Molecular subtypes and single tissue and serum biomarkers for the prediction of platinum sensitivity of muscle-invasive bladder cancer

PhD thesis

Csilla Oláh

Rácz Károly Doctoral School of Clinical Medicine

Semmelweis University

Supervisor: Tibor Szarvas, DSc, Senior researcher

Official reviewers:

Donát Alpár, PhD, Senior researcher

Levente Kuthi, MD, PhD, University lecturer

Head of the Complex Examination Committee:

Csaba Bödör, DSc, University professor

Members of the Examination Committee:

Tamás Beöthe, MD, PhD, Chief physician
Gábor Szabó, MD, PhD, Chief physician

Budapest, 2024

1. Introduction

Bladder cancer (BC) is one of the most common malignancies worldwide, with 550 000 new cases and 200 000 deaths yearly. Approximately 75-80% of newly diagnosed BC cases are non-muscle invasive (≤T1) at first presentation. These patients have a high recurrence rate (60-70%) but a low progression rate to muscle-invasive stage (15%) and a good prognosis with over 90% 5-year survival. The incidence of muscle-invasive bladder cancer (MIBC) represents 20-25% at the first presentation but the 5-year survival rate is considerably lower (50-60%). For patients with muscle-invasive and clinically organ-confined tumors (≥cT2, LN0-LNx, M0) cisplatin-based neoadjuvant chemotherapy (NAC) is the gold standard prior to radical cystectomy (RC). NAC provided a modest 5-8% overall survival (OS) benefit at 5 years compared to RC alone. On the other hand, the delay of RC due to NAC could potentially reduce the life expectancy of patients with chemotherapy-resistant tumors. Therefore, accurately predicting platinum sensitivity before treatment initiation may help to identify patients who would benefit from NAC, preventing ineffective platinum treatment and the delay of RC in resistant patients. Patients with locally advanced tumors (pT3/4) and/or positive lymph nodes (LN+) at radical cystectomy (RC) are recommended to undergo adjuvant chemotherapy (AC) if no neoadjuvant chemotherapy (NAC) was administered preoperatively. Although guidelines recommending NAC, prior RC and subsequent adjuvant platinum treatment remained widespread in clinical practice. Recently, new therapy options became available mostly in second- and third-line settings, such as immune checkpoint inhibitors or the FGFR inhibitor erdafitinib and two antibody-drug conjugates; the nectin-4 targeting Enfortumab vedotin and the TROP-2 targeting Sacituzumab govitecan. The availability of potentially effective second- and third-line treatments has increased the need for chemotherapy prediction in order to improve therapeutic decision-making. Despite numerous attempts to identify chemotherapy predictive markers, none of the identified and investigated markers have been integrated into the clinical routine yet.

Recently, the continuous development of molecular biological methods, such as high-throughput sequencing and gene expression analysis provided a comprehensive insight into the molecular background of MIBC and revealed different molecular patterns among histologically similar urothelial tumors. Numerous studies have performed mRNA sequencing or utilized gene expression chip analyses, revealing that MIBCs can be

stratified into diverse molecular subtypes through cluster analysis of gene expression patterns. Subsequent studies described that the distinct molecular subtypes are associated with different prognoses and sensitivities to chemotherapy and immunotherapy. However, these studies used often different molecular classification systems with only partly overlapping nomenclature, thus a direct comparison of results between different studies is difficult. Therefore, a consensus classification has been suggested following the reanalysis of 1750 previously published transcriptome profiles of MIBCs. This comprehensive study differentiated six molecular subtypes: (1) luminal papillary, (2) luminal non-specified, (3) luminal unstable, (4) stroma-rich, (5) basal, and (6) neuronal-like. Additionally, the study confirmed a more favorable prognosis for the luminal papillary subtype and a poor prognosis for the neuronal subtype. Notably, the authors did not observe any platinum-predictive values for any of the six subtypes, and suggested that the luminal non specified, luminal, and neuronal-like subtypes might benefit from immune checkpoint inhibitor therapy.

The published classifier methods are based on the analysis of thousands of genes, thus the molecular classification could not widespread into the clinical routine yet. To overcome these methodological difficulties, we aimed to develop a simple and cost-effective molecular subtype classifier method, which can be performed on the pathological routine collected formalin-fixed and paraffin-embedded (FFPE) tissue samples and can reproduce the most relevant molecular subtype classification systems.

In addition to molecular subtypes, we aimed to identify additional tissue and serum markers for the prediction of platinum-treatment. For this, we selected potential platinum-markers from the literature and own preliminary research and measured the tissue gene expressions of 12 genes (*CLDN4, ERCC1, BIRC5, HMGA2, MKI67, APOBEC3A, APOBEC3B, APOBEC3G, CDK12, BSG, MMP7, TOP2A),* the tissue protein expression of four gene (*CLDN4, ERCC1, MMP7, SDC1)* and the pretreatment serum concentrations of SDC1 and MMP7.

2. Objectives

2.1 The aim of the present retrospective study was:

1) To develop a cost-effective and simple analytical method for the reproduction of the most relevant gene expression-based molecular subtype classification systems.
2) To assess the prognostic and platinum-predictive value of various molecular subtypes in our own institutional BC patient cohorts.
3) To assess the platinum-predictive value of 12 single genes at the mRNA as well as two genes at the protein level in our own institutional BC patient cohorts.
4) To assess the platinum-predictive value of SDC1 and MMP7 in tissue and serum samples of our institutional BC patient cohorts.

2.2 Required steps / overview of the research

1) Definition of a reduced gene set to differentiate between molecular BC subtypes.
2) *In silico* development of classifier methods for each molecular classification systems (MDA, LundTax, TCGA, and Consensus) using the reduced marker set.
3) *In silico* validation of the newly developed classifiers on publicly available datasets.
4) Application of the above developed marker set and classifier methods to our own RC-treated institutional BC cohorts. For this, we determined the expression of 68 genes in 100 frozen BC tissue samples by using the TaqMan array card quantitative polymerase chain reaction-based method (qPCR).
5) Application of the above developed marker set and classifier methods to our own institutional cohort of patients with pT3/4 or LN-positive BC cases, who did vs. did not receive adjuvant chemotherapy. For this, we determined the mRNA expression of 48 genes in 160 FFPE tissue samples by the NanoString nCounter method.
6) In addition, to subtype-specific genes, we also investigated 12 single genes with potential platinum-predictive value.
7) Validation of the chemotherapy predictive value of the selected single genes in an independent data set.
8) Analysis of the chemotherapy predictive value of SDC1 and MMP7 in serum samples by ELISA method and in FFPE tissue samples by immunohistochemistry.

3. Methods

3.1 Molecular subtype classification

To investigate the prognostic and predictive values of molecular subtypes, we classified two different MIBC cohorts into molecular subtypes based on the gene expression patterns of the preselected reduced gene sets with 68 (RT-qPCR) and 48 (NanoString) genes. The gene expression analyses were performed by TaqMan Gene Expression Assay using the 364-well TaqMan Array Card platform on QuantStudio™ 7 Flex Real-Time PCR System (Applied Biosystems, Life Technologies, Thermo Fisher Scientific, Waltham, MA, USA) and by NanoString nCounter Analysis System (NanoString Technologies, Seattle, WA, USA). The first cohort included 100 fresh-frozen MIBC samples, while the second cohort included 160 FFPE MIBC samples in approximately 1:1 ratio with adjuvant chemotherapy-treated *vs.* non-treated patients.

We developed an *in silico* classifier method, tailored to the reduced gene set using a step-by-step rule set-based classification approach. The method is able to determine the different molecular subtypes according to various classification systems within a properly sized patient cohort. The step-by-step classifier was established on publicly available transcriptome datasets and was optimized until the highest overlap has been reached between our rule set-based and the original transcriptome-based classifications. Then, the so optimized rule set-based classifiers were validated on independent publicly available gene expression datasets with available transcriptome-based subtype class information. We established rule set-based classifiers for the TCGA, MDA, LundTax, and Consensus classification systems, and identified molecular subtypes in our institutional cohorts. Then, the so defined subtypes were correlated with clinicopathological variables and OS data.

3.2 Gene expression analyses of 12 single marker candidates

The mRNA expression of 12 additional single genes was measured by the NanoString nCounter technology in 160 FFPE samples of RC-treated MIBC patients with histological findings of pT3-4 or LN+. These 12 potential chemotherapy-predictive genes were selected (1) based on literature data, (2) former results of our research group in a chemotherapy-treated MIBC patient cohort, and (3) functional analysis with cisplatin resistant BC cell lines. Gene expression results were then correlated with

clinicopathological variables and OS data. The most promising genes were further investigated in an external MIBC transcriptome dataset (received from the Lund University) including NAC-treated and non-treated MIBC samples (n=125 and n=161).

3.3 Immunohistochemical analysis of CLDN4 and ERCC1 proteins

Immunohistochemical (IHC) analyses were performed with two potential chemotherapy predictive markers in a postoperatively chemotherapy-treated MIBC cohort. For CLDN4 immunostaining a mouse monoclonal antibody (clone 3e2c1, dilution 1:1000, Invitrogen, Waltham, Massachusetts, USA) was used. CLDN4 staining intensity was scored as 1, 2, or 3, equivalent to negative, moderate, and strong intensities. High CLDN4 expression was considered as intensity ≥1.

To perform ERCC1 staining a mouse monoclonal antibody (Mob 336-05, clone: 8F1, dilution 1:100, Diagnostic Biosystems, CA, USA) was used. ERCC1 staining intensity was scored as 1, 2, or 3, equivalent to negative, moderate, and strong intensities, while a percentage score was defined as 0–10%—0 Pts., 11–50%—1 Pts., 51–80% —2 Pts., and 81–100%—3 Pts.. Then, an IHC-score was calculated by multiplying the intensity and percentage scores. High ERCC1 expression was considered as IHC-score >4.
The protein expression results were correlated with OS data.

3.4 Serum ELISA analyses of SDC1 and MMP7

Former studies suggested SDC1 and MMP7 to be involved in resistance to chemotherapy in non-urological cancers. Therefore, SDC1 and MMP7 serum levels were determined in baseline samples from 52 patients treated with postoperative chemotherapy using Enzyme-Linked Immunosorbent Assay (ELISA) (SDC1 ELISA kit: Diaclone CD138, Gene-Probe, San Diego, CA, USA; Cat. Nr.: 950.640.096) (MMP7 ELISA kit: Quantikine ELISA kit from R&D Systems, Wiesbaden, Germany; Cat. Nr.: DMP700). The cut-off value of SDC1 for dichotomization was set at the upper 25th percentile (180 ng/mL), while of MMP7 for dichotomization was set at the median.

3.5 Immunohistochemical analysis of SDC1 and MMP7 proteins

IHC evaluations of SCD1 (SDC1/CD138 (clone MI15, dilution 1:100, Dako/Agilent, Santa Clara, CA, USA) and MMP7 (JL07, dilution 1:75, Santa Cruz Biotechnology, Dallas, Texas, USA) were performed in 72 FFPE tissue samples of patients who

underwent later postoperative chemotherapy. SDC1 and MMP7 staining intensities were scored as 1, 2, or 3, equivalent to negative, moderate, and strong intensities. A percentage score was also defined as 0–10%—0 Pts., 11–20%—1 Pt., 21–30%—2 Pts., 31–40%—3 Pts., 41–50%—4 Pts., 51–60%—5 Pts., 61–70%—6 Pts., 71–80%—7 Pts., 81–90%—8 Pts., 91–100%—9 Pts. Then, an IHC-score was calculated by multiplying the intensity score and percentage score. Weak SDC1 expression was considered as a score <4, moderate ≥4 and <10, and strong expression was considered as a score ≥10. SDC1 expression was evaluated separately for cell membrane, cytoplasm, and stroma. High MMP7 expression was considered as IHC-score >3. The serum levels and protein expression results of SDC1 and MMP7 were correlated with OS data.

4. Results

4.1 Molecular subtype analysis

We developed a classifier method using a reduced gene set and a step-by-step rule set for the identification of molecular subtypes. The rule set was established *in silico* on the original transcriptome-based datasets and was then validated on independent sample cohorts. Using our classification method, we were able to classify samples into the TCGA, MDA, LundTax, and Consensus subtypes from the original studies with 78%, 81%, 67% and 75% accuracy. Then, we applied the newly developed subtype classification method to our own institutional cohort of 100 frozen RC MIBC tissue samples (Figure 1). The OS analysis revealed that patients with neuronal subtype tumors have poor OS (p=0.002). However, in contrast to the TCGA study; we could not confirm the favorable prognosis of luminal-papillary tumors.

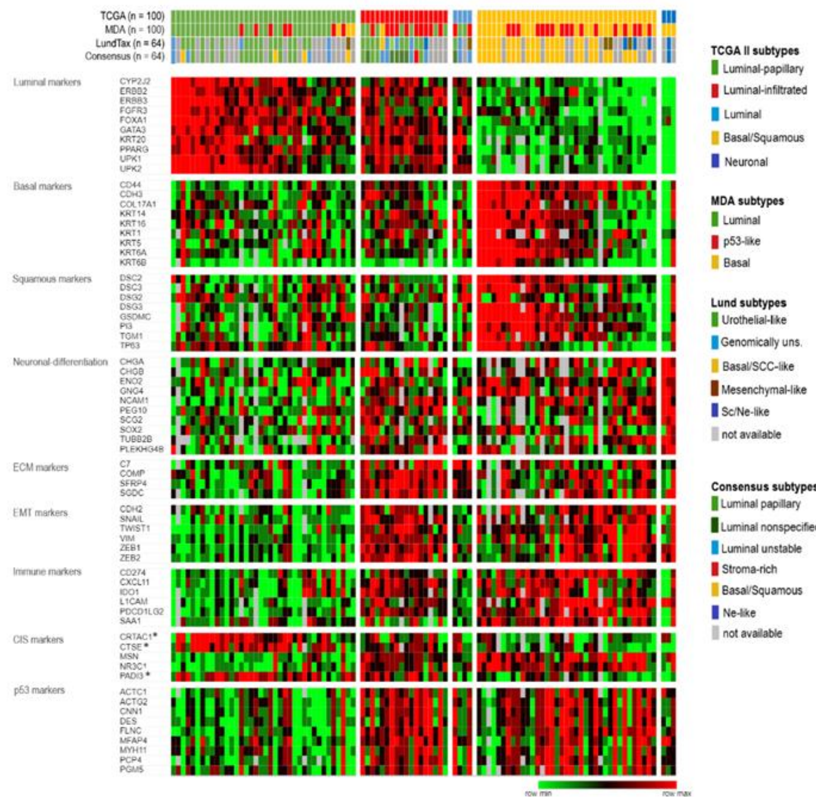

Figure 1. Molecular subtypes determined by rule set-based classifiers and their distinct gene expression profile in our institutional cohort visualized by heatmap. Genes that are downregulated* in CIS. CIS: carcinoma in situ. MDA: MD Anderson, TCGA: The Cancer Genome Atlas. Figure 1 has been published in the article: *Olah C, Hahnen C, Nagy N, et al. A RT-qPCR based method for molecular subtype classification of urinary bladder cancer-stromal gene expressions show higher prognostic values than intrinsic tumor genes. Int J Cancer. 2022;150(5):856-867. doi:10.1002/ijc.33809*

As a second step, we further optimized the marker set and reduced it from 68 to 48 genes. Some genes were excluded based on the experiences of the first study, and in addition, some neuronal genes with lower specificity were removed and substituted by other neuronal genes. The above detailed 48-gene set was applied to the NanoString nCounter method, which is also applicable for low amount and quality RNA samples, which is frequently true for RNA samples purified from FFPE tumor material. The TCGA, MDA, and Consensus classifiers by the 48 gene set reached similar overlaps with the original transcriptome-based classifiers as our formerly used 68 gene set (TCGA: 78 *vs*. 76%, MDA: 81 *vs*. 82%, Consensus: 75 *vs*. 74%), while the LundTax classifier reached higher overlap by the 48-gene marker set (67 *vs*. 73%).

We determined the expression values of the 48 genes in our institutional cohort of 160 FFPE MIBC samples from patients present with an indication for adjuvant chemotherapy (pT3-4 and/or LN+ finding at RC) and identified the molecular subtypes. The cohort included 81 adjuvant-chemotherapy treated patients, and 79 non-treated patients (who refused or were ineligible to platinum treatment). Our results showed that tumors with luminal-papillary subtypes (according to TCGA and Consensus classifications; p=0.036 and 0.009, respectively) and urothelial-like subtype (according to LundTax classification; p=0.001) have benefit from adjuvant chemotherapy by significant longer OS, while in contrast basal tumors had similar OS rates in the chemo and non-chemo groups. These suggest that tumors with luminal subtypes rather than basal subtypes are benefiting from adjuvant platinum therapy.

4.2 mRNA expression of single genes for chemotherapy prediction

We selected 12 additional potentially platinum-predictive genes based on own, yet unpublished research (*APOBEC3A, APOBEC3B, APOBEC3G, TOP2A, BSG, MMP7*) as well as on literature data (*BIRC5, CDK12, CLDN4, ERCC1, HMGA2,* and *MKI67*). The gene expressions were measured by the NanoString method in our institutional cohort of 160 FFPE MIBC samples (n=81 chemotherapy-treated, n=79 non-treated). The OS rate was directly compared between chemotherapy-treated and untreated patient cohorts within marker low and high subgroups. Based on the results, the 12 markers can be divided into three groups: 1) favorable factors with improved OS in the chemotherapy cohort (*APOBEC3G; p=0.002, CLDN4; p=0.004, ERCC1; p=0.003*) , 2) risk-factors

with poor OS in the chemotherapy cohort (*BIRC5; p=0.007, HMGA2; p=0.005, MKI67; p=0.002*) , 3) factors with no association with the outcomes in either cohort (*APOBEC3A, APOBEC3B, CDK12, BSG, MMP7, TOP2A*).

Next, we aimed to validate the chemotherapy predictive value of the above identified three protective and three risk genes on an independent publicly available transcriptome dataset from patients who received NAC (n=125) or upfront RC without chemotherapy (n=161). We could confirm the chemotherapy-predictive values of three of the six genes. The high gene expression values of *APOBEC3G* and *CLDN4* identified patients with improved OS in the NAC (p=0.026 and p=0.025, respectively), but not in the upfront RC cohort (p=0.576 and p=0.493, respectively), which confirmed the chemotherapy predictive value of these two genes. In addition, the high *APOBEC3G* expression was significantly correlated with higher pathological complete response rate in the validation cohort (p=0.028). The low gene expression value of *BIRC5* identified patients with favorable OS after chemotherapy in the Lund validation cohort (p=0.032). High *ERCC1* and low *HMGA2* and *MKI67* gene expression levels were associated with improved OS in our institutional cohort. Additionally, high *HMGA2* and *MKI67* gene expression values tended to be associated with worse OS in the NAC-treated cohort, but these associations did not reach the significance level (p=0.059 and p=0.345, respectively).

4.3 Tissue CLDN4 and ERCC1 levels for chemotherapy prediction

The predictive value of CLDN4 could also be confirmed by IHC in an independent institutional FFPE cohort treated with chemotherapy (n=88). The protein expression of CLDN4 was consistent with the gene expression results, indicating that high CLDN4 expression is associated with enhanced OS in chemotherapy-treated patients (p=0.017). On the other hand, the results of ERCC1 protein expression were inconsistent with the gene expression results, as high ERCC1 protein expression tended to associate with shorter OS (n=75, p=0.061). This pattern aligns with the results observed in the validation dataset at the gene expression level.

4.4 Serum SDC1 and MMP7 levels for chemotherapy prediction

We examined the baseline serum levels of SDC1 and MMP7 in 52 chemotherapy-treated MIBC patients. High serum SDC1 and MMP7 levels were associated with shorter OS (p=0.004 and p=0.033, respectively).

4.5 Tissue SDC1 and MMP7 levels for chemotherapy prediction

We evaluated the protein expression of SDC1 and MMP7 in FFPE tissue samples of patients who underwent postoperative chemotherapy (n=72). Of note, there was no overlap between the cohorts with available serum and tissue samples, therefore a direct comparison between tissue IHC and serum ELISA results was not possible. MMP7 staining was localized in the cytoplasm of tumor cells and often showed higher expression at the tumor-normal interface. High MMP7 tissue expression was also associated with worse outcomes (p=0.017). Consequently, the protein expression results confirmed the findings at gene expression level and underscored the chemotherapy-predictive value of MMP7. We assessed the membrane, cytoplasmic and stromal SDC1 protein expression separately, but found no significant difference in patient outcomes between high and low SDC1 staining in either compartment (membranous SDC1: p=0.918, cytoplasmic SDC1: p=0.802, stromal SDC1: p=0.452).

5. Conclusions

1. We developed a molecular subtype classification method using a reduced gene set that was able to accurately reproduce the most common mRNA-based classification systems (TCGA, MDA, LundTax, Consensus).
2. In a fresh-frozen patient cohort with 100 MIBC samples, we found that the neuronal and luminal subtypes had the worst prognosis.
3. In a FFPE MIBC cohort (all pT3/4 or LN+ at RC) with 81 chemotherapy-treated and 79 non-treated patients, we observed improved OS for platinum-treated compared to untreated patients in the luminal-papillary (according to TCGA and Consensus classifiers), and urothelial-like subtypes (according to LundTax classifier), while no OS difference was observed in the basal subtype groups.
4. In the same patient cohort, we assessed the gene expression levels of 12 single markers with platinum predictive value and the genes with confirmed predictive values were further validated in an independent transcriptome dataset of patients with or without neoadjuvant chemotherapy. These analyses confirmed a positive platinum predictive value of high *APOBEC3G*, *CLDN4*, and low *BIRC5* when using OS as an endpoint. Moreover, high *APOBEC3G* expression was significantly associated with the pathological complete response rates in the validation cohort.
5. The protein expression results of CLDN4 were consistent with the findings at gene expression level and confirmed the chemotherapy-predictive value of CLDN4.
6. We identified high baseline serum levels of SDC1 and MMP7 as independent predictors of poor OS in platinum-treated MIBC patients.
7. High tissue expression of MMP7 but not SDC1, proved to be associated with shorter OS in platinum-treated MIBC patients.

Bibliography of the candidate's publications

Related to the thesis:

1. Szarvas T, **Oláh C**, Riesz P, Géczi L, Nyirády P. A húgyhólyag urothelialis daganatainak molekuláris alcsoportbeosztása és annak klinikai vonatkozásai [Molecular subtype classification of urothelial bladder cancer and its clinical relevance]. *Orv Hetil.* 2019;160:1647-1654. Hungarian. doi: 10.1556/650.2019.31559. **IF: 0.540**

2. **Olah C**, Tschirdewahn S, Hoffmann MJ, Krafft U, Hadaschik B, Nyirady P, Szendröi A, Módos O, Csizmarik A, Kovalszky I, Reis H, Szarvas T. Soluble Syndecan-1 Levels Are Associated with Survival in Platinum-Treated Bladder Cancer Patients. *Diagnostics*. 2020;10:864. doi: 10.3390/diagnostics10110864. **IF: 3.706**

3. Szarvas T, Hoffmann MJ, **Olah C**, Szekely E, Kiss A, Hess J, Tschirdewahn S, Hadaschik B, Grotheer V, Nyirady P, Csizmarik A, Varadi M, Reis H. MMP-7 Serum and Tissue Levels Are Associated with Poor Survival in Platinum-Treated Bladder Cancer Patients. *Diagnostics* (Basel). 2020;11:48. doi: 10.3390/diagnostics11010048. **IF: 3.706**

4. **Olah C**, Hahnen C, Nagy N, Musial J, Varadi M, Nyiro G, Gyorffy B, Hadaschik B, Rawitzer J, Ting S, Sjödahl G, Hoffmann MJ, Reis H, Szarvas T. A quantitative polymerase chain reaction based method for molecular subtype classification of urinary bladder cancer-Stromal gene expressions show higher prognostic values than intrinsic tumor genes. *Int J Cancer*. 2022 Mar 1;150(5):856-867. doi: 10.1002/ijc.33809. **IF: 7.316**

5. **Olah C**, Reis H, Hoffmann MJ, Mairinger F, Ting S, Hadaschik B, Krafft U, Grünwald V, Nyirady P, Varadi M, Győrffy B, Kiss A, Szekely E, Sjödahl G, Szarvas T. Predictive value of molecular subtypes and APOBEC3G for adjuvant chemotherapy in urothelial bladder cancer. *Cancer Med*. 2023;12:5222-5232. doi: 10.1002/cam4.5324. **IF: 4.711**

6. Koll FJ, Döring C, **Olah C**, Szarvas T, Köllermann J, Hoeh B, Chun FK, Reis H, Wild PJ. Optimizing identification of consensus molecular subtypes in muscle-invasive bladder cancer: a comparison of two sequencing methods and gene sets using FFPE specimens. *BMC Cancer*. 2023;23:504. doi: 10.1186/s12885-023-11016-9. **IF: 3.8**

7. **Olah C**, Szarvas T. A Panel-Based Method for the Reproduction of Distinct Molecular Subtype Classifications of Muscle-Invasive Urothelial Bladder Cancer. *Methods Mol Biol*. 2023;2684:27-43. doi: 10.1007/978-1-0716-3291-8_2. IF: **1.4**