
Prior intravesical therapy							
BCG	2	2	1	1	1	4	1
Chemotherapy	4	4	3	4	1	4	
Size of lesions							
<3 cm	39	35	29	35	10	36	0.9
≥3 cm	72	65	54	65	18	64	
Number of lesions							
≤3	49	44	37	45	12	43	0.9
>3	62	56	46	55	16	57	
CIS (prior to revision)	7	6	3	4	4	14	0.06
CIS (post-revision)	31	28	17	20	14	50	0.003
T1 substage[#] (n=91)							
extensive invasion	52	57	40	60	12	50	0.4
microinvasion	39	43	27	40	12	50	
Lymphovascular invasion	2	2	0	0	2	7	0.01
Variant histology	9	8	7	8	2	7	0.9
Presence of detrusor	91	82	66	80	25	89	0.1
Re-TURBT							
Yes – before BCG ind.	35	32	23	28	12	43	0.3
Yes – after BCG ind.	66	59	53	64	13	46	
No	10	9	7	8	3	11	
BCG							
Induction (≥5/6)	111	100	83	100	28	100	
Adequate BCG[%]	38	34	33	40	5	18	0.03

Abbreviations: BC: bladder cancer; CIS: carcinoma in situ; ind.: induction; HG: high grade; LG: low grade; NMIBC: non-muscle-invasive bladder cancer; ROIs: regions of interest.

*Comparison between progressors vs. non-progressors by chi-squared test. &Primary T1: no prior NMIBC; Secondary T1: prior Ta and/or Tis NMIBC; #Lamina propria invasion depth cut-off 0.5mm. %Adequate BCG as defined by the FDA: least 5/6 doses of induction plus 2/3 doses of maintenance, or at least 5/6 doses of induction plus at least 2/6 doses of a second induction course.

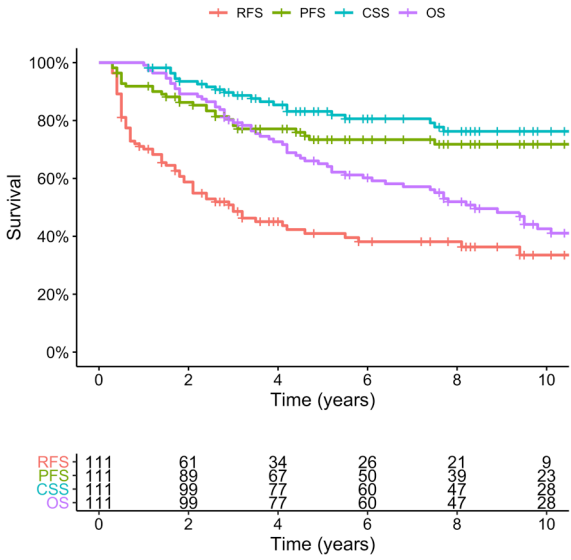


Figure 2: A combined survival plot for all outcomes of interest for all patients (n=111): recurrence-free survival (RFS), progression-free survival (PFS), cancer-specific survival (CSS), and overall survival (OS).

7.3.2 Differences in IHC markers (Figure 3; Supplementary Table 1)

7.3.2.1 Progressors vs. non-progressors

We assessed whether there were differences between progressors vs. non-progressors in stromal, tumoral or combined (stromal and tumoral) expression of all IHC markers in T1, Ta, CIS, or all ROIs combined. FAP expression was significantly higher in progressors in T1-only and in all ROIs combined, but not in Ta-only and CIS-only. Stromal CD163 expression was significantly higher in progressors in T1-only, but not in Ta-only, CIS-only, or all ROIs combined. The

other IHC markers (CD8, CD8-GZMB, FOXP3, PD-L1 SP142/SP263, GATA3, CK5) did not differ significantly between progressors vs. non-progressors. Additionally, the ratio of activated CD8 cells (CD8-GZMB/CD8) did not differ significantly between progressors vs. non-progressors.

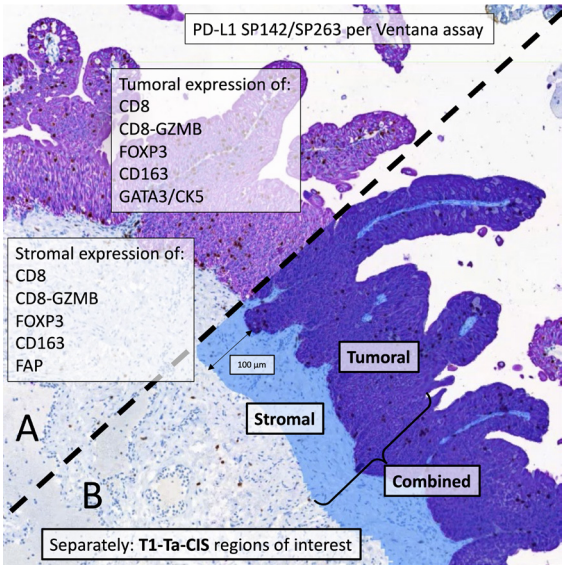


Figure 3: Illustrative figure of a PanCK stain (2.5x magnification) on assessment of immunohistochemical (IHC) markers. A: Stromal and/or tumoral (or epithelial) expression of IHC markers were assessed depending on relevance. PD-L1 (clones SP142 and SP263) was scored per Ventana assay. B: Using a PanCK stain, we differentiated between stromal and tumoral (or epithelial) expression of IHC markers, or a combination of both, within the Visiopharm software. Additionally, we differentiated expression of IHC markers for T1, Ta, and CIS regions of interest.

7.3.2.2 Regions of interest: T1 vs. Ta vs. CIS

Second, we assessed whether there were differences between T1, Ta, and CIS ROIs in stromal, tumoral, or combined (stromal and tumoral) expression of all IHC markers. FAP expression was significantly higher in T1 compared to both Ta and CIS. CD8 tumor expression was significantly higher in CIS compared to Ta, but not in T1 compared to Ta or CIS. Tumor expression of CD8-GZMB, and FOXP3 was significantly higher in both T1 and CIS compared to Ta. CD163 tumor expression was significantly higher in Ta compared to T1. CD8, CD8-GZMB, FOXP3, and CD163 combined (stromal and tumoral) expression was

significantly higher in both T1 and CIS compared to Ta. PD-L1 SP 142 immune cell score was significantly higher in T1 compared to CIS. The other IHC markers (PD-L1 SP263, GATA3, CK5) did not differ significantly between T1, Ta, and CIS ROIs.

7.3.2.3 Stroma vs. tumor

Finally, we assessed whether there were differences between stromal and tumoral expression of the IHC markers (CD8, CD8-GZMB, FOXP3, CD163) in T1, Ta, CIS, or all ROIs combined. Stromal expression of CD8, CD8-GZMB, FOXP3, and CD163 was significantly higher compared to tumoral expression. The latter was also the case in a subgroup analysis of only progressors and only non-progressors.

7.3.3 Immune phenotypes

Immune phenotype distribution was: inflamed 24% (n=27), excluded 52% (n=58), and desert 24% (n=26). Immune phenotypes correlated significantly with PFS (p=0.03; **Figure 4**), and there was a significant difference in progression between immune phenotypes (p=0.01): inflamed 41% (n=11), excluded 14% (n=8), and desert 35% (n=9). There were no significant differences in clinicopathological variables between immune phenotypes except for presence of CIS (p=0.001): inflamed 56% (n=15), excluded 21% (n=12), and desert 15% (n=4). The expression of CD163, FOXP3, CD8-GZMB was significantly higher in inflamed immune phenotypes, compared with excluded and desert immune phenotypes. The latter was true for stromal, tumoral, and combined (stromal and tumoral) expression of these markers. FAP expression did not differ between immune phenotypes.

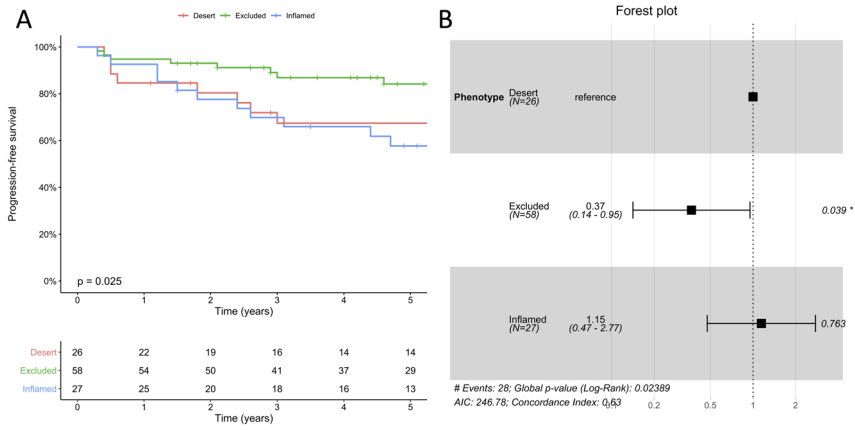


Figure 4: A: Progression-free survival and immune phenotypes correlate significantly with progression-free survival (PFS) on log-rank test (p=0.03). On pairwise comparison corrected by Bonferroni PFS, the inflamed vs. excluded immune phenotypes differed significantly (p=0.03), whilst excluded vs. desert (p=0.1) and inflamed vs. desert (p=1) immune phenotypes did not differ significantly. B: A forest plot of PFS stratified by immune phenotype reporting hazard ratio's with as reference the desert immune phenotype.

7.3.4 Progression-free survival correlated with clinicopathological and IHC variables

All clinicopathological variables (**Table 1**), IHC expression data, and immune phenotypes were correlated with PFS as continuous or categorical variables in univariable Cox proportional hazard models. Of the clinicopathological variables, the presence of CIS (p=0.004), and receipt of adequate BCG (p=0.03) correlated significantly with PFS. FAP expression correlated significantly with PFS when assessed in all ROIs combined (p=0.0009), in T1-only (p<0.0005), and in Ta-only (p=0.003). Combined (stromal and tumoral) CD163 expression in T1-only correlated significantly with PFS (p=0.02), while the other immune cell markers (CD8, CD8-GZMB, FOXP3, PD-L1 SP142, and PD-L1 SP263) did not. Expression of GATA3 in CIS alone (p=0.04), and CK5 in CIS alone (p=0.04) correlated significantly with PFS. Immune phenotypes correlated significantly with PFS (p=0.03). We assessed all variables for inclusion in a multivariable Cox proportional hazard model using a stepwise selection of variables. The only variable retained after selection was FAP expression (either in all ROIs combined, only T1, or only Ta). We next constructed a non-parametric survival

tree that included all clinicopathological variables, IHC expression data, and immune phenotypes with as clinical outcome PFS (**Figure 5A**). FAP expression in all ROIs (≤0.023 vs. >0.023) and presence of CIS (absent vs. present) were included in the survival tree as the most prognostic for PFS using this approach. Three subgroups or end nodes were created: FAP low with CIS absent, FAP low with CIS present, and FAP high. These subgroups correlated significantly with PFS ($p<0.0001$; **Figure 5B**). To visualize and illustrate these findings in combination with immune phenotypes and IHC expression data, we created an illustrative heatmap (**Figure 6**).

7.3.5 Subgroup analyses of the tumor immune microenvironment (Supplementary data)

We performed several additional survival analyses to further identify correlations between FAP, immune phenotypes, and immune cell activation related to PFS. Survival curves with results of the log-rank tests are provided in the supplementary data.

Immune phenotypes: FAP

In a subgroup analysis of each immune phenotype, we assessed correlation of FAP expression with PFS. FAP expression in all ROIs correlated significantly with PFS in the excluded and desert immune phenotypes, but in the excluded immune phenotype only FAP expression in T1-only correlated significantly with PFS.

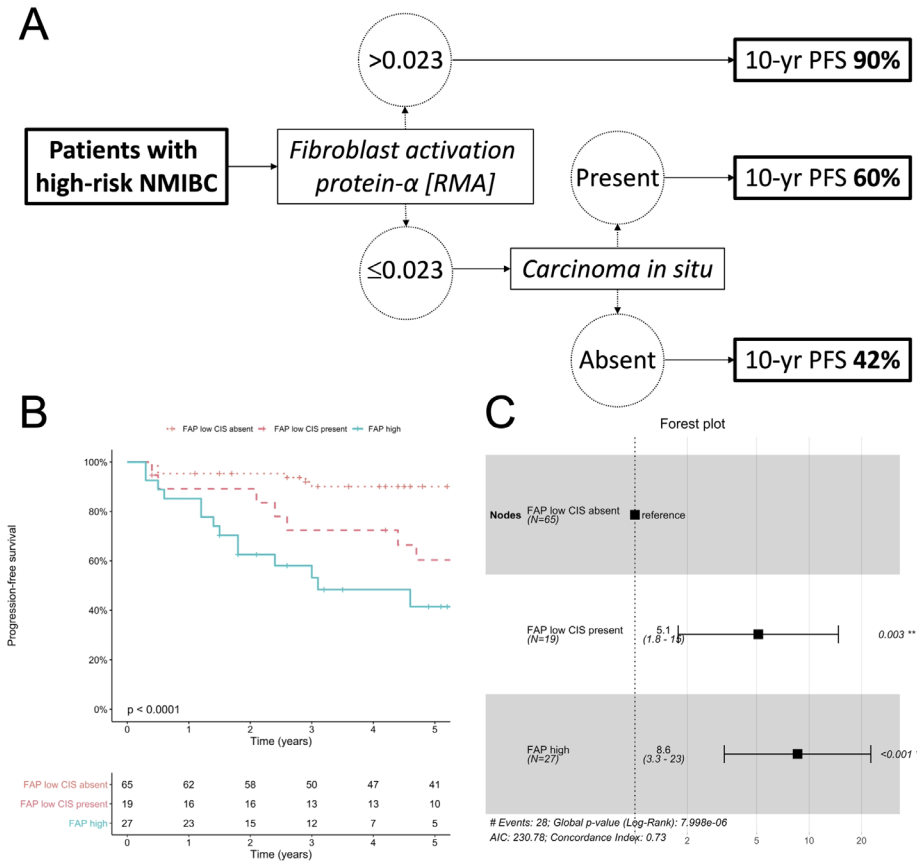


Figure 5: Tree-structured survival model using conditional inference with the outcome progression-free survival (PFS). **A**: Our cohort is first split by FAP expression in all regions of interest (T1, Ta, and CIS). Cut-off at 0.023 is chosen as most optimal by the model. The FAP ≤0.023-group is split by presence of carcinoma in situ: absent vs. present. 10-yr PFS: FAP high group 42% (95%CI: 25-69); FAP low CIS present-group 53% (95%CI: 33-84); FAP low CIS absent group 90% (95%CI: 83-98). **B**: Progression-free survival stratified by the nodes of the survival tree: FAP low with CIS absent, FAP low with CIS present, and FAP high. PFS differed significantly between groups (p<0.0001). **C**: A forest plot of PFS stratified by the terminal nodes reporting hazard ratio's with as reference the FAP low with CIS absent-group. Abbreviations: CI: confidence interval; CIS: carcinoma in situ; FAP: fibroblast activation-protein-α; NMIBC: non-muscle invasive bladder cancer; PFS: progression-free survival; RMA: relative marker area.

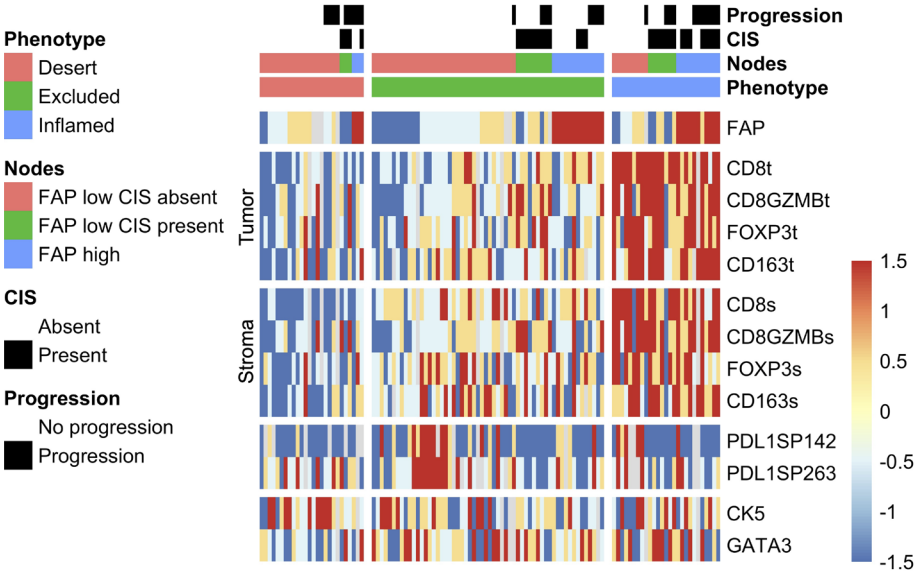


Figure 6: Illustrative (non-clustered) heatmap of all patients in which immunohistochemical (IHC) expression data was incorperated which were ordered manually: firstly by presence of carcinoma in situ (yes vs. no), secondly by progression (yes vs. no), thirdly by terminal nodes (of the non-parametric survival tree), and lastly by immune phenotypes. IHC expression data is visualized per quartile of each IHC variable to illustrate relative differences for each IHC variable between patients more accurately. Stromal or tumoral expression of a IHC marker was differentiated using the 's' or 't' behind the markers name. Abbreviations: CIS: carcinoma in situ; FAP: fibroblast activation protein-α.

Immune phenotypes: CD8-GZMB

In a subgroup analysis of each immune phenotype, we assessed correlation of CD8-GZMB expression with PFS. CD8-GZMB expression correlated significantly with PFS in the excluded immune phenotype, but did not in the inflamed and desert immune phenotype.

FAP high vs. low: CD8 and CD8-GZMB

In a subgroup analysis of patients stratified by FAP expression (≤median vs. >median), we assessed correlation of both CD8 and CD8-GZMB expression with PFS. CD8 expression did not correlate significantly with PFS in the FAP low subgroup and FAP high subgroup. However, CD8-GZMB expression correlated

significantly with PFS in the FAP high subgroup, but not in the FAP low subgroup.

Immune phenotypes with FAP high: CD8-GZMB

In a subgroup analysis of each immune phenotype separately but only in those patients with high FAP expression (>median), we assessed correlation of CD8-GZMB expression with PFS. CD8-GZMB expression did not correlate significantly with PFS in any of the immune phenotypes with high FAP expression while it did correlate with PFS in all patients in the aforementioned subgroup analysis.

CD8 high or CD8GZMB high: PD-L1

In a subgroup analysis of patients with high CD8 (>median) or with high CD8-GZMB (>median), we assessed correlation of PD-L1 SP142 IC score (in all ROIs and T1-only; positive vs. negative and ≤median vs. >median), and PD-L1 SP263 CPS (in all ROIs and T1-only; positive vs. negative and ≤median vs. >median) with PFS. PD-L1 scores did not correlate significantly with PFS in the CD8 high subgroup. However, in the CD8-GZMB high subgroup PD-L1 SP142 in T1-only and PD-L1 SP263 in T1-only correlated significantly with PFS, but not when assessed in all ROIs.

Adequate BCG: FAP

In a subgroup analysis of patients that had either adequate or inadequate BCG, we assessed the correlation of FAP expression with PFS. Both FAP in all ROIs and FAP in T1-only correlated significantly with PFS in patients that had inadequate BCG. However, in patients that had adequate BCG, FAP in T1-only correlated significantly with PFS while FAP in all ROIs did not correlate significantly.

7.4 Discussion

Using a pair-matched nested case-control approach we aimed to identify differences in patients with high-risk T1 NMIBC in terms of clinicopathological variables, components of the tumor immune microenvironment, and immune phenotypes. FAP expression was significantly higher in progressors which was mainly driven by expression of FAP in T1 ROIs. Additionally, CD163 expression in the peritumoral stroma was significantly higher in the T1 ROIs of the

progressors group. FAP expression correlated significantly with PFS, but stromal CD163 expression did not. Of all clinicopathological variables, presence of CIS correlated significantly with PFS. Interestingly, immune phenotypes also correlated significantly with PFS with a better prognosis for the excluded immune phenotype (**Figure 4**). Lastly, according to our survival tree analysis, FAP expression was found as the most important predictor for PFS. We constructed an easily applicable survival tree in which three subgroups correlated significantly with PFS: FAP low with CIS absent, FAP low with CIS present, and FAP high (**Figure 5**).

In earlier work, we identified FAP as a strong prognostic marker for clinical outcome in T1 NMIBC (6). These findings were confirmed in our current study. Additionally, FAP is more prognostic for PFS as presence of carcinoma in situ in our cohort, when analysed in a non-parametric survival tree based on conditional inference. This stresses the prognostic value of FAP in high-risk NMIBC where there is unmet need for prognostic biomarkers (14). Most immune cell markers did not correlate significantly with PFS on univariable analysis. However, we defined three immune phenotypes, of which the excluded phenotype had the best clinical outcome in terms of PFS (**Figure 4**). There is limited data regarding immune phenotype distribution in NMIBC, especially with regards to definitions based on spatial distribution of cytotoxic T cells in the tumor or peritumoral stroma. Our approach with a quantified cut-off for each immune phenotype using digital image analysis software, has the advantage that our findings can be applied to other cohorts of patients with high-risk NMIBC. This could help to identify optimal cut-offs for immune phenotypes, to identify the optimal candidates for BCG therapy. Surprisingly, the excluded immune phenotype had the best clinical outcome in terms of PFS. We hypothesize that this could be attributed to an improved recruitment of immune cells during BCG therapy, despite being excluded beforehand. Additionally, peritumoral stromal CD163 expression was significantly higher in progressors in our cohort. Tumor associated macrophages (TAM) can be either activating (M1), or suppressive (M2), and play an important role in the tumor immune microenvironment (15). CD163-positive tumor associated macrophages (TAM) can attract circulating monocytes, and prime these macrophages to turn into pro-tumoral TAMs which can produce cytokines such as TGF-β, which in their turn play a role in epithelial mesenchymal transition (EMT) (10). However, expression of peritumoral stromal CD163 did not correlate with PFS in our cohort which limits its use a prognostic biomarker.

A multi-omics analysis on NMIBC identified a class of tumors based on transcriptomic analysis with high immune cell infiltration, and significant more presence of CIS which had the highest risk of progression compared to the other classes (16). This class might correspond with our inflamed immune phenotype, in which we observed a higher immune cell infiltration (CD8, CD8-GZMB, CD163, FOXP3), more presence of CIS, and also an impaired PFS. GATA3 and CK5 as surrogate markers for luminal and basal differentiation were not prognostic in our cohort, which is in line with a recent study focusing on subcategorization of T1 NMIBC (17). Another transcriptomic study in T1 disease-only identified two classes with desert and inflamed immune phenotypes (18). Epithelial mesenchymal transition (EMT) was present in two classes of tumors with enriched hallmarks for TGF-β signaling. TGF-β is a pivotal inducer for fibroblast activation which can also be linked to FAP expression (19). Most disease recurrences were seen in the class with a higher EMT presence. The T1BC classification applies more directly to our cohort, as our cohort also consisted of mostly T1 disease treated with BCG. The findings of these transcriptomics studies stress the importance of CAFs, and immune cell infiltration, which both seem to play an important role in tumorigenesis and in clinical outcome of patients with NMIBC (18,20).

We performed several subgroup analyses to further explore the prognostic value of combinations of components of the tumor immune microenvironment. Interestingly, we found that FAP expression is prognostic for PFS in patients with the excluded and desert phenotype, but not in the inflamed phenotype. We hypothesize that FAP has an immune suppressive effect in the peritumoral stroma thus preventing immune cell infiltration into the tumor in both the excluded and desert phenotypes (21). Additionally, CD8-GZMB expression, which are activated cytotoxic T cells, only had prognostic value in the excluded phenotype. These patients have a better PFS compared to the other immune phenotypes, which could perhaps partially be explained by the importance of activated cytotoxic T cells, which can play a role in a more adequate response to BCG. GZMB itself has an important role in cytotoxic lymphocyte activity and can be used to identify activated cytotoxic T cells (22). However, activated cytotoxic T cells had a counterintuitive prognostic value in patients with high FAP expression. Patients with high FAP expression had a significantly impaired PFS if CD8-GZMB expression was higher. This could indicate that high FAP expression is able to suppress the effects of activated cytotoxic T cells. The role of CD8-GZMB becomes more interesting, as we also identified that patients with high CD8-GZMB expression and high PD-L1 expression, had a better PFS.

Lastly, we identified that FAP expression in T1-only was most prognostic in patients that had adequate BCG therapy, as FAP expression all ROIs did not impact PFS in patients with adequate BCG therapy. Based on these hypothesis generating subgroup analyses, we conclude that FAP expression, immune cell activation by CD8-GZMB with or without PD-L1 activation, and the combination with immune phenotypes seem to play a major role in clinical outcome of patients. FAP expression could potentially be linked to resistance to BCG, highlighting the importance of stromal components in the outcome of patients with high-risk NMIBC. The combination of FAP expression with CD8-GZMB expression, could be the reason that a large proportion of the inflamed tumors do not respond to BCG. The quality of CD8 expression (activation with CD8-GZMB), in combination with stromal components, could be key in determining response to BCG.

Our study is not devoid of limitations. Firstly, our cohort consisted initially of T1 disease-only but after single pathologists revision on the last cut H&E slide, 83% of patients (n=92) were confirmed to have T1 disease on new cut slides. Nonetheless, the non-T1 patients could be classified as high-risk NMIBC based on other disease characteristics. These patients were retained for analyses, as identifying differences in expression of IHC markers between ROIs (T1 vs. Ta vs. CIS) was one of the objectives of the current study. Secondly, after single pathologists revision we identified a significant difference in presence of CIS (p=0.003) between progressors and non-progressors. This higher yield of presence of CIS after revision can be attributed to several reasons: 1. inter-pathologist variability; 2. the scope of the current revision, with annotation of often multiple ROIs (T1 vs. Ta vs. CIS) within one patient; 3. the fact that T1 disease can actually be a mixed disease entity (with e.g. papillary disease or CIS), which is not differentiated anymore after the diagnosis of T1 disease. Additionally, CIS correlated with PFS which could have influenced our survival tree analysis. However, FAP expression was still more prognostic in our survival tree despite these differences in CIS, which stress the clinical relevance of FAP expression. Thirdly, we performed several subgroup analyses regarding components of the tumor immune microenvironment in smaller subgroups of patients, which could impair the clinical significance of these findings. However, we feel that these analyses are hypothesis generating, and are of value for additional research. Fourthly, we developed a survival tree, which are known to overfit datasets. To tackle this problem, we used a method of unbiased recursive partitioning which overfits less likely and is generally more reliable (23). Lastly, it is difficult to differentiate whether the identified

prognostic biomarkers are actually prognostic for progression, or either predictive for response to BCG, due to the fact that all patients in our cohort are treated with BCG. This issue could be tackled with a prospective study design, in which tissue of patients with T1 disease will be assessed before and after BCG treatment, and also in a non-BCG treated control arm (e.g. transurethral resection of T1 disease followed by early cystectomy). Such a study design would also enable use to identify shifts between immune phenotypes before and after BCG treatment, and their possible interactions with FAP.

We believe that our study has several strengths. First, we used full tissue slides which differs from the tissue microarrays (TMA), used in most studies in this field. TMAs have less tissue, less histological variation and offer no guarantee that T1 disease is still present. This is illustrated by the fact that only 83% of the initial T1 patients in our cohort still had presence of T1 disease on the last cut tissue slide, a percentage which will inevitably much lower with TMAs. Second, we have analysed IHC expression of several markers of the tumor immune microenvironment in a unique and detailistic approach in which we were able to quantify marker expression, and differentiate for each patient between tumoral, and peritumoral stromal expression, and additionally between T1, Ta, and CIS expression. Most studies in this field use TMAs which lack histological accuracy compared to full tissue slides, and do not dissect T1 disease as meticuliously in our approach. Third, we developed a survival tree which incorporated both FAP expression and presence of CIS. If prospectively validated, this survival tree is easy to use for clinical decision making. Lastly, we have identified FAP as a promising IHC marker, which will be easy to implement for pathologists in clinical practice, if validated in future studies. Therefore, we aim to develop a FAP scoring tool for pathologists that correlates with FAP expression as assessed by digital image analysis software, to facilitate clinical implementation. If validated and available, pathologists and clinicians should be able to easily stratify patients with high-risk NMIBC using a calculated FAP score and assessment of presence of CIS.

This study provides evidence for a prognostic value of FAP expression in high-risk NMIBC, and indirect evidence for a predictive value of FAP expression to BCG response. Immune cells solely were not prognostic, but immune phenotypes correlated with PFS in our cohort. It is known that CAFs can induce changes in the extracellular matrix of the peritumoral stroma to prevent migration of infiltrating leukocytes, and based on this study and FAP data in other tumor types, FAP might have such an immunosuppressive function

(24, 25). Future studies in larger and prospective patiënt cohorts are needed to validate FAP as a prognostic and predictive marker, to further establish the relation between FAP and immune cells and to verify our immunosuppressive hypothesis in NMIBC.

7.5 Acknowledgments

Steven Joniau is a senior clinical researcher of the Research Foundation of Flanders (FWO).

7.6 Supplementary data

Immune phenotypes: FAP

In a subgroup analysis of each immune phenotype, we assessed correlation of FAP expression with PFS. FAP expression in all ROIs correlated significantly with PFS in the excluded and desert immune phenotypes (**Figure S1**), but FAP expression in T1-only did only correlate significantly with PFS in the excluded immune phenotype (**Figure S2**).

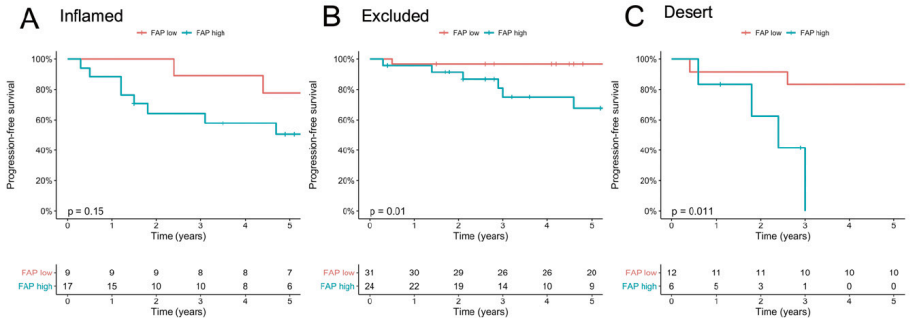


Figure S1: Progression-free survival stratified by FAP high vs. low (cut-off median) for each immune phenotype; **A**: inflamed (n=26); **B**: excluded (n=55); **C**: desert (n=18).

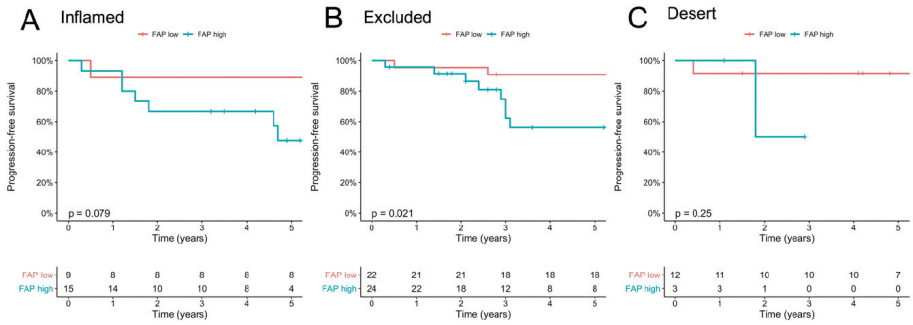


Figure S2: Progression-free survival stratified by FAP high vs. low (cut-off median) for each immune phenotype in T1 ROIs; **A**: inflamed (n=24); **B**: excluded (n=46); **C**: desert (n=15).

Immune phenotypes: CD8-GZMB

In a subgroup analysis of each immune phenotype, we assessed correlation of CD8-GZMB expression with PFS. CD8-GZMB expression correlated significantly with PFS in the excluded immune phenotype, but did not in the inflamed and desert immune phenotype (**Figure S3**).

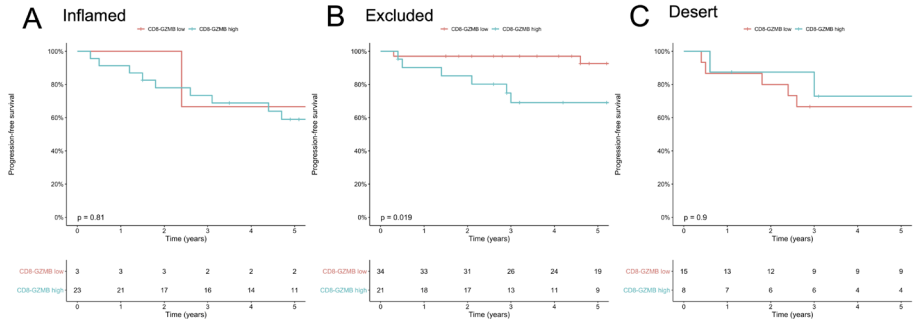


Figure S2: Progression-free survival stratified by CD8-GZMB high vs. low (cut-off median) for each immune phenotype; **A**: inflamed (n=26); **B**: excluded (n=55); **C**: desert (n=23).

FAP high vs. low: CD8 and CD8-GZMB

In a subgroup analysis of patients stratified by FAP expression (≤median vs. >median), we assessed correlation of both CD8 and CD8-GZMB expression with PFS. CD8 expression did not correlate significantly with PFS in the FAP low subgroup and FAP high subgroup. However, CD8-GZMB expression correlated

significantly with PFS in the FAP high subgroup, but not in the FAP low subgroup **(Figure S4)**.

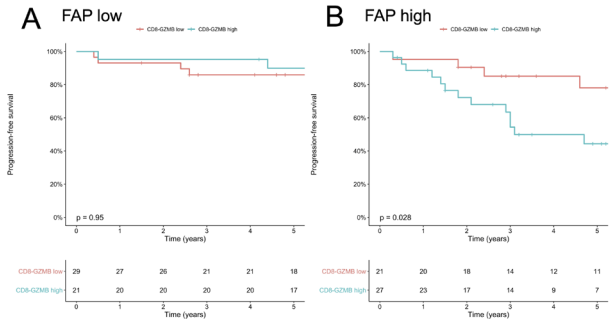


Figure S2: Progression-free survival stratified by CD8-GZMB high vs. low (cut-off median) for FAP low (≤median) and FAP high (>median); **A**: FAP low (n=50); **B**: FAP high (n=48).

Immune phenotypes with FAP high: CD8-GZMB

In a subgroup analysis of each immune phenotype separately but only in those patients with high FAP expression (>median), we assessed correlation of CD8-GZMB expression with PFS. CD8-GZMB expression did not correlate significantly with PFS in any of the immune phenotypes with high FAP expression while it did correlate with PFS in all patients in the aforementioned subgroup analysis.

CD8 high or CD8GZMB high: PD-L1

In a subgroup analysis of patients with high CD8 (>median) or with high CD8-GZMB (>median), we assessed correlation of PD-L1 SP142 IC score (in all ROIs and T1-only; positive vs. negative and ≤median vs. >median), and PD-L1 SP263 CPS (in all ROIs and T1-only; positive vs. negative and ≤median vs. >median) with PFS. PD-L1 scores did not correlate significantly with PFS in the CD8 high subgroup. However, in the CD8-GZMB high subgroup PD-L1 SP142 in T1-only and PD-L1 SP263 in T1-only correlated significantly with PFS, but not when assessed in all ROIs.

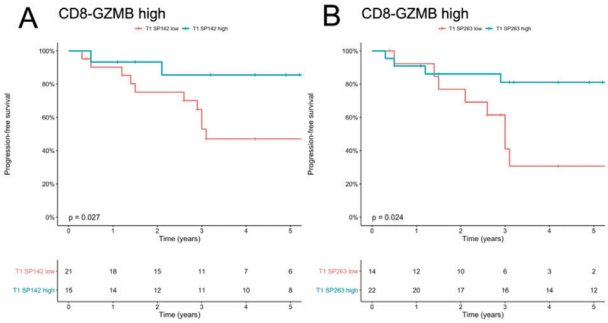


Figure S3: Progression-free survival stratified by PD-L1 in T1-only high vs. low (cut-off median) for CD8-GZMB (>median); **A**: PD-L1 SP142 score assessed in T1-only (n=36); **B**: PD-L1 SP263 score assessed in T1-only (n=36).

Adequate BCG: FAP

In a subgroup analysis of patients that had either adequate or inadequate BCG, we assessed correlation of FAP expression with PFS. Both FAP in all ROIs and FAP in T1-only correlated significantly with PFS in patients that had inadequate BCG (**Figure S4B and S4D**). However, in patients that had adequate BCG, FAP in T1-only correlated significantly with PFS (**Figure S4C**) while FAP in all ROIs did not correlate significantly (**Figure S4A**).

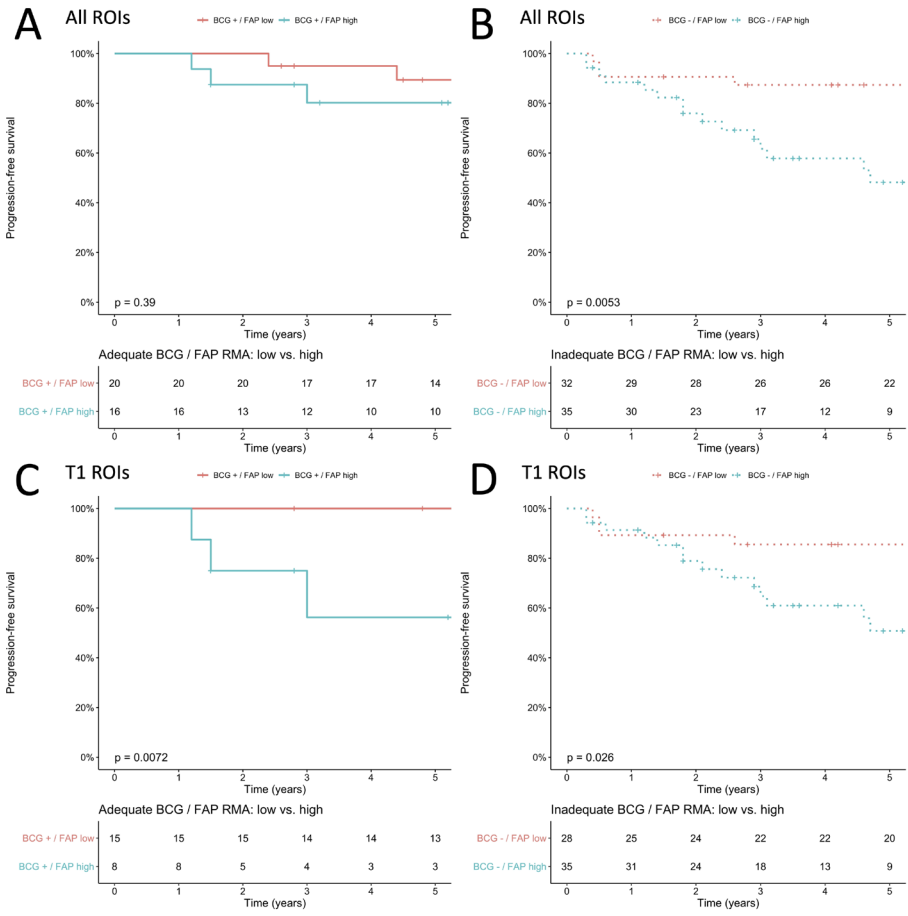


Figure S4: Progression-free survival stratified by inadequate vs. adequate BCG for each immune phenotype; **A**: Adequate BCG subgroup stratified by FAP (high vs. low) in all ROIs (n=36); **B**: Inadequate BCG subgroup stratified by FAP (high vs. low) in all ROIs (n=67); **C**: A: Adequate BCG subgroup stratified by FAP (high vs. low) in T1 ROIs (n=23); **D**: Inadequate BCG subgroup stratified by FAP (high vs. low) in T1 ROIs (n=63).

7.7 References

1. van den Bosch S, Alfred Witjes J. Long-term cancer-specific survival in patients with high-risk, non-muscle-invasive bladder cancer and tumour progression: a systematic review. Eur Urol. 2011 Sep;60(3):493–500.
2. Shariat SF, Karakiewicz PI, Palapattu GS, Lotan Y, Rogers CG, Amiel GE, et al. Outcomes of radical cystectomy for transitional cell carcinoma of the bladder: a contemporary series from the Bladder Cancer Research Consortium. J Urol. 2006 Dec;176(6 Pt 1):2414–22; discussion 2422.
3. Hautmann RE, de Petriconi RC, Pfeiffer C, Volkmer BG. Radical cystectomy for urothelial carcinoma of the bladder without neoadjuvant or adjuvant therapy: long-term results in 1100 patients. Eur Urol. 2012 May;61(5):1039–47.
4. Fernandez-Gomez J, Madero R, Solsona E, Unda M, Martinez-Piñeiro L, Gonzalez M, et al. Predicting nonmuscle invasive bladder cancer recurrence and progression in patients treated with bacillus Calmette-Guerin: the CUETO scoring model. The Journal of urology. 2009 Nov;182(5):2195–203.
5. Sylvester RJ, Rodríguez O, Hernández V, Turturica D, Bauerová L, Bruins HM, et al. European Association of Urology (EAU) Prognostic Factor Risk Groups for Non–muscle-invasive Bladder Cancer (NMIBC) Incorporating the WHO 2004/2016 and WHO 1973 Classification Systems for Grade: An Update from the EAU NMIBC Guidelines Panel. European Urology. 2021 Apr 1;79(4):480–8.
6. Muilwijk T, Akand M, Daelemans S, Marien K, Waumans Y, Kockx M, et al. Stromal marker fibroblast activation protein drives outcome in T1 non-muscle invasive bladder cancer. PLOS ONE. 2021 Sep 15;16(9):e0257195.
7. Calvete J, Larrinaga G, Errarte P, Martín AM, Dotor A, Esquinas C, et al. The coexpression of fibroblast activation protein (FAP) and basal-type markers (CK 5/6 and CD44) predicts prognosis in high-grade invasive urothelial carcinoma of the bladder. Hum Pathol. 2019 Sep;91:61–8.
8. Chen DS, Mellman I. Elements of cancer immunity and the cancer–immune set point. Nature. 2017 Jan 19;541(7637):321–30.
9. Binnewies M, Roberts EW, Kersten K, Chan V, Fearon DF, Merad M, et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. Nat Med. 2018 May;24(5):541–50.

10. Rubio C, Munera-Maravilla E, Lodewijk I, Suarez-Cabrera C, Karaivanova V, Ruiz-Palomares R, et al. Macrophage polarization as a novel weapon in conditioning tumor microenvironment for bladder cancer: can we turn demons into gods? Clin Transl Oncol. 2019 Apr;21(4):391–403.
11. Dadhania V, Zhang M, Zhang L, Bondaruk J, Majewski T, Siefker-Radtke A, et al. Meta-Analysis of the Luminal and Basal Subtypes of Bladder Cancer and the Identification of Signature Immunohistochemical Markers for Clinical Use. EBioMedicine. 2016 Oct;12:105–17.
12. Guo CC, Bondaruk J, Yao H, Wang Z, Zhang L, Lee S, et al. Assessment of Luminal and Basal Phenotypes in Bladder Cancer. Sci Rep. 2020 Jun 16;10(1):9743.
13. Echarti A, Hecht M, Büttner-Herold M, Haderlein M, Hartmann A, Fietkau R, et al. CD8+ and Regulatory T cells Differentiate Tumor Immune Phenotypes and Predict Survival in Locally Advanced Head and Neck Cancer. Cancers (Basel). 2019 Sep 19;11(9):1398.
14. Bessa A, Maclennan S, Enting D, Bryan R, Josephs D, Hughes S, et al. Consensus in Bladder Cancer Research Priorities Between Patients and Healthcare Professionals Using a Four-stage Modified Delphi Method. European Urology. 2019 Aug;76(2):258–9.
15. Mantovani A, Sozzani S, Locati M, Allavena P, Sica A. Macrophage polarization: tumor-associated macrophages as a paradigm for polarized M2 mononuclear phagocytes. Trends Immunol. 2002 Nov;23(11):549–55.
16. Lindskrog SV, Prip F, Lamy P, Taber A, Groeneveld CS, Birkenkamp-Demtröder K, et al. An integrated multi-omics analysis identifies prognostic molecular subtypes of non-muscle-invasive bladder cancer. Nat Commun. 2021 Apr 16;12(1):2301.
17. Budina A, Farahani SJ, Lal P, Nayak A. Subcategorization of T1 Bladder Cancer on Biopsy and Transurethral Resection Specimens for Predicting Progression. Archives of Pathology & Laboratory Medicine [Internet]. 2021 Dec 6 [cited 2021 Dec 13]; Available from: https://doi.org/10.5858/arpa.2021-0175-OA
18. Robertson AG, Groeneveld CS, Jordan B, Lin X, McLaughlin KA, Das A, et al. Identification of Differential Tumor Subtypes of T1 Bladder Cancer. European Urology. 2020 Oct 1;78(4):533–7.
19. Puré E, Blomberg R. Pro-tumorigenic roles of fibroblast activation protein in cancer: back to the basics. Oncogene. 2018 Aug 9;37(32):4343–57.
20. Hedegaard J, Lamy P, Nordentoft I, Algaba F, Høyer S, Ulhøi BP, et al. Comprehensive Transcriptional Analysis of Early-Stage Urothelial Carcinoma. Cancer Cell. 2016 Jul 11;30(1):27–42.

21. Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, et al. TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. Nature. 2018 Feb 14;554(7693):544.
22. Nowacki TM, Kuerten S, Zhang W, Shive CL, Kreher CR, Boehm BO, et al. Granzyme B production distinguishes recently activated CD8+ memory cells from resting memory cells. Cell Immunol. 2007 May;247(1):36–48.
23. Hothorn T, Hornik K, Zeileis A. Unbiased Recursive Partitioning: A Conditional Inference Framework. Journal of Computational and Graphical Statistics. 2006 Sep 1;15(3):651–74.
24. Kaur A, Ecker BL, Douglass SM, Kugel CH, Webster MR, Almeida FV, et al. Remodeling of the Collagen Matrix in Aging Skin Promotes Melanoma Metastasis and Affects Immune Cell Motility. Cancer Discov. 2019 Jan;9(1):64–81.
25. Redelman-Sidi G, Glickman MS, Bochner BH. The mechanism of action of BCG therapy for bladder cancer: a current perspective. Nature reviews Urology. 2014;11(3):153–62.

Chapter 8:

Discussion and Conclusion



Chapter 8:

Discussion and Conclusion

8.1 Discussion

Risk stratification is standard of care for patients with NMIBC, and is especially important in those patients at high-risk for disease progression or cancer-related mortality. Management of BC should thus become patient-tailored, and in line with clinical guidelines. In this thesis, I investigated the prognostic or predictive value of immunohistochemical expression of the tumor immune microenvironment and molecular subtypes in NMIBC. Moreover, we investigated guideline adherence in NMIBC, and risk stratification of patients with MIBC after radical cystectomy. The principal findings of my thesis are:

- Compliance to QCIs differed significantly between centers, and compliance to two QCIs (MMC in low-risk NMIBC, and BCG in high-risk NMIBC) correlated with clinical outcome of patients (**Chapter 3**).

- The AJCC staging system outperforms the COBRA score, and should therefore remain standard of care for risk stratification of patient with MIBC after radical cystectomy (**Chapter 4**).

- CK5 and presence of HG disease correlate inversely in patients with Ta NMIBC which suggest a different molecular biology of basal and luminal markers for NMIBC and MIBC (**Chapter 5**).

- FAP expression is strongly prognostic for PFS in a cohort of T1 NMIBC patients while luminal and basal IHC markers were not (**Chapter 6**).

- FAP expression and presence of CIS are the strongest drivers of PFS in cohort of high-risk NMIBC. Additionally, immune phenotypes correlate significantly with PFS, which underline the importance of tumor immune microenvironment in BC (**Chapter 7**).

We will discuss the main findings of **Chapter 3 to 7** in the following paragraphs, and end this chapter with concluding remarks and future perspectives.

8.1.1 Quality control indicators correlate with clinical outcome

Based on a previous review, we identified several promising QCIs in the management of NMIBC (64). However, most QCIs were not validated and there was no evidence to support them. Since the EAU Guidelines recommend

stratification of patients with NMIBC and tailor treatment to this stratification, it is important to assess whether urologists actually adhere to these recommendations. To accurately capture QCIs, we first checked the quality of reporting on these QCIs in our TURBT registry. We found that quality of reporting on some QCIs was suboptimal, which was attributed to not fully capturing all pathological variables and outcome data. Therefore, we introduced a new eCRF that is currently used in all participating hospitals. Interestingly, we identified significant differences in QCI compliance between centers, reflecting differences in clinical practice. In addition, compliance of two QCIs (postoperative SI of MMC and re-TURBT) increased significantly over time after implementation of the eCRF and initiation of data collection in the TURBT registry. Urologists are reminded of the possibility to administer a SI of MMC based on the clinical data they enter in the eCRF, which could explain an increase in this QCI. Participation in an TURBT registry creates opportunities to provide feedback to urologists and residents on their compliance to QCIs. However, we should avoid burdening urologists and residents with cumbersome administrative tasks. During the implementation of our eCRFs, we experienced that eCRFs should above all be easy to use, should not take up (too much) extra time and should only collect the most important information that is mandatory. The surgery report that is automatically generated after a TURBT procedure is an example of data collection that translates into immediate benefit for the user. Standardized collection makes surgery reports of patients within hospitals more uniform between urologists and enables an objective history taking and treatment plan. Moreover, adherence to the QCIs, postoperative instillation of MMC, and initiation of BCG induction therapy correlated with patients' RFS. If urologists find that their adherence to QCIs correlates with patients' clinical outcome, urologists become motivated to change their clinical practice.

8.1.2 The COBRA score is outperformed by the AJCC staging system

Patient with MIBC are at high risk for cancer-specific mortality after radical cystectomy. Adjuvant treatment options are currently not standard of care, but may be considered in the future, especially in those patients at greatest risk of progression. Therefore, we aimed to improve stratification of these patients by validating the COBRA score. Surprisingly, the COBRA score did not outperform the AJCC staging system. Based on our findings, we conclude that the AJCC staging system should remain the current standard for stratifying patients after radical cystectomy and used for clinical trial design. However, we still believe

that improving risk stratification in patients with high-risk MIBC is of importance as the 10-year CSS may drop to 23-31% for patients with T4 disease, and 14-25% for lymph node positive disease (13,14). Adjuvant treatment options will hopefully improve the outcome of these patients, especially in those patients at the highest risk of recurrence. Patient selection will become crucial to avoid overtreatment in this cohort of patients, as most are already frail before their treatment due to their age and significant comorbidities (29).

8.1.3 Luminal and basal markers have a different role in Ta NMIBC compared to MIBC

There is limited data on the relevance of luminal and basal IHC markers in NMIBC, especially in Ta NMIBC. Transcriptomic data have revealed prognostic and predictive molecular subtypes, with the basal or squamous molecular subtype being the most prominent in a consensus classification (65). Therefore, we assessed the expression patterns of different luminal (GATA3, CK20) and basal (CK5, P40, P63) IHC markers in a group of patients with Ta NMIBC. Interestingly, we identified an inverse correlation between CK5 expression and the percentage of HG disease. We confirmed with qPCR that CK5 positivity at IHC correlated with upregulation at RNA expression level. However, CK5 expression did not correlate with RFS in this cohort of patients. In addition, we identified a positive correlation between CK20 and the percentage of HG disease, and an inverse correlation between CK5 and CK20, which makes sense as both are opposite to the percentage of HG disease. Conversely, GATA3 was omnipresent in Ta disease and did not correlate with any clinicopathological characteristics or clinical outcome in this cohort, nor did P40 and P63. Our findings suggest a role for CK5 and CK20 as markers for urologists and pathologists to differentiate between HG and LG Ta NMIBC.

8.1.4 FAP expression correlates strongly with PFS in T1 NMIBC

Using a detailed approach in which we individually analyzed T1 regions within each patient, we identified FAP as strongly prognostic biomarker for clinical outcome in T1 NMIBC. Other clinicopathological features did not correlate with clinical outcome. Additionally, we also investigated luminal (GATA3) and basal (CK5) markers, as they have been suggested as surrogate for molecular subtypes (46,50). However, both GATA3 and CK5 did not differ between progressors and non-progressors and did not correlate with PFS. FAP is a

promising biomarker, and can be targeted for diagnosis and treatment. FAP-specific PET scans are being investigated, as is targeted radioligand therapy (66). In addition, multiple avenues are being explored to target FAP with drugs, such as peptide-drug complexes, FAP inhibitors and FAP antibodies. Thus, FAP could serve as an easily applicable prognostic biomarker for risk-stratifying patients with HG T1 NMIBC for whom accurate risk stratification is paramount, and could provide additional diagnostic and treatment options in the future.

8.1.5 FAP remains the strongest driver for PFS in an in depth analysis of the tumor immune microenvironment in high-risk NMIBC

In a more in-depth characterization of the tumor immune microenvironment, we confirmed the prognostic role of FAP in a cohort of high-risk NMIBC. Our approach to assess the tumor immune microenvironment was unique and detailed, especially compared to most studies using tissue micro arrays, which lack histological accuracy compared to full tissue slides, and do not dissect regions of T1, Ta and CIS as meticulously as our approach. In addition, we defined immune phenotypes that significantly correlated with PFS with a better prognosis for the excluded immune phenotype. In our survival tree analysis, FAP expression remained the main predictor of PFS. In addition, our subgroup analyses suggest an important role in the quality of CD8 expression. The activity of CD8 when co-expressed with GZMB had a favorable prognostic value in the excluded phenotype. Furthermore, patients with high CD8-GZMB expression and high PD-L1 expression had a better PFS compared to patients who did not. The combination of FAP expression with CD8-GZMB expression could be the reason that a large proportion of inflamed tumors do not respond to BCG. The quality of CD8 expression (activation with CD8-GZMB), in combination with stromal components, could be key in determining response to BCG.

8.2 Concluding remarks and future perspectives

This thesis provides an addition to the field of risk stratification of BC, which is of importance especially in those patients with BC at risk for disease progression and ultimately cancer-related mortality.

Our results stress the need for guideline adherence, as QCIs correlate with clinical outcome of patients. Ten Belgian centers are currently actively recruiting patients in our TURBT registry, and recording >1000 TURBT

procedures per year. Improvements that were inspired on this primary analysis of QCIs, have led to the possibility to automatically generate QCIs using live data from our EMR. It will be of interest to further investigate the effects of introducing a feedback loop for participating centers, and to assess its effect on compliance to QCIs.

The field in MIBC and metastatic BC is rapidly evolving with the advent of CPIs and other targeted therapies. Risk stratification will therefore become more important, and can help identify candidates for these adjuvant treatment options. We found that the COBRA score did not outperform the AJCC staging system. Staging itself is also currently changing with the advent of FDG-PET scans. FDG-PET performs more accurately compared to conventional imaging in this setting, which also directly correlates with patient survival, as well as influencing clinical decision-making during multidisciplinary team meetings (67,68). Therefore, clinical trials will play a critical role to define our management based on this more accurate imaging modalities, as tumor stage migration may occur, which may affect the clinical outcome of patients (69,70).

This thesis further refined our knowledge on the heterogeneity of the tumor immune microenvironment in NMIBC and its prognostic value. Our analyses regarding the tumor immune microenvironment were primarily based on IHC expression, which is readily available and has relatively low costs. However, IHC is not standardized worldwide, and quantifying results is often difficult. We used image analysis software to objectively quantify and compare IHC expression between patients but also between tumor regions within one patient, to fully capture tumor heterogeneity. Future research should investigate the effect of BCG therapy on FAP expression, to further differentiate whether FAP can be used as a predictive biomarker for BCG or a prognostic biomarker for PFS in general. In addition, immune phenotypes may potentially shift from an excluded state to an inflamed state after BCG therapy, potentially predicting response to therapy or prognosticating clinical outcome.

Assessment of tumors using the transcriptome, or more sophisticated multi-omics techniques, will become more and more available, and will further enhance our understanding of the tumor immune microenvironment. Examples of such novel techniques are single-cell transcriptome sequencing, which assesses expression of genes on individual cell level. As BC is a heterogeneous disease, this technique will provide additional information regarding differences and interplay between immune cells and cancer-associated

fibroblasts. Furthermore, single-cell spatial transcriptomics will allow us to combine information of histological sections with presence of cell types and their cell-specific transcriptome.

It is paramount that these advances translate into clinical trial design, which is the only way to validate their clinical use. The biomarker-driven multi-arm adaptive design of the BISCAY and ATLANTIS trials are examples of such clinical trial designs with the goal of delivering personalized medicine (71,72). Although the BISCAY trial was negative, it provided us with information that such a study design is feasible. For example, a FAP biomarker-driven, or immune phenotype biomarker-driven trial design could become a reality, if the results of this thesis are validated. Using IHC or transcriptomic analysis as patient selection tools, these trials may enable true patient-tailored medicine, which can be based on objective risk stratification.

At the end of this thesis, I would like to make some final remarks on the future of the management of BC in general. We are currently at an exciting point with multiple new agents becoming available for patients with NMIBC and MIBC. However, as a urological community, we need to be critical when reviewing approvals for these agents. For example, control arms in the BCG unresponsive setting should be adequate, and should not include substandard treatment options for these patients, such as intravesical chemotherapy. The most optimal control arm would of course be a radical cystectomy in most of these cases, however most pharmaceutical companies will not be interested in comparing their therapy with such a control arm. Therefore, as urologists, we should be motivated to design trials in both the BCG unresponsive NMIBC, MIBC, and also metastatic setting, including surgical experimental or control arms such as cystectomy or metastasis-directed therapy. It has been helpful to the pharmaceutical industry that there were multiple BCG shortages in the recent past, with little to no interest in stepping up production of BCG. New drugs are more interesting for the pharmaceutical industry because they can potentially generate more revenue. We must therefore remain vigilant that expansive and newly approved agents do not replace current treatment options due to availability issues, with comparable efficacy. Additionally, side effects of new drugs must also be critically assessed and weighed against the alternatives. Most patients with MIBC or metastatic BC are of advanced age and have significant comorbidities. New treatments for patients with NMIBC should not have too many serious adverse events with risk of death, and should be compared with the mortality rate after cystectomy. Based on all these findings,

it will be important to consider the patient as a whole and discuss pros and cons of these treatment options with our patients.

Taking these considerations into account, I believe that we as urologists will play an important role in the future. We will be the physicians who will educate patients about these new available options, and perhaps even on biomarker-driven treatments. It will therefore be important to always discuss these decisions in multidisciplinary teams. As urologists, we have a responsibility to deliver the best care to our patients, and I believe that this can only be done in a multidisciplinary way.

Chapter 9:

Scientific Summary



Chapter 9:

Scientific Summary

9.1 English summary

Patients with bladder cancer (BC) are at risk for disease progression which encompasses significant treatment-related and cancer-specific morbidity and mortality. Therefore, stratification of patients is important to identify patients that are at risk for progression of disease, and for cancer-related mortality. However, current risk stratification lacks accuracy to identify which patients will progress and ultimately succumb to their disease. This thesis provides an addition to risk stratification of patients with non-muscle invasive BC (NMIBC) and muscle invasive BC (MIBC). Furthermore, we assessed whether hospitals adhered to risk stratification of patients, and whether this adherence correlated with clinical outcome.

To assess adherence to risk stratification, we used a large prospective registry of transurethral resection of bladder tumors in three hospitals (**Chapter 3**). Using this registry, we were able to successfully assess six quality control indicators (QCI) in the management of NMIBC. Interestingly, compliance to these QCIs differed significantly between hospitals, and compliance to two QCIs correlated with clinical outcome of patients. These QCIs both directly translate to risk stratification of patients as they were a single postoperative instillation of Mitomycin C in low-risk NMIBC, and start of Bacillus Calmette-Guerin (BCG) induction in high-risk NMIBC.

Patients with MIBC are treated with a radical cystectomy, and are afterwards at high risk for disease recurrence and cancer-related mortality. Therefore, we validated the Cancer of the Bladder Risk Assessment (COBRA) score which predicts cancer-specific mortality after radical cystectomy. However, the AJCC staging system outperformed the COBRA score, and should therefore remain standard of care for risk stratification in this setting (**Chapter 4**).

Molecular subtypes and the tumor immune microenvironment are potential prognostic and predictive biomarkers for NMIBC. In Ta NMIBC we identified immunohistochemical (IHC) markers CK5 and CK20 useful to differentiate between low and high-grade disease which suggested a different molecular biology of basal and luminal markers for NMIBC and MIBC (**Chapter 5**). Next, we assessed several IHC markers for molecular subtypes (basal vs. luminal), and the tumor immune microenvironment in T1 NMIBC. This was done meticulously with digital image analysis software to fully capture the heterogeneity of NMIBC. In T1 NMIBC we identified fibroblast activation protein-α (FAP)

expression as a strong prognostic marker for progression. FAP is a marker for cancer-associated fibroblasts, which play an important role in the tumor microenvironment. However, luminal and basal IHC markers were not prognostic (**Chapter 6**). Additionally, we identified FAP expression and presence of CIS as the strongest drivers of progression in cohort of high-risk NMIBC. Interestingly, immune phenotypes correlated significantly with progression in this cohort (**Chapter 7**).

To conclude, this thesis provides additional information to stratify patients with BC, and confirms the important role of tumor immune microenvironment. Researchers should focus on implementing important advances that are being made in the field of BC, by designing biomarker-driven randomized controlled trials. This way, personalized medicine will move forward, and stratification of patients can really occur based on each patients clinicopathological and molecular characteristics.

9.2 Nederlandse samenvatting

Patiënten met blaaskanker hebben een groot risico op ziekteprogressie, dat voor hen een significante behandelingsgerelateerde en kankerspecifieke morbiditeit en mortaliteit inhoudt. Risicostratificatie is daarom belangrijk, om juist de patiënten te identificeren die het hoogste risico lopen op ziekteprogressie en kankergerelateerde mortaliteit. De huidige risicostratificatie is echter te onnauwkeurig om te kunnen bepalen welke patiënten progressie van hun ziekte zullen hebben en hier uiteindelijk aan zullen bezwijken. Dit proefschrift biedt een aanvulling op de risicostratificatie van patiënten met niet-spierinvasieve blaaskanker (NMIBC) en spierinvasieve blaaskanker (MIBC). Daarnaast hebben we onderzocht of ziekenhuizen zich aan de risicostratificatie van patiënten hielden, en hebben dit gecorreleerd met de klinische uitkomst van patiënten.

Om de naleving van het gebruik van risicostratificatie in de behandeling van patiënten te beoordelen, hebben we een groot prospectief register van transurethrale resectie van blaastumoren in drie ziekenhuizen gebruikt (**Hoofdstuk 3**). Met dit register waren we in staat om zes kwaliteitsindicatoren (QCI) in de behandeling van NMIBC te beoordelen. De naleving van deze QCI's verschilde significant tussen ziekenhuizen, en de naleving van twee QCI's correleerde met de klinische uitkomst van patiënten. Deze QCI's waren beide direct gerelateerd aan de risicostratificatie van patiënten: een eenmalige

postoperatieve instillatie van mitomycine C bij laag-risico NMIBC en het opstarten van de inductie van Bacillus Calmette-Guerin (BCG) bij hoog-risico NMIBC.

Patiënten met MIBC worden behandeld met een radicale cystectomie en hebben daarna een grote kans op terugkeer van de ziekte en kankergerelateerde mortaliteit. Daarom hebben we de Cancer of the Bladder Risk Assessment (COBRA)-score extern gevalideerd. Het AJCC-stadiëringssysteem presteerde echter beter dan de COBRA-score en zou daarom de standaard moeten blijven voor risicostratificatie van patiënten met MIBC na radicale cystectomie **(Hoofdstuk 4**).

Moleculaire subtypes en de tumor immuun micro-omgeving zijn potentiële prognostische en predictieve biomarkers voor NMIBC. In Ta NMBC hebben we de immunohistochemische (IHC) markers CK5 en CK20 gebruikt om te kunnen differentiëren tussen laag- en hooggradige ziekte en ontdekten we een andere moleculaire biologie van deze basale en luminale markers in NMIBC vergeleken met MIBC (**Hoofdstuk 5**). In T1 NMIBC hebben we vervolgens verschillende IHC markers bekeken voor moleculaire subtypes (basaal vs. luminaal) en de tumor immuun micro-omgeving. Dit werd minutieus gedaan met digitale beeldanalysesoftware om de heterogeniteit van NMIBC volledig vast te leggen. In T1 NMIBC identificeerden we fibroblast activatie eiwit-α (FAP) expressie als een sterke prognostische marker voor progressie. FAP is een marker voor met kanker geassocieerde fibroblasten, die een belangrijke rol spelen in de tumor micro-omgeving. Luminale en basale IHC markers waren echter niet prognostisch (**Hoofdstuk 6**). Bovendien identificeerden we FAP-expressie en aanwezigheid van CIS als de sterkste aanjagers van progressie in cohort van hoog-risico NMIBC. Interessant is dat immuunfenotypes ook significant correleerden met progressie in deze cohort patiënten (**Hoofdstuk 7**).

Concluderend biedt dit proefschrift aanvullende informatie om patiënten met blaaskanker te stratificeren en bevestigt het de belangrijke rol van de tumor immuun micro-omgeving. Klinische toepassing van de grote ontwikkelingen die worden gemaakt op het gebied van BC is van uiterst belang. Dit kan enkel worden gerealiseerd met biomarker-gestuurde gerandomiseerde gecontroleerde onderzoeken. Alleen op deze manier zal gepersonaliseerde geneeskunde stappen vooruit kunnen zetten en kan de risicostratificatie echt plaatsvinden op patiënt-niveau door gebruik te maken van patiënt-specifieke klinisch-pathologische en moleculaire eigenschappen.

Chapter 10:
Final statements



Chapter 10:

Final statements

10.1 Scientific acknowledgement and Personal Contribution

This thesis was written by Tim Muilwijk and corrected by Steven Joniau, Thomas Gevaert, and Maarten Albersen.

We thank Murat Akand as co-first author of **Chapter 4 and 5**.

We thank the urologists and urology residents working at University Hospitals Leuven, AZ Sint-Blasius, and AZ Groeninge for their willingness for filling in eCRFs after each TURBT procedure at their institution, which were used in **Chapter 3**.

We thank Annouschka Laenen (statistician, Leuven Biostatistics and Statistical Bioinformatics Centre, KU Leuven) for her statistical advice regarding the analyses in **Chapter 3, 4, 6, and 7**.

We thank Rieta Van Bree (lab assistant, unit Woman and Child, KU Leuven) for her technical assistance with the qPCR experiments of **Chapter 5**.

We thank Mark Kockx, and CellCarta (Antwerp, Belgium) for performing all IHC stains and for the opportunity to use the digital image analysis software Visiopharm for **Chapter 6 and 7**.

We thank Sofie Daelemans, Koen Marien, Yannick Waumans, and Mark Kockx (CellCarta, Antwerp, Belgium) for their help and advice on using the image analysis software (Visiopharm) for the analyses of **Chapter 6 and 7**. We also thank CellCarta for the validated IHC stains, and the development of applications within the Visiopharm software which were used and adopted for further analyses.

10.2 Conflict of Interest

Sofie Daelemans, Koen Marien, Yannick Waumans, and Mark Kockx are employees of CellCarta.

The remaining authors declare no conflict of interest.