의학 박사 학위논문

Gene expression characteristics of non-muscle invasive upper tract urothelial carcinoma

-in association with CK5/6 and CK20 immunostaining-

비근육침윤성 상부요로상피암의 유전자 발현 연구

-CK5/6과 CK20 단백발현에 관련하여-

2020년 2월

서울대학교 대학원

의학과 병리학전공

정 민 선

의학 박사 학위논문

Gene expression characteristics of non-muscle invasive upper tract urothelial carcinoma

-in association with CK5/6 and CK20 immunostaining-

비근육침윤성 상부요로상피암의 유전자 발현 연구

-CK5/6과 CK20 단백발현에 관련하여-

2020년 2월

서울대학교 대학원

의학과 병리학전공

정 민 선

Gene expression characteristics of non-muscle invasive upper tract urothelial carcinoma

-in association with CK5/6 and CK20 immunostaining-

지도교수 문 경 철

이 논문을 의학박사 학위논문으로 제출함

2019년 10월

서울대학교 대학원

의학과 병리학전공

정 민 선

정민선의 의학박사 학위논문을 인준함

2020년 1월

위 원 장 조 정 연 (인)

부위원장 문 경 철 (인)

위 원 정 경 천 (인)

위 원 구 자 현 (인)

위 원 류 영 준 (인)

Abstract

Gene expression characteristics of non-muscle invasive upper tract urothelial carcinoma

-in association with CK5/6 and CK20 immunostaining-

Minsun Jung
Department of Pathology
The Graduate School
Seoul National University

Introduction: Urothelial carcinoma is a molecularly heterogeneous tumor that consists of various genetic subtypes. The basal and luminal subtypes, which are distinctly enriched in basal and luminal genes and proteins, are the two main clusters of these subtypes, showing a unique pathobiology, prognosis, and treatment response. For example, patients with basal type muscle-invasive bladder carcinoma (MIBC) have a poor prognosis and may benefit from chemotherapy more than patients with the other subtypes. The enrichment of CK5/6, CD44 and CK20 reflects the basal or luminal molecular subtype of urothelial carcinoma. In contrast, non-muscle invasive bladder carcinoma (NMIBC) is enriched in basal-like markers, and patients in some studies have shown a favorable prognosis for unknown reasons. Despite recent advances in the

molecular landscape of MIBC, the genetic characteristics of subtypes of non-muscle invasive upper tract urothelial carcinoma (NMIUTUC) have not been studied. Therefore, we aimed to investigate the clinicopathological, prognostic, and transcriptional characteristics of papillary NMIUTUC associated with immunohistochemical (IHC) staining for basal and luminal markers.

Material and Methods: IHC staining for CK5/6, CD44, and CK20 was conducted on 211 patients with papillary NMIUTUC using a tissue microarray (prognosis cohort). Staining was classified as negative, positive or normal pattern, and its prognostic significance was analyzed. In addition, we characterized the gene expression profiles of subgroups of papillary NMIUTUC classified by IHC staining for CK5/6 and CK20 (GEP cohort) as follows: group 1 (high-grade CK5/6-high/CK20-low), group 2 (high-grade CK5/6-high/CK20-high), group 3 (high-grade CK5/6-low/CK20-high), group 4 (low-grade CK5/6-high/CK20-low), and group 5 (low-grade CK5/6-high/CK20-high). IHC staining for P-cadherin, E-cadherin, and Ki-67 was performed to validate the genetic alterations among the high-grade GEP cohorts.

Results: We found that CK5/6-negative, CD44-negative and CK20-positive tumors were distinctly high-risk subgroups that were associated with a high WHO grade (CK5/6-negative, $p < 0.001$; CD44-negative, $p < 0.001$; CK20-positive, $p = 0.017$) and frequent intravesical recurrence (CK5/6-negative, $p = 0.002$). Using Kaplan-Meier and log-rank tests, we found that these IHC subgroups were correlated with poor progression-free survival (CK5/6-negative, $p = 0.001$; CD44-negative, $p = 0.009$; CK20-positive, $p = 0.031$) and cancer-specific survival (CK5/6-negative, $p = 0.009$). Furthermore, CK5/6 negativity was identified as an independent prognostic factor for short progression-free survival ($p = 0.009$) and cancer-specific survival ($p = 0.045$). CK5/6

improved Harrell's C-indices for progression-free (0.68 to 0.77, p = 0.029) and cancer-specific (0.59 to 0.77, p <0.001) survival. When these markers were combined, the luminal-like (CK5/6-negative/CK20-positive) subtypes showed distinctively poor prognoses. A transcriptional analysis revealed 308 differentially expressed genes (DEGs) across the high-grade GEP cohorts. Functional analyses of the genes identified cell adhesion as a common process differentially enriched in group 3 compared to groups 1 and 2. Furthermore, late cell cycle/proliferation signatures were also enriched in group 3 and in some of the other groups, which could explain the high-risk phenotype of group 3. Group 2, characterized by low levels of genes associated with mitogen-activated protein kinase and tumor necrosis factor signaling pathways, was hypothesized to represent the least cancerous subtype considering its normal urothelium-like IHC pattern. Compared to this, low-grade GEP subgroups 4 and 5 had 24 DEGs that were enriched in the cAMP-kinase pathway. Analyses of the grade-related genetic differences demonstrated that CK5/6-high/CK20-low tumors had 52 DEGs that indicated altered apoptosis/necroptosis between high- and low-grade tumors. On the other hand, CK5/6-high/CK20-high papillary NMIUTUC had as little as eight DEGs related to grade difference. Finally, IHC staining for P-cadherin, E-cadherin, and Ki-67 revealed that CK5/6-negative and/or CK20-positive papillary NMIUTUC had marginally stronger E-cadherin expression and a much higher Ki-67 proliferative index than the other immunophenotypes, which validated the enrichment of adhesion and late cell cycle/proliferation signatures, respectively. P-cadherin was reactive in a similar fashion to CK5/6 staining as a basal subtype marker.

Conclusion: IHC staining for CK5/6 and CK20 classified papillary NMIUTUC into subgroups relevant not only to their prognosis but also to genotypic-phenotypic traits. Clinicopathological

characteristics of NMIUTUC relative to this stratification were shared by NMIBC but not by MIBC, suggesting that CK5/6 and CK20 expression in early urothelial carcinoma requires distinction from the conventional basal or luminal paradigm of MIBC. Luminal-like (CK5/6-high/CK20-low) papillary NMIUTUC was a distinct aggressive subtype that may be ascribed to altered adhesive and proliferative functions. Adhesion molecules or Ki-67 may be used as a prognostic biomarker complementary to CK5/6 and CK20. Understanding the genetic uniqueness of papillary NMIUTUC would aid in deciphering the genetic evolution of cancer progression and personalizing treatment for NMIUTUC.

Keyword: Upper Track Urothelial Carcinoma; Papillary Urothelial Carcinoma; Carcinoma, Transitional Cell; RNA, messenger; Gene Expression Profiling; High-Throughput Nucleotide Sequencing; Immunohistochemistry; Prognosis

Student Number: 2016-33938

Table of Contents

Chapter 1. Introduction

1.1. Study Background

Upper tract urothelial carcinoma (UTUC) is a rare disease that accounts for 5% of all urothelial carcinomas [1]. UTUC has a high mortality rate, with a 5-year survival rate <50% for muscle invasive tumors and <10% for pT4 stage disease [2]. Recently, the molecular landscape of urothelial carcinoma has become clearer than before [3]. Although vast information on the genetic characteristics of urothelial carcinoma has enlightened the molecular pathogenesis of urothelial carcinoma and made treatment approaches based on individualized cancer genetics more likely, most studies have focused on urinary bladder urothelial carcinoma (UBC) (either muscle-invasive bladder cancer (MIBC) or non-muscle invasive bladder cancer (NMIBC)) [4]. Compared to UBC, upper track urothelial carcinoma (UTUC) has been much less evaluated. Moreover, the heterogeneity of the genetic profiles of urothelial carcinoma makes its clinical utilization challenging, including prognostication or patient selection for neoadjuvant and adjuvant treatments.

1.2. Genomic Landscape of Urothelial Carcinoma

Molecular Oncogenesis of Urothelial Carcinoma

Urothelial carcinoma has a unique tumorigenesis that comprises two divergent pathways that differ in tumor behavior and molecular alterations [5]. Low-grade non-invasive papillary urothelial carcinoma is the most common type of urothelial carcinoma arising from hyperplastic lesions, including papilloma and papillary urothelial neoplasm with low malignant potential, by acquiring *HRAS* and/or *FGFR3* activating mutation or

overexpression [5]. *FGFR3* mutation, which occurs in 70% of low-grade papillary tumors, constitutively activates downstream Ras, PI3K, and STAT signaling pathways [5]. On the other hand, high-grade muscle-invasive urothelial carcinoma that frequently stems from carcinoma in situ (CIS) lesions harbors *TP53* and *RB* inactivations, which are less observed in low-grade papillary lesions [5]. In partial agreement with this observation, Lindgren et al proposed two genomic circuits of urothelial carcinomagenesis: *FGFR3*/*CCND1*/*CDKN2A* and *E3F3*/*RB1* alterations. However, *TP53* alterations have been invariably observed in advanced urothelial carcinoma derived from both circuits [6]. Furthermore, chromosomal alterations and genetic mutations are distinct among the gene expression-based subtypes [6]. Therefore, the genomic characterization of urothelial carcinoma that has been oriented to its carcinogenesis may be closely associated with distinct innate subtypes of urothelial carcinoma.

Muscle-Invasive Bladder Carcinoma

Gene-expression profiling has defined common subtypes of UBC across seminal studies [7]. In the initial period, Lidgren et al of the Lund group [8] found two different intrinsic subtypes (MS1 and MS2) by analyzing the gene-expression profiles (GEP)s of 144 specimens consisting of both MIBC and NMIBC. They showed that the two molecular subtypes had different prognoses, mutation patterns, and copy number alterations [8]. Then, the same group, by incorporating more samples (n = 308) and additional methods, including immunohistochemical (IHC) staining, expanded the subtypes to five genetically defined hierarchical clusters: urothelial-like A (UroA), genomically unstable (GU), infiltrated, urothelial-like B (UroB), and squamous cell carcinoma-like (SCC-like) (Table 1) [9-11]. Although these subtypes did not completely overlap pathological parameters, Ta/T1 and grade 1/2 tumors were predominantly the UroA or GU subtypes [9]. In line with this, the

UroA subtype showed good outcomes, the GU and infiltrated subtypes showed intermediate outcomes, and the UroB and SCC-like subtypes showed the worst outcomes [9]. Functionally, gene expression of the immune response, extracellular matrix, myofibroblast, early/late cell cycles, cytokeratin and keratinization, receptor tyrosine kinases, and cell adhesion was differentially enriched between the subtypes [9]. UroA is enriched in early cell cycle genes (*CCND1, ID*), *KRT20*, *UPK*, *FGFR3* signatures, and certain adhesion genes (*CLDN1*, *TJP2*, *TJP3*, *CDH1*, *CDH3*, *CDH23, PKP4, DSG4, GJD3, LAMA3, ITGA6,* and *COL17A1*) and shows a papillary architecture [9]. Because UroB tumors have similar genomic denominators but present with more advanced stages than UroA tumors, the UroB subtype is believed to be a progressed form of the UroA subtype [9]. The GU, UroB, and SCC-like subtypes are enriched in late cell cycle genes (*CCNA*, *CCNB*, and *CCNE*), which makes these genes potential biomarkers of high-risk urothelial carcinoma [9, 12]. In addition, immunologic and extracellular matrix/myofibroblast-related genes are highly expressed in the infiltrated subtype, *TP53* mutations with genetic rearrangements are frequently observed in the GU subtype, and keratinization-related genes (*KRT6* and *KRT14*) and a squamous histology are observed in the SCC-like subtype, as the names of each subtype indicate [9]. Immune/mesenchymal/myofibroblast signatures may result from a stromal component that challenges the existence of the infiltrated subtype as an intrinsic molecular subgroup [13]. Aside from the functional suggestions made in the above studies, the aggressive subtype of urothelial carcinoma characterized by basal cell/squamous-like gene expression was first indicated and subsequently refined by other studies [14]. Then, the intrinsic subtypes of MIBC were alanyzed (Table 1). Briefly, Damrauer et al (the University of North Carolina group) [15], Choi et al (the MD Anderson group) [16], Rebouissou et al [17], and The Cancer Genome Atlas (TCGA) [18] illustrated transcriptional

profiles with/without integrating DNA and chromosomal alterations. Although the classifications of MIBC were named differently by the respective research groups, common luminal and basal molecular subtypes were shared. Luminal-type tumors express genes enriched in urothelial differentiation or those of luminal breast cancer, such as *KRT20*, *PPARG*, *GATA3*, *FOXA1*, *CD24*, *ERBB2*, *ERBB3*, *XBP1*, *SNX31*, and *UPKs* [2]. The basal type is enriched in basal/squamous/stem cell signatures, including *KRT5*, *KRT6*, *KRT14*, *CDH3*, *TGM1*, *DSC3*, *PI3*, *EGFR*, and *CD44* [2]. A set of BASE47 genes were developed to predict the luminal and basal subtypes, which showed distinct clinical outcomes: the basal type had a worse prognosis than the luminal type [15]. In addition, the MDA group added the p53-like subtype, which resembles the luminal type, but it was also characterized by wild-type *TP53*-associated genes [16]. However, the silhouette score of this subtype was low which indicated that p53-like subtype may be not distinct from other subtypes [16]. Rebouissou et al [17] independently established the basal-like type in their repository, or CIT, and validated it in six other MIBC datasets. This basal-like tumor is associated with an advanced stage and short overall survival, and presents an activated epidermal growth factor receptor (EGFR) pathway with sensitivity to EGFR-targeted therapy [17]. Biton et al [19], by applying Independent Component Analysis to the CIT and other datasets, were also able to find basal-like components in MIBC. Finally, the TCGA group delineated the integrated genomic characteristics of MIBC in 2014 [18] and 2017 [4] (Table 1). First, they established the four mRNA-based clusters (I-IV) [18]. Cluster 1 shows a papillary histology, *FGFR3* activation, HER2 overexpression and active ER signaling. Cluster II is also enriched in the HER2 and ER pathways. These features are shared with luminal A-type breast cancer. Cluster III highly expresses *KRT14* and *CD44*, and is similar to basal-like breast cancer and squamous cell carcinoma at other sites. Finally, Cluster

IV shows similarities with Cluster III but also expresses immune/mesenchymal cell-related genes, which may indicate this subtype as a target of immune checkpoint inhibitors [7]. Later, in 2017, the TCGA expanded the previous study into the full TCGA cohort (n=408) [4], where the two major subtypes, basal and luminal, were confirmed. Moreover, they were able to make a deeper discrimination, yielding five subclassifications in total (luminal, luminal-papillary, luminal-infiltrated, basal-squamous, and neuronal). Luminal signatures are high, and basal/squamous/stem cell signatures are similarly low in the luminal, luminal-papillary, and luminal-infiltrated subtypes. However, wild-type *TP53* and CIS-related genes are lower in luminal-papillary type than in the other two, which implies that luminal-papillary tumors develope from papillary NMIBC. The luminal-infiltrated type is characterized by a prominent infiltration of immune cells and fibroblasts. The basal-squamous subtype shows strong expression of CIS signature genes and a loss of SHH signaling, indicating that this type may originate from CIS lesions [4]. The neuronal type is enriched in neuroendocrinal markers and/or histologic differentiation and characterized by a mutational loss of *TP53* and *RB1*. Clinically, the luminal-papillary subtype shows good overall survival; the luminal-infiltrated, luminal, and basal-squamous types show intermediate survival, and the neuronal group shows poor overall survival [4]. Likewise, the Lund group refined their prior five classifications (UroA, GU, infiltrated, SCC-like, and UroB) into six subtypes (Uro, GU, epithelial-infiltrated, SCC-like/mesenchymal-infiltrated, SCC-like/UroB, small cell/neuroendocrine-like), incorporating not only MIBC but also NMIBC (Table 1), which reiterated those of the TCGA report [11].

Table 1. Major studies on the molecular subtypes of urothelial carcinoma

Group	Primary subject	Method	Subtype
Lund [9]	n=308 (95 MIBC + 213 NMIBC; GSE32894)	Transcription	UroA, GU, Inf, UroB, SCCL
Lund [11]	n=307 (243 MIBC + 57 NMIBC)	Transcription, IHC,	Uro, GU, Epi-Inf, SCCL/Mes-Inf, SCCL/UroB, Sc/NE
UNC [15]	n=262 MIBC (high grade; public; GSE13507, GSE31684, GSE32894, GSE5287)	Transcription, Mutation	Luminal, Basal-like
MDA [16]	n=73 MIBC (public; GSE13507, GSE32894, GSE1827)	Transcription, Mutation, IHC, microRNA	Luminal, Basal, p53-like
CIT [17]	n=85 MIBC	Transcription, Mutation, IHC, Western blot, RT-PCR, CNV	Basal-like
TCGA [18]	n=131 MIBC (high grade)	Transcription, Mutation, CNV, Translocation, Methylation, microRNA, RPPA	Cluster I-IV
TCGA [4]	n=412 (378 MIBC + 34 NMIBC)	Transcription, Mutation, CNV, Translocation, Methylation, microRNA, lncRNA, RPPA	Luminal, Luminal-papillary, Luminal-infiltrated, Basal-squamous, Neuronal
UROMOL [20]	n=460 NMIBC (345 Ta + 112 T1 + 3 CIS)	Transcription, Mutation (based on RNA-seq)	Class 1-3

MIBC, muscle-invasive bladder cancer; NMIBC, non-muscle invasive bladder cancer; UroA, urothelial-like A; GU, genomically unstable; Inf, Infiltrated; UroB, urothelial-like B; SCCL, squamous cell carcinoma-like; IHC, immunohistochemistry; Epi-Inf, epithelial-infiltrated; Mes-Inf, mesenchymal-infiltrated; Sc/NE, small-cell/neuroendocrine-like; UNC, University of North Carolina; MDA, MD Anderson; CIT, Cartes d' Identite des Tumeurs; RT-PCR, reverse transcriptase polymerase chain reaction; CNV, copy number variant; TCGA, The Cancer Genome Atlas; RPPA, reverse phase protein array; RNA-seq, RNA sequencing

Direct comparisons among different subtypes identified by each research group showed significant overlap. The luminal and p53-like subtypes of the MDA group are very similar to Clusters I and II of the initial TCGA report, respectively and to the luminal and luminal-infiltrated subtypes of the later report, which mostly fall into the luminal subtype defined by the UNC study and into the UroA, UroB, GU, and some of infiltrated subtypes suggested by the Lund group [4, 13, 21]. Likewise, there is a strong correlation among the basal type of the UNC or MDA group, Cluster III or IV and basal-squamous of the TCGA, and UroB and SCC-like of the Lund group [4, 7, 9, 21]. The simplest explanation for the existence of intrinsic subtypes (basal versus luminal) may be the development of different cells of origin: basal tumors from basal cells and luminal tumors from intermediate or umbrella cells [21]. The urothelium consists of hierarchically organized differentiated cells [22]. It would be reasonable to speculate that cancers that develope from such hierarchically differentiated cells differ in their pathophysiology [22]. Basal cells of the urothelium contain many features of cancer stem cells, including the enrichment of CD44, CK14, CK5, LRC, 67LR, ALDH1A1, and p63, which may predict the aggressive behavior of urothelial carcinoma [22]. The primitive differentiation state of the cell-of-origin of urothelial carcinoma, together with certain genetic alterations (e.g., STAT3 activation), may lead to the development of basal-type MIBC originating from CIS [22]. The expression profile of CKs is differentially regulated by the differentiation state of urothelial carcinoma as follows: CK14-positive/CK5-positive/CK20-negative, CK14-

negative/CK5−positive/CK20−negative, and CK14−negative/CK5−negative/CK20−positive expression was revealed to indicate basal, intermediate, and differentiated states of urothelial carcinoma [23]; CD44, CD90, and CD49f positivity also represents the primitive/basal phenotype of urothelial carcinoma [23]. In addition, high CK14 expression was associated with poor outcomes in various urothelial carcinoma cohorts [23]. Further investigation across GEPs of subtypes mentioned above showed that these primitive/basal tumors corresponded to the basal type of the MDA group, Clusters III/IV of the TCGA reports, and the UroB/SCC−like subtypes of the Lund cohort [7].

In an attempt to define the intrinsic subtypes of MIBC, a joint meeting was held in Spain in 2015 [7]. The intrinsic subtype should represent stable, inherent characteristics of the tumor type. The p53−like and luminal subtypes are unstable compared to the basal subtype and frequently shows molecular switching [7, 16]. Instead, tumors with high CK5/6 and CK14 but low FOXA1 and GATA3 expression by IHC staining were found among studies and shared poor prognosis and thus named the basal−squamous−like (BASQ) subtype [7]. Further studies are required to make a consensus on the definition of other molecular subtypes although several experts agreed that there are other subtypes of MIBC [7].

Non−Muscle Invasive Bladder Carcinoma

In addition, a few studies have investigated the genetic characteristics of NMIBC (Table 2). According to the findings reported in the Lund studies [9, 11], low−grade Ta NMIBC is most similar to the UroA subtype, and high−grade T1 NMIBC is most similar to the UroB/SCC−like or the GU or infiltrated subtype (Table 2). As previously described, the UroA, GU, and infiltrated subtypes generally correspond to the luminal type of MIBC, with retained urothelial differentiation signals, which can be identified at both the mRNA and tumor cell protein expression levels [11, 24].

Cell cycle signatures, early for UroA and late for GU/infiltrated, are some of the key discriminating features among these luminal subtypes [24]. Hedegaard and colleagues [20] exclusively focused on NMIBC using comprehensive transcriptional, mutational, and clinical data of 460 patients who were involved in the UROMOL project (Tables 1–2). They delineated three genetically defined classes (classes 1–3) of NMIBC which showed distinct GEPs and survival rates [20]. Regarding luminal vs basal classification, class 1 and class 2 demonstrate luminal–like gene expression (e.g., *UPK* and *KRT20*) but class 3 expresses basal–like genes (e.g. *KRT5*, *KRT15*, and *CD44*) [20]. Class 1 tumors exhibit high expression of early cell cycle genes (*CCND1, CCND2, CCND3, RBL2, ID1, ID2, ID3,* and *WEE1*), class 2 is enriched with late cell cycle signatures (*CDK1, CDK4, CDK2, CCNE1, CDC20, CCNB2, CCNB1, CCNA2, BUB1, CDC25A, CCNE2, MYBL2, FOXM1,* and *PLK1*), which are targets of many transcription factors (*E2F1, E2F4, MYC, MYCN, KDM5B, NFYA,* and *IRF3*), and class 3 is characterized by repressed protein–coding genes, the high expression of histone/chromatin modification signatures (*ERG1, HNF4A, YAP1, FLI1, KDM5B, CREB1, IRF3, H3K79me2, H4K20me1, H3K4me1, H3K36me3,* and *H3K27ac*), and long non–coding RNAs [20]. *KRT14*, as well as stem cell markers other than CD44 (*ALDH1A1*, *ALDH1A2*, *CD133, NES,* and *CD90)* were primarily found in both classes 2 and 3 [20]. At the same time, however, urothelial differentiation is observed in class 2 at the genetic level [20]. Furthermore, class 2 tumors are defined by the activation of EMT–related transcription factors and stem cell markers [20]. Interestingly, class 2 has higher rates of high–grade tumors, high–stage tumours, concurrent CIS, high European Organization for Research and Treatment of Cancer (EORTC) scores, and progression to MIBC than class 1, despite elevated levels of luminal type markers, including KRT20, which indicates that class 2 may represent a molecular shift toward muscle–invasive carcinoma by

taking the carcinoma in situ (CIS) pathway [20]. In addition, class 3, a basal type-like NMIBC, is not a precursor of the basal type but may represent an inactive state [20]. A 117-gene classifier was proposed for the prediction of the three classes [20]. Consensus clustering according to other taxonomies demonstrated that the GU (Lund) and basal (BASE47) subtypes made a large concord with class 2 and class 3, respectively [20]. As the authors applied the 117-gene predictor to metachronous tumor sets, they found that high-risk tumors were classified as class 2 or demonstrated progression, taking a class 2-like state [20]. Consistent with these findings, the TCGA MIBC cohorts gave rise to class 1 and class 2 but not class 3 according to the classifier, all of which belong to the luminal type; or TCGA cluster 1 and cluster 2 overlap with class 1 and class 2, respectively [20]. Predicted mutations based on RNA sequencing (RNA-seq) associated with class 2 tumors are associated with the DNA damage response, *TP53, ERCC2*, APOBEC mutation signatures, MAPK/ERK, and ERBB pathways, but those with class 1 and 3 include *FGFR3* [20].

Table 2. Transcriptional subtypes of non-muscle invasive bladder carcinoma and their clinicopathological and molecular characteristics

Group	Subtype	Number (%)	Clinicopathological features	Molecular features
Lund [9, 10]	UroA	122 (57)	Papillary, TaLG, Good Px IHC: CK5 (+) basal, CK20/UPK3 (+) luminal or aberrant, CK14 (−), P-Cad (+) basal, Ki-67 low	Early cell cycle, KRT20/UPK, KRT5, FGFR3 (mut), PIK3CA (mut), intact adhesion
	GU	55 (26)	T1HG, Poor Px IHC: CK5/14 (−) or low, CK20 aberrant, P-Cad low, Ki-67 high, HER2 high	Late cell cycle, KRT20/UPK, p53 (mut), loss of adhesion
	UroB	10 (5)	Ta/T1HG IHC: CK5/14 (+) basal and intermediate, CK20 (−) or aberrant, Ki-67 intermediate	Late cell cycle, KRT5/6/14, FGFR3 (mut) PIK3CA (mut), p53 (mut)
	Inf	23 (11)	Rich stromal cells	Immune/stromal cells
	SCCL	4 (2)	Squamoid	Late cell cycle, KRT5/6/14
UROMOL [20]	Class 1	97 (21)	Papillary, TaLG, Good Px,	Early cell cycle, differentiated; UroA-like
	Class 2	232 (51)	Papillary + other, Ta/T1HG, CIS, Poor Px,	Late cell cycle, diff, KRT20, KRT14, CSC, EMT; GU or Inf-like
	Class 3	129 (28)	Papillary, TaLG, Good-mod Px	Dormant state, KRT5/14/15, CD44; UroA or basal-like

Uro, urothelial-like; LG, low grade; Px, prognosis; IHC, immunohistochemical staining; (+), positive; (−), negative; P-Cad, P-cadherin; mut, mutation; GU, Gemonically Unstable; HG, high grade; Inf, Infiltrated; SCCL, squamous cell cancer-like; diff, differentiation; CSC, cancer-stem cell; EMT, epithelial-mesenchymal transition

Upper Tract Urothelial Carcinoma

Compared to UBC, UTUC is more likely to present a muscle-invasive stage [25], and is enriched in microsatellite instability and promoter methylation [26]. Sfakianos et al [26], using targeted massive parallel sequencing, revealed that although the two share a similar mutational landscape, UTUC has more frequent *FGFR3*, *HRAS*, and *CDKN2B* but less frequent *TP53*, *ARID1A*, and *RB1* mutations than UBC; in particular, *RB1* mutation is absent in UTUC. In addition, *FGFR3* mutation is more frequently observed in low-grade than high-grade tumors, but it is concordant between areas with different grades, stages, and even metastatic regions within a single patient [26]. A comparison of the transcriptomes of UTUC and UBC was performed by Sanford et al [27], who concluded that urothelial carcinoma of the two locations share similar GEPs overall, with the exceptions of stage-stratified tyrosine kinase and apoptotic signaling pathways that were enriched in pT3 UTUC compared to pT3 MIBC [1, 27]. Moreover, a pathway analysis identified that HGF and TNF cascades are downregulated in UTUC compared with UBC [27]. Finally, the BASE47 classifier clusters most UTUCs and only a few UBCs into the luminal subtype [27].

1.3. Clinical Implications of the Molecular Subtypes of Urothelial Carcinoma

The features inherent to the molecular subtypes may aid in understanding the pathogenesis and mechanisms of disease progression of urothelial carcinoma and in applying tailored therapy. Several genetic determinants of the intrinsic subtypes have been proposed. Basal type tumors are enriched in RB1 pathway alterations (*RB1* mutation, *CCND1*/*CCNE1*/*E2F3* amplification) and p63 activation, whereas luminal type tumors are enriched in activating *FGFR3* mutation, the estrogen receptor (ER) pathway,

HER2, and PPARγ [15, 16]. FOXA, GATA, PPARG, ELF3, and IRF1 are involved in urothelial differentiation and may mediate the development of the luminal subtype [7]. Furthermore, the p53-like subtype shows resistance to neoadjuvant chemotherapy [16]. These findings suggest that different molecular subtypes have a distinct pathogenesis and targetable genetic alterations. According to a retrospective investigation on the response to cisplatin-based neoadjuvant chemotherapy (NAC), patients with basal tumors may benefit from NAC the most, but those with luminal type MIBC have no improved survival with NAC [28]. The infiltrated subtype according to the Lund definition or TCGA cluster II has only a limited benefit from NAC. The molecular subtypes are also associated with different responses to immune-checkpoint inhibitors. Cluster II tumors show the most significant response to Atezoluzumab, but cluster III tumors are the most responsive to Nivolumab [29, 30].

Because IHC staining on formalin-fixed paraffin embedded (FFPE) tissue is invariably and widely used in pathological practice, phenotypical characteristics of the molecular subtypes were investigated using IHC staining. The Lund group tried to validate their molecular subtypes by conducting IHC staining for a set of antibodies (CCNB1, CCND1, CCNE1, CDH1, CDH2, CDH3, CDKN2A, DSC2/3, E2F3, EGFR, ERBB2, FGFR3, CK14, CK20, CK5, CK6, Ki-67, RB1, p63, and UPK3) on 237 NMIBC and MIBC specimens obtained from their previous study consisting of the UroA, UroB, GU, and SCC-like subtypes genetically [10]. These proteins showed a strong correlation between IHC staining quantity and mRNA expression level with the exception of E-cadherin (*CDH1*), which had a weak ($r < 0.3$) correlation [10]. UroA tumors frequently exhibit a urothelial-like papillary histology, CK5-positive and P-cadherin-positive basal cells, a layer of luminal to basal cells that is positive to E-cadherin, CK20-positive luminal cells with frequent aberrant expression in the interior portion of the tumor as well as

ki-67 expression at the tumor-stromal interface [10]. UroB tumors express CK5 that is not restricted to the basal cell layer and levels of CK14 and Ki-67 expression that are higher than those of UroA tumors [10]. SCC-like tumors present low levels of IHC staining for FGFR3, p63, CK20, and E-cadherin but higher levels of CK5, Ki-67, and P-cadherin than UroA or UroB tumors [10]. GU subtype tumors show mostly invasive growth patterns, and only a few exhibit a urothelial-like histology [10]. The GU subtype is characterized by low FGFR3, CCND1, p63, P-cadherin, CK5, and CK14 expression [10]. Furthermore, two genomic circuits operating urothelial carcinoma pathogenesis, *FGFR3*/*CCND1* alteration and *p16* deletion versus *E2F3* amplification/*RB1* deletion and high p16 expression were materialized by IHC staining for such markers: the former is operative in UroA and UroB tumors, and the latter is active in GU tumors [10]. This classification was suggested to be reproduced by using four variables, a urothelial-like histology, grade, CCNB1 >17%, and CK5 expression score >0.57, with a high (0.88) accuracy [10]. Although this classification successfully predicted the outcomes of each subtype correctly (Uro had a good outcome, GU had an intermediate outcome, and SCC-like had a poor outcome), there was no reliable way to differentiate between UroB and UroA using such histopathological parameters [10]. Moreover, a urothelial-like histology defined by a smooth interface between the tumor and the stroma depicted by SMA immunostaining should indicate papillary tumors that usually present non-invasive or T1 stage. Therefore, this classifier needed to be verified within early-stage papillary tumors. The same strategy was used to predict the risk of progression of 167 patients with primary T1NxM0 NMIBC; 48 Uro, 87 GU, and 14 SCC-like tumors were predicted, excluding patients with discordant expression between the duplicate cores [12]. Concurrent to the previous study, GU tumors showed lower expression levels of CK5 and P-cadherin than Uro tumors, which were confined to the basal cells [12]. Tumors classified as the GU

subtype, which are associated with high frequencies of multifocality, the presence of CIS, and lymphovascular invasion, have higher 3-year recurrence rates than Uro tumors [12]. However, a urothelial-like papillary histology may be evaluated from the non-invasive component, which hinders the molecular characterization specific to T1 NMIBC [12]. Regarding the latest Lund classifications [11], BASQ markers are limited to the SCC-like/mesenchymal-infiltrated or SCC-like/UroB subtypes, and the expression of their counterpart (CK5-negative, CK14-negative, GATA3-positive, and FOXA1-positive) is frequently observed in the Uro, GU, and epithelial-infiltrated subtypes. Additionally, typical BASQ-type tumors showed progressed immunophenotypes, with low EPCAM/E-cadherin but high P-cadherin, from the Uro or GU subtypes [11]. Although regulators of urothelial differentiation, such as *RXRA*, *PPARG*, *FOXA1*, *GATA3*, and *ELF3* show coordinated upregulation in the Uro, GU, and epithelial-infiltrated subtypes but downregulation in the SCC-like/mesenchymal-infiltrated and SCC-like/UroB subtypes, CK20 and UPK3 expression is very heterogeneous and not limited to urothelial differentiation, in the cellular aspects [11]. Consequently, the authors proposed FGFR3 positivity, CCND1 positivity, RB1 positivity, and p16 negativity as markers for the Uro subtype, FGFR3 negativity, CCND1 negativity, RB1 negativity, and p16 positivity for the GU subtype, CK5 positivity, CK14 positivity, FOXA1 negativity, and GATA3 negativity for the basal/SCC-like subtype, vimentin positivity, ZEB2 positivity, CDH1 negativity, EPCAM negativity for the mesenchymal-like subtype, and TUBB2 positivity, EPCAM positivity, CDH1 negativity, and GATA3 negativity for the small cell/neuroendocrine-like subtype, which were based on IHC staining [11]. Although these working definitions procreated the GEP with high sensitivity and specificity (>75%), the authors concluded that UroA, UroB, and some GU tumors were not identifiable using IHC staining [11]. Considering

that NMIBC mainly consists of the UroA, UroB, and GU subtypes, these classifications may not be applicable to early urothelial carcinoma. However, MIBC may undergo pseudo differentiation, making differentiation-related markers expressed in the tumors in a way that is distinct from that in NMIBC [11]. Furthermore, Uro tumors defined by IHC staining are the most heterogeneous group that can harbor GU-like or basal/SCC-like genotypes and they probably take different progression pathways [11]. Regarding the basal versus luminal subtypes and their correlation with protein expression, basal-like tumors express high CK5/6 and low FOXA1 at both mRNA and IHC expression levels [17]. IHC staining for CK5/6 and FOXA1 reliably identified a basal-like group with 89% sensitivity and 95.5% specificity [17]. Luminal- and basal-type MIBCs in the TCGA dataset, when they are classified based on MDA clusters, exhibited differentially expressed proteins, including GATA3, E-cadherin, HER2/3, Rab-25, and Src in the luminal type and annexin-1, CD49, cyclin B1, and EGFR in the basal type [31]. IHC staining for GATA3, CK18, CK20, uroplakin 2, cyclin D1, GER2, CK5/6, CK14, p63, p16, bcl2, SMA, myosin, calponin, and desmin was used to discriminate the luminal, basal, and p53-like molecular subtypes of MIBC [31]. As a result, IHC staining for CK5/6, CK14, GATA3, CK20, and uroplakin 2 showed differential expression between the luminal and basal types and a strong positive correlation with their mRNA expression levels [31]. However, p63, cyclin D1, and CK18 expression showed significant overlap between the two subtypes, and moreover, p53-like tumors were revealed to more likely be a stromal-contaminated non-intrinsic group [31]. Further investigation revealed that 20% of the cut-off positivity value best segregated the luminal vs basal molecular subtype and the use of CK5/6 and GATA3 as two surrogate markers can classify the molecular classifications with 91% accuracy; however, some basal subtype and double-negative tumors exhibited luminal type IHC staining [31]. However, the authors used only 10, 20, 30, and

40% cut-offs for classification, which hindered the detailed discrimination of the expression of these proteins.

In addition to histopathological characteristics, which are well-established prognostic factors, several approaches have been made for the risk prediction of urothelial carcinoma, including one or combinations of the expression of a few genes, epigenetics, and IHC markers [12]. These tissue biomarkers of UTUC were proposed before and after the advent of the molecular subtypes of urothelial carcinoma [25]. Much of them are directly or indirectly related to transcriptional subtypes and their molecular characteristics (e.g., molecules involved in the cell cycle, cell signaling, and urothelial differentiation) [1, 11, 25, 32-34]. Sikic et al [33] risk stratified UTUC into subgroups by using CK5 and CK20, which are prognostic in MIBC. However, the prognosis of UTUC subgroups was in opposition to what was demonstrated in MIBC; UTUC, showing both CK5 negativity and CK20 positivity, was associated with a poor prognosis. Interestingly, NMIBC shows a similar association between CK5-negative/CK20-positive tumours and adverse outcomes [20]. As previously described, the class 2 of the UROMOL datasets showed a poor prognosis despite elevated levels of luminal type markers, including *KRT20* [2, 20]. In accordance with this finding, NMIBC, with high CK20 and/or with low CK5 or CD44 expression, is associated with high-risk clinicopathological characteristics and poor outcomes [32, 35, 36]. Nevertheless, because both non-muscle invasive UTUC (NMIUTUC) and muscle invasive UTUC (MIUTUC) were included in a previous prognostic investigation [33], it is still unclear whether NMIUTUC can be risk stratified according to IHC-based subgroups and whether the subgroups have similar prognostic characteristics as those of NMIBC [1].

Papillary urothelial carcinoma is a molecularly and architecturally distinct tumor that comprises the majority of early urothelial carcinomas [1]. A normal pattern IHC staining is one of

the unique immunoprofiles of this type of urothelial carcinoma that resembles the reactivity of the normal urothelium, showing basal cells that are positive for CK5/6 and CD44 and apical cells that are positive for CK20. These normal patterns have been associated with a favorable prognosis in NMIBC patients, irrespective of the percent positivity [1, 35, 37-42] These results suggest that recognizing a normal pattern of IHC staining in papillary NMIUTUC is necessary. However, the relationship between a normal patterned IHC staining and the prognosis or GEPs of UTUC patients is still unclear.

1.4. Purpose of the Research

We aimed to describe the prognostic significance of luminal/basal subtype-defining markers in papillary NMIUTUC using IHC staining and to investigate the transcriptional characteristics of the prognostically significant IHC-based subgroups of NMIUTUC. To this end, we first analyzed the expression of CK5/6, CD44, and CK20 in 211 patients with papillary NMIUTUC and correlated the expression levels/patterns with the clinicopathological characteristics and clinical outcomes of the patients [1]. Furthermore, we analyzed the GEPs of papillary high-grade NMIUTUC, which was segregated into subgroups according to the IHC expression of CK5/6 and CK20. The transcriptional differences between the subgroups were validated using IHC staining. In addition, we applied the same approach to low-grade tumors to determine whether similar genetic discriminations can be made. Fourth, the transcriptomes of high-grade and low-grade NMIUTUC were compared for us to investigate the genetic background orchestrating the grade differences and its relationship with CK5/6 and CK20 immunoprofiles.

Chapter 2. Materials and Methods

2.1. Patients

The prognostic significance of IHC staining for CK5/6, CD44, and CK20 was evaluated in 211 patients with papillary NMIUTUC (prognosis cohort). Patients underwent radical nephroureterectomy (n = 202) or ureterectomy (n = 9) between 1998 and 2012 at Seoul National University Hospital. Overall, 170 men and 41 women were included in this cohort. The median age at diagnosis was 67 years (range, 29-95 years). The mean diameter of the primary tumor was 3.5 cm (standard deviation [s.d.], 2 cm). Most patients were in the low TNM stage: 72 patients were in stage 0a and 138 patients were in stage I. However, one patient with a T1 high-grade tumor in the ureter was in TNM stage IV because this patient had regional lymph node metastasis (N1) at the time of diagnosis. For statistical purposes, the patient was included in the stage ≥I group. Distant metastasis was not observed. According to the WHO/ISUP grading system, 118 tumors were low grade, and 93 were high grade.

Gene-expression profiling was conducted on 23 radical nephroureterectomy fresh specimens consisting of 15 high-grade (high-grade GEP cohort) and eight low-grade (low-grade GEP cohort) papillary NMIUTUCs that were diagnosed at the same hospital between January 2016 and June 2018. These GEP cohort tumors were selected according to IHC staining for CK5/6 and CK20 from a total of 50 tumors that were harvested and divided into two. One half was snap-frozen and stored for RNA-seq, and the other half was made into an FFPE block. The protein expression of frozen tissue was evaluated based on the IHC staining of its counterpart FFPE block. Every hematoxylin and eosin slide was reviewed, and the stage and grade were reclassified according to

the Cancer Staging Manual of the American Joint Committee on Cancer, 8th edition [43] and the 2004 World Health Organization (WHO)/International Society of Urological Pathology (ISUP) grading system [44], respectively. A papillary growth pattern was determined on the basis of the gross and microscopic findings. Tumors showing divergent differentiation were excluded because of their potential aberrant reaction to CK20 [34]. Clinical data were obtained from medical records. This study was approved by the Institutional Review Board of Seoul National University Hospital (07 Nov 2018, IRB No. H-1810-148-983).

2.2. Tissue Microarray

Three cores (2 mm in diameter) per patient included in the prognosis cohort were embedded in new recipient paraffin blocks using a trephine apparatus (Superbiochips Laboratories, Seoul, Republic of Korea) for tissue microarray (TMA) construction. Tumor areas showing a papillary architecture were mainly obtained. Each core was derived from different tumor areas with the highest grade to overcome the issue of regional heterogeneity. For multifocal tumors, the area with the highest stage and grade was chosen.

2.3. Immunohistochemistry

IHC staining was conducted on 4-μm-thick FFPE sections with a Benchmark autostainer (Ventana, Tucson, AZ, USA), according to the manufacturer's instructions. Mouse monoclonal antibodies against CK5/6 (1:100; D5/16 B4; Dako, Glostrup, Denmark), CD44 (1:100; 156-3C11; ThermoFisher, Rockford, IL, USA), CK20 (1:50; Ks 20.8; Dako), CK14 (1:300; LL002; Cell Marque, Rocklin, CA, USA), GATA3 (L50-823; 1:500; Cell

Marque), FOXA1 (1:500; PA5-27157; ThermoFisher), p53 (1:1000; DO7; Dako), P-cadherin (1:300; 610228; BD Biosciences, Franklin Lakes, NJ), E-cadherin (1:25; 36B5; Leica Biosystems, Wetzlar, Germany), and Ki-67 (1:100; MIB-1; Dako) were used. A normal pattern for CK5/6, CD44, CK20, and P-cadherin was defined when more than two-thirds of the papillary tumor area showed a normal-pattern: CK5/6, CD44, and P-cadherin reactivity restricted to basal cells and CK20 reactivity confined to umbrella cells with occasional weak reaction to intermediate cells (Figure 1A) [35, 42]. Then, a semiquantitative assay of positive versus negative was performed on the remaining specimens (Figure 1B-C). The staining extent (1, 0-30%; 2, 30-50%; 3, 50-70%; 4, 70-100%) and intensity (0, absent; 1, weak; 2, moderate; 3, strong) of E-cadherin expression were evaluated and multiplied. The cut-off values were 20% for CK5/6, CD44, and P-cadherin [31], 1% for CK20 [35], 0% for CK14, and the multiplied score of ≥6 for E-cadherin. As basal IHC markers [16], we applied the same cut-off values for CK5/6, CD44, and P-cadherin. For CK20, the cut-off value of 1% discriminated patients' progression better than the other values tested (10%, 20%, and 50%) [1]. For CK5/6, CK20, and P-cadherin expression, total tumor areas of three TMA cores were analyzed as a whole. Moreover, full-section staining for CK5/6, CD44, and CK20 was performed on 19 randomly selected specimens that included more than three cases per every IHC results, and results that were concordant with those from the TMA were obtained. Quantification of the Ki-67 proliferative index and the H-scores of GATA3 and FOXA1 (3*strong staining nuclei (%) + 2*moderately staining nuclei (%) + 1*weakly staining nuclei (%)) was performed using QuPath (version 0.1.2) [45]. Moreover, the positive tumor cell proportion to IHC staining that was scored by a 10% scale was evaluated for all IHC staining in the GEP cohort [2].

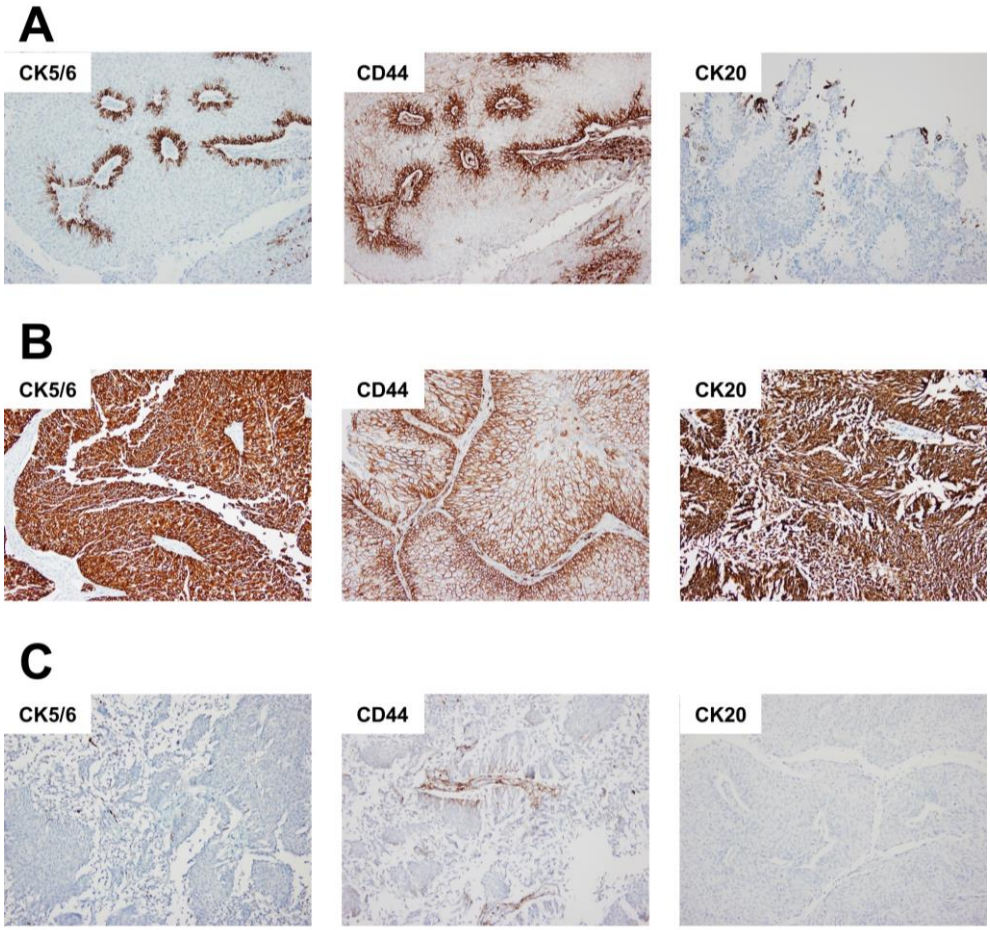

Figure 1. Representative images of immunohistochemical staining for CK5/6, CD44, and CK20. **(A)** Normal pattern staining. **(B)** Positive staining. **(C)** Negative staining. Stromal cells show positivity for CD44.

2.4. Statistical Analysis

Comparison analyses of clinicopathological characteristics and IHC results were performed using Pearson's chi-square test, Fisher's exact test, Mann-Whitney U test, or Kruskal-Wallis H test. Progression-free survival (PFS) was defined as the interval between surgery and upper urinary tract recurrence or distant

metastasis, or the last follow-up visit for patients who did not show any of these. Lower urinary tract recurrence was analyzed separately as intravesical recurrence (IVR), except for the patients who had already undergone cystectomy before the UTUC operation (missing number = 4). Cancer-specific survival (CSS) was defined as the interval between surgery and cancer-related death or the last follow-up visit. Kaplan-Meier analysis and log-rank tests were used to compare survival rates. A Cox proportional hazards regression model was used for univariate and multivariate analyses. A two-tailed P-value of <0.05 was considered statistically significant. Harrell's C-index was used to quantify the prognostic discrimination of independent prognostic factors. All statistical analyses were performed with SPSS STATISTICS 23 (IBM, Armonk, NY, USA), STATA 13 (StataCorp, College station, TX, USA), or R 3.4.3 statistical environment.

2.5. Transcription Analysis and the Identification of Differentially Expressed Genes

mRNA was extracted using a punch with a plunger (3 mm in diameter) from the pure papillary tumor area (tumor cellularity >90%) of frozen tissue that was marked on a hematoxylin and eosin-stained frozen slide. RNA-seq was performed as previously described [46] using a barcode-based Illumina Truseq Access RNA Kit [47, 48] on the Illumina HiSeq 2500 platform (Illumina, San Diego, CA, USA). Paired demultiplexed fastq files were generated, and initial quality control was performed using FastQC (Phred quality score >30) (Babraham Bioinformatics). Adapter trimming was conducted using the Trimmomatic tool [49] followed by mapping to the human genome (UCSC hg19) using Hisat2 [50]and Bowtie2 [51]. Known transcripts were assembled with the StringTie tool [52]. ComBat was used to standardize the different

cohorts and minimize the batch effect. Quantile normalization was performed on log2 (FPKM+1). To fit a linear model for each gene based on the three-group study design, the function of lmFit was used (limma) [53]. Then, the eBayes function was used to calculate the empirical Bayes moderated t-statistic.

2.6. Functional Enrichment Analysis

The functional enrichment of differentially-expressed genes (DEGs) was determined from the Gene Ontology (GO) [54] and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway [55] databases. An FDR <0.05 was used as the cut-off. A Gene Set Enrichment Analysis (GSEA) of all genes was performed to interpret their functions from the public database [56]. Biological signatures were acquired from the respective publications. The integrated score for the progression of NMIBC was calculated by subtracting the total value of *COL4A3BP, MBNL2, NEK1, FABP4*, and *SKAP2* from that of *KPNA2, BIRC5, UB2C, CDC25B, COL4A1, MSN*, and *COL18A1* [57].

2.7. TCGA Data Preprocessing and Survival Validation

From the TCGA database, 399 samples of UBC that contained mRNA expression and survival data were collected. Most tumors were WHO high-grade (n = 375) and MIBC (n = 396). The median follow-up period was 1735 days, during which 109 patients died. Preprocessing of the mRNA expression data was performed in R 3.4.3 statistical environment: the mRNA expression level quantified by RSEM raw counts was normalized using the limma-voom, which applied linear modeling to voom-transformed read counts [58]. We determined the median value of gene expression as

a cut-off point to divide patients into high and low expression groups. Kaplan-Meier and log-rank tests were applied for survival validation.

Chapter 3. Results

3.1. Prognostic Significance of Immunohistochemical Staining for CK5/6, CD44, and CK20

Initially, the prognostic implication of IHC staining for CK5/6, CD44 and CK20 was analyzed within the prognosis cohort with a three-tier classification as positive, negative or normal pattern (Figure 2A). According to the Kaplan-Meier and log-rank tests, CK5/6-negative (p = 0.002, PFS; p = 0.032, CSS), CD44-negative (p = 0.018, PFS), and CK20-positive (p = 0.076, PFS) tumors showed distinct outcomes that were worse than those of any other subgroup (Figure 2A). In other words, favorable-to-intermediate prognoses were similarly found in either positive staining and normal pattern (CK5/6 and CD44) or negativity and normal pattern (CK20) of each marker. Therefore, a simplified two-tier classification system consisting of negative versus other (CK5/6 and CD44) and positive versus other (CK20) was adopted for further prognostic evaluation.

The median follow-up period was 74 months (range, 0-261 months). During this time, 19 patients (9.0%) experienced disease progression and six (4.3%) died of the disease. CK5/6 negativity, CD44 negativity and CK20 positivity were observed in 98 (46.4%), 99 (46.9%) and 118 (55.9%) patients, respectively. Kaplan-Meier analysis and log-rank tests (Figure 2B) demonstrated that CK5/6 negativity (p = 0.001), CD44 negativity (p = 0.009) and CK20 positivity (p = 0.031) were significantly associated with short PFS. In addition, CK5/6 negativity was associated with short CSS (p <0.009). The univariate Cox regression analysis revealed that patients with CK5/6-negative [hazard ratio (HR) = 6.54, p = 0.003], CD44-negative (HR = 3.58, p = 0.014), CK20-positive (HR = 3.16, p = 0.041) or high-grade (HR = 5.25, p = 0.003)

tumors had an increased risk of disease progression (Table 3). Patients with CK5/6−negative (HR = 9.46, p = 0.034) or multifocal (HR = 5.52, p = 0.016) tumors showed an increased risk of cancer−related death (Table 3). Multivariate analyses on each IHC marker were performed with the other variables that were statistically significant in univariate analyses. Consequently, CK5/6 negativity was identified as an independent prognostic factor for short PFS (HR = 4.63, p = 0.017) and CSS (HR = 8.42, p = 0.045) after adjusting for tumor grade and multifocality, respectively (Table 4). The other independent prognostic factors included high grade for short PFS (HR = 3.56, p = 0.027) and tumor multifocality for shorter CSS (HR = 4.43, p = 0.036), after adjusting for CK5/6 (Table 4). The addition of CK5/6 to these baseline models (i.e., grade for PFS and multifocality for CSS) significantly improved the C−indices for predicting PFS (from 0.68 to 0.77, p = 0.029) and CSS (from 0.59 to 0.77, p <0.001). Furthermore, the results for CD44 showed significant correlations with those for CK5/6 (p <0.001) and CK20 (p = 0.008) (Table 5). CK5/6-CK20 and CD44-CK20 combinations stratified patient outcomes in more detail: CK5/6 negativity/CK20 positivity and CD44 negativity/CK20 positivity indicated particularly poor PFS (p <0.001 and p = 0.018, respectively) and CSS (p = 0.007 and p = 0.349, respectively) (Figure 3).

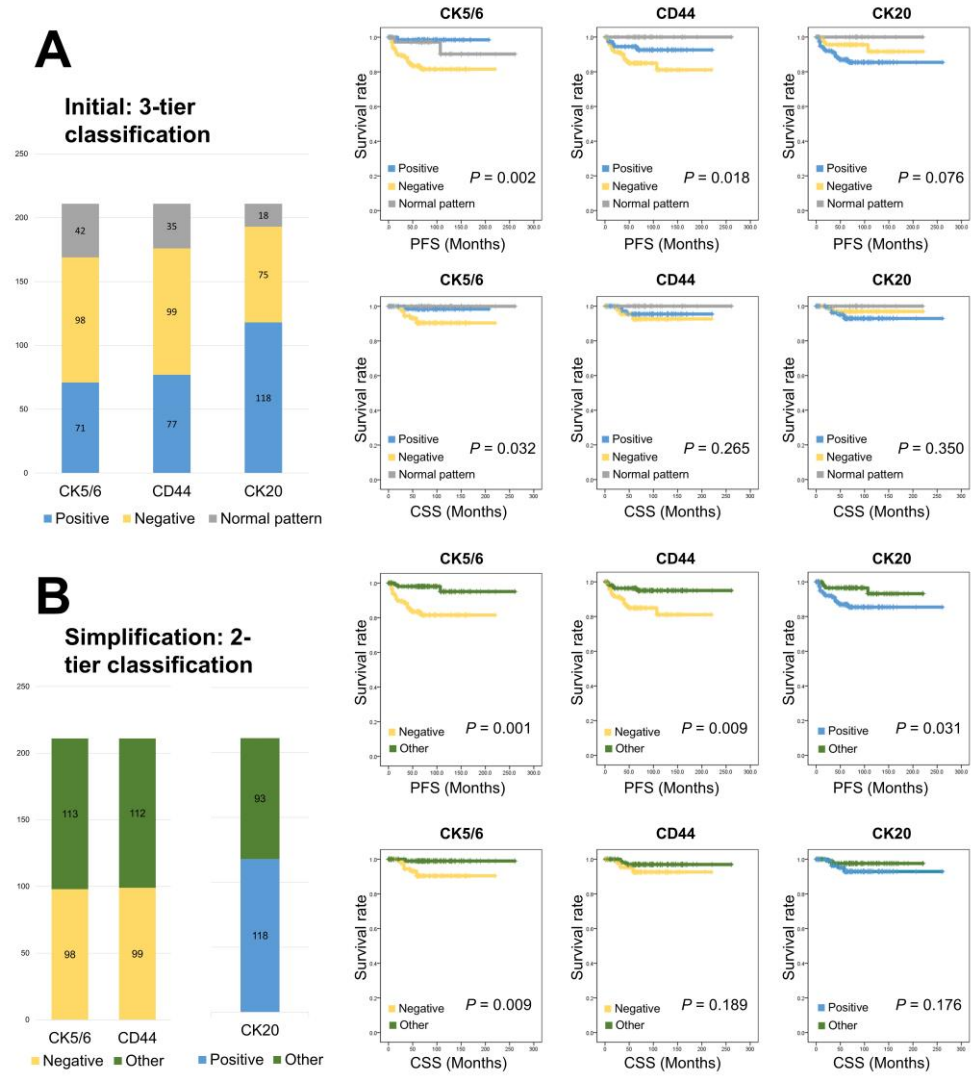

Figure 2. Survival analyses for patients with papillary non-muscle-invasive papillary upper tract urothelial carcinoma. **(A)** Three-tier classification with a positive, negative or normal pattern of IHC staining for CK5/6, CD44, and CK20, and its survival prediction of progression-free survival (PFS) and cancer-specific survival (CSS). **(B)** Simplified two-tier classification and its survival analyses of PFS and CSS. 'Other' subgroups stand for combined 'normal pattern' and 'positive' groups of CK5/6 and CD44 or 'normal pattern' and 'negative' groups of CK20 in the three-tier classification.

Table 3. Univariate Cox proportional regression survival analysis of the prognosis cohort

	Progression-free survival			Cancer-specific survival		
Variable	HR	95% CI	p	HR	95% CI	p
Age (continuous)	0.98	0.95-1.03	0.432	0.99	0.93-1.05	0.727
Sex						
Female vs Male	0.80	0.23-2.76	0.726	0.54	0.08-4.30	0.559
Size (continuous)	0.98	0.78-1.22	0.828	0.82	0.56-1.21	0.319
Multifocality						
Present vs Absent	2.97	0.99-8.96	0.053	5.52	1.38-22.05	0.016
TNM stage						
≥ I vs 0	1.48	0.53-4.11	0.452	1.84	0.38-8.87	0.446
Grade						
High vs Low	5.25	1.74-15.82	0.003	1.64	0.44-6.11	0.461
CK5/6						
Neg vs Other*	6.54	1.91-22.48	0.003	9.46	1.18-75.63	0.034
CD44						
Neg vs Other*	3.58	1.29-9.96	0.014	0.41	0.10-1.63	0.205
CK20						
Pos vs Other†	3.16	1.05-9.52	0.041	2.81	0.58-13.55	0.197

* Other subgroups of CK5/6 and CD44 indicate positive and normal-pattern staining.

† Other subgroup of CK20 indicated negative and normal-pattern staining.

HR, hazard ratio; CI, confidence interval; Neg, negative; Pos, positive

Table 4. Multivariate Cox proportional regression survival analysis of the prognosis cohort

Variable	Comparison detail	HR	95% CI	p
PFS				
CK5/6 and grade				
	CK5/6 Neg vs Other*	4.63	1.31-16.36	0.017
	Grade High vs Low	3.56	1.15-11.02	0.027
CD44 and grade				
	CD44 Neg vs Other*	2.51	0.88-7.17	0.085
	Grade High vs Low	4.13	1.33-12.80	0.014
CK20 and grade				
	CK20 Pos vs Other†	2.58	0.85-7.82	0.095
	Grade High vs Low	4.66	1.54-14.13	0.007

Variable	Comparison detail	HR	95% CI	p
CSS				
CK5/6 and multifocality				
	CK5/6 Neg vs Other*	8.42	1.05–67.78	0.045
	Multifocality Present vs Absent	4.43	1.10–17.80	0.036

* Other subgroups of CK5/6 and CD44 indicate positive and normal pattern staining.

† Other subgroup of CK20 indicate negative and normal pattern staining.

HR, hazard ratio; CI, confidence interval; Neg, negative; Pos; positive

Table 5. Cross–correlation between CK5/6, CD44, and CK20 expression

		CD44			CK20		
		Neg	Other*	p‡	Pos	Other†	p‡
CK5/6	Neg	60	38	<0.001	58	40	0.406
	Other*	39	74		60	53	
CD44	Neg				65	34	0.008
	Other*				53	59	

* Other subgroups of CK5/6 and CD44 indicate positive and normal pattern staining.

† Other subgroup of CK20 indicate negative and normal pattern staining.

‡ Statistics analyzed by Pearson's χ^2 test

Neg, negative; Pos, positive

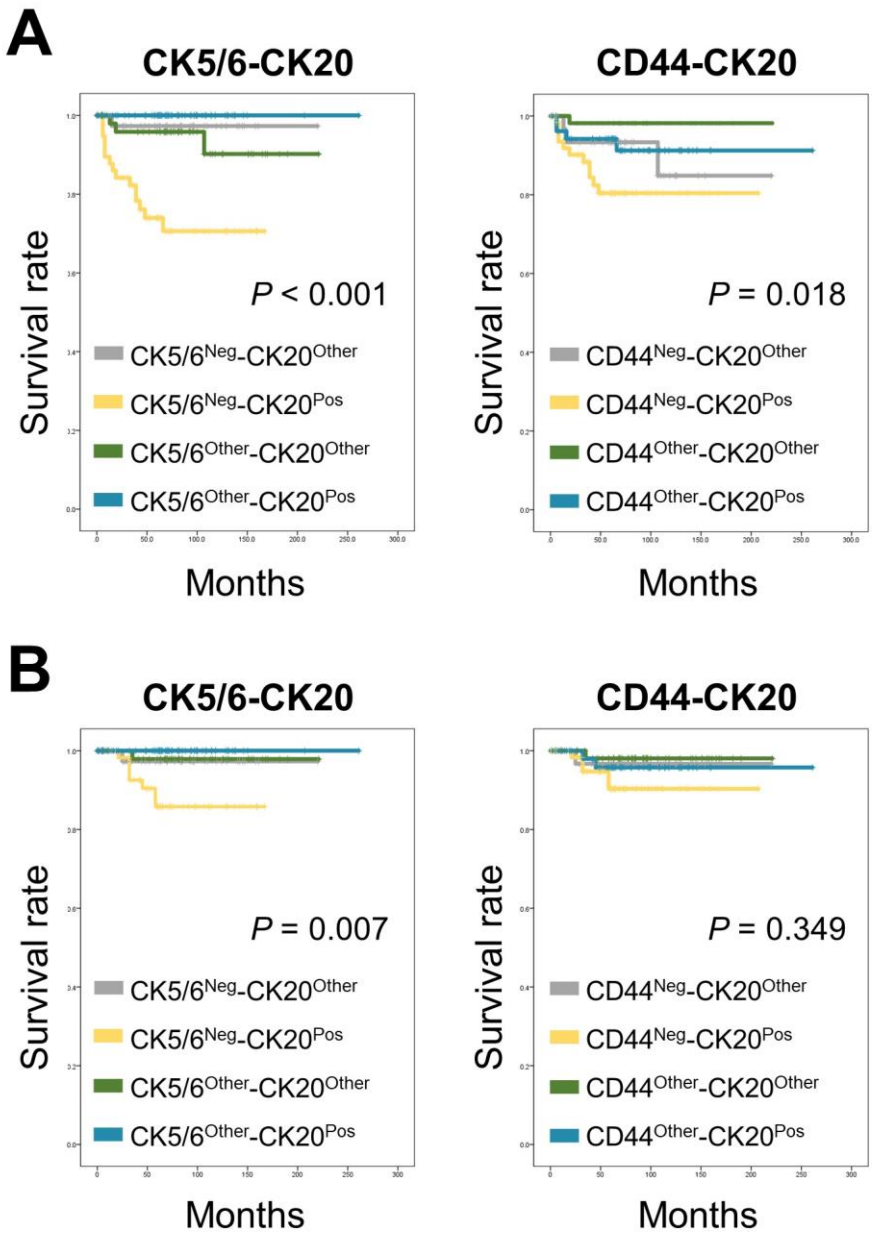

Figure 3. Kaplan-Meier analyses on subgroups according to the combinations of basal (CK5/6 or CD44) and luminal (CK20) markers based on the two-tier classification. **(A)** Progression-free survival. **(B)** Cancer-specific survival.

3.2. Clinicopathological Association of Basal and Luminal Marker Expression

The clinicopathological association with basal and luminal marker expression was evaluated in both the TMA and whole tissue slides. Consistent with their prognostic results, negativity to CK5/6 or CD44 and positivity to CK20 were significantly associated with a high grade in the prognosis cohort (CK5/6−negative, p <0.001; CD44−negative, p <0.001; CK20−positive, p = 0.017) (Table 6). In addition, patients with CK5/6−negative tumors had a significantly more frequent IVR rate (p = 0.002) (Table 6). The expression levels of the other BASQ markers, CK14, GATA3, and FOXA1, as well as their association with CK5/6, CD44, and CK20 expression were also investigated (Table 6). Positivity to CK14 was significantly associated with negativity to CK5/6 (p = 0.002) and positivity to CK20 (p = 0.001). GATA3 and FOXA1 expression was significantly higher in CK20−positive tumors (p = 0.004 and 0.038, respectively).

Table 6. Clinicopathological details associated with CK5/6, CK20, and CD44 expressions of the prognosis cohort

	CK5/6			CD44			CK20		
Variable	Neg	Other*	p	Neg	Other*	p	Pos	Other †	p
Age (year)			0.053			0.215			0.581
<67	41	63		44	60		56	48	
≥67	57	50		55	52		62	45	
Sex			0.383			0.385			0.165
Male	76	94		77	93		91	79	
Female	22	19		22	19		27	14	
Location			0.012			0.207			0.230
Pelvis	69	65		67	67		69	65	
Ureter	19	42		23	38		39	22	
Both	10	6		9	7		10	6	
Size (cm)			0.890			0.406			0.267
<3.4	52	62		50	64		68	46	
≥3.4	46	51		49	48		50	47	

	CK5/6			CD44			CK20		
Variable	Neg	Other*	p	Neg	Other*	p	Pos	Other †	p
Multifocality			0.340			0.637			0.630
Absent	87	105		89	103		106	86	
Present	11	8		10	9		12	7	
TNM stage			0.244			0.885			1.000
0	29	43		33	39		32	40	
≥ I ‡	69	70		66	73		61	78	
Grade			<0.001			<0.001			0.017
Low	40	78		39	79		57	61	
High	58	35		60	33		61	32	
IVR §			0.002			0.080			0.384
Absent	50	83		55	78		76	57	
Present	44	29		40	33		37	36	
CK14 ¶			0.002			0.097			0.001
Negative	43	75		57	61		72	47	
Positive	30	18		30	18		41	7	
GATA3 (mean) ¶	43.4	41.2	0.982	43.6	41.0	0.746	49.4	28.4	0.004
FOXA1 (mean) ¶	29.8	36.5	0.740	27.3	38.0	0.058	35.9	27.7	0.038

* Other subgroups of CK5/6 and CD44 indicate positive and normal pattern staining.

† Other subgroup of CK20 indicate negative and normal pattern staining.

‡ Including a patient with stage IV disease.

§ IVR was assessed for 205 patients.

¶ CK14, GATA3, and FOXA1 was assessed for 166, 184, and 205 patients, respectively.

Comparison of GATA3 and FOXA1 scores were carried out by Mann–Whitney U test. Other statistics were based on Pearson's chi test.

Neg, negative; Pos, positive; IVR, intravesical recurrence

Furthermore, a similar clinicopathological association of CK5/6, CD44, and CK20 was verified in the independent cohort using whole tissue sections (n = 50; Table 7). A high WHO/ISUP grade was significantly related to CK5/6 (p = 0.004) and CD44 (p = 0.048) and marginally related to CK20 (p = 0.083) and p53 (p = 0.051) (Figure 4). However, GATA3 (p = 0.339) and FOXA1 (p = 0.778) were not significantly different, which were positive in >80% of 95% and 76% of evaluable specimens, respectively.

Table 7. Sample selection for the GEP cohorts

No	RNA*	Group	Org	St	Gr	CIS	CK5/6	CK20	CK14	GATA3	FOXA1	CD44	p53
15	Y	1	Pel	a	H	−	3	1	1	10	10	2	10
20	Y	1	Pel	1	H	+	7	1	3	1	2	8	9
74	Y	1	Pel	1	H	−	8	1	1	9	9	8	2
137	Y	1	Ure	1	H	+	9	1	1	10	10	8	9
166	Y	1	Pel	1	H	−	5	1	1	10	1	4	n.a.
25	Y	2	Ure	1	H	+	10	9	1	10	10	7	5
45	Y	2	Pel	a	H	−	6	9	1	10	10	2	9
105	Y	2	Ure	a	H	−	4	10	1	10	10	2	7
130	Y	2	Ure	1	H	−	9	6	1	10	10	9	4
163	Y	2	Ure	a	H	−	8	8	1	10	10	7	n.a.
21	Y	3	Ure	1	H	+	1	10	1	10	9	1	1
89	Y	3	Pel	a	H	−	1	10	1	10	9	2	6
95	Y	3	Pel	1	H	−	1	10	1	10	9	7	3
135	Y	3	Pel	1	H	+	1	10	1	10	5	1	9
150	Y	3	Ure	a	H	−	1	10	1	10	10	4	n.a.
42	Y	4	Ure	1	L	−	10	1	1	10	10	7	6
83	Y	4	Pel	a	L	−	6	1	1	10	9	5	5
86	Y	4	Ure	a	L	−	10	1	1	10	10	10	6
125	Y	4	Pel	a	L	−	6	1	1	10	9	6	2
131	Y	4	Pel	1	L	−	9	1	1	10	7	4	2
120	Y	5	Ure	a	L	−	7	9	1	10	10	9	7
133	Y	5	Ure	a	L	−	7	6	1	10	10	5	4
138	Y	5	Pel	a	L	−	6	7	1	10	10	4	1
4	N	n.a.	Pel	a	L	−	2	1	1	10	2	8	2
14	N	n.a.	Pel	a	L	−	1	3	1	10	1	1	2
16	N	n.a.	Ure	a	H	−	2	4	1	10	n.a.	2	5
22	N	n.a.	Pel	1	H	−	1	4	1	10	n.a.	1	2
23	N	n.a.	Ure	a	L	−	3	10	1	10	n.a.	9	1
36	N	n.a.	Ure	1	H	−	4	5	1	10	n.a.	10	5
49	N	n.a.	Pel	a	H	−	1	1	1	9	n.a.	4	2
53	N	n.a.	Ure	1	H	+	1	10	1	10	n.a.	1	10
61	N	n.a.	Ure	1	H	+	2	1	1	9	n.a.	8	10
67	N	n.a.	Pel	1	H	−	2	9	1	10	n.a.	1	4

No	RNA*	Group	Org	St	Gr	CIS	CK5/6	CK20	CK14	GATA3	FOXA1	CD44	p53
70	N	n.a.	Ure	1	H	−	1	3	1	10	n.a.	7	10
76	N	n.a.	Pel	1	H	+	1	10	1	10	n.a.	2	2
78	N	n.a.	Ure	1	H	−	2	1	1	10	n.a.	3	6
118	N	n.a.	Pel	1	L	−	2	1	1	10	n.a.	7	3
122	N	n.a.	Ure	a	H	−	1	1	1	10	n.a.	5	10
123	N	n.a.	Pel	a	L	−	1	1	1	9	n.a.	10	4
139	N	n.a.	Pel	1	H	−	1	8	1	9	n.a.	5	7
148	N	n.a.	Ure	1	H	−	2	2	1	10	n.a.	n.a.	n.a.
160	N	n.a.	Ure	1	H	+	1	1	1	2	n.a.	n.a.	1
161	N	n.a.	Ure	a	H	+	3	3	1	10	n.a.	n.a.	1
162	N	n.a.	Pel	a	L	−	10	7	1	10	n.a.	n.a.	n.a.
164	N	n.a.	Pel	1	H	−	1	1	n.a.	n.a.	n.a.	n.a.	n.a.
165	N	n.a.	Pel	1	H	−	1	9	n.a.	n.a.	n.a.	n.a.	n.a.
167	N	n.a.	Pel	1	H	−	10	10	n.a.	n.a.	n.a.	n.a.	n.a.
168	N	n.a.	Pel	1	L	−	10	7	n.a.	n.a.	n.a.	n.a.	n.a.
169	N	n.a.	Ure	1	H	−	1	2	n.a.	n.a.	n.a.	n.a.	n.a.
170	N	n.a.	Pel	a	L	−	10	1	n.a.	n.a.	n.a.	n.a.	n.a.

*Samples subjected to RNA−seq

No, case number; Org, organ; St, T stage; Gr, WHO/ISUP grade; CIS, carcinoma in situ; Y, yes; Pel, pelvis/calyx; H, high; −, absent; +, present; Ure, ureter; L, low; N, no

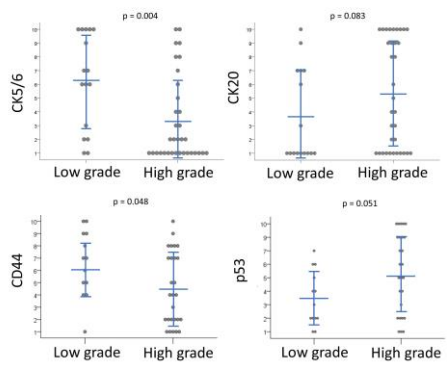

Figure 4. IHC staining expression scores of CK5/6, CK20, CD44, and p53 and WHO/ISUP grade of the GEP cohort. CK5/6-low and CK20-high IHC staining is related to high grade. Blue bars indicate the mean value ± standard deviation.

3.3. GEP Cohort Assembly and Subgroups Based on Immunohistochemical Staining for CK5/6 and CK20

Because of the prognostic significance and the distribution of reactive samples, we focused on gene-expression details associated with CK5/6 and CK20 expression as basal and luminal markers, respectively. First, the transcriptional investigation was limited within the same grade tumors to remove grade-related genetic diversity [3]. Fresh tissue with high (IHC score ≥6) or low (IHC score = 1) CK5/6 and CK20 expression was subjected to RNA-seq to enhance gene-protein expression relevance. Consequently, five subgroups were established as follows (Table 7): group 1, high-grade CK5/6-high/CK20-low; group 2, high-grade CK5/6-high/CK20-high; group 3, high-grade CK5/6-low/CK20-high; group 4, low-grade CK5/6-high/CK20-low; and group 5, low-grade CK5/6-high/CK20-high (Figure 5). Papillary low-grade NMIUTUC that was highly reactive only to CK20 was not observed. CK immunoreaction was diffusely stained in the whole layer of tumors in groups 1, 3, and 4, without noticeable compartmentalization (Figure 5). Although group 2 and 5 tumors were positive for both CK5/6 and CK20 in ≥50% of tumor cells, the expression of CK5/6 and CK20 was accentuated in basal and luminal cells, leaving at least one cell layer of the basal and luminal portion, respectively, in all specimens (Figure 5).

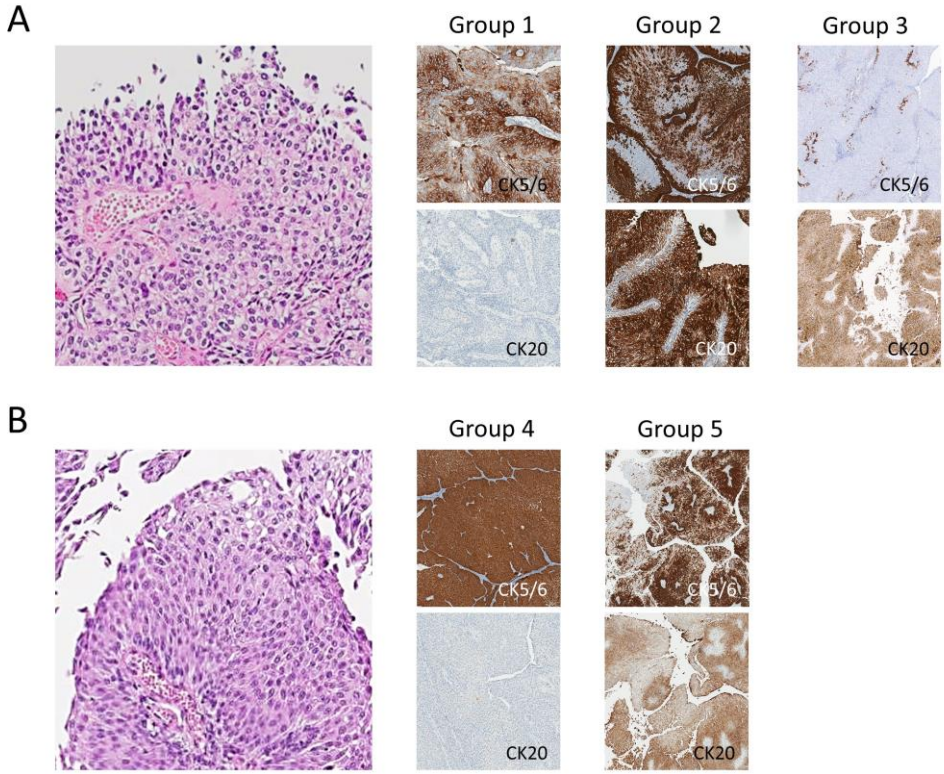

Figure 5. Representative images of IHC staining for CK5/6 and CK20 across subgroups of the papillary NMIUTUC of the GEP cohort. **(A)** High-grade GEP cohorts. **(B)** Low-grade GEP cohorts.

Clinicopathological data of patients in the high-grade GEP cohort are summarized in Table 8. The median age of the patients was 69 years (range, 56-84) and the male/female sex ratio was 2:1. The tumors measured 3.6 ± 3.24 cm (mean ± s.d.) in maximal diameter. Six (40%) patients were in stage pTa, and the other 9 (60%) patients were in pT1. CIS was observed in 5 patients (33.3%). There were no significant differences in clinicopathological parameters or IHC profiles among the subgroups, except for CK5/6 and CK20 expression (Table 8). One sample that belonged to group 1 showed 20%-30% positivity for CK14. With the exception of this patient, IHC staining for GATA3 and FOXA1

showed diffuse staining in all samples. In addition, the median age of the low-grade GEP patients was 59.5 years (range, 42-81 years) and the male/female sex ratio was 3:1. The tumors measured 2.3 ± 1.10 cm (mean ± s.d.) in maximal diameter. Six (75%) patients were in stage pTa. CIS was absent. Only CK20 expression was significantly higher in group 5 (p = 0.017). The other IHC markers were similarly expressed in both low-grade subgroups.

Table 8. Clinicopathological characteristics of high-grade GEP subgroups

Variables	Group 1	Group 2	Group 3	p
Number (%)	5 (100)	5 (100)	5 (100)	
Age				1.000
≥69	3 (60)	3 (60)	3 (60)	
<69	2 (40)	2 (40)	2 (40)	
Sex				1.000
Male	3 (60)	4 (80)	3 (60)	
Female	2 (40)	1 (20)	2 (40)	
Size				0.725
≥3.6	2 (40)	1 (20)	0 (0)	
<3.6	3 (60)	4 (80)	5 (100)	
Organ				0.301
Pelvis/calyx	4 (80)	1 (20)	3 (60)	
Ureter	1 (20)	4 (80)	2 (40)	
Stage				0.800
T1	4 (80)	2 (40)	3 (60)	
Ta	1 (20)	3 (60)	2 (40)	
CIS				1.000
Present	2 (40)	1 (20)	2 (40)	
Absent	3 (60)	4 (80)	3 (60)	
IHC score (mean ± s.d.)*				
CK5/6	6.4 ± 2.41	7.4 ± 2.41	1 ± 0.00	0.001
CK20	1.0 ± 0.00	8.4 ± 1.52	10 ± 0.00	<0.001
CD44	6.0 ± 2.83	5.4 ± 3.21	3.0 ± 2.55	0.162
p53	7.5 ± 3.70	6.3 ± 2.22	4.8 ± 3.55	0.428
CK14	1.4 ± 0.89	1.0 ± 0.00	1.0 ± 0.00	1.000
GATA3	8.0 ± 3.94	10.0 ± 0.00	10.0 ± 0.00	0.286
FOXA1	6.4 ± 4.51	10.0 ± 0.00	8.4 ± 1.95	0.066

*Semiquantitative scores from 1 (<10%) to 10 (90-100%) by a 10% scale.

CIS, carcinoma in situ; IHC, immunohistochemistry; s.d., standard deviation

Table 9. Clinicopathological characteristics of low-grade GEP subgroups

Variables	Group 4	Group 5	p
Number (%)	5 (100)	3 (100)	
Age			0.143
≥59.5	1 (20)	3 (100)	
<59.5	4 (80)	0 (0)	
Sex			0.464
Male	3 (60)	3 (100)	
Female	2 (40)	0 (0)	
Size			0.143
≥2.3	4 (80)	0 (0)	
<2.3	1 (20)	3 (100)	
Organ			1.000
Pelvis/calyx	3 (60)	1 (33.3)	
Ureter	2 (40)	2 (66.7)	
Stage			0.464
T1	2 (40)	0 (0)	
Ta	3 (60)	3 (100)	
CIS			1.000
Present	0 (0)	0 (0)	
Absent	5 (100)	3 (100)	
IHC score (mean ± s.d.)*			
CK5/6	8.2 ± 2.0	6.7 ± 0.6	0.536
CK20	1.0 ± 0.0	7.3 ± 1.5	0.017
CD44	6.4 ± 2.3	6.0 ± 2.6	0.763
p53	4.2 ± 2.0	4.0 ± 3.0	1.000
CK14	1.0 ± 0.0	1.0 ± 0.0	1.000
GATA3	10.0 ± 0.0	10.0 ± 0.0	1.000
FOXA1	9.0 ± 1.2	10.0 ± 0.0	0.168

*Semiquantitative scores from 1 (<10%) to 10 (90-100%) by a 10% scale.
CIS, carcinoma in situ; IHC, immunohistochemistry; s.d., standard deviation

The prognostic implications for the IHC criteria used in the high-grade GEP cohort were validated in the TMA samples. Of 93 patients with high-grade tumors among the prognosis cohort, 39 met these IHC staining criteria: high (≥50%) or low (<10%), comprising 13 CK5/6-high/CK20-low, 4 CK5/6-high/Ck20-high, and 22 CK5/6-low/CK20-high tumors. As a result, CK5/6-low/CK20-high expression tended to show the worst PFS (p = 0.071) (Figure 6). Because of the universally favorable prognosis, this analysis was not performed among low-grade tumors,

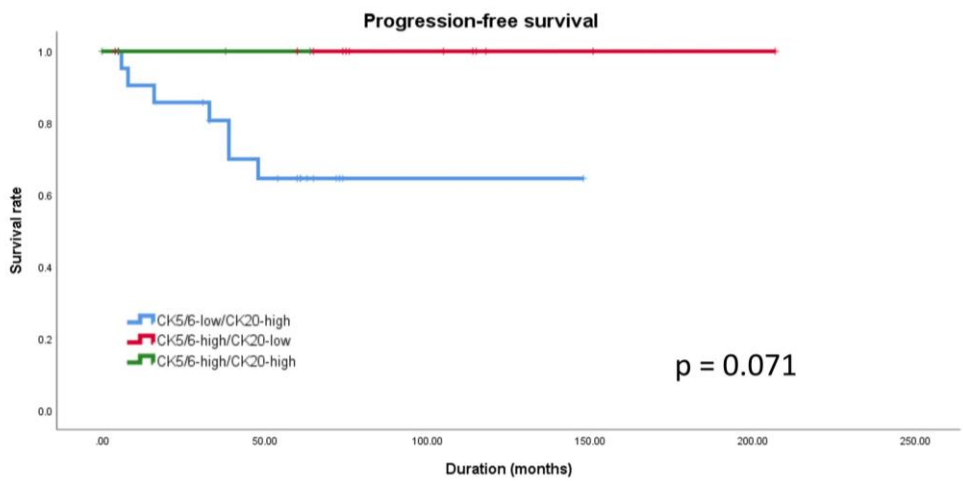

Figure 6. PFS of papillary high-grade NMIUTUC that was included in the prognosis cohort. Tumors having similar IHC staining profile with the Group 3, CK5/6-low/CK20-high expression, was marginally associated with shorter PFS period.

3.4. Identification of Differentially Expressed Genes Among the High-Grade GEP Subgroups

Across the 15 high-grade GEP cohort samples, 21,780 genes were identified. Using a p-value <0.01 and a |fold change| >2 as the cut-offs, we identified 132, 115, and 143 DEGs between groups 1 and 2, groups 1 and 3, and groups 2 and 3, respectively

(Figure 7A–B and Table 10). Group 1 had 58 upregulated and 74 downregulated genes compared to group 2. Group 3, relative to groups 1 and 2, harbored increased levels of 57 and 58 genes and decreased levels of 52 and 91 genes, respectively. Consequently, a total of 308 DEGs across all comparisons were found. These genes were analyzed with unsupervised hierarchical clustering, which demonstrated the congregation of the subgroups defined by IHC staining (Figure 7C). Principal component analysis of the DEGs revealed that tumors with similar CK5/6 and CK20 expression profiles generated clusters that separated from one another in 2D space (Figure 7D).

Table 10. Top 10 DEGs with most significant p-values of each comparison

Comparison	DEG	p-value	Fold change (log_2)
Group 1 vs 2	*SHC1*	2.58E-05	1.22
	OR9K2	2.79E-05	-1.03
	IL1R1	3.44E-05	1.62
	ANO5	4.65E-05	-1.83
	ATP5IF1	8.52E-05	-1.34
	PTPDC1	9.34E-05	-1.25
	SPATA18	1.62E-04	-1.64
	MVB12B	1.63E-04	-1.41
	FSIP2	2.09E-04	-2.44
	PRR15	2.28E-04	-1.31
Group 1 vs 3	*CLCA2*	1.04E-04	4.84
	ERP27	2.01E-04	-1.93
	KRCC1	2.13E-04	-2.91
	SLC24A3	2.16E-04	2.22
	AK3	2.74E-04	-1.11
	CTSZ	2.93E-04	-1.41
	SLC35D2	3.26E-04	-1.35
	RASL11A	3.70E-04	-1.51
	COL7A1	3.80E-04	3.00
	C11orf49	3.80E-04	-1.10
Group 2 vs 3	*NSG1*	1.05E-04	2.79
	TTC39A	1.08E-04	-1.48

Comparison	DEG	p-value	Fold change (log_2)
	MAPK3	1.19E-04	-1.25
	CLCA2	1.19E-04	4.76
	ANO5	1.46E-04	1.62
	WNT10A	1.92E-04	2.74
	SNORA57	2.14E-04	2.48
	SH3BP5	2.21E-04	-1.46
	LRRC69	2.41E-04	1.24
	EVC2	2.56E-04	2.53
Group 5 vs 4	*DTNA*	1.71E-04	9.33
	BTNL9	4.68E-04	2.01
	CDH3	1.19E-03	-2.86
	SNCG	2.00E-03	-3.68
	ITGB5	2.09E-03	-2.76
	ANKRD22	2.53E-03	-3.51
	WDR66	2.62E-03	-2.55
	HCAR3	3.47E-03	-4.80
	ANP32C	4.70E-03	2.82
	RAET1L	5.48E-03	-4.04
Group 4 vs 1	*MIR196A1*	8.24E-05	7.96
	GGT6	1.23E-04	3.26
	CTSZ	1.95E-04	2.87
	CD151	2.64E-04	2.4
	C11orf49	3.47E-04	2.08
	NEB	5.44E-04	-8.37
	ALDH3B1	6.87E-04	2
	TMEM181	1.01E-03	2.16
	ARL14	1.20E-03	2.59
	HIST2H2AA3	1.30E-03	-2.57
Group 5 vs 2	*SNORD116-17*	6.58E-04	19.35
	TMEM51	3.53E-03	2.54
	C3orf67	4.34E-03	-3.94
	AMIGO2	4.61E-03	2.99
	OR5P2	4.71E-03	6.53
	SNORD116-19	6.20E-03	-17.42
	PCDHGA9	6.27E-03	-2.49

DEG, differentially-expressed gene

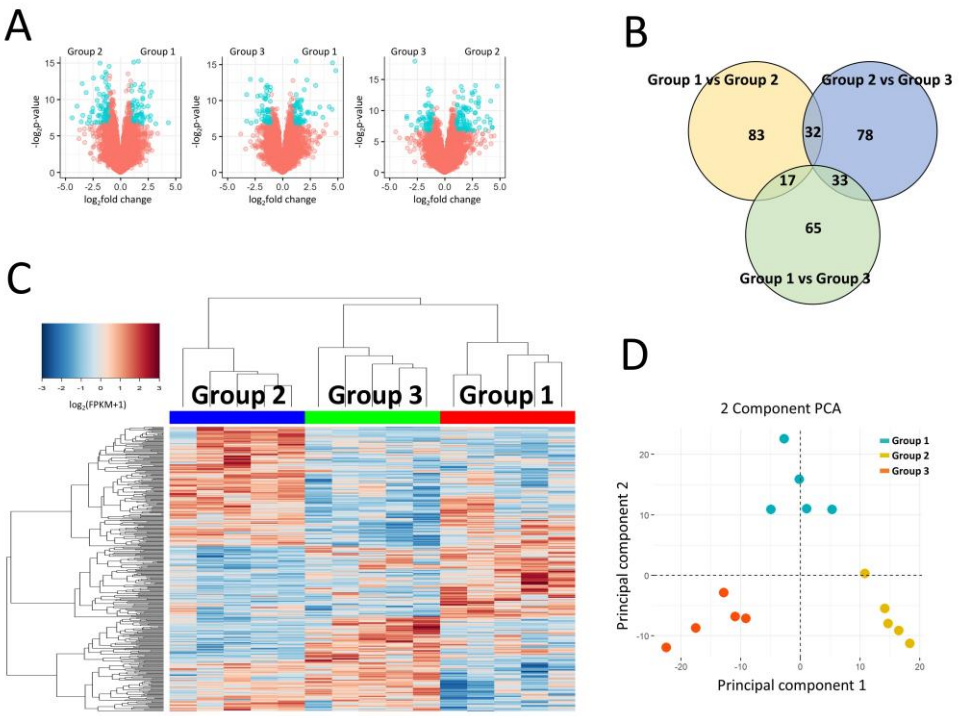

Figure 7. Identification of DEGs between the subgroups of high-grade GEP cohort. **(A)** Volcano plots of the DEGs between each subgroup. Genes that meet the threshold are in blue. **(B)** Venn diagram of all comparisons demonstrates 308 DEGs across the subgroups. **(C)** Unsupervised hierarchical clustering of all DEGs for all samples. **(D)** Principal component analysis of DEGs. Groups 1, 2, and 3 form distinct clusters.

3.5. Functional Analysis of the High-Grade GEP Subgroups

Figure 8, Figure 9, and Figure 10 show the top 10 significant (false discovery date [FDR] <0.05) results of GO and KEGG pathway analyses of DEGs between the subgroups.

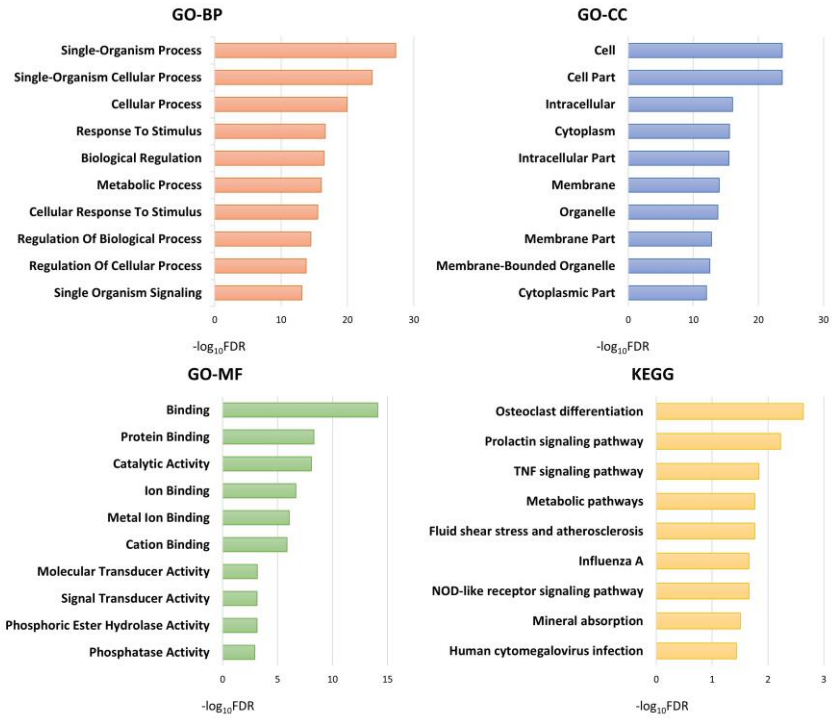

Figure 8. Top 10 significantly enriched biologic process (BP), cellular component (CC), and molecular function (MF) terms of GO and KEGG analyses of DEGs between Group 1 and Group 2.

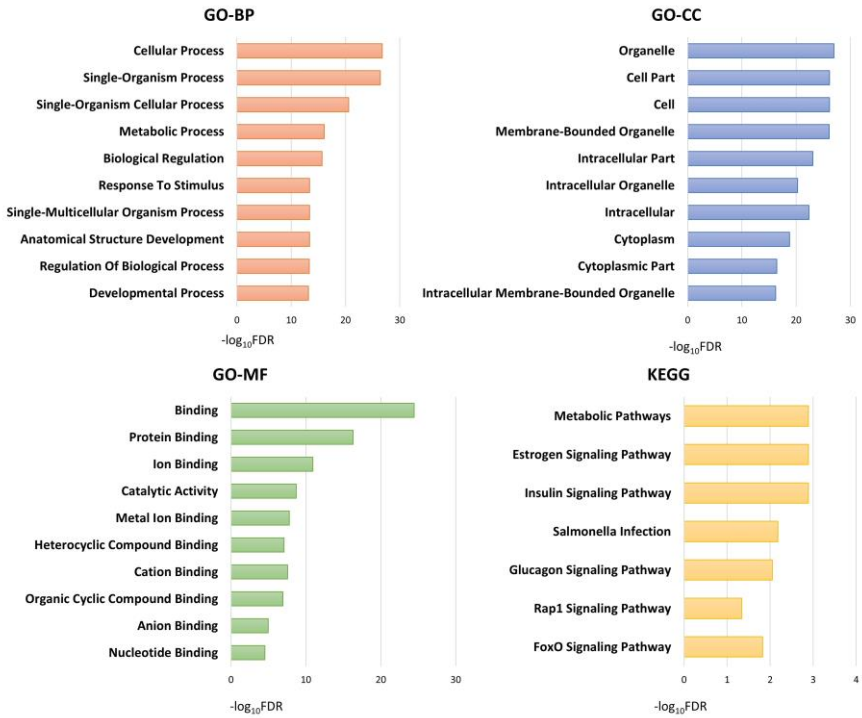

Figure 9. Top 10 significantly enriched biologic process (BP), cellular component (CC), and molecular function (MF) terms of GO and KEGG analyses of DEGs between Group 1 and Group 3.

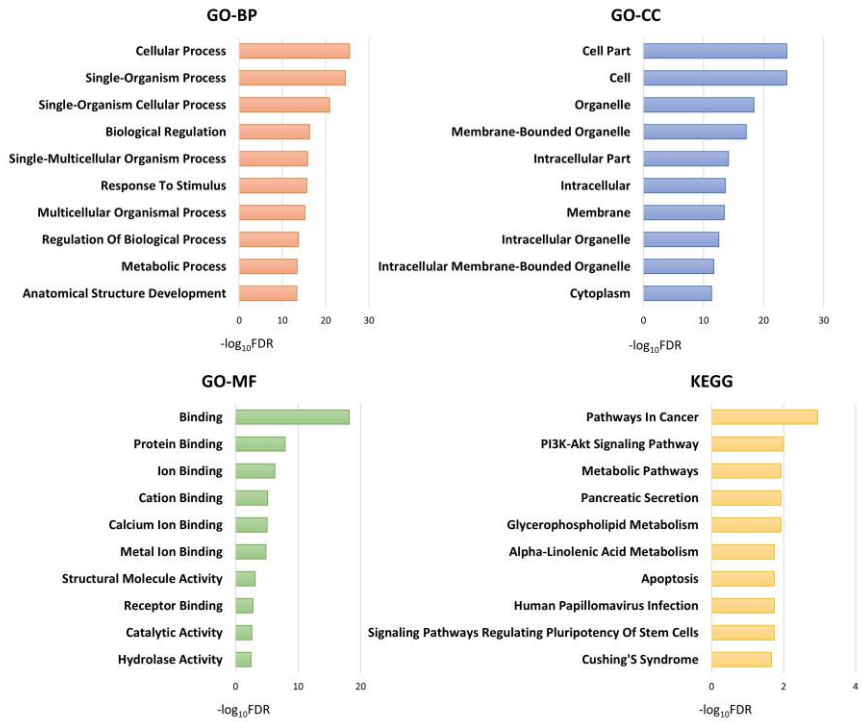

Figure 10. Top 10 significantly enriched biologic process (BP), cellular component (CC), and molecular function (MF) terms of GO and KEGG analyses of DEGs between Group 2 and Group 3.

Group 3 is Enriched in Processes Related to Cellular Adhesion and Motility

Notably, the DEGs between groups 3 and 1 or between groups 3 and 2 were enriched in cell adhesion and motility in GO analyses, including 'cell adhesion' (FDR <0.001), 'biological adhesion' (FDR <0.001), 'cell-cell adhesion' (FDR <0.001), 'regulation of cell adhesion' (FDR = 0.015), and 'cell junction' (FDR = 0.021) between groups 3 and 1 and 'cell adhesion' (FDR <0.001), 'biological adhesion' (FDR <0.001), 'cell-cell adhesion' (FDR <0.001), 'cell-cell adhesion via plasma-membrane adhesion

molecules' (FDR = 0.028), 'cell motility' (FDR = 0.045), and 'regulation of cell adhesion' (FDR = 0.046) between groups 3 and 2. These processes involved several downregulated genes (e.g., *CASK, MYO10, SPINK5, FAT2, CLCA2, CELSR1, SHC1, LGALS8, COL7A1, WNT5A, LAMA2, CDH8, NRXN3,* and *LY6D*) and upregulated genes (e.g., *WWC1, PKN1, FREM2, XBP1, CLDN4, PKP2, ANG,* and *PFN1*) in group 3 compared to the other subgroups (Figure 11). GSEA confirmed the alteration in the cellular binding/junction/migration signatures in group 3, including the diminished function of binding and the enhanced function of cell migration (Figure 11). Furthermore, GO analysis of the DEGs between groups 3 and 1 showed an enrichment in junctional complexes, such as 'adherens junction' (FDR = 0.005)' and 'anchoring junction' (FDR = 0.006). Finally, we validated the correlation of these DEGs related to cellular adhesion and motility with the prognoses of patients with urothelial carcinoma in the TCGA database. As a result, the expression levels of *CASK* (p = 0.024), *LAMA2* (p = 0.040), *LY6D* (p = 0.018), and *CLDN4* (p = 0.043) were significantly associated with the overall survival of patients with urinary bladder carcinoma (Figure 12). Finally, KEGG analysis of the DEGs between groups 3 and 2 showed an enrichment in 'pathways in cancer' (FDR = 0.001) and 'PI3K−Akt signaling pathway' (FDR = 0.010) (Figure 10).

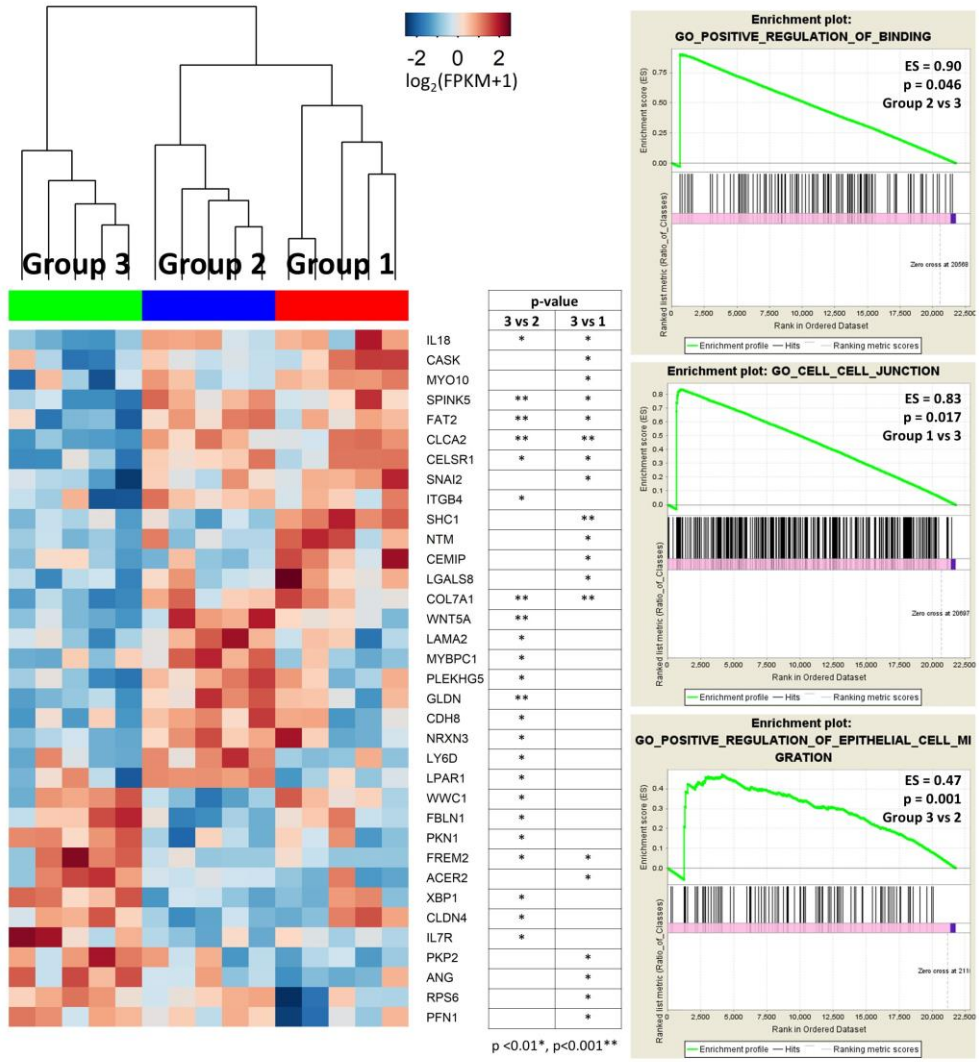

Figure 11. Heatmaps of DEGs enriched in cellular adhesion/motility processes and related GSEA results. Group 3 was enriched with biologic themes of cellular adhesion and motility ES, enrichment score.

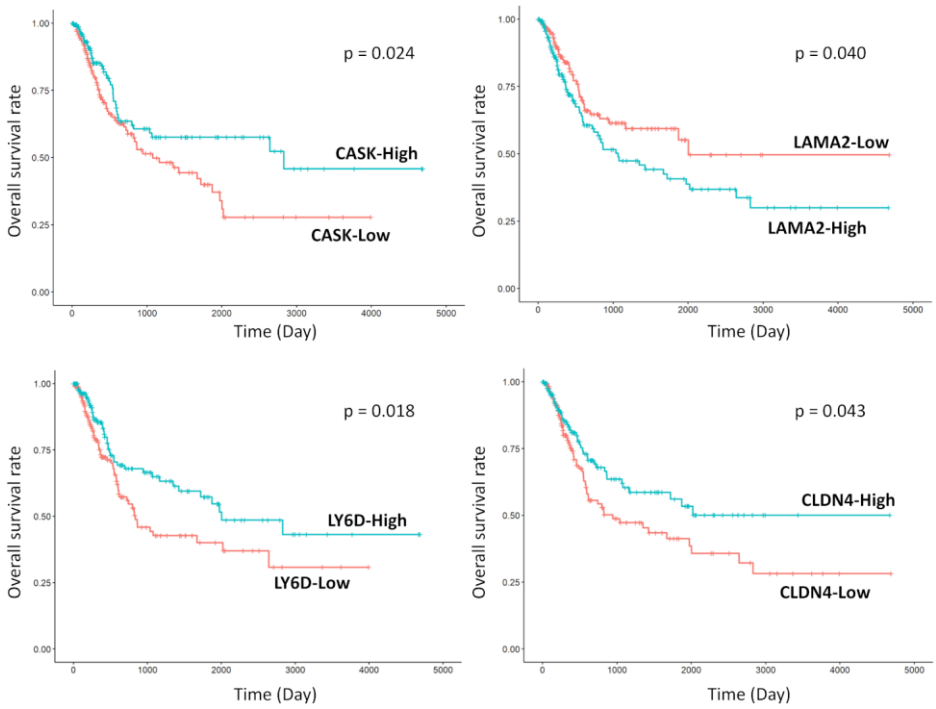

Figure 12. Survival analyses of DEGs involved in cellular adhesion/motility in the TCGA database. Four genes were significantly associated with the overall survival of the patients with UBC.

Mitogen-Activated Protein Kinase (MAPK) and Tumor Necrosis Factor (TNF) Signaling Pathways are Downregulated in Group 2 Tumors

Group 2 was enrichment in the MAPK process in in GO analysis compared to group 1 ('MAPK cascade', FDR = 0.044) and group 3 ('activation of MAPK activity', FDR = 0.009; 'regulation of MAP kinase activity', FDR = 0.012; 'MAPK cascade', FDR = 0.018; 'regulation of MAPK cascade', FDR = 0.040). In addition, KEGG pathway analysis revealed that the DEGs between groups 1 and 2 were enriched in the TNF signaling pathway (FDR = 0.014) (Figure 8). Most DEGs associated in these pathways were downregulated in group 2 tumors (Figure 13). GSEA also indicated that group 2

tumors were less reactive to MAPK and TNF signaling cascades than group 1 tumors (Figure 13).

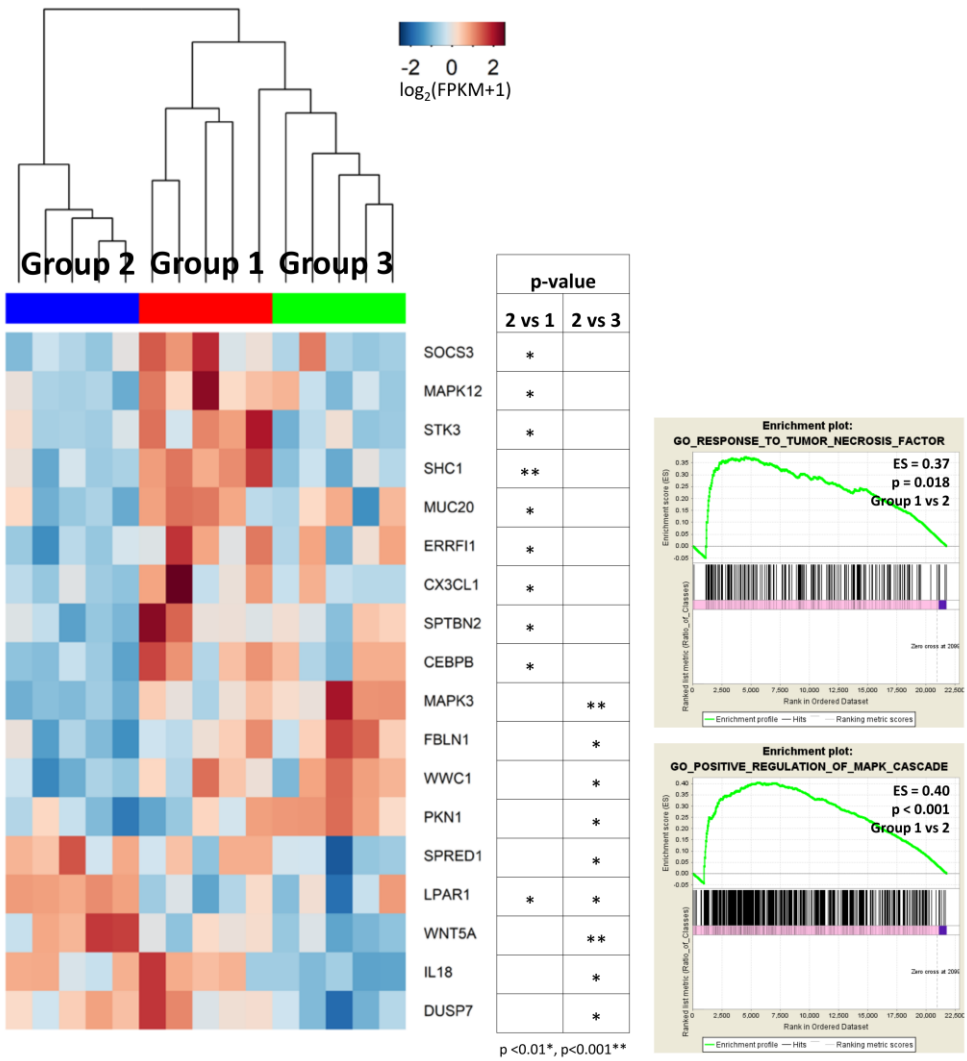

Figure 13. Heatmaps of DEGs enriched in MAPK or TNF signaling pathways, along with related GSEA results. Group 2 was characterized by low levels of genes involved in MAPK or TNF signaling pathways, except for a few genes that play regulatory roles in these pathways. ES, enrichment score.

3.6. Expression of Biologic Signature Genes of the High-Grade GEP Subgroups

To further delineate the properties of the IHC-based high-grade GEP subgroups, we evaluated the expression of known biologic markers (Figure 14) [9, 15, 17, 19, 20, 59]. Each subgroup was enriched in different types of keratin: *KRT5*, *KRT6*, *KRT14*, and *KRT15* in group 1, *KRT18* and *KRT20* in group 3, and all in group 2. The expression patterns of basal and luminal type markers clustered moderately with group 1 and group 3, respectively. Group 2 showed a modest enrichment in both subtype markers. Urothelial differentiation markers, which share some of the gene sets with those of the luminal type, showed moderate overexpression in group 3. Gene expression of the late cell cycle/proliferation signature overlapped substantially with group 3 and moderately with some specimens in the other groups, which was also differentially enriched between groups 3 and 1 ('cell proliferation', FDR = 0.042) and between groups 2 and 1 ('regulation of cell proliferation', FDR = 0.012; 'cell proliferation', FDR = 0.025) in GO analyses. Consistent with these results, the expression of genes related to the progression of early urothelial carcinoma [57] was higher in group 3 and 1 than in group 2 (Mann-Whitney U test, p = 0.028), showing a significant correlation with that of late cell cycle/proliferation genes, in that it was high in group 3 and in two samples of group 1. The two group 1 tumors with elevated levels of cell cycle/proliferation genes tended to express *KRT14* more than the other tumors, and the tumor with the highest *KRT14* level was the only one positive for CK14 by IHC staining. Gene expression of the stem cell markers known in urothelial carcinoma other than *KRT5* was not significantly different among the groups 1-3, although *CD44* and *FOXM1* were marginally overexpressed in the groups 1 and 3, respectively (Figure 15). The expression levels of

early cell cycle and epithelial-to-mesenchymal transition did not significantly overlap with the subgroups based on IHC staining.

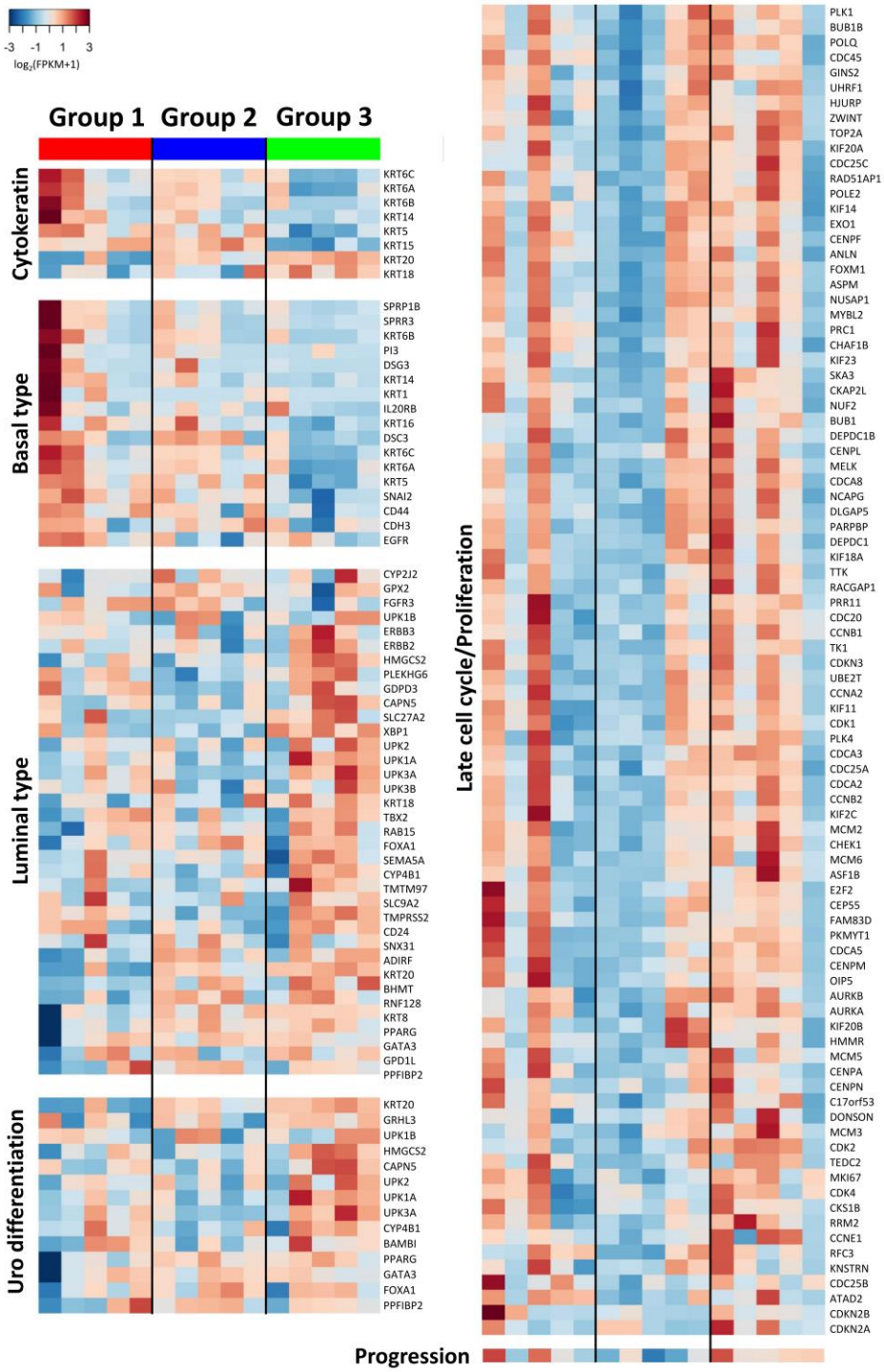
-3 -1 1 3
log2(FPKM+1)
Group 1
Group 2
Group 3
Cytokeratin
KRT6C
KRT6A
KRT6B
KRT14
KRT5
KRT15
KRT20
KRT18
Basal type
SPRP1B
SPRR3
KRT6B
PI3
DSG3
KRT14
KRT1
IL20RB
KRT16
DSC3
KRT6C
KRT6A
KRT5
SNAI2
CD44
CDH3
EGFR
Luminal type
CYP2J2
GPX2
FGFR3
UPK1B
ERBB3
ERBB2
HMGCS2
PLEKHG6
GDPD3
CAPN5
SLC27A2
XBP1
UPK2
UPK1A
UPK3A
UPK3B
KRT18
TBX2
RAB15
FOXA1
SEMA5A
CYP4B1
TMTM97
SLC9A2
TMPRSS2
CD24
SNX31
ADIRF
KRT20
BHMT
RNF128
KRT8
PPARG
GATA3
GPD1L
PPFIBP2
Uro differentiation
KRT20
GRHL3
UPK1B
HMGCS2
CAPN5
UPK2
UPK1A
UPK3A
CYP4B1
BAMBI
PPARG
GATA3
FOXA1
PPFIBP2
Late cell cycle/Proliferation
PLK1
BUB1B
POLQ
CDC45
GINS2
UHRF1
HJURP
ZWINT
TOP2A
KIF20A
CDC25C
RAD51AP1
POLE2
KIF14
EXO1
CENPF
ANLN
FOXM1
ASPM
NUSAP1
MYBL2
PRC1
CHAF1B
KIF23
SKA3
CKAP2L
NUF2
BUB1
DEPDC1B
CENPL
MELK
CDCA8
NCAPG
DLGAP5
PARPBP
DEPDC1
KIF18A
TTK
RACGAP1
PRR11
CDC20
CCNB1
TK1
CDKN3
UBE2T
CCNA2
KIF11
CDK1
PLK4
CDCA3
CDC25A
CDCA2
CCNB2
KIF2C
MCM2
CHEK1
MCM6
ASF1B
E2F2
CEP55
FAM83D
PKMYT1
CDCA5
CENPM
OIP5
AURKB
AURKA
KIF20B
HMMR
MCM5
CENPA
CENPN
C17orf53
DONSON
MCM3
CDK2
TEDC2
MKI67
CDK4
CKS1B
RRM2
CCNE1
RFC3
KNSTRN
CDC25B
ATAD2
CDKN2B
CDKN2A
Progression

Figure 14. Heatmaps of known biological signature genes extracted from previous reports, with samples arranged in each column and subgroups divided by black vertical lines. Progression (right bottom) was visualized according to the integrated score (red, high; blue, low). Uro differentiation, urothelial differentiation.

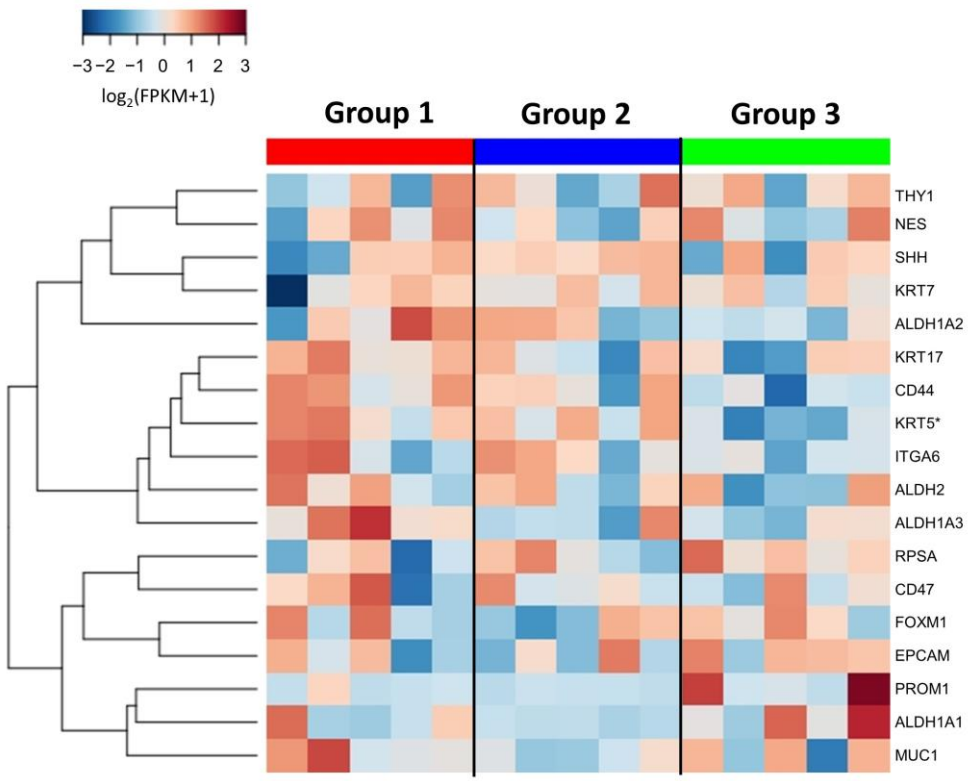

Figure 15. Stem cell marker expression of the groups 1−3 with samples arranged in the same order of the figure 14 (*:DEG).

Finally, hierarchical clustering analysis using the classifiers of MIBC (BASE47) and NMIBC (117-gene) that were previously published was applied to the present tumors (Figure 16) [15, 20]. Although they yielded gene-expression clusters that were similar to their original designs, they did not overlap with the present IHC-defined subgroups.

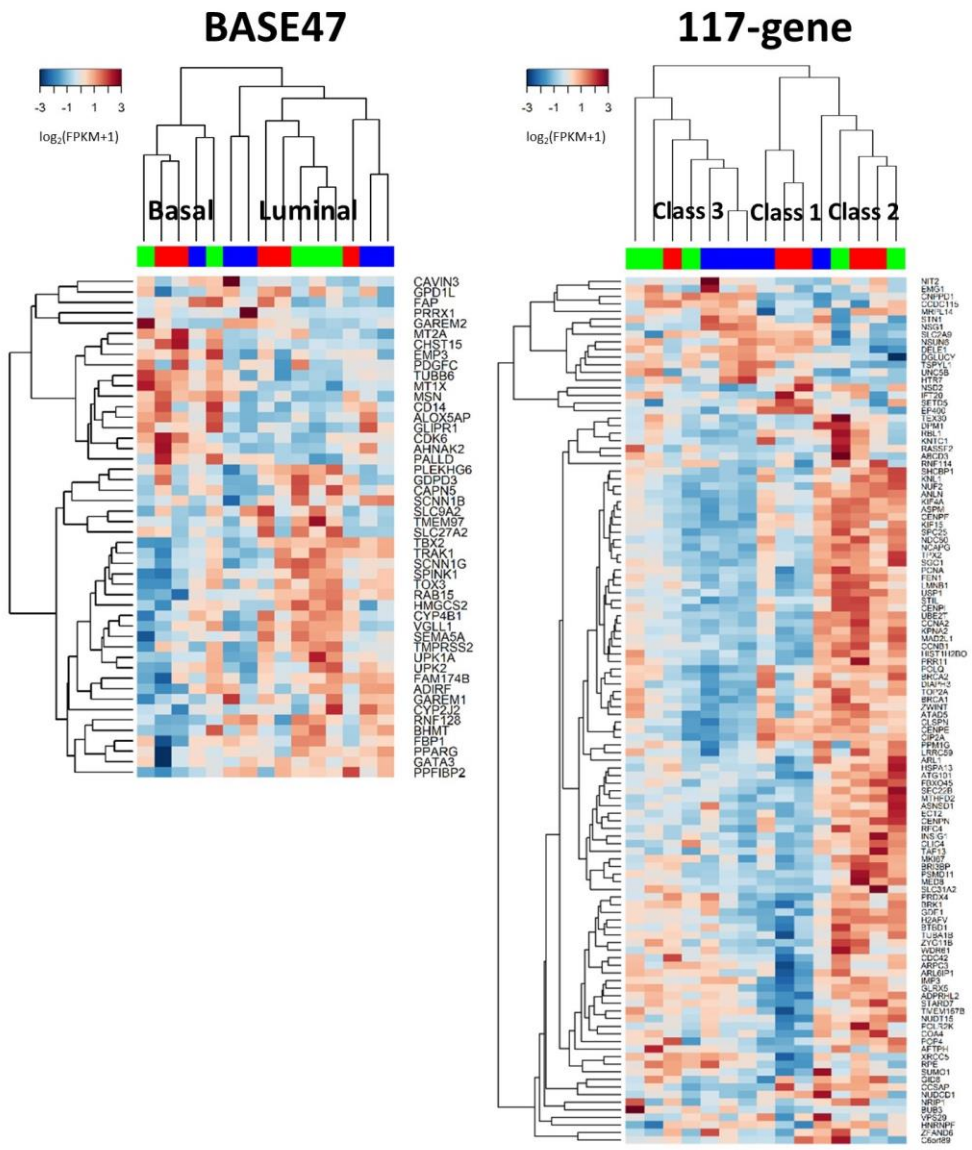

Figure 16. Unsupervised hierarchical clustering of BASE47 (left) and 117-gene (right) classifiers. All of BASE47 and 115 of 117 genes were present in the samples of this study. Color bars on the top indicate IHC-defined groups: red (group 1), blue (group 2), and green (group 3).

3.7. Genetic Features of the Low-Grade GEP Subgroups

We next determined whether low-grade tumors shared the genetic characteristics of high-grade GEP cohort showing similar immunostaining patterns. For robust statistical calculation, three samples of group 4 (samples #83, 125, and 131; Table 7) were selected and compared with group 5 tumors. The same IHC cut-off value of the high-grade GEP cohort was applied to define the low-grade tumor group. We identified 24 DEGs from a total of 22556 genes found in groups 4 and 5 (Table 10 and Figure 17). Functional analysis of these genes indicated the cyclic AMP (cAMP) signaling pathway (FDR = 0.039), which involved the downregulation of *HCAR1, HCAR2, and HCAR3* G-protein coupled receptor genes in group 5. The accumulation of intracellular cAMP activates AMP-dependent kinase (AMPK) via STK11-mediated phosphorylation. AMPK is a master regulator of cellular metabolism and tumor proliferation [60]. Notably, in group 5 tumors, *STK26* was upregulated ($p = 0.009$) but *PKIA*, a repressor of AMPK activity, was downregulated ($p = 0.008$). In addition, group 5 tumors exhibited higher *ANP32C* expression ($p = 0.005$) and lower *CD82* expression ($p = 0.008$) than group 4. *CDH3* ($p = 0.001$) and *ITGB5* ($p = 0.002$) were significantly enriched in group 4 tumors.

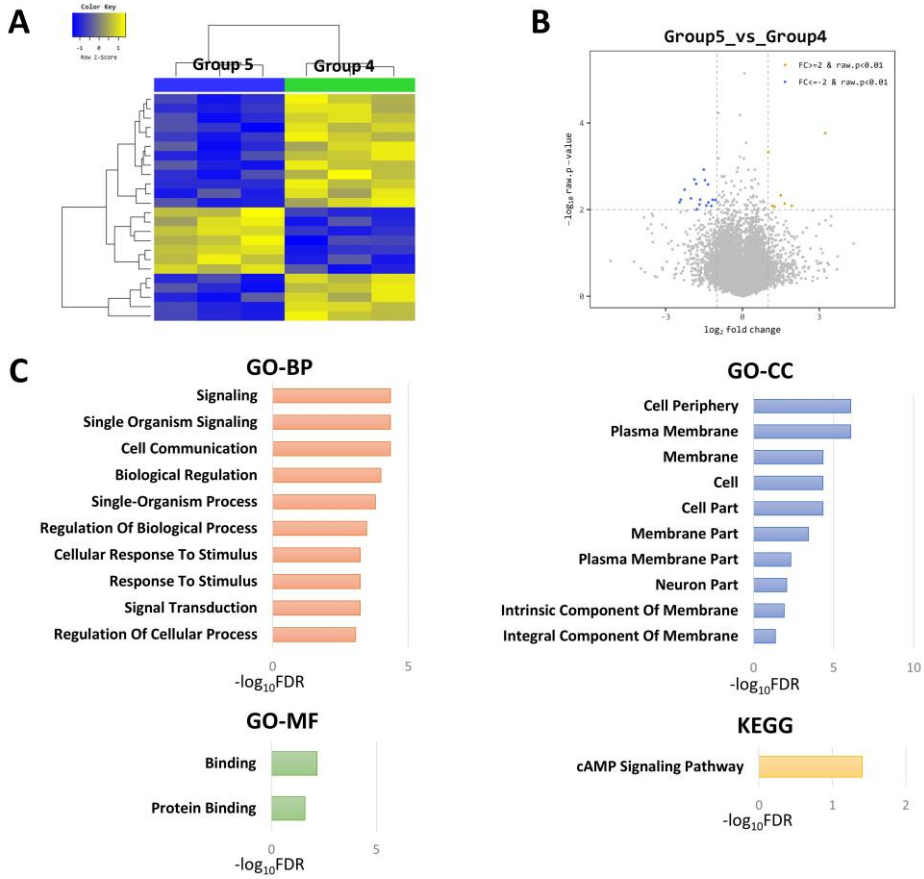

Figure 17. DEGs and functional enrichment between the low-grade GEP subgroups. **(A)** Unsupervised hierarchical clustering of the low-grade groups 4 and 5. **(B)** Volcano plots depicts 24 DEGs which are upregulated in group 5 (blue) or group 4 (yellow). **(C)** Top 10 significantly enriched biologic process (BP), cellular component (CC), and molecular function (MF) terms of GO and KEGG analyses of DEGs between group 4 and group 5.

3.8. Grade-Related Genetic Differences and Their Relationship to CK5/6 and CK20 Expression

We aimed to determine whether grade-related genetic differences were differentially affected by CK5/6 and CK20 expression. Genetic comparisons between groups 1 and 4 (n = 5 for each; Figure 18) and groups 2 and 5 (n = 3 for each; Figure 19)

were performed. As a result, 59 DEGs were found between CK5/6-high/CK20-low groups 1 and 4, which gave rise to the functional enrichment of apoptosis-associated GO-BP ('positive regulation of apoptotic signaling pathway', FDR = 0.023; 'positive regulation of extrinsic apoptotic signaling pathway', FDR = 0.031; 'extrinsic apoptotic signaling pathway', FDR = 0.034) and 'necroptosis' in KEGG pathway (FDR <0.001; Figure 18C). These themes were associated with several DEGs, including *PYCARD* (p = 0.003), *SLC25A6* (p = 0.008), *HIST1H2AE* (p = 0.006), *HIST2H2AA3* (p = 0.001), *TLR4* (p = 0.009), *STK3* (p = 0.005), and *ITM2C* (p = 0.007). On the other hand, only eight DEGs were found between groups 2 and 5, the CK5/6-high/CK20-high expressing groups (Table 10), with no significant GO or KEGG functional indications (Figure 19).

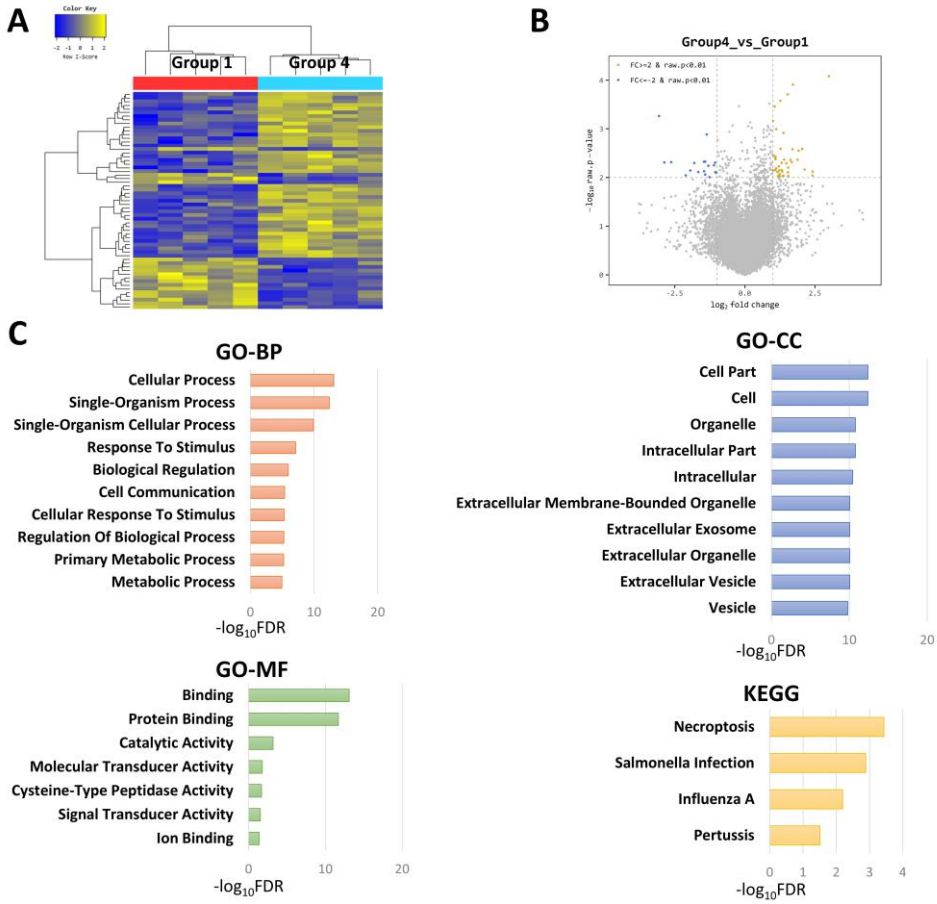

Figure 18. DEGs and functional enrichment between groups 1 and 4 tumors. (A) Unsupervised hierarchical clustering. (B) Volcano plots depicts 59 DEGs which are upregulated in group 4 (blue) or group 1 (yellow). (C) Top 10 significantly enriched biologic process (BP), cellular component (CC), and molecular function (MF) terms of GO and KEGG analyses of DEGs between group 1 and group 4.

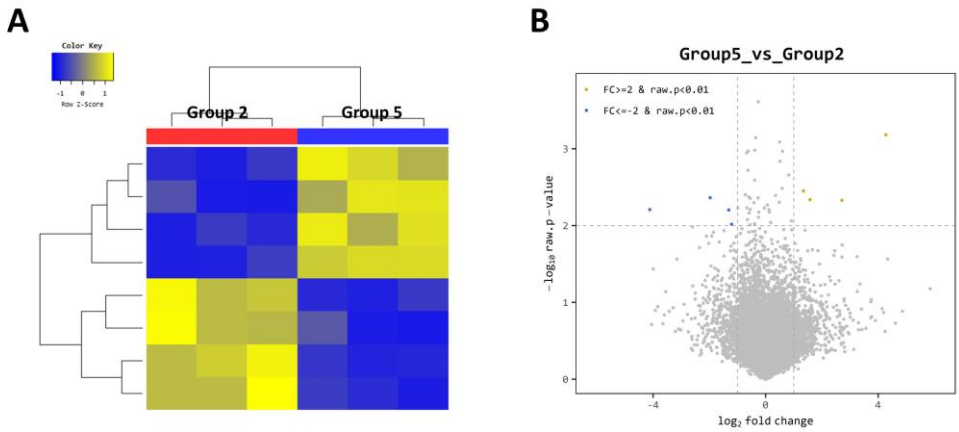

Figure 19. DEGs and functional enrichment between groups 2 and 5 tumors. **(A)** Unsupervised hierarchical clustering. **(B)** Volcano plots depicts eight DEGs which are upregulated in group 5 (blue) or group 2 (yellow).

3.9. Verification of the genetic signatures related to basal and luminal protein expression

As we discovered that the signatures of adhesion and late cell cycle/proliferation were significantly different among groups 1–3, we aimed to validate these functions in the prognosis cohort using IHC staining for P-cadherin (n = 170), E-cadherin (n = 172), and Ki-67 (n = 185). P-cadherin positivity (Figure 20A) and a normal pattern (Figure 20B) were observed in 25 (14.7%) and 33 (19.4%) patients, respectively, both of which made up the 'Other' group. Consequently, negative staining for P-cadherin was associated with negativity to CK5/6 (p = 0.006) and CD44 (p = 0.076) (Figure 20C). In addition, P-cadherin showed an interrelationship with the CK5/6-CK20 classification (p = 0.029) (Figure 20D). However, the semiquantitative measurement of P-cadherin, leaving the pattern out of consideration, showed no correlation with CK5/6, CD44, and CK20 expression (Figure 20E). On the other hand, E-cadherin was diffuse-positive in most

patients with occasional accentuation in the luminal cell layer (Figure 21A, B). Although no significant association with individual CK5/6, CD44, and CK20 expression was found in E-cadherin staining (Figure 21C), high E-cadherin expression tended to be observed in CK5/6-negative/CK20-positive tumors (Figure 21D). Furthermore, tumors showing CK5/6 negativity or CK20 positivity, either assessed separately ($p < 0.001$, CK5/6; $p = 0.012$, CK20; Figure 22A) or together ($p < 0.001$; Figure 22B), had a higher proliferation index than the other tumors.

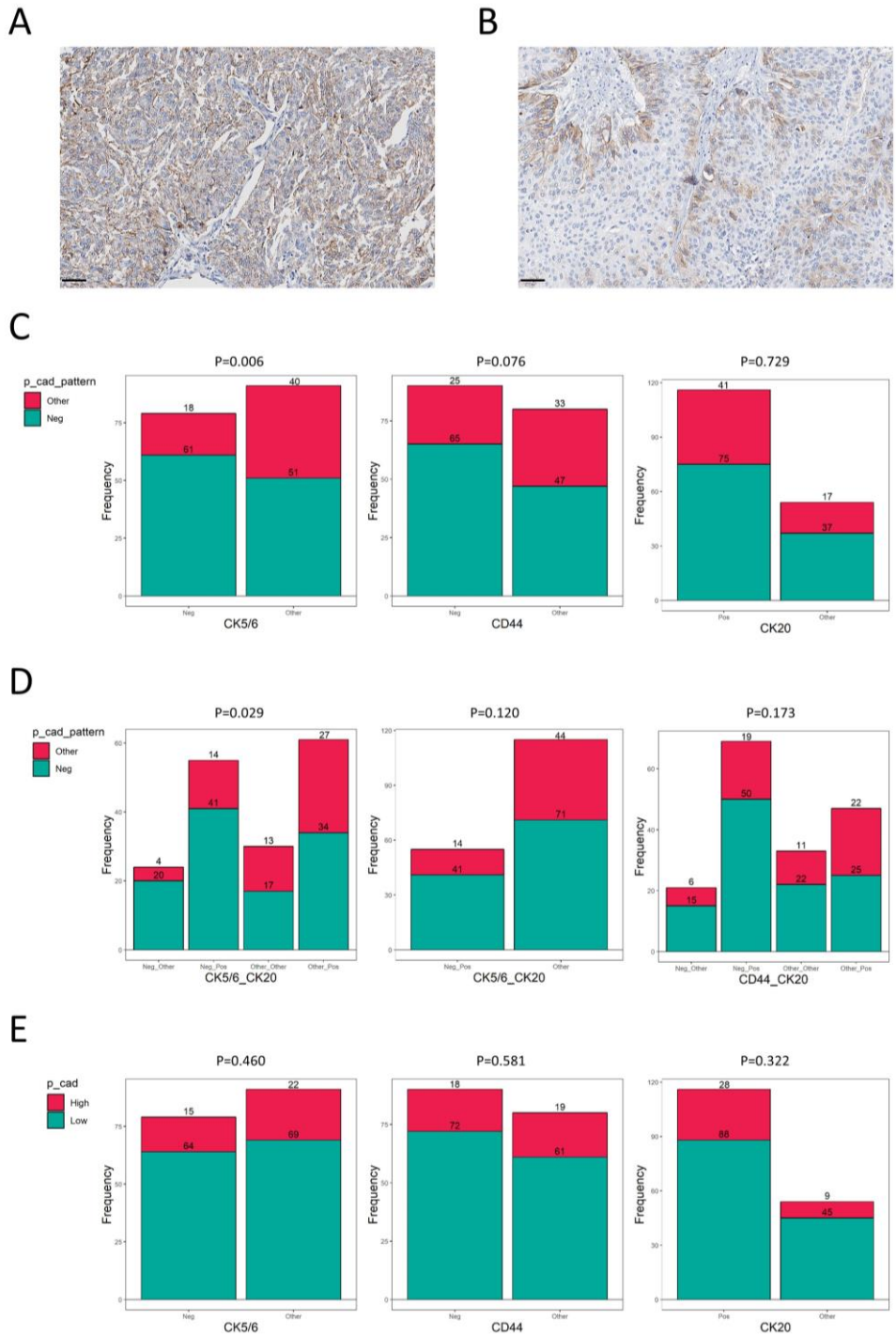

Figure 20. P-cadherin expression associated with CK5/6, CD44, and CK20. **(A)** Positive and **(B)** normal-pattern of P-cadherin expression was regarded as 'Other' category (scale bar = 50 ㎛). **(C-D)** Pattern-wise P-cadherin

categories associated with two-tier CK5/6, CD44, and CK20 expression and their combinations. CK5/6-Neg/CK20-Pos (luminal-like phenotype) was separated from the others in the middle. **(E)** Cutoff-based measurement of P-cadherin expression was not statistically related to CK5/6, CD44, and CK20 expression. Statistical analyses were performed by Rex (Version 3.0.3, RexSoft Inc., Seoul, Korea).

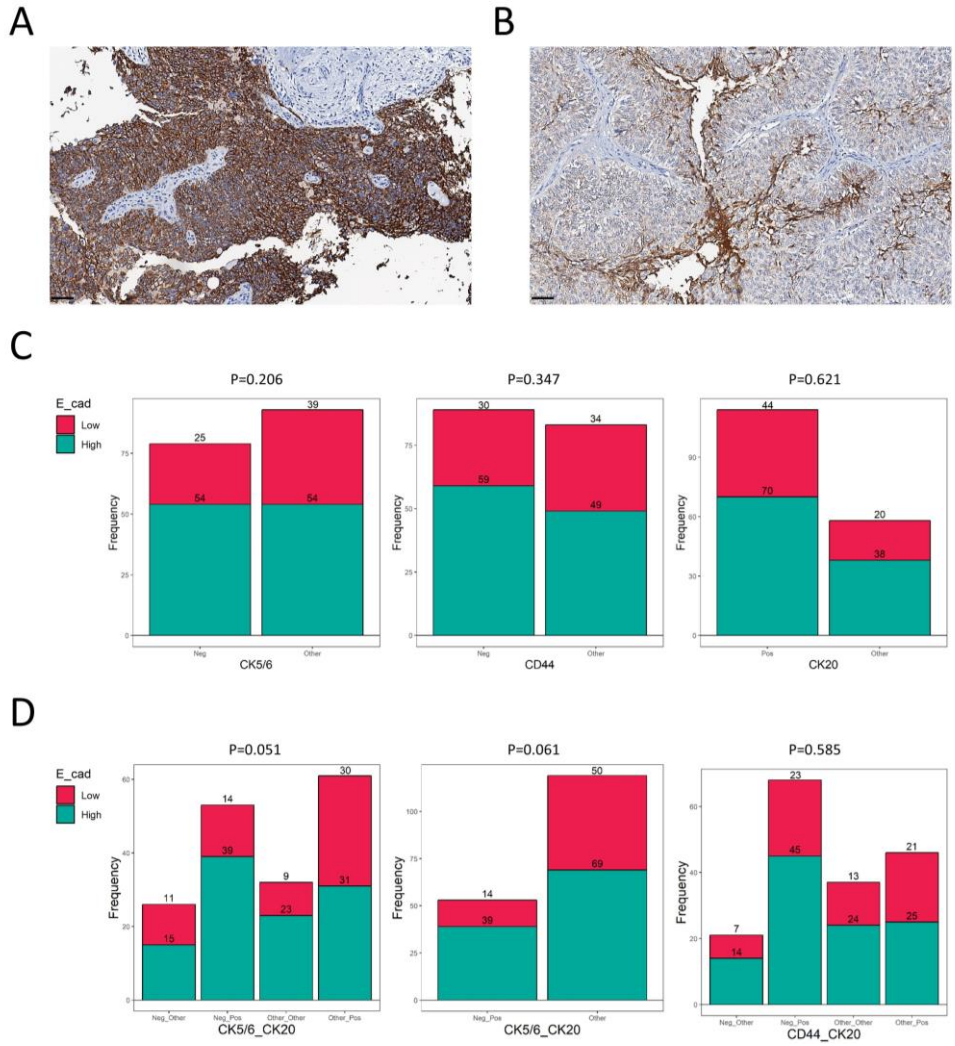

Figure 21. E-cadherin expression associated with CK5/6, CD44, and CK20. **(A)** 70-100% of tumor (score = 4) was strongly (score = 3) stained for E-cadherin. **(B)** 50-70% of tumor (score = 3) expressed E-cadherin with moderate (score = 2) intensity with luminal side accentuation (scale bar = 50 ㎛). **(C-D)** Semiquantitative expression of E-cadherin associated with two-tier CK5/6, CD44, and CK20 expression and their combinations. CK5/6-Neg/CK20-Pos (luminal-like phenotype) was separated from the others in the middle. Statistical analyses were performed by Rex (Version 3.0.3, RexSoft Inc., Seoul, Korea).

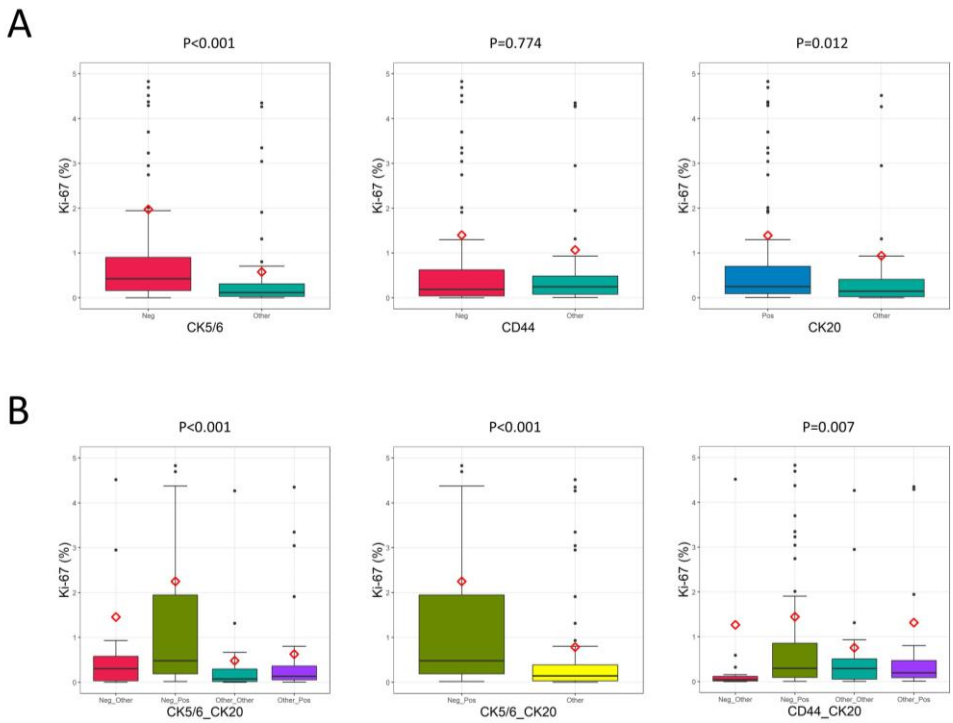

Figure 22. Ki-67 proliferation index of NMIUTUC. Expression in the prognosis cohort divided by the **(A)** two-tier classifications **(B)** or combined criteria. CK5/6-Neg/CK20-Pos (luminal-like phenotype) was compared from the others in the middle of the lower row. P-values calculated by Mann-Whitney U test in **(A)** or by Kruskal-Wallis test in **(B)**. Statistical analyses were performed by Rex (Version 3.0.3, RexSoft Inc., Seoul, Korea).

3.10. Differential Protein Expression between Early and Advanced Urothelial Carcinoma

We further analyzed the differential expression of basal and luminal markers, cadherins, and Ki-67 between NMIBC and MIBC using data published by the Lund group [9, 10]. The intrinsic subtypes of NMIBC exhibited distinct immunophenotypes of these proteins from those of MIBC (Figure 23). For example, the GU subtype of NMIBC appeared to express lower CK5/6 and higher CK20 than the UroA subtype of NMIBC. The UroB and SCCL

subtypes of NMIBC showed comparable levels of CK20 to the other subtypes. CK14 was highly expressed in the UroB subtype of NMIBC, but it was similar among the other subtypes. However, the expression of these basal and luminal markers was different among the genetic subtypes of MIBC (Figure 23). Likewise, P-cadherin (*CDH3*) and E-cadherin (*CDH1*) were relatively stable among the genetic subtypes of NMIBC compared to MIBC. Nevertheless, the Ki-67 index was markedly high in the GU and SCCL subtypes of both NMIBC and MIBC (Figure 23). Because the UroA and GU subtypes accounted for 92% of NMIBCs of the Lund cohort, a direct comparison between the two subtypes was conducted: GU had a higher Ki-67 proliferative index than UroA ($p < 0.001$).

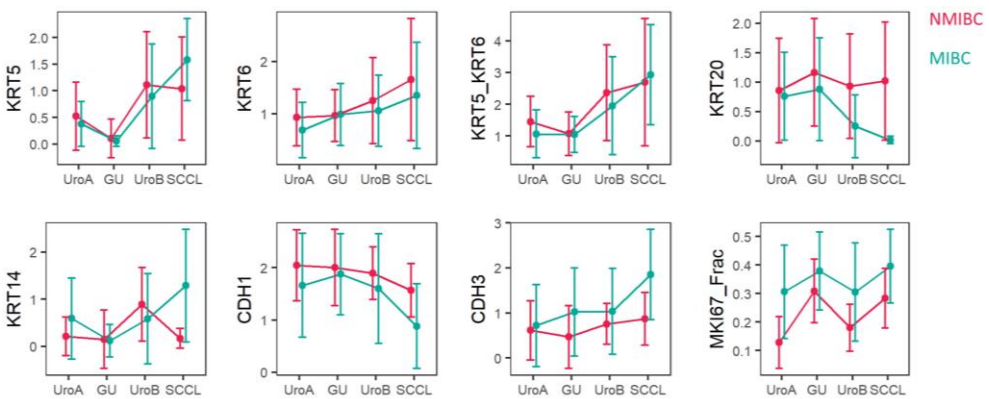

Figure 23. Immunohistochemical staining for selected proteins across the Lund molecular subtypes divided by muscle-invasiveness. Mean (dot) with standard deviation (bars) of the expression levels of each marker is separately drawn as red (NMIBC) and green (MIBC). All IHC staining was measured as a multiplied intensity (0-3) and extent (0-100% by 10%), except for Ki-67 that was recorded with positive tumor cell fraction. The y-axis of KRT5_KRT6 is the sum of KRT5 and KRT6 values. Statistical analyses were performed by Rex (Version 3.0.3, RexSoft Inc., Seoul, Korea).

Chapter 4. Discussion

CK5/6, CD44, and CK20 are differentially expressed among the molecular subtypes of MIBC [9, 17, 18, 21] and NMIBC [20], and their prognostic values have been demonstrated in the urinary bladder [10, 15, 16, 35]. In this study, we classified papillary NMIUTUC using IHC staining for CK5/6 and CK20 as surrogate markers of the basal and luminal-like subtype, respectively, and investigated the genetic characteristics of the subgroups. To minimize molecular divergence between low- and high-grade urothelial carcinoma [3], transcriptional analyses were conducted separately on low and high-grade tumors. To ensure the protein expression of the RNA-seq samples, the expression levels of CK5/6 and CK20 were determined based on IHC staining on adjacent tissue with strict cut-offs [11]. Consequently, the mRNA levels of basal/intermediate and luminal keratins overall concurred with the IHC staining profiles of CK5/6 and CK20 in the high-grade GEP cohort [61]. The prognostic significance of the IHC criteria applied to the GEP cohort was reviewed in the prognostic cohort: CK5/6-low/CK20-high high-grade NMIUTUC patients had marginally grave PFS.

We first focused on the risk stratification of papillary NMIUTUC using basal (CK5/6 and CD44) and luminal (CK20) markers in the prognosis cohort. This cohort included 211 patients which is the largest number of patients among studies on papillary NMIUTUC to date. We demonstrated that IHC staining for CK5/6, CD44 and CK20 was significantly associated with PFS. CK5/6 expression was also associated with CSS. In combination with IHC markers, the luminal-like subtypes (CK5/6-negative/CK20-positive and CD44-negative/CK20-positive) showed distinctly poor outcomes. These results contradicted the prognosis of such subtypes of MIBC [4, 9, 10, 12, 15, 17, 59]. Moreover, CK5/6

negativity was revealed as a strong independent prognostic factor for shorter PFS and CSS, and its addition improved the prognostic prediction of the baseline models. Because of the limited number of patients and survival-related events, CD44 negativity and CK20 positivity may have indicated poor CSS, although the prognostic differences were not great enough to be statistically significant. Consistent with the prognosis, we found that luminal-like CK5/6, CD44, and CK20 expression was associated with high-risk clinicopathological characteristics, including WHO/ISUP grade and IVR, in the prognosis and GEP cohorts. These conflicting prognostic results of non-muscle invasive urothelial were previously reported in the upper and lower urinary tracts [4, 20, 32]. High *KRT20* mRNA expression is associated with high tumor grade, and low *KRT5* mRNA expression is indicative of a high Ki-67 index, poor relapse-free survival, poor PFS, and poor CSS of pT1 NMIBC [32]. An increase in IHC staining for CK20 was predictive of the recurrence of pTa/T1 papillary NMIBC [35]. Previously, Sikic et al demonstrated the prognostic role of CK5 and CK20 expression in UTUC with all T stages [33], showing that the CK5-negative/CK20-positive luminal-like subtype was an independent poor prognostic factor for CSS in UTUC [33]. Therefore, we hypothesize that papillary NMIUTUC may have subtypes that are clinicopathologically and prognostically akin to those of NMIBC and that these subtypes may have similar genotype-phenotype correlations [12, 20, 35]. It was found that UTUC shares similar genetic expression profiles and genetic alterations with UBC, especially in early stages [27]. MIBC has been suggested to be able to arise either from CIS or from papillary lesions with CIS-like molecular progression [20, 22]. As the high-risk molecular subtypes of NMIBC were enriched in CK20 [20], high CK20 expression and low CK5/6 expression were also characteristics of CIS in the urinary bladder. We hypothesize that papillary high-grade NMIUTUC can progress with the acquisition of CIS-related

CK20 and the loss of a differentiation-related normal pattern of CK5/6. Accordingly, CK5/6 negativity and CK20 positivity may characterize the high-risk subgroups of papillary NMIUTUC that have CIS-like phenotypes and possess the potential for disease progression. Furthermore, this implies that the molecular switch to a CIS-like pathway may precede invasion in papillary NMIUTUC. CD44 is assumed to be representative of cancer stem cell and tumor-initiating abilities, and is associated with unfavorable outcomes in patients with MIBC [22, 23].

Basal and luminal subtypes, reminiscent of those of breast cancer, are the two most common subtypes shared by many genetic studies, although they are named diversely [9-13, 15, 16, 18, 62]. It has been demonstrated that the MIBC subtypes differ in their prognosis [8, 9, 15-17, 23], response to chemotherapy or immunotherapy [16, 28, 30, 63], and targetable gene alterations [9, 17, 18, 64], emphasizing their practical implications. Similarly, the genetic subtypes of NMIBC show distinct prognoses and mutation profiles of genes involved in structural organization, chromatin handling, the MAPK-ERK pathway or the ERBB pathway, which may be directed to subtype-based treatments [20]. The simple approach (basal versus luminal subtypes) also reflects the differentiation status and the cellular origin of the tumor [15, 18, 23, 65]. Therefore, we further investigated the molecular details of papillary NMIUTUC associated with CK5/6 and CK20 expression. CK5/6 and CK20 were adopted as surrogate markers of basal and luminal proteins, respectively, by IHC staining. CK immunostaining is a widely accessible and well-established technique. In addition, CK expression may indicate the differentiation status of urothelial carcinoma [66] and in turn represent the molecular classifications of NMIUTUC [4, 16, 32-34]. For example, CK5/6, which is enriched in the basal and progenitor cells of the normal urothelium, is conserved in basal type urothelial carcinoma that supposedly obtains a basal cell-like, undifferentiated phenotype [22, 23]. On

the other hand, high levels of CK20 expression in luminal type urothelial carcinoma, as normally observed in terminally differentiated urothelium, are suggested to indicate a neoplasm in well-differentiated states [22]; therefore, IHC staining for these proteins can be a useful surrogate marker pertinent to the molecular subtypes of urothelial carcinoma [9, 15, 16, 18]. For instance, genetically defined basal subtype MIBC is highly reactive to CK5/6 and CD44 but minimally reactive to CK20 in IHC staining, and luminal subtype tumors show the opposite IHC result [16]. In addition, CK5/6 and CK20 IHC expression is negatively correlated across pT3 UBC specimens [10]. In addition to CK5/6 and CK20, other proteins have also been suggested to be genetically correlated using IHC staining. For example, CK14-positive, CK5/6-positive, GATA3-negative, and FOXA1-negative expression indicates the BASQ subtype of MIBC [7]. However, we showed that this was not readily applicable to papillary NMIUTUC because of the scant positivity for CK14 and negativity for GATA3 or FOXA1. Most papillary NMIUTUC exhibited consistently high GATA3 and FOXA1 expression, both in the prognosis and GEP cohorts. CK14 expression may be the earliest indicator of the undifferentiated phenotype of urothelial carcinoma [23]. As we specified CK14 expression <10% into a more detailed scale in the prognosis cohort, CK14 positivity which was defined by >0% was significantly associated with CK5/6-negative and CK20-positive tumors, which directly contradicted the previous basal/luminal paradigm of CK5/6 and CK20 expression. Therefore, the usefulness of BASQ markers in papillary NMIUTUC needs further verification, but CK14 may indicate advanced disease formation, even in papillary NMIUTUC. At the mRNA level, CK14 was highly expressed in one of the high-grade GEP cohort samples and was one of the characteristics of the SCC-like subtype [10]; therefore, CK14 immunostaining could be used for subclassifying CK5/6-high/CK20-low tumors.

Transcriptional analysis across the high-grade GEP subgroups revealed 308 differentially expressed genes. Functional analyses of the genes identified cell adhesion as a common process differentially enriched in group 3 compared to the other groups, which could explain its high-risk phenotype. Late cell cycle/proliferation signatures were also enriched in group 3 and in some of the other groups, which may be used as prognostic biomarkers complementary to CK5/6 and CK20. Group 2, characterized by low levels of genes associated with the MAPK and TNF signaling pathways, was hypothesized to represent the least cancerous subtype considering its normal urothelium-like IHC pattern. Previous gene-expression classifiers of urothelial carcinoma assigned the present specimens in such a way that was independent of their IHC staining patterns [20, 31], which implies that the molecular frameworks determining the CK5/6 and CK20 immunophenotypes and their associated clinical behaviors are not limited to the basal versus luminal paradigm in NMIUTUC. Consistent with their correlations in MIBC, we demonstrated that early papillary urothelial carcinomas showing CK5/6-high/CK20-low and CK5/6-low/CK20-high immunoreactions were enriched in basal type and luminal type/urothelial differentiation signature genes, respectively, in this study and in the previous study [9]. Cell adhesion and motility, in agreement with other reports [9, 10], was the distinct function differentially enriched between groups 3 and 1 and between groups 3 and 2. Of note, several genes involved in major cell adhesion complexes, including the adherens junction (*CDH8* and *CLCA2*) [67-69], the desmosome (*SPINK5*) [70], focal adhesion (*MYO10* and *SHC1*) [71, 72], and the basement membrane and their interaction with it (*LGALS8*, *LAMA2*, *COL7A1,* and *ITGB4*) [73, 74], with the exception of the tight junction (*CLDN4*) [75], are downregulated in group 3 tumors and they would promote the migration and invasion of cancer cells. As suggested by Sjodahl et al [9], the GU subtype, the major high-risk subtype of NMIBC,

has low expression levels of genes engaged in major adhesion structures, such as the adherens junction, the desmosome, and gap junctions, with the exception of the tight junction (*CLDN*s), compared to the UroA subtype. The GU subtype showed IHC staining for CK5 and CK20, similar to that of group 3 in this study [9, 10]. In line with this finding, functional enrichment analyses and GSEA supported the premise that the loss of non-apical cell-to-cell and cell-to-matrix adhesions is a major determinant of the high-risk group 3 subtype [68]. In addition, the non-canonical Wnt/PCP pathway (*FAT2*, *CELSR1*, and *WNT5A*), which is known to suppress early-stage cancers by regulating cell adhesion or migration, was downregulated in group 3 [76]. This was also substantiated by the results showing that alterations in the genes observed in group 3, including the downregulation of *CLCA2* [69], *SPINK5* [77], *SHC1* [72], *LGALS8* [74], *COL7A1* [73], *FAT2* [76], *WNT5A* [76], and *NRXN3* [78] and the upregulation of *CLDN4* [75], *PKN1* [79], *FREM2* [80], *XBP1* [81, 82], *PKP2* [83], *ANG* [84], and *PFN1* [85], have been reported to promote invasion of and to be associated with the poor prognoses of urothelial carcinoma and other various malignancies. We also aimed to validate the prognostic values of these genes in the independent TCGA urinary bladder carcinoma cohort. However, their prognostic consequences were inconsistent, and a few of the genes altered in group 3 even predicted longer survival in the TCGA database. This result, in conjunction with those from previous studies [67, 73, 76, 85], suggests that adhesion molecules may act differently in advanced disease, which accounted for 99.2% of the TCGA cohort. In the treatment of NMIUTUC, adhesion molecules differentially expressed among the present subgroups serve as candidates for targeted therapy [76, 86], which warrants deeper investigation. Cadherin is one of the major adhesion molecules in the epithelium and the differential expression of P- and E-cadherins, or cadherin switching has been suggested to regulate urothelial cancer

progression [67, 87]. P-cadherin is normally expressed in the basal layer of the urothelium and is one of the basal markers of breast cancer and UBC [13, 16] Likewise, the P-cadherin expression pattern shows a clear association with CK5/6 and a trend with CD44. Although robust statistical analysis was hindered by the patient number, we showed a tendency of high IHC staining for E-cadherin in CK5/6-negative/CK20-positive tumors. Cadherin switching is a hallmark of the EMT process, which accompanies increased P-cadherin and decreased E-cadherin expression that is associated with the advanced states of MIBC [67]. Low reactivity to P-cadherin and CK5/6 may together indicate the progression of papillary NMIUTUC in a way different from that of EMT. In short, P- and E-cadherins are associated with the basal-like and luminal-like phenotypes, respectively, which may be linked to the adhesion signatures differentially enriched in groups 1-3 of the GEP cohort. Furthermore, we hypothesize that cadherins play roles in papillary NMIUTUC distinctively from those in advanced tumors.

Late cell cycle/proliferation signature genes that were predominantly upregulated in most group 3 tumors, as well as in a few samples of the other subgroups, may also mediate the high-risk phenotypes of these tumors. This result was consistent with the molecular subtypes of NMIBCs with poor prognoses suggested in previous reports: the GU, UroB, and SCC-like subtypes share elevated levels of these signature genes compared to the UroA subtype [9, 10, 20, 59]. High levels of late cell cycle/proliferation markers have also been observed in aggressive subtypes of MIBC [20, 59]. In addition, IHC staining for cyclin B1, as a late cell cycle marker, when it was combined with CK5 expression and tumor histology, stratified the risk of the progression of patients with T1 NMIBC [10, 12]. Moreover, we showed that the enrichment pattern of late cell cycle/proliferation correlated closely with that of the NMIBC progression signature [57]. Therefore, it is feasible to

postulate that the enhanced cell proliferation of group 3 supports its aggressive phenotype, which is possibly obtained by bypassing cell cycle checkpoints [9]. In addition, the different molecular subtypes of NMIBC that showed diverse clinical behaviors (UroA, UroB, and SCC-like) exhibited CK5/6-high/CK20-low expression [9, 10, 12], which was representative of group 1 tumors. Therefore, group 1 is suggested to contain prognostically heterogeneous tumors. Considering the remarkable overexpression of these signature genes observed in certain tumors of group 1, genes related to late cell cycle/proliferation are expected to be used as risk substratification markers within group 1 tumors. Likewise, they would also be useful for group 2 tumors that show modest discrepancies in the level of the late cell cycle/progression signature within the subgroup. This seemed analogous to the UroA and UroB subtypes, as both subtypes can be positive for CK5/6 and CK20, similar to group 2 tumors [9, 10]. Using IHC staining for Ki-67, we confirmed high proliferative activities in CK5/6-negative and CK20-positive tumors, and those tumors met both characteristics. The Ki-67 index relative to the genetic subtypes was stable in NMIBC and MIBC of the Lund cohort, which further underlies its usefulness. Moreover, CD44 expression failed to show a close association with Ki-67 index, which implies that CK5/6 is a more useful basal cell marker than CD44. As above mentioned, CD44 expression represents a basal/stem cell population of urothelial carcinoma and it was confined to the basal-like class 3 of NMIBC [23]. Other cancer stem cell markers (e.g., *ALDH1A1*, *ALDH1A2*, *PROM1*, *NES*, and *THY1*) were specifically highly expressed in the luminal-like class 2 tumors [23]. On the other hand, we identified that *CD44* and *FOXM1* was modestly upregulated in the group 1 and group 3, respectively. However, expression of other genes related to urothelial carcinoma stem cells, including *THY1*, *NES*, *ALDH1A1*, *ALDH1A2*, *ALDH1A3*, *ALDH2*, *PROM1* was rather evenly distributed across the present groups 1-

3. These results suggest that a cancer stem cell signature may not be specifically regulated by the groups of papillary NMIUTUC determined by CK5/6 and CK20 expression.

We separated tumors with normal pattern IHC staining from those with positive and negative patterns in the prognosis cohort. Interestingly, normal pattern staining for CK5/6, CD44, or CK20, irrespective of the staining quantity, indicated a favorable prognosis in accordance with previous studies on papillary urothelial carcinoma [35]. A normal pattern staining was previously described in UBC [35, 38]. A low-risk group of T1 UBC identified by Patschan et al [12] showed CK5 reactivity limited to basal cells, which corresponds to a normal pattern of CK5/6 expression in this study. Desai et al [35] revealed that normal patterns of CD44 and CK20 staining were associated with a low grade in early-stage UBC. Consistent with these results, we found a normal pattern staining in papillary NMIUTUC as a distinct subgroup with an intermediate-to-low risk. According to the staining proportion, most normal pattern tumors would have been placed in the high-risk groups (negativity to CK5/6 and CD44 or positivity to CK20). Therefore, recognizing a normal pattern of IHC staining in papillary NMIUTUC is necessary to complement approaches based on the staining quantity. Reminiscent of the normal staining pattern, accentuated IHC patterns with intact CK stratification were observed in group 2 of high-grade GEP cohort [35]. Similar IHC staining for CK5/6 and CK20 was also reported in the low-risk UroA subtype, which was well differentiated at the gene-expression level [9, 10, 12]. According to the transcriptional investigation, genes involved in urothelial differentiation as well as in basal and luminal type signatures were moderately enriched in group 2. Moreover, most DEGs enriched in the MAPK or TNF pathway were expressed at lower levels in group 2 than in the other groups, except for *SPRED1, DUSP7*, and *WNT5a*, which downregulate the MAPK and TNF signaling cascades, respectively

[88, 89]. MAPK activation is an important mediator of cellular transformation induced by constitutively activated *FGFR3*, which plays crucial roles in the tumorigenesis of papillary urothelial carcinoma [5]. *FGFR3* mutations, which are found in 60–70% of non–invasive papillary urothelial carcinomas, and structural aberrations of *FGFR3*, including *FGFR3–TACC3* fusion, activate the MAPK pathway, which causes urothelial hyperplasia and eventually the early carcinogenesis of this tumor [24]. Relative under–expression of *MAPK3* and *MAPK12* genes in group 2, together with enrichment in the oncogenic pathways, the MAPK pathway, pathways in cancer, and the PI3K–Akt signaling cascade, supports that group 2 may have downregulated oncogenic signatures. In addition, TNF, a major inflammatory cytokine, participates in many steps of carcinogenesis and tumor progression [90]. As a result, we hypothesize that group 2 represents the least cancerous subtype that maintains the molecular structures of the normal urothelium, as exemplified by the expression of *KRT*s; the expression of CK5/6 and CK20 in this group is not entirely representative of the basal and luminal types, respectively [91]. This dual–positive expression pattern of basal (CK5/6) and luminal (CK20) markers was only partially characterized in a prior report of urothelial carcinoma subtypes [10]. Although the UroA subtype of the Lund classification was reported to be reactive to CK20 aberrantly as well as to CK5, CK20 immunoreactivity was located in the interior portion of this subtype, which would not entirely correspond to the normal–like staining observed in our group 2 and group 5 tumors [10]. The present study applied a novel approach of gene–expression profiling in association with pattern–wise CK5/6 and CK20 expression as well as quantitative expression of these markers. However, it is still not clear whether this subgroup is a precursor of progressed lesions (e.g., group 1 or group 3), or a distinct subtype with a different pathogenesis [20].

The genetic comparison between low-grade groups 4 versus 5 yielded much fewer DEGs than those between groups 1 and 2. The absence of CK14-expressing samples in group 4 compared to group 1, which might be linked to the SCC-like subtype, substantiated that group 1 tumors have a wider genetic spectrum than group 4, which would yield more DEGs [9]. Sjodahl et al showed that the molecular subtypes of UBC were present across tumors in different stages and grades and that they stably expressed subtype-defining genes [9]. However, almost all Ta samples were of the UroA subtype and rarely of the UroB or SCC-like subtype [9]. Another possible explanation is that IHC staining for CK5/6 and CK20 is only loosely associated with the molecular subtypes, in case of low-grade tumors. In other words, the phenotypic difference shared by the high-grade (1 vs 2) and low-grade (4 vs 5) GEP cohorts may be not entirely identical at the genetic level. Nevertheless, as we demonstrated that the AMPK-related gene, *ANP32C*, an oncogene [92], and *CD82,* a tumor-suppressor gene [93], were differentially enriched in group 5 versus group 4, it is implied that IHC staining for CK5/6 and CK20 can still classify papillary low-grade NMIUTUC in coordination to tumor biology.

Finally, we demonstrated that apoptotic and necroptotic processes were differentially enriched between high- and low-grade tumors expressing only CK5/6, not CK20. Both apoptosis and necroptosis are programmed cell death pathways and are mostly driven by receptor interacting kinase 3 [94]. Many cancers modify these cellular machineries and develop resistance to cell death [95]. The regulation of necroptosis affects tumor-inhibitory inflammatory reactions, tumor progression, and metastasis [95]. For example, *PYCARD*, which is an apoptosis-associated tumor suppressor gene frequently downregulated in urothelial carcinoma [96], was upregulated in group 4 compared with group 1 tumors. Interestingly, we revealed that group 2 and 5 tumors had similar

transcriptional properties. As previously described, CK5/6− and CK20−expressing tumors could represent well−differentiated carcinomas maintaining normal−like immunoprofiles. Microscopically, group 2 tumor cells showed a relatively preserved cellular orientation and had bland nuclei, which led to locational heterogeneity in tumor grade, even though we attempted to select the highest−grade samples of group 2 for the comparison. Therefore, high−grade papillary NMIUTUC showing IHC staining for both CK5/6 and CK20, as we believe, is not entirely high−grade in the molecular aspects, and a disturbance in normal CK expression represents the early changes of tumor progression.

Taken together, our findings indicate that NMIUTUC and NMIBC share a similar transcriptional landscape. However, subtype−related IHC markers defined in MIBC would be expressed differently in NMIBC, as we found from the Lund cohort [9, 10], in which IHC staining for basal and luminal markers as well as cadherins across molecular subtypes varied between NMIBC and MIBC. For example, CK20 expression discriminated four major Lund subtypes of MIBC; however, it was only slightly high in the GU subtype among NMIBCs. In addition, although IHC staining for CK14 was the highest in the SCC−like subtype among MIBCs, such a subtype of NMIBCs expressed relatively low levels of CK14. It is still unclear whether the genetic subtype identified in NMIUTUC remains in muscle−invasive cancer. It would be reasonable to speculate that, at certain points, genetic and phenotypic basal−like changes are acquired by some high−risk luminal−like (CK5/6−negative/CK20−positive) tumors. To this end, we plan to analyze genetic changes as the stage advances, in terms of CK5/6 and CK20 expression.

This study had some limitations. The retrospective analysis hinders the generalization of the findings of this study. The number of patients subjected to RNA−seq was constrained by the strict cut−off criteria of IHC staining. Because of the short follow−up

duration of the patients subjected to RNA-seq, their exact outcome was unevaluable, even though it was validated in the prognosis cohort. Finally, MAPK or TNF signaling pathway genes, suggested as potential biological keys in the present study and in previous studies, need external validation for confirmation. However, this study has an advantage in terms of the large number of patients included with papillary NMIUTUC only. Moreover, this study had merits of investigating the transcriptional characteristics of subgroups defined by some of the routine IHC staining (CK5/6 and CK20), through which we may have provided a new, practical approach to investigate the intrinsic characteristics of prognostically significant subgroups of NMIUTUC. This would facilitate the application of the accumulating genetic-phenotypic information of urothelial carcinoma in practice.

In conclusion, we demonstrated that IHC staining for CK5/6, CD44 and CK20 was significantly associated with the clinicopathological characteristics and prognosis of patients with papillary NMIUTUC. By combining IHC staining, luminal-like CK5/6-negative/CK20-positive papillary NMIUTUC showed the worst prognosis. Subgroups of papillary high-grade NMIUTUC were clustered in relevance to the molecular characteristics that were distinct from previous gene-expression subtypes established in MIBC or even in NMIBC. Group 2 tumors were postulated to represent a less cancerous subtype, which had similar transcriptional profiles with low-grade group 5 tumors. This information will be valuable in practice, including predicting patients' outcomes and treatment responses.

Bibliography

1. Jung, M., B. Kim, and K.C. Moon, *Immunohistochemistry of cytokeratin (CK) 5/6, CD44 and CK20 as prognostic biomarkers of non-muscle-invasive papillary upper tract urothelial carcinoma.* Histopathology, 2019. **74**(3): p. 483-493.
2. Jung, M., et al., *Transcriptional Analysis of Immunohistochemically Defined Subgroups of Non-Muscle-Invasive Papillary High-Grade Upper Tract Urothelial Carcinoma.* Int J Mol Sci, 2019. **20**(3).
3. Audenet, F., K. Attalla, and J.P. Sfakianos, *The evolution of bladder cancer genomics: What have we learned and how can we use it?* Urol Oncol, 2018. **36**(7): p. 313-320.
4. Robertson, A.G., et al., *Comprehensive molecular characterization of muscle-invasive bladder cancer.* Cell, 2017. **171**(3): p. 540-556 e25.
5. Wu, X.R., *Urothelial tumorigenesis: A tale of divergent pathways.* Nat Rev Cancer, 2005. **5**(9): p. 713-25.
6. Lindgren, D., et al., *Integrated genomic and gene expression profiling identifies two major genomic circuits in urothelial carcinoma.* PLoS One, 2012. **7**(6): p. e38863.
7. Lerner, S.P., et al., *Bladder cancer molecular taxonomy: summary from a consensus meeting.* Bladder Cancer, 2016. **2**(1): p. 37-47.
8. Lindgren, D., et al., *Combined gene expression and genomic profiling define two intrinsic molecular subtypes of urothelial carcinoma and gene signatures for molecular grading and outcome.* Cancer Res, 2010. **70**(9): p. 3463-72.
9. Sjodahl, G., et al., *A molecular taxonomy for urothelial carcinoma.* Clin Cancer Res, 2012. **18**(12): p. 3377-86.

10. Sjodahl, G., et al., *Toward a molecular pathologic classification of urothelial carcinoma.* Am J Pathol, 2013. **183**(3): p. 681-91.
11. Sjodahl, G., et al., *Molecular classification of urothelial carcinoma: global mRNA classification versus tumour-cell phenotype classification.* J Pathol, 2017. **242**(1): p. 113-125.
12. Patschan, O., et al., *A molecular pathologic framework for risk stratification of stage T1 urothelial carcinoma.* Eur Urol, 2015. **68**(5): p. 824-32; discussion 835-6.
13. Aine, M., et al., *Biological determinants of bladder cancer gene expression subtypes.* Sci Rep, 2015. **5**: p. 10957.
14. Eich, M.L., L. Dyrskjot, and G.J. Netto, *Toward personalized management in bladder cancer: the promise of novel molecular taxonomy.* Virchows Arch, 2017. **471**(2): p. 271-280.
15. Damrauer, J.S., et al., *Intrinsic subtypes of high-grade bladder cancer reflect the hallmarks of breast cancer biology.* Proc Natl Acad Sci U S A, 2014. **111**(8): p. 3110-5.
16. Choi, W., et al., *Identification of distinct basal and luminal subtypes of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy.* Cancer Cell, 2014. **25**(2): p. 152-65.
17. Rebouissou, S., et al., *EGFR as a potential therapeutic target for a subset of muscle-invasive bladder cancers presenting a basal-like phenotype.* Sci Transl Med, 2014. **6**(244): p. 244ra91.
18. The Cancer Genome Atlas Research Network, *Comprehensive molecular characterization of urothelial bladder carcinoma.* Nature, 2014. **507**(7492): p. 315-22.
19. Biton, A., et al., *Independent component analysis uncovers the landscape of the bladder tumor transcriptome and reveals insights into luminal and basal subtypes.* Cell Rep, 2014. **9**(4): p. 1235-45.

20. Hedegaard, J., et al., *Comprehensive transcriptional analysis of early-stage urothelial carcinoma.* Cancer Cell, 2016. **30**(1): p. 27-42.
21. Choi, W., et al., *Intrinsic basal and luminal subtypes of muscle-invasive bladder cancer.* Nat Rev Urol, 2014. **11**(7): p. 400-10.
22. Ho, P.L., A. Kurtova, and K.S. Chan, *Normal and neoplastic urothelial stem cells: Getting to the root of the problem.* Nat Rev Urol, 2012. **9**(10): p. 583-94.
23. Volkmer, J.P., et al., *Three differentiation states risk-stratify bladder cancer into distinct subtypes.* Proc Natl Acad Sci U S A, 2012. **109**(6): p. 2078-83.
24. Sanli, O., et al., *Bladder cancer.* Nat Rev Dis Primers, 2017. **3**: p. 17022.
25. Kammerer-Jacquet, S.F., et al., *Genomics in upper tract urothelial carcinoma.* Curr Opin Urol, 2017. **27**(1): p. 35-40.
26. Sfakianos, J.P., et al., *Genomic characterization of upper tract urothelial carcinoma.* Eur Urol, 2015. **68**(6): p. 970-7.
27. Sanford, T., S. Porten, and M.V. Meng, *Molecular analysis of upper tract and bladder urothelial carcinoma: Results from a microarray comparison.* PLoS One, 2015. **10**(8): p. e0137141.
28. Seiler, R., et al., *Impact of molecular subtypes in muscle-invasive bladder cancer on predicting response and survival after neoadjuvant chemotherapy.* Eur Urol, 2017. **72**(4): p. 544-554.
29. Rosenberg, J.E., et al., *Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial.* Lancet, 2016. **387**(10031): p. 1909-20.
30. Sharma, P., et al., *Nivolumab in metastatic urothelial carcinoma after platinum therapy (CheckMate 275): a*

multicentre, single-arm, phase 2 trial. Lancet Oncol, 2017. **18**(3): p. 312-322.

31. Dadhania, V., 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. **12**: p. 105-117.
32. Breyer, J., 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. **470**(3): p. 267-274.
33. Sikic, D., et al., *Immunohistochemiocal subtyping using CK20 and CK5 can identify urothelial carcinomas of the upper urinary tract with a poor prognosis.* PLoS One, 2017. **12**(6): p. e0179602.
34. Eckstein, M., et al., *mRNA-expression of KRT5 and KRT20 defines distinct prognostic subgroups of muscle-invasive urothelial bladder cancer correlating with histological variants.* Int J Mol Sci, 2018. **19**(11).
35. Desai, S., et al., *Relationship of cytokeratin 20 and CD44 protein expression with WHO/ISUP grade in pTa and pT1 papillary urothelial neoplasia.* Mod Pathol, 2000. **13**(12): p. 1315-23.
36. Christoph, F., et al., *Quantitative detection of cytokeratin 20 mRNA expression in bladder carcinoma by real-time reverse transcriptase-polymerase chain reaction.* Urology, 2004. **64**(1): p. 157-61.
37. Alsheikh, A., et al., *Comparison of the WHO/ISUP classification and cytokeratin 20 expression in predicting the behavior of low-grade papillary urothelial tumors. World/Health Organization/Internattional Society of Urologic Pathology.* Mod Pathol, 2001. **14**(4): p. 267-72.

38. van Oers, J.M., et al., *FGFR3 mutations and a normal CK20 staining pattern define low-grade noninvasive urothelial bladder tumours.* Eur Urol, 2007. **52**(3): p. 760-8.
39. Bertz, S., et al., *Combination of CK20 and Ki-67 immunostaining analysis predicts recurrence, progression, and cancer-specific survival in pT1 urothelial bladder cancer.* Eur Urol, 2014. **65**(1): p. 218-26.
40. Harnden, P., N. Mahmood, and J. Southgate, *Expression of cytokeratin 20 redefines urothelial papillomas of the bladder.* Lancet, 1999. **353**(9157): p. 974-7.
41. Harnden, P., et al., *Cytokeratin 20 expression by non-invasive transitional cell carcinomas: potential for distinguishing recurrent from non-recurrent disease.* Histopathology, 1995. **27**(2): p. 169-174.
42. Ramos, D., et al., *Cytokeratin expression patterns in low-grade papillary urothelial neoplasms of the urinary bladder.* Cancer, 2003. **97**(8): p. 1876-83.
43. Paner, G.P., et al., *Updates in the Eighth Edition of the Tumor-Node-Metastasis Staging Classification for Urologic Cancers.* Eur Urol, 2018. **73**(4): p. 560-569.
44. Roupret, M., et al., *European association of urology guidelines on upper urinary tract urothelial carcinoma: 2017 update.* Eur Urol, 2018. **73**(1): p. 111-122.
45. Bankhead, P., et al., *QuPath: Open source software for digital pathology image analysis.* Sci Rep, 2017. **7**(1): p. 16878.
46. Martin, J.A. and Z. Wang, *Next-generation transcriptome assembly.* Nat Rev Genet, 2011. **12**(10): p. 671-82.
47. Sultan, M., et al., *A simple strand-specific RNA-Seq library preparation protocol combining the Illumina TruSeq RNA and the dUTP methods.* Biochem Biophys Res Commun, 2012. **422**(4): p. 643-6.
48. Van Nieuwerburgh, F., et al., *Quantitative bias in Illumina TruSeq and a novel post amplification barcoding strategy for*

multiplexed DNA and small RNA deep sequencing. PLoS One, 2011. **6**(10): p. e26969.

49. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: A flexible trimmer for Illumina sequence data.* Bioinformatics, 2014. **30**(15): p. 2114-20.
50. Kim, D., B. Langmead, and S.L. Salzberg, *HISAT: A fast spliced aligner with low memory requirements.* Nat Methods, 2015. **12**(4): p. 357-60.
51. Langdon, W.B., *Performance of genetic programming optimised Bowtie2 on genome comparison and analytic testing (GCAT) benchmarks.* BioData Min, 2015. **8**(1): p. 1.
52. Pertea, M., et al., *StringTie enables improved reconstruction of a transcriptome from RNA-seq reads.* Nat Biotechnol, 2015. **33**(3): p. 290-5.
53. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic Acids Res, 2015. **43**(7): p. e47.
54. Ashburner, M., et al., *Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet, 2000. **25**(1): p. 25-9.
55. Kanehisa, M., et al., *KEGG: New perspectives on genomes, pathways, diseases and drugs.* Nucleic Acids Res, 2017. **45**(D1): p. D353-D361.
56. Subramanian, A., et al., *Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15545-50.
57. Dyrskjot, L., et al., *Analysis of molecular intra-patient variation and delineation of a prognostic 12-gene signature in non-muscle invasive bladder cancer; technology transfer from microarrays to PCR.* Br J Cancer, 2012. **107**(8): p. 1392-8.

58. Law, C.W., et al., *voom: Precision weights unlock linear model analysis tools for RNA-seq read counts.* Genome Biol, 2014. **15**(2): p. R29.
59. Tan, T.Z., et al., *Molecular subtypes of urothelial bladder cancer: results from a meta-cohort analysis of 2411 tumors.* Eur Urol, 2018.
60. Shackelford, D.B. and R.J. Shaw, *The LKB1-AMPK pathway: metabolism and growth control in tumour suppression.* Nat Rev Cancer, 2009. **9**(8): p. 563-75.
61. Schaafsma, H.E., et al., *Distribution of cytokeratin polypeptides in human transitional cell carcinomas, with special emphasis on changing expression patterns during tumor progression.* Am J Pathol, 1990. **136**(2): p. 329-43.
62. Perou, C.M., et al., *Molecular portraits of human breast tumours.* Nature, 2000. **406**(6797): p. 747-52.
63. McConkey, D.J., et al., *A prognostic gene expression signature in the molecular classification of chemotherapy-naive urothelial cancer is predictive of clinical outcomes from neoadjuvant chemotherapy: A phase 2 trial of dose-dense methotrexate, vinblastine, doxorubicin, and cisplatin with bevacizumab in urothelial cancer.* Eur Urol, 2016. **69**(5): p. 855-62.
64. McConkey, D.J., et al., *Therapeutic opportunities in the intrinsic subtypes of muscle-invasive bladder cancer.* Hematol Oncol Clin North Am, 2015. **29**(2): p. 377-94, x-xi.
65. Adam, R.M. and D.J. DeGraff, *Molecular mechanisms of squamous differentiation in urothelial cell carcinoma: a paradigm for molecular subtyping of urothelial cell carcinoma of the bladder.* Urol Oncol, 2015. **33**(10): p. 444-50.
66. Southgate, J., P. Harnden, and L.K. Trejdosiewicz, *Cytokeratin expression patterns in normal and malignant urothelium: a review of the biological and diagnostic implications.* Histol Histopathol, 1999. **14**(2): p. 657-64.

67. Bryan, R.T., *Cell adhesion and urothelial bladder cancer: the role of cadherin switching and related phenomena.* Philos Trans R Soc Lond B Biol Sci, 2015. **370**(1661): p. 20140042.
68. Koivunen, J., et al., *Protein kinase C alpha/beta inhibitor Go6976 promotes formation of cell junctions and inhibits invasion of urinary bladder carcinoma cells.* Cancer Res, 2004. **64**(16): p. 5693-701.
69. Ramena, G., et al., *CLCA2 interactor EVA1 is required for mammary epithelial cell differentiation.* PLoS One, 2016. **11**(3): p. e0147489.
70. Ishida-Yamamoto, A., et al., *Clinical and molecular implications of structural changes to desmosomes and corneodesmosomes.* J Dermatol, 2018. **45**(4): p. 385-389.
71. Jacquemet, G., H. Hamidi, and J. Ivaska, *Filopodia in cell adhesion, 3D migration and cancer cell invasion.* Curr Opin Cell Biol, 2015. **36**: p. 23-31.
72. Zheng, Z., et al., *TNF-alpha inhibits the migration of oral squamous cancer cells mediated by miR-765-EMP3-p66Shc axis.* Cell Signal, 2017. **34**: p. 102-109.
73. Brunner, A. and A. Tzankov, *The role of structural extracellular matrix proteins in urothelial bladder cancer (review).* Biomark Insights, 2007. **2**: p. 418-27.
74. Kramer, M.W., et al., *Decreased galectin-8 is a strong marker for recurrence in urothelial carcinoma of the bladder.* Urol Int, 2011. **87**(2): p. 143-50.
75. Szekely, E., et al., *Expression of claudins and their prognostic significance in noninvasive urothelial neoplasms of the human urinary bladder.* Journal of Histochemistry & Cytochemistry, 2011. **59**(10): p. 932-941.
76. Wang, Y., *Wnt/Planar cell polarity signaling: A new paradigm for cancer therapy.* Mol Cancer Ther, 2009. **8**(8): p. 2103-9.
77. Zhao, C., et al., *An integrated methylation and gene expression microarray analysis reveals significant prognostic*

biomarkers in oral squamous cell carcinoma. Oncol Rep, 2018. **40**(5): p. 2637-2647.

78. Sun, H.T., et al., *FoxQ1 promotes glioma cells proliferation and migration by regulating NRXN3 expression.* PLoS One, 2013. **8**(1): p. e55693.
79. O'Sullivan, A.G., E.P. Mulvaney, and B.T. Kinsella, *Regulation of protein kinase C-related kinase (PRK) signalling by the TPalpha and TPbeta isoforms of the human thromboxane A2 receptor: Implications for thromboxane- and androgen-dependent neoplastic and epigenetic responses in prostate cancer.* Biochim Biophys Acta Mol Basis Dis, 2017. **1863**(4): p. 838-856.
80. Nagaishi, M., et al., *Amplification of the STOML3, FREM2, and LHFP genes is associated with mesenchymal differentiation in gliosarcoma.* Am J Pathol, 2012. **180**(5): p. 1816-23.
81. Fang, D. and H. Kitamura, *Cancer stem cells and epithelial-mesenchymal transition in urothelial carcinoma: possible pathways and potential therapeutic approaches.* Int J Urol, 2018. **25**(1): p. 7-17.
82. Sun, Y., et al., *XBP1 promotes tumor invasion and is associated with poor prognosis in oral squamous cell carcinoma.* Oncol Rep, 2018. **40**(2): p. 988-998.
83. Takahashi, H., et al., *Up-regulation of plakophilin-2 and Down-regulation of plakophilin-3 are correlated with invasiveness in bladder cancer.* Urology, 2012. **79**(1): p. 240 e1-8.
84. Miyake, H., et al., *Increased angiogenin expression in the tumor tissue and serum of urothelial carcinoma patients is related to disease progression and recurrence.* Cancer, 1999. **86**(2): p. 316-24.
85. Frantzi, M., et al., *Silencing of Profilin-1 suppresses cell adhesion and tumor growth via predicted alterations in*

integrin and Ca2+ signaling in T24M-based bladder cancer models. Oncotarget, 2016. **7**(43): p. 70750-70768.

86. Kowalski, M., et al., *A Phase I study of an intravesically administered immunotoxin targeting EpCAM for the treatment of nonmuscle-invasive bladder cancer in BCGrefractory and BCG-intolerant patients.* Drug Des Devel Ther, 2010. **4**: p. 313-20.
87. Bryan, R.T. and C. Tselepis, *Cadherin switching and bladder cancer.* J Urol, 2010. **184**(2): p. 423-31.
88. Delire, B. and P. Starkel, *The Ras/MAPK pathway and hepatocarcinoma: Pathogenesis and therapeutic implications.* Eur J Clin Invest, 2015. **45**(6): p. 609-23.
89. Kim, J.H., et al., *Wnt5a attenuates the pathogenic effects of the Wnt/beta-catenin pathway in human retinal pigment epithelial cells via down-regulating beta-catenin and Snail.* BMB Rep, 2015. **48**(9): p. 525-30.
90. Sethi, G., B. Sung, and B.B. Aggarwal, *TNF: a master switch for inflammation to cancer.* Front Biosci, 2008. **13**: p. 5094-107.
91. Van Batavia, J., et al., *Bladder cancers arise from distinct urothelial sub-populations.* Nat Cell Biol, 2014. **16**(10): p. 982-91, 1-5.
92. Buddaseth, S., et al., *Dysregulation of cell cycle control caused by overexpression of the oncogene pp32r1 (ANP32C) and the Tyr>His mutant pp32r1Y140H.* Biochim Biophys Acta, 2013. **1833**(5): p. 1212-21.
93. Wu, J., et al., *DeltaNp63alpha activates CD82 metastasis suppressor to inhibit cancer cell invasion.* Cell Death Dis, 2014. **5**: p. e1280.
94. Tonnus, W., et al., *The pathological features of regulated necrosis.* J Pathol, 2019. **247**(5): p. 697-707.
95. Krysko, O., et al., *Necroptotic cell death in anti-cancer therapy.* Immunol Rev, 2017. **280**(1): p. 207-219.

96. Sacristan, R., et al., *Molecular classification of non-muscle-invasive bladder cancer (pTa low-grade, pT1 low-grade, and pT1 high-grade subgroups) using methylation of tumor-suppressor genes.* J Mol Diagn, 2014. **16**(5): p. 564-572.

Table contents

Figure contents

국문 초록

비근육침윤성 상부요로상피암의 유전자 발현 연구

-CK5/6과 CK20 단백발현에 관련하여-

정 민 선
서울대학교 대학원
의학과 병리학전공

서론: 요로상피암 (urothelial carcinoma)은 분자유전적으로 이질적인 다양한 아형으로 구성되어 있다. 대표적으로 기저형 (basal subtype)과 내강형 (luminal subtype)을 들 수 있으며 둘은 고유한 유전자 발현, 단백질 발현, 병태생리학적, 예후적, 그리고 임상적 특징을 보인다. 예를 들어, 기저형 근육침윤성 방광암 (muscle-invasive bladder carcinoma)은 다른 종류에 비교하여 불량한 예후를 보이고 항암화학요법에 이득을 볼 가능성이 높다고 알려져 있다. CK5/6, CD44, CK20 와 같은 단백질 발현은 요로상피암의 분자 아형을 대표하는 표지자로 사용될 수 있다. 이에 반하여 기저형 단백질을 발현하는 비근육침윤성 방광암 (non-muscle invasive bladder carcinoma)은 오히려 좋은 예후를 보인다는 보고가 있으나 그 원인은 밝혀지지 않았다. 최근 근육침윤성 방광암의 분자유전학적 특성은 심도 있게 연구되고 있지만 비근육침윤성 상부요로상피암 (non-muscle invasive upper tract urothelial carcinoma)의 분자적 아형에 대해서는 알려진 바가 없다. 따라서 본 연구에서 유두형 비근육침윤성 상부요로상피암의 기저형과 내강형 단백질 발현에 따른 임상병리학적, 예후적, 분자적 차이에 대해 연구하였다.

방법: CK5/6, CD44, CK20 에 대한 면역조직화학염색 (immunohistochemical staining)을 211 명의 유두형 비근육침윤성 상부요로상피암환자의 조직 마이크로어레이 (tissue microarray)에 시행하였다 (예후연구군). 면역 염색은 음성, 양성, 그리고 정상패턴으로 분류하고 예후 차이를 연구하였다. 또한 CK5/6 과 CK20 의 면역염색결과에 따라 유두형 비근육침윤성 상부요로상피암을 구분하여 유전자 발현의 차이를 분석하였다 (유전자분석군). 구분한 방식은 다음과 같으며, 그룹 1 (고등급, 높은 CK5/6 및 낮은 CK20 발현), 그룹 2 (고등급, 높은 CK5/6 및 높은 CK20 발현), 그룹 3 (고등급, 낮은 CK5/6 및 높은 CK20 발현), 그룹 4 (저등급, 높은 CK5/6 및 낮은 CK20 발현), 그리고 그룹 5 (저등급, 높은 CK5/6 및 높은 CK20 발현)이다. P-cadherin, E-cadherin, 그리고 Ki-67 에 대한 면역조직화학염색을 시행하여 고등급 유전자분석군의 유전자 발현 차이를 확인하였다.

결과: 우리는 CK5/6-음성, CD44-음성, 그리고 CK20-양성인 종양이 높은 등급 (CK5/6-음성, p <0.001; CD44-음성, p <0.001; CK20-양성, p = 0.017)과 잦은 방광내 재발 (CK5/6-음성, p = 0.002)과 연관된 고위험 종양이라는 것을 밝혔다. 카플란-마이어 분석을 통해 이 면역염색 결과는 더 짧은 암진행기간 (progression-free survival; CK5/6-음성, p = 0.001; CD44-음성, p = 0.009; CK20-양성, p = 0.031)과 더 짧은 암사망기간 (cancer-specific survival; CK5/6-음성, p = 0.009)을 가리킨다는 사실을 알았다. 게다가 음성의 CK5/6 염색은 짧은 암진행기간 (p = 0.009)과 암사망기간 (p = 0.045)에 대한 독립적 예후인자였으며 일치성지수 (Harrell's concordance index)를 계산하였을 때 각각의 예후 예측을 유의하게 향상시켰다. 표지자를 함께 분석하였을 때 내강형과 유사한 발현 (CK5/6-음성/CK20-양성)을 보이는 경우 확연하게 불량한 예후를 보였다. 전사체 분석을 통해 고등급 유전자분석군사이에 308 개의 차별발현유전자를 발견하였다. 이 유전자들은 세포접합과정 (cell adhesion)에 관여하며 그룹 1 과 2 에 비해서 그룹 3 에 더 관여되는 것으로 나타났고 또한 세포 분열 관련 유전자도 그룹 3 에 유의하게 높게 발현되었다. 그래서 이러한 기능의

변화가 그룹 3 과 같은 종양의 나쁜 예후를 일으킬 수 있음을 제안하였다. 그룹 2 는 mitogen-activated protein kinase (MAPK)와 tumor necrosis factor (TNF) 신호 전달 체계가 낮게 유지되는 것을 발견하였고 정상 요로상피와 유사한 면역염색형태를 보이는 점으로 미루어 보았을 때 악성화가 가장 덜 진행된 것으로 생각된다. 이에 비교하여 저등급 유전자분석군 (그룹 4 와 5) 사이에서는 24 개의 차별발현유전자를 발견하였고 이 유전자들은 cAMP-kinase (AMPK) 신호 전달 체계와 관련을 보였다. 또한 등급에 따른 유전자 발현 차이를 분석하였는데, 높은 CK5/6 과 낮은 CK20 발현을 보이는 종양에서 고등급 종양과 저등급 종양은 세포자멸 및 괴사자멸 (apoptosis/necroptosis)와 관련된 52 개의 차별발현 유전자를 가지고 있었다. 반면에, CK5/6 과 CK20 이 둘 다 높은 종양의 경우에는 고등급과 저등급 종양 사이에 단 8 개의 차별발현유전자만 발견되었다. 마지막으로, P-cadherin, E-cadherin, 그리고 Ki-67 에 대한 면역염색 결과 CK5/6-음성 혹은 CK20-양성 유두형 비근육침윤성 상부요로상피암의 경우 비교적 높은 E-cadherin 발현과 유의하게 높은 Ki-67 활성도를 보였고 이를 통해 세포 접합과 분열에 관련된 특징을 재확인할 수 있었다. P-cadherin 발현의 경우 기저형 표지자로서 CK5/6 면역염색과 비슷한 결과를 보였다.

결론: CK5/6 과 CK20 면역조직화학염색을 통해 유두형 비근육침윤성 상부요로상피암을 예후에 따라 또한 유전자 및 표현형 발현에 따라 나눌 수 있다. 이 분류의 임상병리학적 특성은 비근육침윤성 방광암과 비슷하였고 이는 근육침윤성 방광암의 그것과는 상반된 것이어서, 초기 요로상피암의 CK5/6, CK20 발현은 근육침윤성 방광암에서 기존에 알려진 기저형, 내강형 분류법과는 구별되어야 할 것이다. 이 차이를 이해하는 것은 요로상피암의 진행에 따른 분자유전학적 발달 과정을 해석하는데 중요할 것이며 비근육침윤성 상부요로상피암 환자의 맞춤치료에 유용하게 사용될 수 있다.

주제어: 상부요로상피암; 유두형 요로상피암; 이행세포암; 메신저리보핵산; 유전자발현분석; 차세대염기서열분석; 염기서열분석; 면역조직화학염색; 예후