Artificial Intelligence Assisted Drop Pattern Analysis and RNAseq Profiling for Early Diagnosis and Follow-up of Bladder Cancer

by

Ramiz Demir

A Dissertation Submitted to the

Graduate School of Health Sciences in Partial

Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in

Cellular and Molecular Medicine

KOÇ ÜNİVERSİTESİ

March 1, 2023

Artificial Intelligence Assisted Drop Pattern Analysis

and RNAseq Profiling for Early Diagnosis

and Follow-up of Bladder Cancer

Koc University

Graduate School of Health Sciences

This is to certify that I have examined this copy of a doctoral dissertation by

Ramiz Demir

and have found that it is complete and satisfactory in all respects

and that any and all revisions required by the final

examining committee have been made

Committee Members:

_____

Prof. Dr. Devrim Gözüaçık (Advisor)

_____

Prof. Dr. Çiğdem Gündüz Demir

_____

Assoc. Prof. Dr. İlker Tinay

_____

Assoc. Prof Dr. Özlem Kutlu

_____

Assist. Prof. Dr. Gözde Korkmaz

Date: 01.03.2023

ABSTRACT

Artificial Intelligence Assisted Drop Pattern Analysis

and RNAseq Profiling for Early Diagnosis

and Follow-up of Bladder Cancer

Ramiz Demir

Doctor of Philosophy in Cellular and Molecular Medicine

March 1, 2023

Bladder cancer is one of the most common cancer types of cancer in the urinary system. Current bladder cancer diagnosis and follow-up techniques are time-consuming, expensive, and invasive. The gold standard for the diagnosis of bladder cancer in clinical practice is invasive biopsy followed by histopathological analysis. In recent years, costly tests involving bladder cancer biomarkers were developed, but these tests have high false-positivity and false-negativity rates, limiting their reliability. Hence, there is an urgent need for the development of novel and practical diagnostic tests.

In this thesis, by analyzing droplet patterns of blood and urine samples from patients, we developed a deep learning- and artificial intelligence-assisted quick, cheap, and reliable diagnosis method. Droplet pattern analysis of evaporated blood or urine deposits was performed using patient and normal control samples. Our proposed AI-assistant solution (ResNet-18 pre-trained ImageNet) can be systematically applied across droplets, enabling comparisons to reveal shared spatial behaviors and underlying morphological patterns, which precisely differentiate patient-derived samples from controls with high accuracy. The innovative diagnostic method has been presented based on the recognition and classification of complex patterns formed by dried urine or blood drops under different conditions. Our results indicate that AI-based systems have a great potential for a non-invasive and accurate diagnosis of bladder cancer.

RNA sequencing was also performed to identify potential candidate markers for bladder cancer in different gene classes, including up and down-regulated genes, gradient-increased or decreased genes, case-specific genes, and secreted genes from bladder cancer. According to the results, various novel genes have been found to be

candidate markers that can be used for bladder cancer diagnosis. For example, the OVOL2 gene was found to be a disease-free marker for bladder cancer. In addition, eleven secreted genes (C4ORF48, TNFSF15, VEGFA, PRSS22, SDC1, DHRS11, ENTPD6, SEMA3F, CCDC134, LYPD6, SDC4) were found to be potential secreted candidates from bladder tumors. Also, the integration of blood or urine droplet patterns with these secreted genes was performed to investigate the possible contribution of secreted genes to droplet patterns. It was found that expression levels of the genes were differentiated among patients' blood and urine droplet patterns may be an alternative perspective to determine patients who have bladder tumor variations.

Two genes (LGALS3 and SH3D21) were also chosen from RNAseq outputs for molecular analysis. LGALS3 gene-coded Gal-3 protein was knocked out (KO) by CRISPR/Cas9 system, have shown that Gal-3 KO T24 bladder cell lines increased spheroid diameter in a three-dimensional cell culture system. Another gene, SH3D21, was also evaluated for mRNA expression level in tumor samples, and the expression level of the gene was found to be increased in tumors.

In conclusion, the innovative diagnostic method has been presented based on recognizing and classifying complex patterns formed by dried urine or blood drops under different conditions. Our results indicate that AI-based systems have great potential for a non-invasive and accurate diagnosis of bladder cancer. Determined candidate genes from RNA sequencing will be expected to understand the molecular biology of bladder cancer and discover new therapeutic perspectives in bladder cancer management.

Key Words: Artificial Intelligence, Machine Learning, Bladder Cancer, Diagnosis, Follow-up, RNA sequencing

ÖZETÇE

Mesane Kanserinin Erken Tanı ve Takibinde Yapay Zeka Destekli Damla Motif Analizi ve RNAseq Profilleme

Hücresel ve Moleküler Tıp, Doktora

1 Mart, 2023

Mesane kanseri, üriner sistemde en yaygın görülen kanser türlerinden biridir. Mevcut mesane kanseri tanı ve takip teknikleri zaman alıcı, pahalı ve invazivdir. Klinik uygulamada, mesane kanseri tanısında kullanılan altın standard, invaziv bir yöntem olan biyopsi alımı ve bunu takiben histopatolojik analizlerdir. Son yıllarda, mesane kanseri biyobelirteçlerini içeren maliyetli testler geliştirildi, ancak bu testlerin yüksek yanlış pozitiflik ve yanlış negatiflik oranları, güvenilirliklerini sınırlıyor. Bu nedenle, yeni ve pratik teşhis testlerinin geliştirilmesine acil bir ihtiyaç vardır.

Bu tezde hastalardan alınan kan ve idrar örneklerinin damlacık motiferini analiz ederek derin öğrenme ve yapay zeka destekli hızlı, ucuz ve güvenilir bir teşhis yöntemi geliştirdik. Buharlaşan kan veya idrar birikintilerinin damlacık motif analizi, hasta ve normal kontrol numuneleri kullanılarak yapıldı. Önerilen yapay zeka destekli yardımcısı çözümümüz (ImageNet ile eğitilmiş ResNet-18), damlacıklar arasında sistematik olarak uygulanarak, hastadan türetilen numuneleri yüksek doğrulukla kontrollerden tam olarak ayıran paylaşılan uzamsal davranışları ve altta yatan morfolojik kalıpları ortaya çıkarmak için karşılaştırmalara olanak tanır. Kurutulmuş idrar veya kan damlalarının farklı koşullar altında oluşturduğu karmaşık modellerin tanınması ve sınıflandırılmasına dayalı yenilikçi teşhis yöntemi sunulmuştur. Sonuçlarımız, AI tabanlı sistemlerin mesane kanserinin invaziv olmayan ve doğru teşhisi için büyük bir potansiyele sahip olduğunu göstermektedir.

Yukarı ve aşağı regüle edilmiş genler, gradyan artmış veya azalmış genler, vakaya özgü genler ve mesane kanserinden salgılanan genler dahil olmak üzere farklı gen sınıflarında mesane kanseri için potansiyel aday belirteçleri belirlemek için RNA dizilimi

de yapıldı. Elde edilen sonuçlara göre, çeşitli yeni genlerin mesane kanseri teşhisinde kullanılabilecek aday belirteçler olduğu bulunmuştur. Örneğin, OVOL2 geninin mesane kanseri için hastalıksız bir belirteç olduğu bulundu. Ek olarak, mesane tümörlerinden salgılanan on bir genin (C4ORF48, TNFSF15, VEGFA, PRSS22, SDC1, DHRS11, ENTPD6, SEMA3F, CCDC134, LYPD6, SDC4) potansiyel olarak salgılanan adaylar olduğu bulundu. Ayrıca, salgılanan genlerin damlacık modellerine olası katkısını araştırmak için kan veya idrar damlacık motiflerinin bu salgılanan genlerle entegrasyonu gerçekleştirildi. Genlerin ekspresyon düzeylerinin hastaların kan ve idrar damlacık modelleri arasında farklılaştığı bulundu, bu mesane tümörü varyasyonları olan hastaların belirlenmesi için alternatif bir bakış açısı olabilir.

Moleküler analiz için RNAseq çıktılarından iki gen (LGALS3 ve SH3D21) de seçildi. LGALS3 gen kodlu Gal-3 proteini, CRISPR/Cas9 sistemi tarafından devre dışı bırakıldı (KO), Gal-3 KO T24 mesane hücre hatlarının, üç boyutlu bir hücre kültürü sisteminde uluşturdukları sferoidlerin çapının arttırdığını gösterdi. Diğer bir gen olan SH3D21 de tümör örneklerinde mRNA ifade düzeyi açısından değerlendirilmiş ve tümörlerde genin ifade düzeyinin arttığı saptanmıştır.

Sonuç olarak, farklı koşullar altında kurumuş idrar veya kan damlalarının oluşturduğu karmaşık modelleri tanımaya ve sınıflandırmaya dayalı yenilikçi tanı yöntemi sunulmuştur. Sonuçlarımız, AI tabanlı sistemlerin mesane kanserinin invaziv olmayan ve doğru teşhisi için büyük bir potansiyele sahip olduğunu göstermektedir. RNA dizilemesinden belirlenen aday genlerin, mesane kanserinin moleküler biyolojisini anlaması ve mesane kanseri yönetiminde yeni terapötik perspektifler keşfetmesi beklenmektedir.

Anahtar Kelimeler: Yapay Zeka, Makine Öğrenmesi, Mesane Kanseri, Tanı, Takip, RNA dizileme

ACKNOWLEDGMENTS

Firstly, I would like to express my gratitude to my thesis supervisor Prof. Dr. Devrim Gözüaçık for his guidance during the periods of my Ph.D. thesis. I would like to thank all my thesis jury members; Prof. Dr. Çiğdem Gündüz Demir, Assoc. Prof. Dr. İlker Tinay, Assoc. Prof Dr. Özlem Kutlu, and Assist. Prof. Dr. Gözde Korkmaz for their valuable discussions and critical evaluation of this thesis. I also would like to thank our collaborators in artificial intelligence analysis Prof. Dr. Çiğdem Gündüz Demir; tumor, blood and urine samples collections for Assoc. Prof. Dr. İlker Tinay; bioinformatic analysis for Prof. Dr. Tunahan Çakır.

I would like to special thanks to artificial intelligence analysis for MSc. Soner Koç; and bioinformatic analysis for Büşra Hatice Konuk. I want to express thanks lab. mates in particular to Dr. Yunus Akkoç and Dr. Öyküm Kaplan Arabacı for limitless support to me. I want to express all thank to current and former Gözüaçık laboratory members for their support to me and in this project.

This work was supported by the TÜBİTAK Newton-Katip Çelebi Fund Program grant number 216S915. I also would like to thank to Koç University Research Center for Translational Medicine (KUTTAM) for supporting excellent research environment.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

µg	Microgram
µl	Microliter
2D	Two-dimensional
3D	Three-dimensional
AA	Aromatic amine
AI	Artificial intelligence
ANN	Artificial neural network
APAF1	Apoptotic Protease-activating Factor 1
AUC	area under the curve
BCa	Bladder cancere
BLCA	Bladder Urothelial Carcinoma
cDNA	Complementary DNA
CIS	Carcinoma in situ
CNN	Convolutional neural network
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DAPI	4′,6-diamidino-2-phenylindole
DAPK	Death-associated protein kinase 1
DBC1	Deleted in bladder cancer protein 1
DL	Deep learning
DLBC	Lymphoid Neoplasm Diffuse Large B-cell Lymphoma
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic acid
DT	Decision tree
EGFR	Epidermal growth factor receptor
FBS	Fetal bovine serum
FGFR3	Fibroblast growth factor receptor 3
GO	Gene ontology
HG	High grade
HPA	The Human Protein Atlas

IGFBP3	Insulin-like growth factor-binding protein 3
ISUP	International Society of Urological Pathology
KCl	Potassium chloride
KEGG	Kyoto Encyclopedia of Genes and Genomes
KIRP	Kidney renal papillary cell carcinoma
KO	Knock out
LG	Low grade
LR	Logistic regression
MAPK	Mitogen-activated protein kinase
MDM4	Mouse double minute 4
MgCl2	Magnesium chloride
MIBC	Muscle invasive bladder cancer
ML	Machine learning
ml	Milliliter
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide
NBI	Narrow band imaging
NMIBC	Non-muscle invasive bladder cancer
OCT	Optical coherence tomography
PAH	Polycyclic aromatic hydrocarbon
PCA	Principal component analysis
PDD	Photodynamic therapy
PI3K	Phosphoinositide 3-kinase
PUNLMP	Papillary urothelial neoplasm of low malignant potential
qPCR	quantitative polymerase chain reaction
RD	Random forest
ResNET	Residual neural network
RGB	Red-Green-Blue
RNA	Ribonucleic acid

RNAseq	RNA sequencing
ROC	Receiver operating characteristic
RPMI	Roswell Park Memorial Institute
RT-qPCR	reverse transcription quantitative real-time PCR
SE	Standard Error
SOCS-1	Suppressor of cytokine signaling 1
STAT1	Signal transducer and activator of transcription 1
TCGA	The Cancer Genome Atlas
TIFF	Tag Image File Format
TIMP-3	Tissue inhibitor of metalloproteinase 3
TSC-1	Tuberous sclerosis 1
TURBT	Transurethral Resection of Bladder Tumors
UICC	Union International Contre le Cancer
UMAP	Uniform Manifold Approximation and Projection
uPA	Urokinase-type plasminogen activator
VEGF	Vascular endothelial growth factor
WHO	World Health Organization
WT	Wild type

CHAPTER 1

INTRODUCTION

1.1 Bladder physiology

The bladder is a part of the urinary system containing the kidneys, renal pelvis, ureters, and urethra (**Figure 1.1**). Urine in the kidney travels down two tubes called ureters to the bladder. The bladder stores urine, allowing urination to be infrequent and controlled. The bladder is covered by a muscle tissue line stretching to hold urine. During urination, the muscles are squeezed, allowing urine to flow out via the urethra [1]. Out of the normal physiological function of the bladder, the occurrence of many abnormalities, such as cancer, in the organ leads to significant public health concerns in various populations [2].

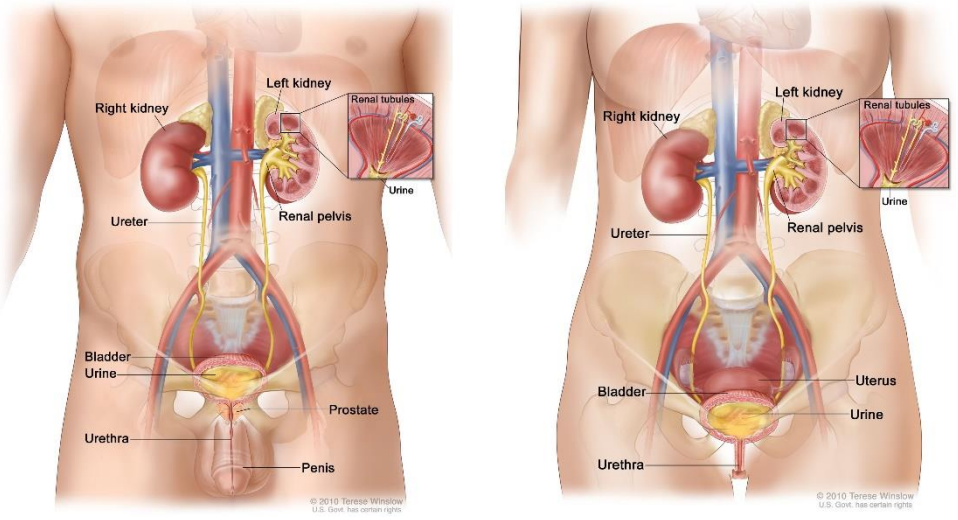

Figure 1.1: Basic anatomy of the male (left) and female (right) urinary system showing the kidneys, renal pelvis, ureter, bladder, and urethra [3].

1.2 Bladder Histology

The urinary bladder is a triangle-shaped hollow organ located in the lower abdomen. The inner surface of the bladder is covered by epithelial cells (urothelium) in that the thickness of the epithelial lining varies from 3 to 7 layers. The urothelium contains three types of cells, including superficial/umbrella cells, intermediate cells, and basal cells [4].

Superficial cells are existed in the deepest surface of the urothelium directly contact with urine, and their size is more extensive shape than others. They contain ample eosinophilic cytoplasm. They also have seen to be binucleated morphology. Secondly, intermediate cells lie under the superficial cells and have elongated and polarized morphology in a vertical arrangement facing the surface of the deepest urothelial cells. Thirdly, basal layer cells present upon the basement membrane, and their size is smaller than intermediate cells. Lower transcriptional activity and condensed chromatin structures have been also observed in these cells [2]. After the urothelial cell layer and basement membrane, an extensive connective tissue with high number of vessels, elastic fibers, nerves, and lymphatics are existed. It is also called "lamina propia" or submucosa [5]. After the lamina propia, a thick muscle layer consisting of longitudinal and circular muscles merges with surrounding perivesical adipose tissue [6] (**Figure 1.2**).

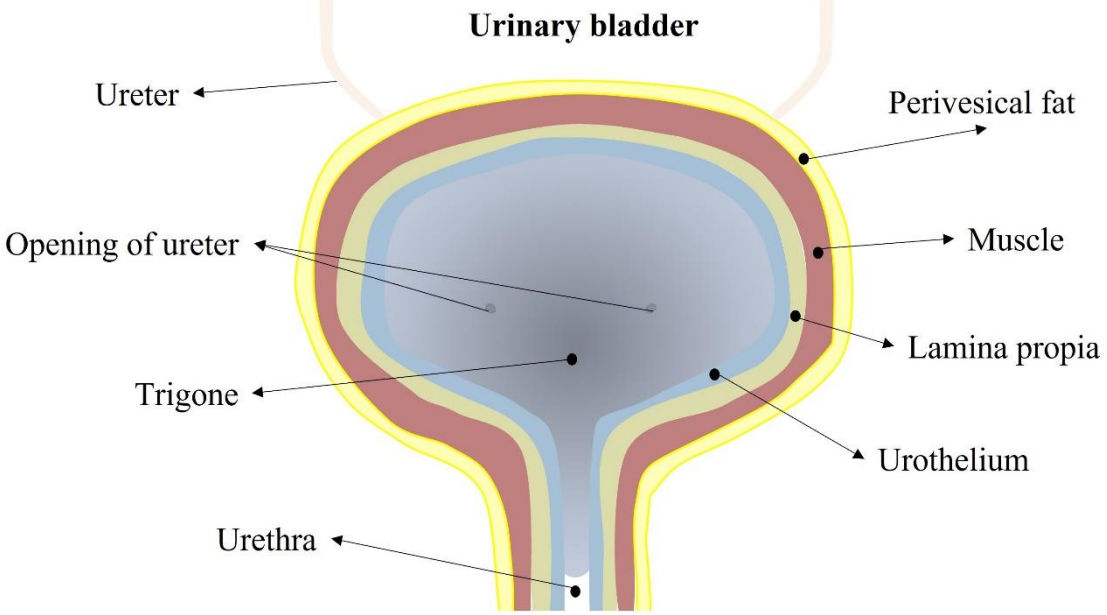

Figure 1.2: Schematic representation of urinary bladder showing the layers and regions of the organ.

1.3 Bladder cancer incidence

Bladder cancer (BCa) is one of the most common malignancies and the second cause of death among genitourinary tumors, with approximately six-hundred-thousand new cases and over two hundred thousand deaths globally [7]. In developed countries, BCa is the fourth diagnosed cancer and the eighth cause of cancer-related death for men, making it a significant public health concern [8]. The incidence rate of BCa is more common in men than in women at four times [9]. It is also estimated that there will be over eighty thousand new cases and seventy thousand death from the disease in 2022 (**Figure 1.3**). In addition, many risk factors have been estimated in relation to bladder cancer, including smoking, exposure to chemicals, gender, age, chronic bladder irritation, ethnicity, infections, genetic background, and family history of other diseases. The short- and long-term outcomes of BCa is predicted by these risk factors for recurrence, progression, and survival [10]. These risk factors related to the occurrence of the BCa appear to reflect the prevalence of tobacco smoking, infections, and occupational exposure to aromatic amines and polycyclic aromatic hydrocarbons in populations [11,12].

	Common Types of Cancer	Estimated New Cases 2022	Estimated Deaths 2022
1.	Breast Cancer (Female)	287,850	43,250
2.	Prostate Cancer	268,490	34,500
3.	Lung and Bronchus Cancer	236,740	130,180
4.	Colorectal Cancer	151,030	52,580
5.	Melanoma of the Skin	99,780	7,650
6.	**Bladder Cancer**	**81,180**	**17,100**
7.	Non-Hodgkin Lymphoma	80,470	20,250
8.	Kidney and Renal Pelvis Cancer	79,000	13,920
9.	Uterine Cancer	65,950	12,550
10.	Pancreatic Cancer	62,210	49,830

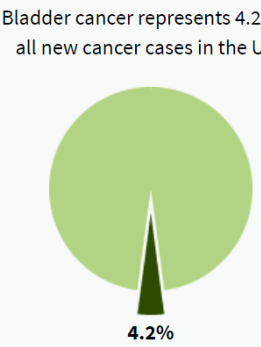

Figure 1.3: Bladder cancer incidence among cancers in TCGA data [3].

According to the Globocan statistics [13] (**Figure 1.4**), the estimated number of new cases is predicted as 5.2 %, with 12.248 cases in Turkey. The BCa is a cancer type that existed in the seventh line considering Turkey's ages and genders. In addition, morbidity and 5-year prevalence are estimated at 3.771 and 35.857 cases, respectively.

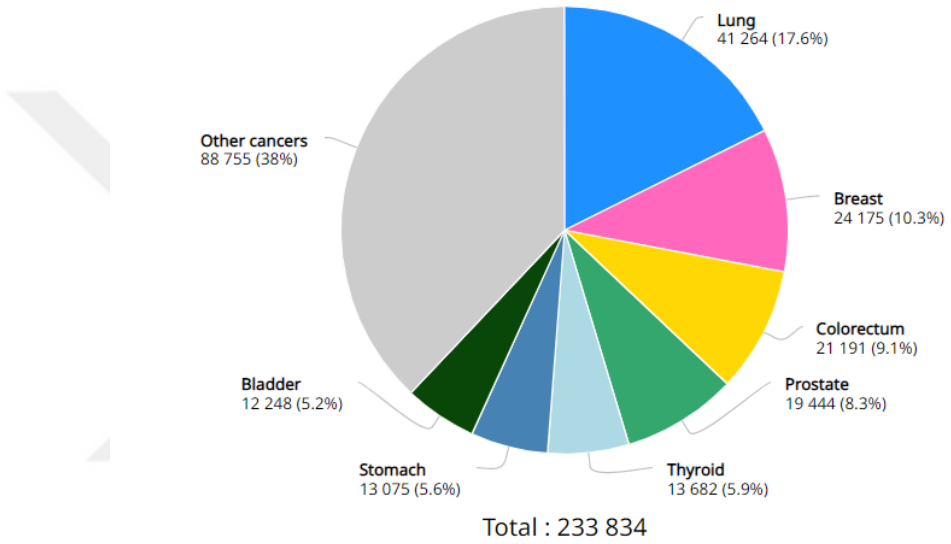

Figure 1.4: Bladder cancer incidence among both sexes and all ages in Turkey [13].

1.4 Bladder Cancer

The BCa generally originates from the urothelial cells that covers the inner surface of the bladder and results in the cancerous character of the normal urothelium (**Figure 1.5**). BCa begins in the urothelium may spread through the lining of the bladder and invade the bladder's muscle layer or spread to nearby organs (such as the prostate) and lymph nodes. According to the invasiveness through the muscle, there are two class of BCa: non-muscle invasive bladder cancer (NMIBC) and muscle-invasive bladder cancer (MIBC) [14]. Approximately 80% of newly diagnosed patients are presented as NMIBC, and 20% are MIBC or metastatic cancer [15].

In short-term, the NMIBC is associated with good survival outcomes [16]. However, due to the high recurrence rate of the NMIBC, patients require life-long follow-up.

The long-term outcomes of patients suffering from BCa highlight the need for increased molecular characteristics of the disease to improve clinical management in diagnosis, prognosis, and therapy [17]. Treatment options of the NMIBC include transurethral resection of the bladder tumor (TURBT), immunotherapy with intravesical instillation of (Bacillus Calmette–Guérin) BCG vaccine, mitomycin uptake, and radical cystectomy for high-grade recurring tumors [18]. Due to the high recurrence of NMIBC, patients are asked to undergo frequent cystoscopy, which is costly, invasive, and requires an experienced user. Cystoscopy is the significant diagnostic test commonly used to screen for BCa patients. However, the characterization of molecular biomarkers highlights BCa remains poorly understood.

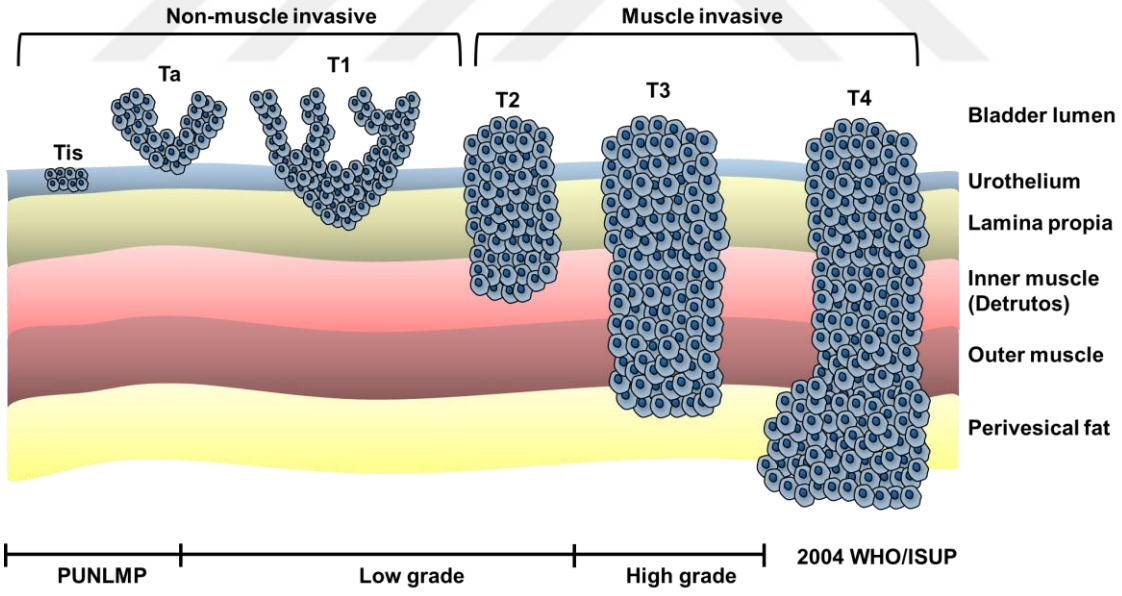

Figure 1.5: Schematic representation of bladder cancer staging and gradin according to the Tumor-Node-Metastasis (TNM) system and 2004 World Health Organization (WHO)/International Society of Urological Pathology (ISUP) criteria. PUNLMP, papillary urothelial malignancy of low malignant potential.

1.5 Staging of bladder cancer

Staging of the bladder tumor is arranged by a TNM classification approved by the Union International Contre le Cancer (UICC) in 2002 [19]. The classification is considering three main behaviours of tumor composed of the invasion (T), the lymph nodes metastasis (N), and the distant metastasis (M). Following the confirmation of the tumor with histopathological investigation, the stage is approved and called as pTNM [20]. TURBT is an important technique to define the tumor expansion of invasion into layers of the bladder (**Figure 1.5**). pTa (non-invasive) is the most superficial tumor restricted with the urothelium. pT1 (superficially invasive) is a tumor that invade the lamina propia. Cis/Tis (carcinoma in situ/tissue in situ) is a flat tumor and restricted to the urothelium. These tumors (pTa, pT1, and pTis) are classified as NMIBC. The tumor invading the muscle layer are classified as pT2. The tumors invading the inner extension of the detrusor muscle is classified as pT2a, and deep extension into the detrusor is classified as pT2b. A tumor is observed in the perivascular fat microscopically (pT3a) or macroscopically (pT3b). If the tumor invades the organs adjacent to the bladder is classified as pT4a, and if it invades the pelvic or abdominal wall is classified as pT4b. Tumors > pT2 are classified as MIBC [20].

1.6 Grading of bladder cancer

According to the 2004 World Health Organization (WHO) classification, non-invasive and invasive bladder tumors are differentiated by molecular analysis [21]. The tumor grading system defines these two main tumor types, showing the differentiation grade of the tumors as low-grade and high-grade (**Figure 1.5**). Low grade tumors defined well-preserved histological architecture with cytological atypia, and few mitoses. While high grade tumors define high cytological atypia, mitosis, highly pleomorphic nuclei with expended nucleoli, and disrupted histological architecture [22,23].

Even after removal and treatment of the tumor, it can return to come back again, which is called a recurrent tumor. In contrast, tumor progression is defined by an irreversible change in a tumor's character that leads to worse outputs, such as complete bladder removal by cystectomy [24].

The development of bladder tumors from normal urothelium can result in the recurrence and progression of the tumor (**Figure 1.6**). According to these two classes of BCa, staging and grading, low stage and low grade tumors tend to be recurrent tumors. In contrast, high grade and high stage tumors mostly lead to metastasis to other organs, such as prostate, lung, and bone tissue [25].

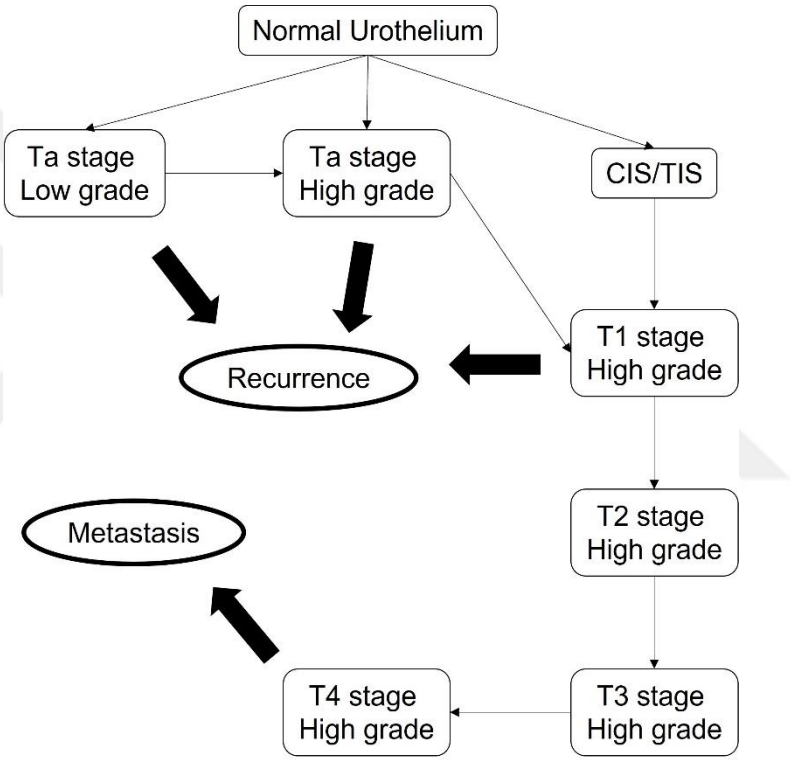

Figure 1.6: Development of bladder cancer from normal urothelium.

1.7 Clinical management of bladder cancer

1.7.1 Bladder carcinogenesis

Carcinogenesis of the urinary bladder primarily derived from exposure to many carcinogens, including aromatic amines (AA), polycyclic aromatic hydrocarbons (PAH), and nitrosamines. In addition, hair dyes and cigarette smoking are strongly associated with increased risk of BCa [26]. Many of these carcinogens cause DNA abnormalities lead chromosomal aberration and mutations in the genome [27,28].

Most prevailing rearrangement and chromosomal aberration are present on chromosome 9 and occurs in more than 50 % of all BCa, including NMIBCs and MIBCs [29]. The candidate tumor suppressor gene in the chromosome are PTCH (Gorlin syndrome gene), DBC1 (deleted in bladder cancer 1 locus), TSC-1 (tuberous sclerosis syndrome gene), CDKN2A (p16), and CDKN2B (p15) [30,31].

Disfunction of these suppressors lead to uncontrollable cell division and proliferation in BCa formation. The p16 specifically binds cyclin dependent protein kinase cdk4 and cdk6 results in repressing the catalytic activity of cyclin D-cdk complex, and lead to arrest cell cycle at the G1/G2 phase. In immunohistochemistry staining, p16 expression was found to be low in 86% (104 BCa cases) compared to control [32]. In addition, the gain of chromosome 7 is a common trisomy causing the increasing number of alleles for epidermal growth factor receptors (EGFR) in BCa [30,31].

Other chromosomal aberrations include chromosome 1, 8, and 11. For instance, amplification of chromosome 1 is associated with the mouse double minute 4 (MDM4) gene homologue that is the regulator of p53 in BCa [33]. CMYC is a candidate oncogene present on chromosome 8, and the alteration of the gene is associated with BCa [34]. The polysomy of chromosome 11 is also associated with bladder cancer, and the chromosome contain putative oncogenes such as cyclin CCND1, EMS1, FGF3, and 4 [35,36].

The different phenotypic characteristics are also observed in both grades of BCa. For example, mutation incidence (60-80%) of fibroblast growth factor receptor 3 (FGFR3) have been suggested as a signature of invasive low-grade tumours [37]. However, p53 mutations are commonly seen in invasive high grade tumors. Mutation in FGFR3 constitutively activates receptor tyrosine kinase led to trigger down-regulation of AKT and cell cycle regulators, and activation of MAP kinase pathway [38].

1.7.2 Diagnosis of bladder cancer

The remarkable diagnostic method of BCa is the combination of urinary cytology and urethra-cystoscopy. The initial assessment is performed by voided urine cytology. The tumor cells are directly in contact with urine and because of their losing cell-cell connections cause tumor cells to leak into the urine [39]. Thereby, investigation of tumor cells directly from voided urine is a significant diagnostic method. However, it has a low sensitivity for low-grade tumors [40].

Alternatively, cystoscopic diagnosis (TURBT) is the main method in the clinic used to confirm the presence and location, numbers, initial detection, and recurrence of the tumor in the bladder. With this technique, flat lession, like carciona in situ (CIS), are hardly detected and are easily escaped from clinical investigation. In addition, papillary lessin (pTa, pT1) and CIS are missed and may cause recurrence after cystoscopy at first follow-up [41]. Invasive character and cost of the technique are other disadvantages [42,43].

Detection problems for small papillary and flat lesions lead to improve the photodynamic diagnosis (PDD) as a special cystoscopy technique. In this cystoscopy, a photoactive reagent is accumulated in neoplastic / abnormal tissue intensely more than normal tissue. This method enhances the differentiation between cancerous and normal tissue. However, it has some false positive outcomes including inflammation and scarring derived from surgical operation [44,45].

Optical coherence tomography (OCT) and Narrow band imaging (NBI) are also other techniques to diagnose urinary bladder cancer. The OCT is based on obtaining cross-sectional images of the bladder wall in high resolution, and NBI is also enhancing the contrast between tissue vascular structures and mucosa for the cancer diagnosis [46,47].

1.7.3 Prognosis of Bladder Cancer

Prognosis and its markers of BCa are the main issues for treatment response, recurrence, survival after TURBT, intravesical therapy, radical cystectomy, radiation therapy, and systemic chemotherapy. Traditional clinical and pathological markers such as tumor grade, stage, extension to lymph node and vascular system, provide significant information but have limitations to metastasis, response to therapy, and survival. Also, no single molecular marker is reliable enough to predict clinical outcome of BCa. However, the literature provides many markers. Ki-67 is a good prognostic marker for prediction of recurrence in NMIBC [48]. FGFR3 gene mutation is also linked to a low grade BCa with low malignant clinical features [49].

In addition, hypermethylation of many genes, including apoptotic protease-activating factor-1 (APAF-1), insulin-like growth factor binding protein-3 (IGFBP3) [50], death-associated protein kinase (DAPK) [51], suppressor of cytokine signalling (SOCS)-1, signal transducer and activator of transcription (STAT)-1, B-cell lymphoma-2 (BCL-2), tissue inhibitor of metalloproteinases (TIMP)-3, and E-cadherin [52] is significantly associated with intravesical recurrence after TURBT. In addition to intravesical recurrence, there are also many progression markers which can help to predict the prognosis of NMIBC to MIBC, such as motility related protein-1 (MRP-1/CD9) [53], intermodulation product 3 (IMP3) [54], gamma catenin, Ki67 (MIB-1) [48], p53 [55], and thrombospondin 1 (TSP-1) [56]. However, difficulties of tissue sampling, selection bias, and defining the positive staining may be limitations for these markers. In addition to these prognostic markers, there are also many candidate markers, including cell cycle inhibitors, tumor suppressors and oncogenes, cell cycle regulators, angiogenic and secreted molecules have been shown to be prognosis in BCa [57].

1.8 Molecular pathways of bladder cancer

Up to date, various experimentally confirmed or candidate molecules as a marker have been identified for BCa. The connection of these markers with each other leads to understanding specific molecular pathways that can be dysregulated in the cancer formation. Copy-number and mutation analysis have shown three main biological pathways or process frequently dysregulated in BCa: (1) cell cycle modulation; (2) PI3K signalling and (3) chromatin remodelling, including histone-modification and nucleosome remodelling complexes [58].

Papillary and invasive cancer of the urothelial bladder appears to show distinct evolving molecular pathways. Low grade papillary tumors frequently show a constitutive activation of RTK-RAS pathway in appearance with activating mutations of human rat sarcoma virus (HRAS) and fibroblast growth factor 3 (FGFR3) genes [59,60]. In contrast, CIS/TIS and invasive tumors frequently show alteration in tumor protein p53 (TP53) and retinoblastoma (RB) gene pathways by activation mutation in epidermal growth factor receptor (EGFR/ERBB2/HER2) [61].

Various genomic alteration (RNA expression changes, mutations, copy number alterations) mainly dysregulates PI3K/AKT/mTOR pathway of the BCa tumor. It is also observed that RTK/RAS pathway of the tumours are found to contain many genomic rearrangement, including genomic amplifications, mutations, and gene fusions [62]. In addition, many pathways contribute to formation of bladder cancer, including RAS-MAPK signal transduction pathway, p53 cell cycle regulation pathway, retinoblastoma pathway, and tumor angiogenesis pathway [63].

The cell cycle is strictly controlled under p53 and RB pathways by receiving extracellular stimulus via MAPK pathway with the MAPK/ERK system [63]. These pathways also suppressed by Ras association domain family 1 protein (RASSF1A) and death-associated protein kinase (DAPK) for activated Ras and ERK, respectively.

The cell cycle progress at G1-S transition is inhibited by p53. It is also controlled through the transcriptional activation of p21, a cyclin dependent kinase inhibitor (CDKI) [64]. In addition, the Mdm2 protein is involved in controlling p53 activity through which binds to p53 and triggers its ubiquitin-mediated degradation in proteosomes [65].

CDKN2A locus highly mutated region in BCa and produces two genes, p14 and p16. Mdm2 gene also transcriptionally inhibited by p14 that is another control mechanism of p53 activity. Methylation level of $p14^{ARF}$ positively corelate with poor prognosis in BCa [66]. RB gene encodes a nuclear phosphoprotein (Rb), which is activated by dephosphorylation and lead to release E2F transcription factor from itself. Hence, Released E2F transcribes DNA synthesis genes and triggers cell proliferation [67,68]. Rb protein is also regulated by p16 protein. T24 cells transfecting with p16 cDNA decreasing of Rb phosphorylation results in decreased cell proliferation. In addition, tumor cells need to oxygen, nutrients, and growth factors for maintenance their high metabolic requirements. Therefore, tumor angiogenesis contributes to provide tumor cells for these supplements [63,69].

The mRNA expression of pro-angiogenic vascular endothelial growth factor (VEGF) is associated with recurrence and progression in superficial tumors [70,71]. Activation of EGFR signal increase VEGF expression, which also induces urokinase-type plasminogen activator (uPA) to degrade extracellular matrix (ECM) to facilitate endothelial cell migration and invasion [72]. In addition, the receptor signal promotes the production of interleukin-8 (IL-8) and matrix metalloproteases (MMPs) via thymidine phosphorylase. IL-8 is mitogenic and chemotactic factor for endothelial cells, and MMPs degrade ECM for endothelial cells to promoting vascularization in BCa [73,74]. Additionally, p53 alteration associated with low thrombospondin-1 (TSP-1), angiogenesis inhibitor, expression in BCa [75] (**Figure 1.7**).

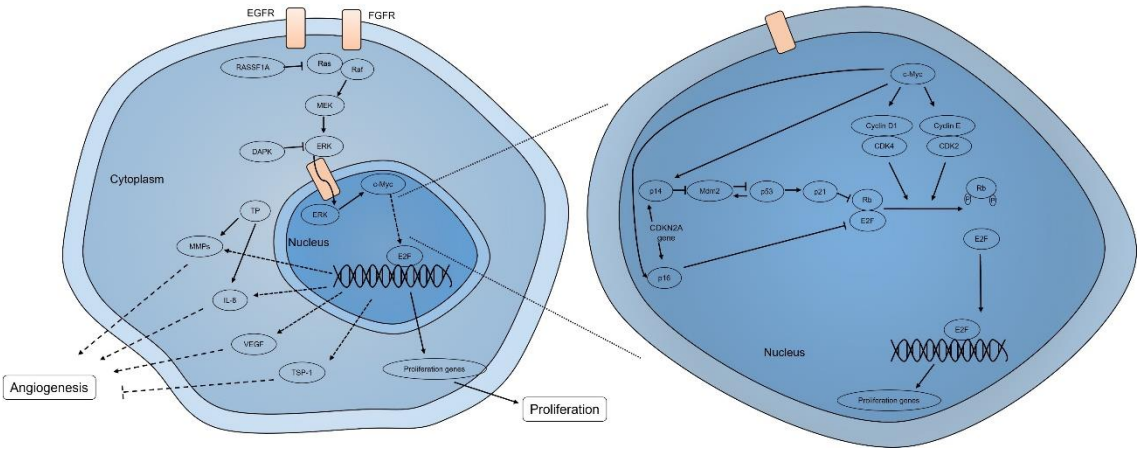

Figure 1.7: Molecular pathway of bladder cancer.

1.9 Artificial Intelligence in bladder cancer

Artificial intelligence (AI) is an algorithm-dependent computer application to modulate increasing data and decrease human-derived error on a crucial medical task. AI-assisted analysis can be a powerful alternative test/assistant method to diagnose numerous diseases, including diabetes, chronic heart disease, tuberculosis, stroke, hypertension, skin and liver disease, Alzheimer's Disease, and cancer [76]. Machine learning (ML) is a subset of AI where computers obtain maximum advantage from data to learn to perform a given issue rather than being obviously programmed with a predetermined set of rules [77]. In this context, deep learning (DL) is a subset of ML and that multiplayer neuronal networks to solve an issue related to a problem in medicine (**Figure 1.8**).

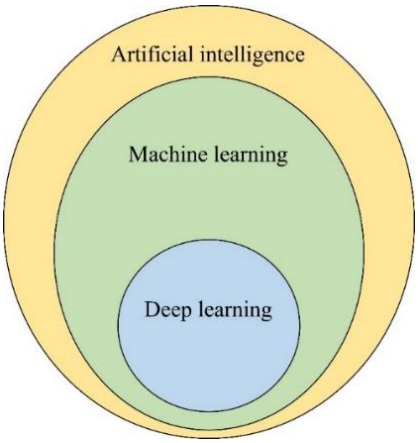

Figure 1.8: Artificial intelligence and its subsets.

Medical data can be obtained in a variety of forms and complexity, including clinical, radiological and pathological images [78], electronical health records [79], data from wearable sensors [80], and omics data [81]. The combination and interpretation of these enormous data sets need to require systematic new perspectives such as AI-based analysis.

The ML process involves a connection map between input and output that produces a model as a solution for an interesting problem. There are five main steps to designing an AI system: 1) getting data, 2) preparation of the data, 3) training of model, 4) testing data/model, and 5) improving the model (**Figure 1.9**). The model can be produced by data that come from medical records, which is also called the learning step of the AI. The vast majority of current AI-based applications can predict a tumor's detection, staging, grading, recurrence, response to chemotherapy, and overall survival prospectively in the BCa [82].

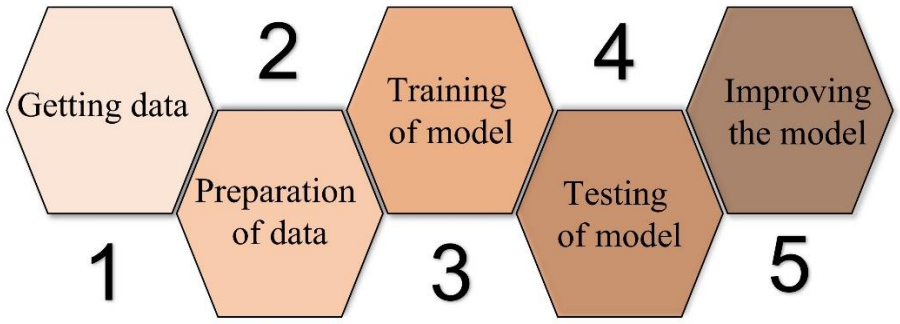

Figure 1.9: Basic steps of artificial intelligence process.

The most common learning method is supervised learning for diagnosis and outcome prediction in BCa. Various learning problems can be solved by using this method, including diagnosis, prognosis, and survival [83]. In supervised learning, the model is trained with labeled data previously. Hence, labels act like a guide for the training process. In the learning process, various solutions are present for learning problems, including regression, classification, and clustering [84]. In classification problems, the output is always a categorical quantity, such as "present" or "not present." In addition, the primary supervised learning algorithm is logistic regression (LR), decision tree (DT), and random forest (RF) [85].

Deep learning models can focus on feature extraction with minimization of external user guidance. The main idea of the DL is to act like functional working of the brain (**Figure 1.10**). DLs automatically identify the critical factors to learn by themselves, especially for image analysis [86]. All decisive factors of an image should be independently defined in ML; but with DL using neural networks, the key features of each image will be automatically detected [87].

In DL, artificial neurons refer to dendrites of the biological neuron are also called perceptrons. Input data from multiple sources are revived and deliver output after processing. Data fed to a perceptron will undergo different functions and transformations and give an output. The perceptrons are connected to an artificial neural network (ANN)[85].

Statistical ML techniques are used in ANN to learn features by a hierarchical structure. This neural network structure has different layers – the input layer receiving all the inputs, hidden layers, and the delivering output layer [87]. The first input is transferred to output layer with extractable patterns, identifying the images lines based on contrast levels. Then the data is transferred to the output layer to deliver the classification[85].

The hidden layer number of the ANN is a critical factor for their classification. Suppose the ANN is constructed with a single layer, called single-layer perception. In contrast, the ANN is constructed with several hidden layers called multilayer perception [85]. In addition, another commonly used algorithm is a convolutional neural network (CNNs), which is helpful for image processing [88]. An image has been able to see a visual input by breaking it down into three color channels (red-green-blue) or gray scale by a computer. Images are classified with their pixels and colour channels. In a CNN, an artificial neuron will partly connect previous layer neurons. Among these applications, convolutional neural networks (CNNs) are the most common AI model to classify medical images [89].

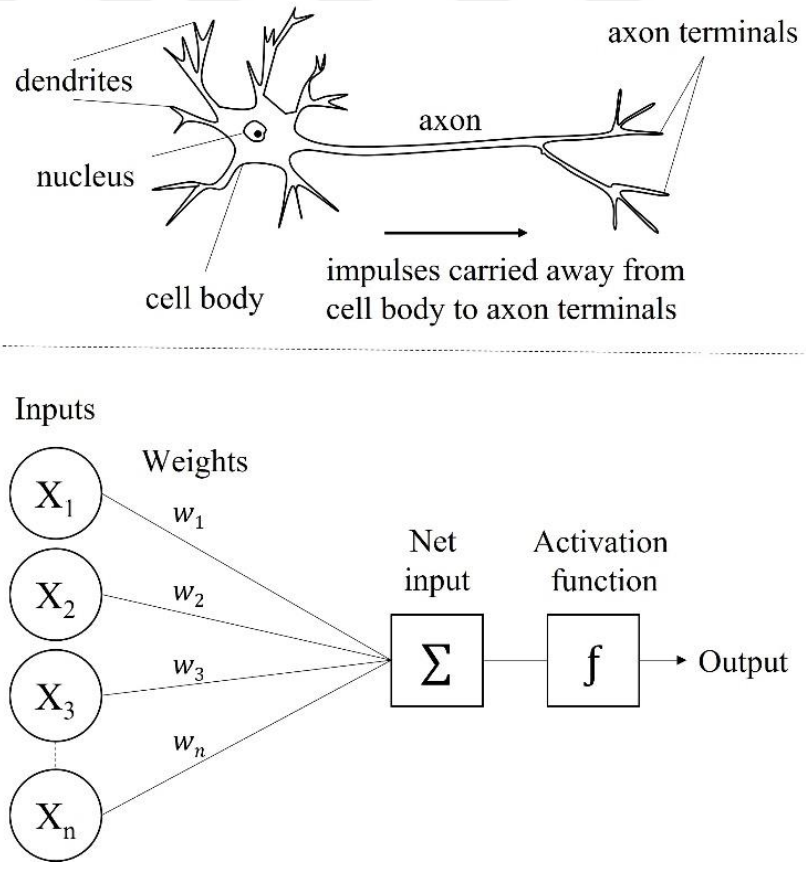

Figure 1.10: Comparation of a biological neuron and artificial intelligence components.

1.10 Clinical importance of blood and urine samples

Blood and urine samples are clinics' most prevalent biological liquids for routine biochemical analysis. They are easily obtained from patients and contain enormous amounts of information relevant to patient healthcare. In addition, many biological factors (physiology, metabolism, and hormones) of patients significantly affect the rheological (study of the flow of matter primarily in a fluid) properties of whole blood (plasma and cells) and urine [90,91]. Morphological properties and biochemical parameters of both biological liquids may be changed in diseases such as BCa. For example, bladder tumors make direct contact with urine, and tumors can leak into the urine by their losing cell-cell contact [39]. Like Nuclear Matrix Protein 22 (NMP-22), the secretion or release of abnormally high levels of specific proteins is positively correlated with BCa [39].

Changes in blood fluidity is determined by rheological factors such as plasma viscosity, erythrocyte aggregation propensity, red blood cell deformability, adhesion properties of platelets and leukocytes, as well as changes in the composition and concentration of plasma components that compose of various proteins, inorganic electrolytes, glucose, and other minor elements. In addition, urine from patients may also contain epithelial cells, erythrocytes, and lymphocytes, as well as other components (sugar, proteins, DNA, and other chemicals) [92]. The drying of deposited drops of blood and urine samples may lead to the formation of complex patterns thanks to these factors affecting their content. During the drying process, the content of both liquids is distributed through evaporation and forms a characteristic pattern derived from specific cracks and distribution of the content [91]. Thereby, whole blood and urine samples obtained from patients with BCa and healthy volunteers are promising opportunities to differentiate patients and control volunteers.

1.11 Transcriptome and RNA sequencing

The transcriptome contains coding or non-coding RNA transcripts of the cells and tissues. These transcripts encode specific proteins that modulate cells' behavior in response to environmental stress conditions. Hence, transcriptomic profile changes are a dynamic process and actively depend on many factors, including developmental stage, environmental condition, and biological process of transcription [93,94]. RNA sequencing (RNA-seq) is the sequencing of complementary DNAs (cDNAs) physically isolated from mRNA transcripts and quantification based on the number of each transcript. This method enables the identification of rare or novel RNA transcript variants in many diseases, such as cancer [94]. Early cancer diagnosis and prognosis are crucial for its response to treatment, diagnosis, prognosis, survival. In addition, biomarkers are found in tissue, blood, and other fluids that are considered a sign of an abnormal condition. In addition, transcriptomic profile can lead to be helpful to understand the alternative new molecular perspectives in many diseases. Therefore, the one of the best ways to discover cancer-specific biomarkers is most probably the transcriptomic analysis of cancer, such as BCa. In addition, tremendous heterogenicity and molecular subtypes of the BCa tumors need to be require uncovering the highly specific markers for BCa diagnosis, prognosis, and follow-up.

1.12 Hypothesis and aims

At least 30 different molecules with BCa biomarker potential have been identified in the literature so far, but few of these markers have been approved for clinical use [14,95]. Available tests are used in clinic, including urine cytology, Nuclear Matrix Protein (NMP-22), BTA TRAK, BTA stat, ImmunoCyt / uCyt +, Fluorescence in situ hybridization (FISH), CxBladder tests [96,97]. The high false positive and false negative rates of these tests make them unreliable diagnostic methods. Moreover, these tests are expensive methods. In addition, there are limited studies leveraging AI models to perform various tasks related to BCa detection, staging, and grading. In this thesis, by analyzing droplet patterns of whole blood and urine samples we aimed to suggested an AI-assisted quick, cheap, and reliable diagnosis method. Droplet pattern analysis of evaporated deposits was performed using blood and urine samples from BCa patient and control volunteers. Our proposed AI-assistant solution (ResNet-18 pre-trained ImageNet) can be systematically applied across those blood and urine droplets, enabling comparisons to reveal shared spatial behaviours and underlying morphological patterns which precisely differentiate cancerous and control samples. The innovative diagnostic method has been presented based on the recognition and classification of complex patterns formed by dried whole blood or urine drops.

Although various markers are suggested to be diagnostic and prognostic signatures for BCa in recurrence and progression, the specific markers and related pathways are rarely documented and still need to be discovered in detail. Recent markers' low sensitivity and specificity are also disadvantages for managing BCa. In this thesis, we comprehensively analyzed and compared the gene expression levels of 10 tumor samples from BCa patients based on RNA seq. We aimed to classify these tumors according to their invasiveness through the bladder wall to uncover the case-specific biomarkers in BCa. In addition, gradients expressed and secreted genes were also identified. In addition, we compared the potential developmental pathways of BCa from NMIBC to MIBC using the gene enrichment analysis. These findings could aid in investigating pathogenesis and mechanisms and provide potential biomarkers and novel therapeutic targets for BCa in the clinic.

In light of the literature evaluation and interpretation, by analyzing droplet patterns of whole blood and urine samples from patients we are planning to develop an artificial intelligence-assisted quick, cheap, and reliable diagnosis method. Droplet pattern analysis of evaporated deposits will be performed using patient and healthy volunteer blood and urine samples. The proposing AI-assistant solution can be systematically applied across those blood and urine droplets, enabling comparisons to reveal shared spatial behaviours and underlying morphological patterns in cancerous and normal samples. The innovative diagnostic method has been presented based on the recognition and classification of complex patterns formed by dried urine or blood drops under different conditions. In addition, to increase the sensitivity and specificity of the AI-based model, we aim to combine mRNA expressions data obtained from bladder tumors with drop patterns to diagnose BCa patients efficiently.

CHAPTER 2
MATERIALS AND METHODS

2.1 Collection of patient samples

2.1.1 Blood and urine samples

The study included 133 BCa patients admitted to the urology department of Marmara University Pendik Training and Research Hospital between 2018 and 2020 in Turkey. The control group was also composed of 64 volunteers who had no BCa diagnosis in their lifetime. After informed consent, the blood samples were taken from BCa patients and the control volunteers before the surgical operation. Urine samples were taken from the same patients and control volunteers after the next day as their first urine **(Table 2.1)**.

Blood and urine samples were taken into the 5 mL blood collection tubes with EDTA, and 50 mL sterile urine containers, respectively, and stored at -80^{o}C until used. All patient and control group samples were obtained according to the ethical rules.

Table 2.1: Number of patients and controls samples.

Samples	Blood	Urine	Tumor
Control	64	36	1
Patient	133	133	10

2.1.2 Tumor samples

The study included 10 BCa tumors and 1 control urinary bladder tissue obtained from the urology department of Marmara University and forensic medicine, respectively. Tumor characteristics were summarized in table 1. The tissue samples of tumors were acquired from BCa patients who underwent transurethral resection of bladder tumor (TURBT). These tumor and control tissue samples were cleaned with sterile normal saline

after the surgical operation and stored at -80 °C until used. According to their invasiveness (T stage) and clinicopathological properties (Grade), they were also classified into 4 cases, including case-1 (pTa-Low grade), case-2 (pTa-High grade), case-3 (pT1-High grade), and case-4 (pT2-High grade) **(Table 2.2)**. All protocols were approved by the Ethical Committee of the Marmara University School of Medicine (Protocol No: 09.2018.367). An opt-out approach was used when obtaining consent from the patients before the study participation.

Table 2.2: Number of patients and controls samples.

Samples	Number of patients	Stage	Grade
Case 1	2	pTa	Low grade
Case 2	3	pTa	High grade
Case 3	3	pT1	High grade
Case 4	2	pT2	High grade

2.2 Preparation of whole blood and urine droplets

The droplet formations were performed with or without two solutions composed of salt mixtures, 1) *KCl* or 2) $KCl+MgCl_2$. Salts were dissolved in deionized water to be used as a stock solution, and the final concentration of each salt chemical was adjusted to 1 Molar. Solution composition selection and optimization steps were previously described [98]. The humidity level was monitored as 32−37% throughout the experiments.

Urine samples were mixed with salt solutions at a 1:1 (volume/volume) ratio. 1 µL urine-salt mixtures or 2 µl blood droplets of were deposited on microscope clear glass slides (Sail Brand, cat. no. 7101) and left to dry at room temperature (22-24°C). Six droplets per patient or control samples were prepared and imaged under the light microscope (Olympus BX53). Dried blood and urine droplets were imaged in adjusted optimum focus and pixel shifts (1360 x 1024 and 4140 x 4096, respectively) for in-depth AI-based

learning analysis. These deposited drops were all imaged in RGB (Red, Green, and Blue) as well as in grayscale color formats. Images were collected and saved in a TIFF file format (**Figure 2.1**).

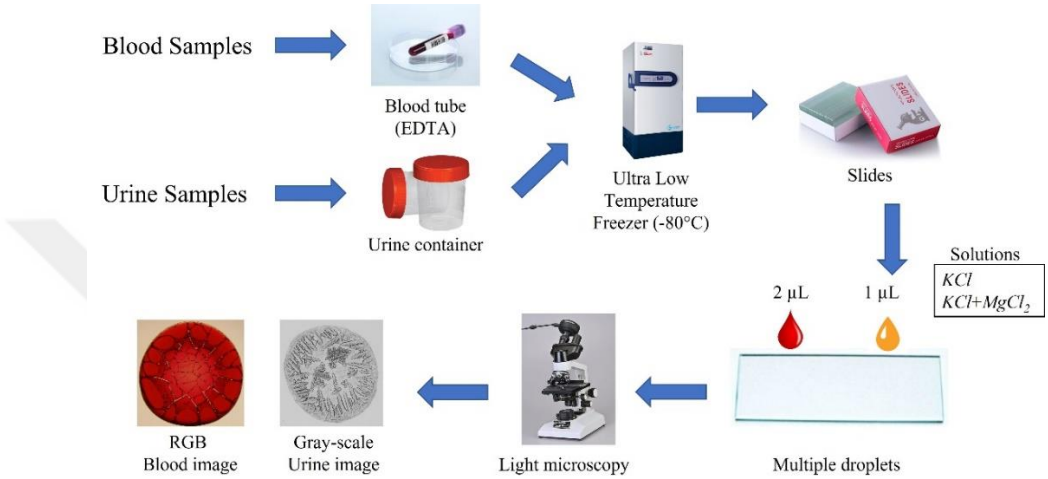

Figure 2.1: Sample collection and imaging steps of blood and urine droplets.

2.3 Investigation of droplet images by AI

Due to its widespread use and success in machine learning and image analysis, ResNet-18 pre-trained ImageNet, a permanent network architecture, was systematically applied across those whole blood and urine droplet images, enabling comparisons to reveal shared spatial behaviours and underlying morphological patterns. ImageNet is also an image database containing hundreds and thousands of images and helps advance computer vision and deep learning processes.

These images of blood and urine samples obtained from BCa patients and control volunteers were categorized into “bladder cancer” and “not bladder cancer.” In these groups, benign blood images were included in the group “not bladder cancer.” The preparation and processing of data were completed in two steps. First, the image data needs to be cleaned to be ready for analysis and AI learning. In the second stage, data was processed online, models were trained, and data analysis performance was optimized.

2.4 CNN architecture and training

We have developed a ResNet-18 CNN-based model for Bca patient/control classification from the blood and urine droplet images [99]. The model used transfer learning on the ResNet architecture with the modified last layer pre-trained on ImageNet, where the network feeds the last fully connected layer of the model into a fully connected layer with 512 neurons. The proposed architecture was given to be used for analysis of whole blood and urine images from Bca patients and control volunteers (**Figure 2.2**). The last fully connected layers of the bladder patient/control model were trained on full-size (1360x1024 and 4140x4096) droplet images to evaluate whole patterns of blood and urine samples, respectively.

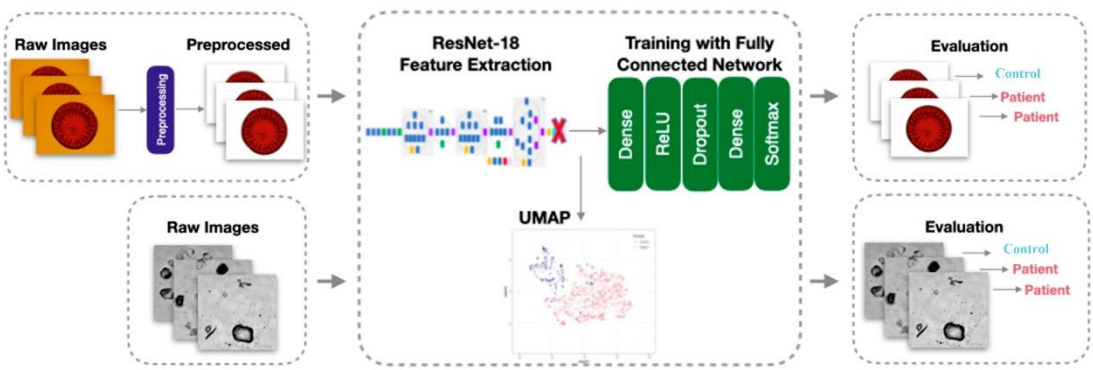

Figure 2.2: Classification workflow for learning of artificial intelligence in blood and urine samples.

The model has been trained with batch-size 64 and epoch 512, where an early stopping method was used to stop training once the model performance stops improving over 20 consecutive epochs on a hold-out 5-fold cross-validation data. The model parameters were optimized via the Adam optimizer with a learning rate of $2x10^{-4}$ with $1x10^{-5}$ l2 weight decay.

For the testing step, 20% of all images from blood and urine samples were excluded. Subsequently, 80% of the images were subjected to a 5-fold cross-validation due to the risk of overfitting. To mitigate the negative effect of a class imbalance, the majority class (bladder patient samples) were under sampled to match the contribution of weights to the minority class (bladder control samples).

2.5 Total RNA isolation and RNA-sequencing analysis

The total RNA of each tumor and control urinary bladder tissue was extracted using the TRIzol Reagent (Invitrogen, 15596018) according to manufacturer's instruction. Briefly, TRIzol reagent added to appropriate amount of tissue was homogenized using a homogenizer. Then, chloroform incubation was performed and obtained RNA was precipitated by isopropanol. Once washing the precipitated RNA with ethanol (75 %), RNA pellet was resuspended in DEPC treated water. RNA samples were quantified by absorbance using the NanoDrop™ spectrophotometer.

RNA samples were sent to BGI genomics (China) for RNA-seq analysis. The RNA sample quantifications (RNA concentration, RNA integrity number value, 28S/18S and the fragment length distribution) were measured by Agilent 2100 Bioanalyzer (Agilent RNA 6000 Nano Kit, Agilent Technologies, Germany). The cDNA library from these samples was prepared in six steps, including 1) mRNA enrichment as oligo dT selection or rRNA depletion, 2) RNA fragment and reverse transcription to double-stranded cDNA by N6 random primers, 3) End repair, A tailing and adaptor ligation, 4) PCR amplification to enrich the purified cDNA templated using PCR primers, 5) Denature and cyclization, and 6) Sequencing by BGISEQ-500 platform (**Figure 2.3**).

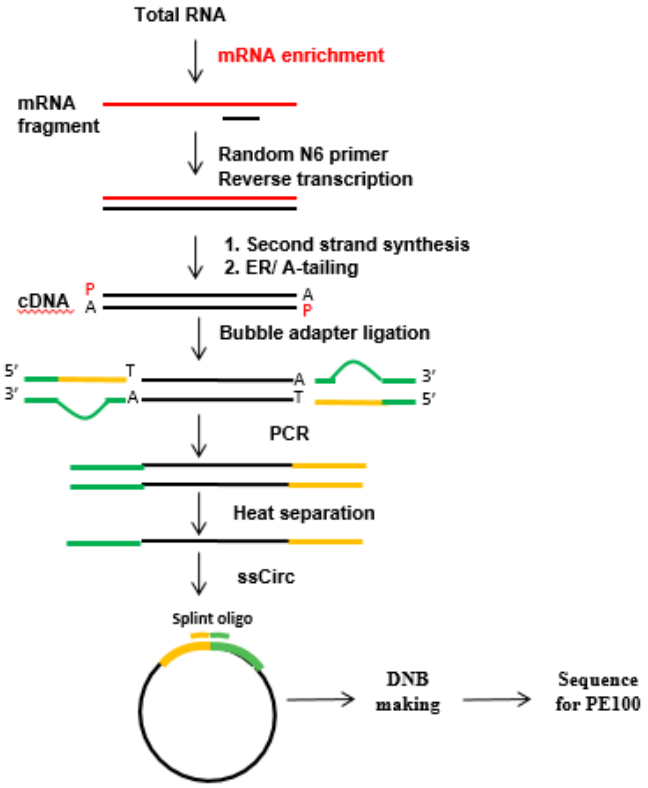

Figure 2.3: BGISEQ-500 transcriptome library workflow.

2.6 Gene expression analysis

The Deseq package of R software was used to normalize for RNA-seq raw count and performed to evaluate differentially expressed genes based on negative binomial distribution [100]. The average expression values of each gene in the tumor cases were divided by expression value of the same gene present in the control sample to calculate the fold change (FC) of each gene separately. The standard error in percent of the mean (SE %) was performed to evaluate genes that have more variable expression values among samples [101]. Thereby, 25 % SE was selected as the threshold for exclusion criteria of more variable genes. After discarding genes with larger than 25 % SE, genes with FC $\leq$ 0.5 and FC $\geq$ 2.0 were assumed to be downregulated and upregulated, respectively.

2.7 Urinary bladder enriched gene analysis

Genes that have high expression in urinary bladder were obtained from The Human Protein Atlas (HPA) [102]. These selected gene lists were compared to previously filtered up and down regulated gene lists of each case separately to identify case-specific genes expressed in the Bca cases.

2.8 Gradient gene expression analysis

Once the gene filtration was performed according to 25 % SE threshold, genes at least 0.2 FC difference between each case pair (case1 and case 2; case 2 and case 3; case 3 and case 4) were selected to determine up / down regulated gradient genes among cancer cases.

2.9 Secreted gene analysis

To detect up regulated and secreted genes in each Bca case, a secreted gene list was occurred by combining gold standard secrete proteins [103] and blood-secreted proteins from the HPA [102]. Then, these two secreted gene lists were combined and used to compare up regulated genes in each Bca case separately.

2.10 The cancer genome atlas (TCGA) expression and survival analysis

GEPIA were used to evaluate pan-cancer, expression/co-expression, and survival analysis of differently expressed genes [104,105]. The TCGA bladder urothelial carcinoma (BLCA) dataset was used to perform gene expression and survival analysis.

2.11 Variant analysis

Somatic variant analysis was performed for the control and tumor samples. First, data quality control was performed using the FastQC tool, and low-quality regions were trimmed using the Trimmomatic tool (v. 0.39). The reference genome (gencode v19) was indexed with the STAR tool (v. 2.7.1a) for alignment. Trimmed reads were aligned to the reference genome with the STAR tool in two-pass mode. The steps between alignment and annotation steps were performed using tools in GATK (v. 4.1.8.1). A reading group was added to the aligned reads with the AddOrReplaceReadGroups tool in GATK, and duplicate reads were marked with the MarkDuplicates tool. Before the variant calling step, the SplitNCigar tool was used to trim the overhanging reads into intronic regions and reassign the mapping qualities for the next step. This step is specific to variant analysis from RNA-Seq data.

Somatic variants were detected with the Mutect2 tool in tumor-only mode. The gnomad file was used as the source of the germline variants while the tool was running. There was an hg38 version of gnomad file (https://console.cloud.google.com/storage/browser/gatk-best-practices/somatic-hg38/af-only-gnomad.hg38.vcf.gz) in gatk cloud. This file was converted to hg19 coordinates with the LiftoverVcf tool. The variant filtration step was performed with the FilterMutectCalls tool. With CancerVar (v. 2020), which was used for the clinical interpretation of somatic variants, variants had been annotated. While running cancervar was specified as tumor-type bladder cancer (BLCA).

Genes that were searched in the literature as mutated in bladder cancer (n=24), upregulated and increased their fold changes from case 1 to case 4 (n=38), and secreted genes for all cases from two sources(n=141). Variants detected in 203 genes determined according to the expression analysis. Moreover, Variants with a read depth of <10 and a mutated read count of <3 were eliminated.

2.12 Pathway enrichment analysis

Downregulated and upregulated genes of each case (case 1, case 2, case 3, and case 4, separately) were used to perform GO enrichment and Pathway enrichment to find out significantly enriched molecular functions, cellular components, biological processes, and pathways. G:profiler (https://biit.cs.ut.ee/gprofiler/gost), a web server for functional enrichment analysis, was used for GO Enrichment and Pathway Enrichment. Terms were listed for upregulated, downregulated, and union of up and downregulated genes for each case.

2.13 Cell culture

T24 cells were cultured in Roswell Park Memorial Institute (RPMI) Medium supplemented with 10% (volume: volume) fetal bovine serum (PAN, P303302), antibiotics (penicillin/streptomycin; Biological Industries, BI030311B) and L-glutamine (Biological Industries, BI030201B) in a 5% CO2 humidified incubator at 37°C.

2.14 Three-dimensional (3D) cell culture

The 3D culture system was performed using commercial Matrigel (BD, 356234) to construct two-layers system containing solid and liquid layers. Liquid Matrigel was spread as bottom solid-layer at 50 µl/well in 96-well cell culture dishes (final concentration, 4 µg/µl). It was waited for 30 minutes for polymerization.

Then, in the upper liquid-layer mixed with cells, 2% Matrigel (final concentration 0.08 µg/µl) was used. T24 cells were prepared with 100 µl medium per 2000 cells, with cell culture medium (Low glucose [1500 mg/L] DMEM, 2% calf serum, Pen/Strep antibiotic and 2% Matrigel) and seeded on the previously polymerized bottom layer. The 3D cultures were monitored for 12 days, with the medium being refreshed in 4 days.

2.15 Establishment of Gal-3 KO T24 cells

CRISPR/Cas9 system was used to generate T24 Gal-3 KO cells. Single guide RNAs (sgRNA) that specifically target to Gal-3 gene were designed by an online tool that was created by Broad institute. After the sequences of BSMBI restriction cutting sites (Thermo, ER0451) have been added into gRNAs, they have been purchased **(Table 2.3)**. The annealed sgRNAs were then cloned into TLCV2 viral vector (Addgene, 87360) **(Figure 2.4)**. The constructed TLCV2 vector containing Gal-3 sgRNAs were transfected concomitantly with pMD2.G (Addgene, 12259) and psPAX2 (Addgene, 12260) vectors into HEK293T cells by calcium phosphate method. After two days later, the cell medium was collected from the transfected cells. The cell media containing virus immediately centrifuged and filtered with a sterile syringe (0.45 µm). T24 cells were transduced by the lentivirus containing media with polybrene (5 µg/ml) (Sigma, H9268). After two days incubation, puromycin selection (1 µg/ml) was performed for five days. Monoclonal selection was performed to obtain individual cell clones, and they are verified by immunoblotting after doxycycline induction.

Table 2.3: Cloning sequences of gRNAs to TLCV2 vector.

gRNAs	Gene and direction	Sequences
gRNA 1	Gal-3 Fwd	CACCGACAAAGCTGGATAATAACTG
	Gal-3 Rev	AAACCAGTTATTATCCAGCTTTGTC
gRNA2	Gal-3 Fwd	CACCGACTTGTTGCAGTACAATCAT
	Gal-3 Rev	AAACATGATTGTACTGCAACAAGTC

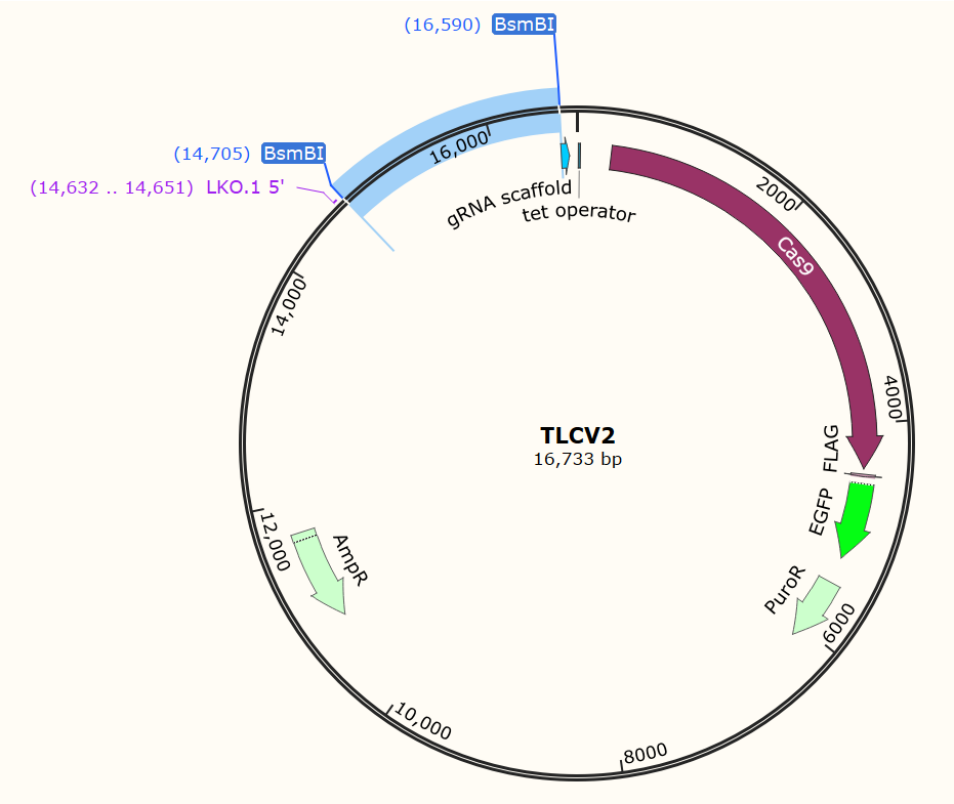

Figure 2.4: Schematic structure and organization of TLCV2 vector. BSMBI, restriction enzyme; AmpR,

2.16 Immunoblot analysis

Total proteins were extracted with RIPA lysis solution (50 mM TRIS-HCl, pH 7.4, 150 mM NaCl, 1% NP40, 0.25% Na-deoxycholate) supplemented with 1 mM phenylmethylsulphonyl fluoride (Sigma-Aldrich, P7626) and protease inhibitor cocktail (Roche, 04693131001). After the protein concentration was measured by nanodrop, total proteins of 50 µg were separated by SDS polyacrylamide gels (12% concentration), and the separated proteins were transferred to nitrocellulose membrane. Following incubation of membrane within 5% non-fat milk solution, membranes were incubated in red solution (3% BSA, pH:7.4 PBS, and 0,05% Tween 20) containing Anti-Galectin 3 antibody (Abcam, ab53082, 1:1000 dilution) or Anti-β-Actin antibody (Sigma-Aldrich, A5441, 1:1000 dilution).

Then, anti-rabbit secondary antibody coupled to horseradish peroxidase for Gal-3 (Jackson Immunoresearch laboratories, 111035144, 1:10,000) and anti-mouse secondary antibody coupled to horseradish peroxidase for β-Actin (Jackson Immunoresearch laboratories, 115035003, 1:10,000) were applied and protein bands were visualized with chemiluminescence.

2.17 Real-time quantitative PCR (RT-qPCR)

TRIzol Reagent (Invitrogen, 15596018) was used to extract total RNA from tumor and control tissue according to manufacturer's instruction mentioned in total RNA isolation section. cDNA was reverse transcribed from the RNA (Dnase treated) by RevertAid enzyme (Thermo Scientific, EP0441) and random primers (Invitrogen, 48190011). SYBR Green master kit (Roche, 04913914001) and a Roche Light Cycler 480 were used to perform for reactions of real time PCR. SYBR Green was activated at 95°C for 10 minutes as an initial step. Fallowing settings were performed for each PCR reaction (total of 40 cycles): 1) 95°C for 15 sec 2) 60°C for 60 second. As a single cycle, thermal denaturation step was performed to generate the dissociation curves for the amplification specificity (95°C for 60 second; 55°C for 60 second; and 80 cycles of 55°C for 10 second). Changes in mRNA level of SH3D21 was quantified using the $2^{-\Delta\Delta CT}$ method using GAPDH (gyceraldehyde-3-phosphate dehydrogenase) mRNA as control gene (**Table 2.4**).

Table 2.4: Primers were used in qPCR experiment.

Genes	Gene and direction	Sequences
SH3D21	Forward	5'-GGGACAGTCAGAAGCTCACC-3'
	Reverse	5'-GGGTTTTGGATCGCTTTCGG-3'
GAPDH	Forward	5'-AGCCACATCGCTCAGACAC-3'
	Reverse	5'-GCCCAATACGACCAAATCC-3'

2.18 Transwell migration

Transwell assays were used to investigate migration ability of T24 cells in 24-well chambers with 8-μm pores (Sarstedt, TC insert, Germany). Enough number of cells (1×10^5 cells/well) were seeded onto upper side of chambers with cell media containing low amount of FBS (1%). Cell media containing high amount of FBS (10%) as a chemo-attractant was also added into lower chambers. Overnight incubation (16 hours) was performed to incubate chambers at 37°C. Then, cotton swab was used to remove non-migrating cells from the chambers. Migrated cell to the high FBS part of transwell membrane successfully were fixed by methanol for 15 minutes. Then, to count the cells under a light microscope, migrated cells were stained with crystal violet (0.1%) for 15 minutes.

2.19 Immunofluorescence staining

Fixation of cells were immediately performed with paraformaldehyde (4%) after removal cell media. Then, permeabilization of cells was performed in PBS containing BSA (0.1%) (Sigma-Aldrich, A4503) and saponin (0.1%) (Sigma-Aldrich, 84510). Immunostaining was performed with anti-galectin-3 (Abcam, ab53082) and anti-LAMP1 antibodies (Abcam, ab25630) for overnight. Then, seconder antibody incubations were performed with anti-rabbit IgG Alexa-Flour-488 (Invitrogen, A11008) and anti-mouse IgG Alexa-Flour-658 (Invitrogen, A11004) for Gal-3 and Lamp1 respectively. DAPI was also used to co-stain cells for 10 minutes (Abcam, ab104139). Mounted slides were then examined under big magnification (63×) by LEICA confocal microscope (DMI8 SP8, Leica, Germany).

2.20 Statistical analysis

Statistical evaluation of the clinical data that may affect blood and urine samples obtained from Bca patient and control volunteers were performed by IBM SPSS Statistics *(Version 20).*

2.21 Ethical approval

This study was approved by the Ethics Committee of Marmara University School of Medicine (Protocol No: 09.2018.367). All procedures were carried out in accordance with the ethical rules and the principles of the Declaration of Helsinki. An opt-out approach was used when obtaining consent from the patients before the study participation.

CHAPTER 3
RESULTS

3.1 Characteristics of patient and control groups

A total of 130 patients with Bca diagnosis (110 male and 20 female) were included to the study. The control cohort group was composed of 64 volunteers (36 male and 28 female) who had no Bca diagnosis in their lifetime. Clinical and pathological characteristics and tumor classifications of cohorts were summarized in **Table 3.1**.

Table 3.1: Clinicopathological distribution of control individuals and bladder cancer patients. *

	Classification	Sample number and gender (Male: Female)	Min. age (Years)	Max. age (Years)	Median age (Years)	Std. deviation (±)
Samples	Control	64 (36:28)	20	83	53	16
	Patient	130 (110:20)	23	89	66	12
Origin of tumor	Primary	96 (82:14)	23	89	66	12
	Recurrence	34 (28:6)	40	88	63	12
Invasiveness	NMIBC	118 (100:18)	23	89	66	12
	MIBC	12 (10:2)	49	86	67	12
Grade	Low grade	61 (53:8)	23	86	63	12
	High grade	68 (56:12)	37	89	68	11
Stage	Cis	2 (2:0)				
	pTa	67 (59:8)	23	88	63	12
	pT1	49 (39:10)	37	89	68	11
	pT2	12 (10:2)	49	86	67	12

***** NMIBC, non-muscle invasive bladder cancer; MIBC, muscle invasive bladder cancer; Min, minimum; Max, maximum; Std, standard

The median age of controls and BCa patients were calculated as 53 ± 16 and 66 ± 12, respectively. All tumors were diagnosed as urothelial cell carcinoma (UCC). The patient cohort was composed of primary (96 cases) or recurrent BCa cases (34 cases). According to invasiveness, patients were categorized as non-muscle invasive (NMIBC, 118 cases) or muscle-invasive (MIBC, 12 cases). Tumor grades were also documented. Tumors were classified as low grade (61 cases) or high grade (68 cases) in the cohort. Tumor grades were determined as Cis (2 cases), pTa (67 cases), pT1 (49 cases), pT2 (12 cases). Detailed information on control cohorts and patients were added as **Appendix A** and **Appendix B** respectively.

3.2 Possible effects of freeze-thaw cycles upon blood samples

Possible effects of freeze-thaw cycles were documented (**Figure 3.1**). During the process, four samples were assessed to investigate evaporated sessile droplet patterns. Total hemolysis was achieved after 3 or more cycles (**Figure 3.2**) and droplet patterns (shadows, cracks, patterns, crystals etc.) were consistent after this treatment. In addition, blood parameter used in clinic were analyzed between control and patients groups (**Appendix C**), and there were major changes were not detected from blood contents.

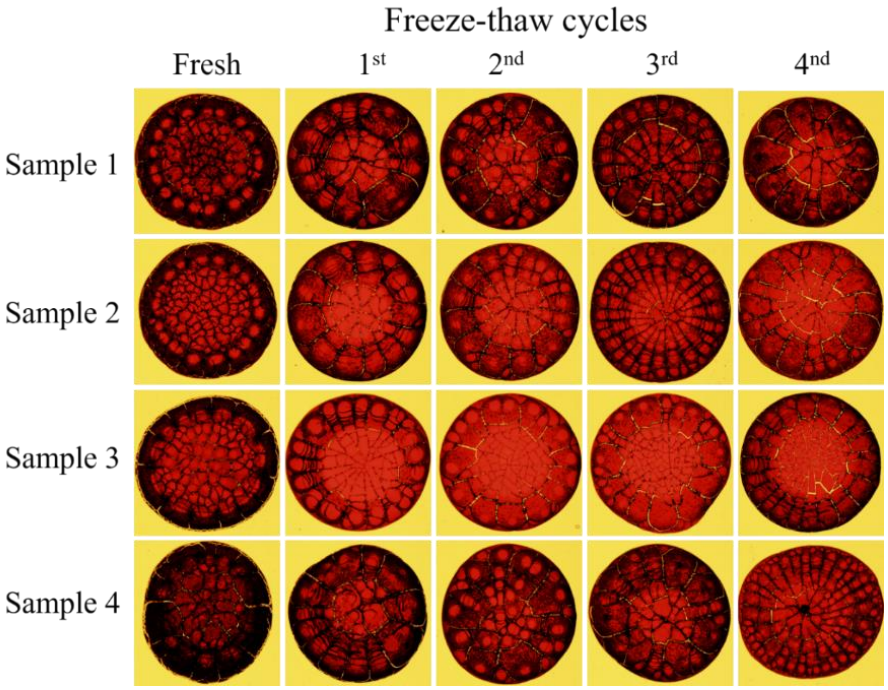

Figure 3.1: Effects of freeze-thaw cycles on the pattern of whole blood droplets.

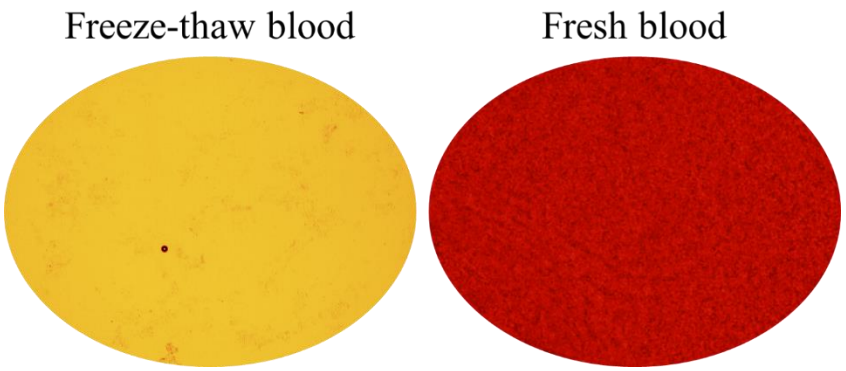

Figure 3.2: Visualization of freeze-thawed and freshly obtained blood samples by hemocytometer under microscopy.

3.3 Imaging of blood droplets

Whole blood samples were collected from BCa patients and control individuals in EDTA tubes before the surgical procedures and samples were frozen min. 2 hrs in -80°C freezers. Droplet patterns were obtained following deposition of 2 µl blood on microscope clear glass slides and drying at room temperature. Images were taken under a light microscope. (**Figure 3.3**). 4-6 droplet images were taken for each case, and a total of 775 and 371 images were captured from patients or controls, respectively. Subsequent machine learning and AI analyses were performed using these image collections.

3.4 Imaging of urine droplets

First morning urine samples were collected from patients or controls and frozen in -80°C freezers. Urine samples were mixed 1:1 (volume: volume) with KCl (1 M) or a KCl (1 M) and $MgCl_2$ (1 M) mixture. Droplet patterns were obtained following deposition of 1 µl urine-salt mixture on microscope clear glass slides and drying at room temperature. and images (4-6 droplet images per case) were analyzed. Images were taken under a light microscope (**Figure 3.3**). 4-6 droplet images were taken for each case, and a total of 779 (KCl mixture), 772 (KCl + $MgCl_2$ mixtures) or 214 (control KCl mixture), 215 (control KCl + $MgCl_2$ mixtures) were taken. Subsequent machine learning and AI analyses were performed using these image collections.

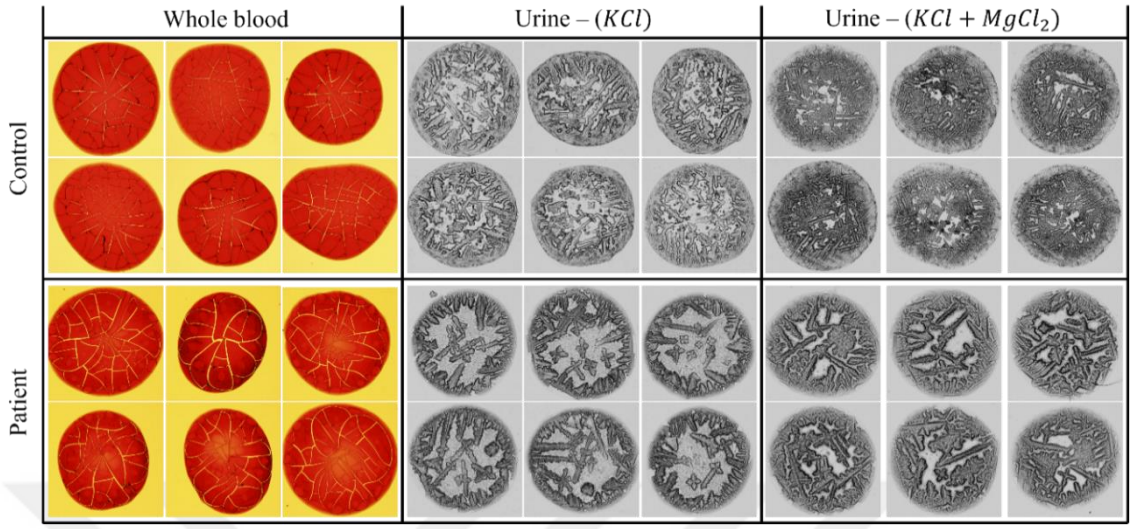

Figure 3.3: Examples of images of whole blood and urine droplet patterns from control individuals and bladder cancer patients. KCl, potassium chloride; $MgCl_2$, magnesium chloride.

3.5 Feature Extraction and CNN Classification

The CNN architecture was used to classify images derived from blood or urine samples of patients or controls using ResNet-18 based neural network that was pretrained on ImageNet (**Figure 2.2**). The last layer of this network was dropped to observe more in-depth behavior of the extracted features from both control and patient droplets. At this step, we have applied a nonlinear dimensionality reduction technique, uniform manifold approximation and projection (UMAP [X]), to analyze high-dimensional image features to find out image-based patterns of droplet samples. The UMAP plots highlighted meaningful organization of sample clusters and improved interpretation of separation behind different samples. UMAP plot of whole blood features extracted from patients and control's images was visualized in **Figure 3.4**. In addition, UMAP plot of urine samples mixed with KCl or KCl+$MgCl_2$ salt solution (urine-KCl or urine-KCl+$MgCl_2$) were also visualized in **Figure 3.5** and **Figure 3.6** respectively.

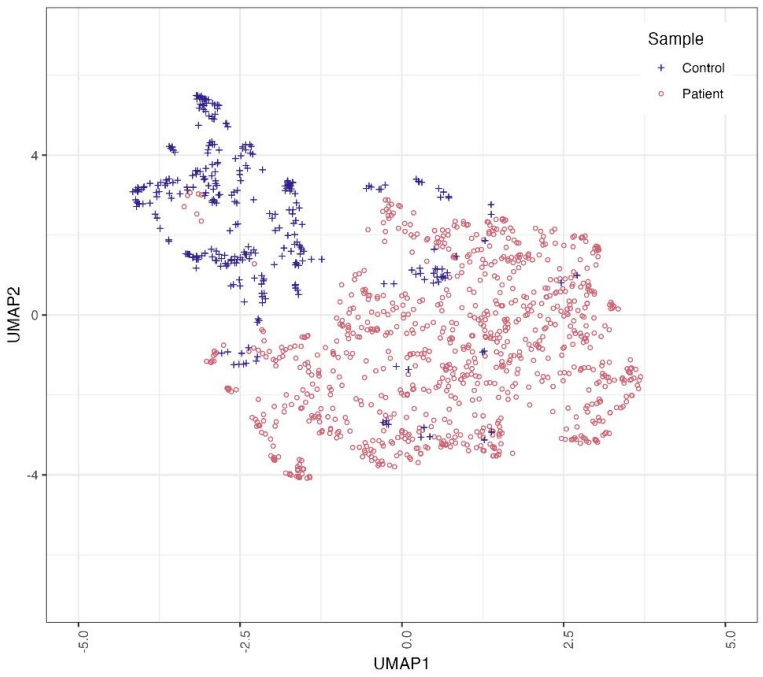

Figure 3.4: Uniform manifold approximation and projection (UMAP) distribution of the whole blood features from control individuals and cancer patients.

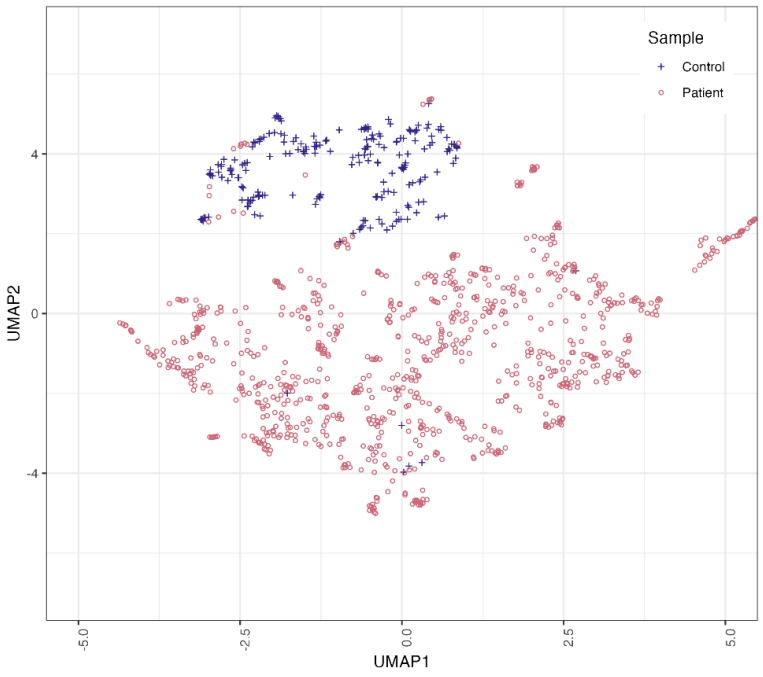

Figure 3.5: Uniform manifold approximation and projection (UMAP) distribution of the urine + KCl features from control individuals and cancer patients.

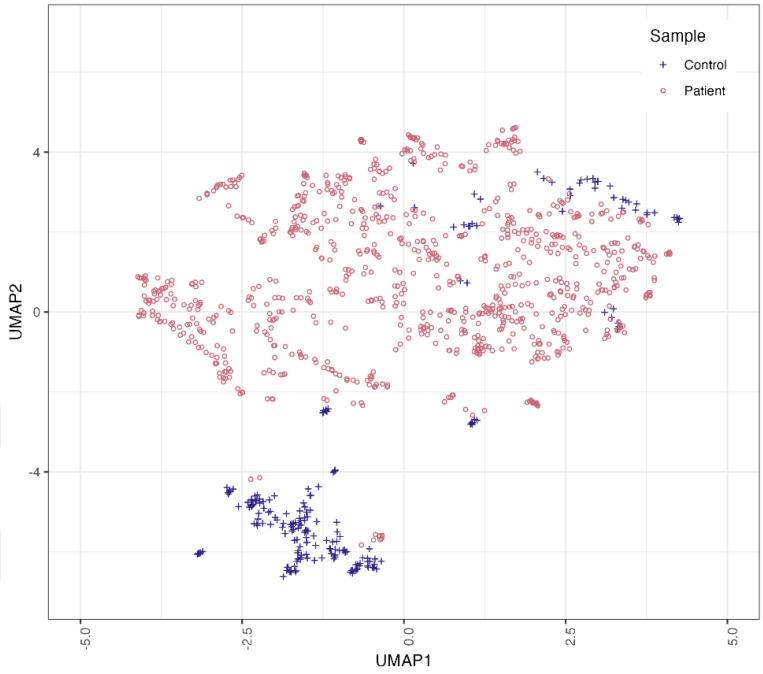

Figure 3.6: Uniform manifold approximation and projection (UMAP) distribution of the urine – KCl + $MgCl_2$ features from control individuals and cancer patients.

The receiver operating characteristic (ROC) curves were plotted as a graphical illustration that shows the discrimination ability of binary classifier system. The ROC curves were performed using true positive (probability of true detection) rate against to false positive rate (probability of false detection). The ROC curve plot of whole blood predictions was shown in **Figure 3.7**. In addition, The ROC curve plots of urine samples mixed with KCl or KCl + $MgCl_2$ salt solutions were shown in **Figure 3.8** and **Figure 3.9**, respectively.

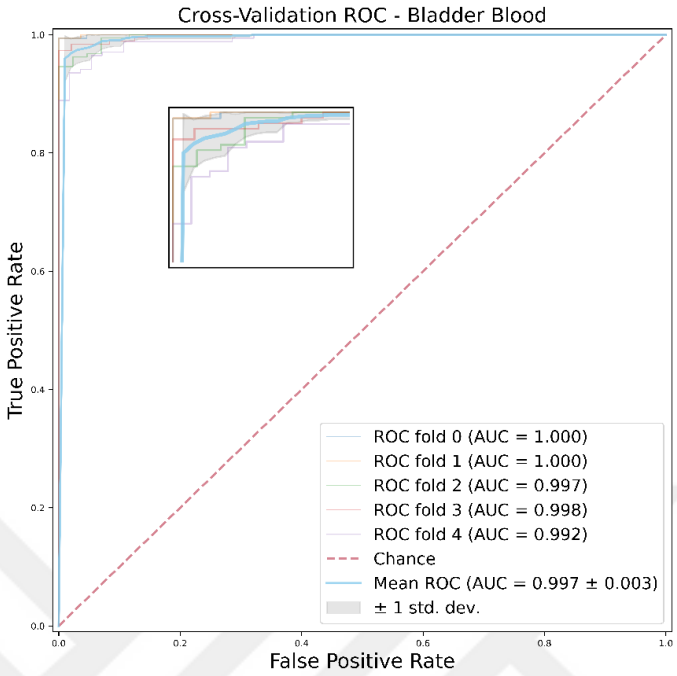

Figure 3.7: Graphical plot of receiver operating characteristic (ROC) curves from whole blood testing outputs by trained artificial intelligence model.

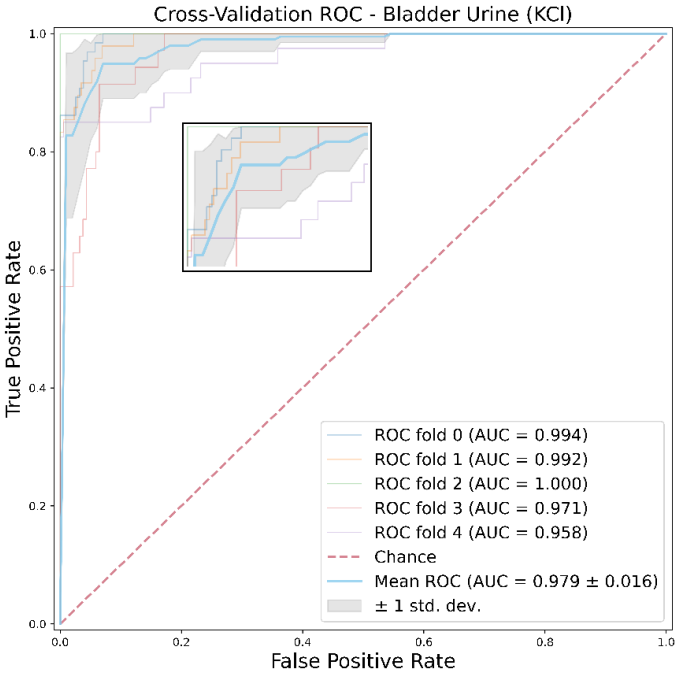

Figure 3.8: Graphical plot of receiver operating characteristic (ROC) curves from urine + KCl testing outputs by trained artificial intelligence model.

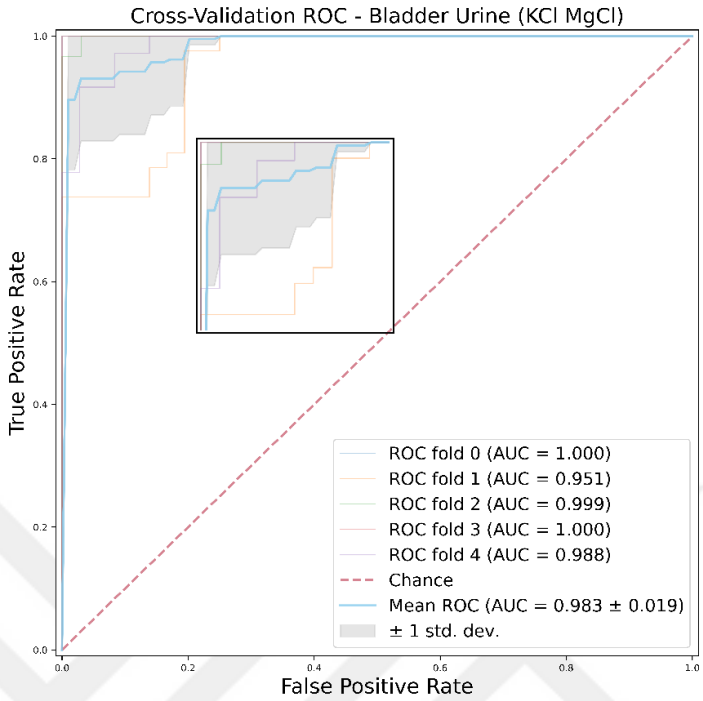

Figure 3.9: Graphical plot of receiver operating characteristic (ROC) curves from urine – KCl + $MgCl_2$ testing outputs by trained artificial intelligence model.

The learning step comes after that behavioral observation step, where 512 extracted features of the ResNet-18 model feeds into a fully connected layer with 512 neurons. This step included two fully connected dense layers with ReLU activation and dropout regularization which ends with softmax layer to make classification between control and patient samples. The model was trained with original size droplet images (1360x1024 for blood and 4140x4096 for urine) to preserve whole pattern in images and to boost the learning without dividing into image patches.

The model training was performed with a batch-size of 64 and epoch 512, where early stopping method was used to stop training once the model performance stops improving over 20 consecutive epochs on a hold-out 5-fold cross-validation data. The model parameters were optimized via the Adam optimizer with a learning rate of 2×10^{-4} with 1×10^{-5} l2 weight decay. For the testing step, 20% of all bladder blood and urine samples were excluded. Subsequently, 80% of the samples were subjected to a 5-fold cross-validation due to the risk of overfitting. To mitigate the negative effect of a class imbalance, the majority class were under sampled to match the contribution of weights to the minority class.

The model was produced by data that obtained from deposited blood and urine images. The network was trained with 80% of the samples and then tested 20% of all bladder blood and urine samples. Given the dataset of images which a patient/control has labeled with "bladder cancer" or "not bladder cancer" in the training step. The network was trained by adjusting its weights so that it tests new unlabeled images that were previously separated from the training samples. The neural network classified these unlabeled images of samples into two categories as "bladder cancer" or "not bladder cancer".

According to receiver operating characteristic curves (ROC) obtained from 20% testing images of the cohort for each blood, urine with KCl or KCl + $MgCl_2$, our proposed model precisely differentiates cancerous patients and control volunteers with mean of AUC (0.997 ± 0.003), accuracy (0.973 ± 0.016), sensitivity (0.977 ± 0.039) and specificity (0.972 ± 0.014) for blood; mean of AUC (0.908 ± 0.066), accuracy (0.953 ± 0.034), sensitivity (0.987 ± 0.119) and specificity (0.829 ± 0.018) for Urine **—** *(*KCl); mean of AUC (0.988 ± 0.021), accuracy (0.748 ± 0.171), sensitivity (0.683 ± 0.386) and specificity (0.882 ± 0.171) for Urine **—** (KCl + $MgCl_2$) (**Table 3.2**).

Table 3.2: Performance outputs of the artificial based model*

	Whole blood	Urine — (KCl)	Urine — (KCl + $MgCl_2$)
AUC	0.997 ± 0.003 (1.000, 1.000, 0.997, 0.998, 0.992)	0.979 ± 0.016 (0.994, 0.992, 1.000, 0.971, 0.958)	0.983 ± 0.019 (1.000, 0.951, 0.999, 1.000, 0.988)
Accuracy	0.973 ± 0.016 (0.996, 0.974, 0.961, 0.979, 0.956)	0.953 ± 0.034 (0.928, 0.955, 1.000, 0.914, 0.968)	0.748 ± 0.171 (0.739, 0.808, 0.635, 1.000, 0.556)
Sensitivity	0.977 ± 0.039 (1.000, 1.000, 0.977, 1.000, 0.911)	0.987 ± 0.119 (0.754, 0.854, 1.000, 0.686, 0.850)	0.683 ± 0.386 (1.000, 0.833, 0.470, 1.000, 0.111)
Specificity	0.972 ± 0.014 (0.994, 0.965, 0.957, 0.973, 0.971)	0.829 ± 0.018 (1.000, 0.983, 1.000, 0.957, 0.994)	0.882 ± 0.171 (0.625, 0.786, 1.000, 1.000, 1.000)

* AUC, area under the receiver operating characteristic curve; KCl, potassium chloride; $MgCl_2$, magnesium chloride; confidence interval of the outputs, 95%.

3.6 RNA sequencing

The number of raw reads of each sample were approximately 47 and 43 million (M), and total clean base of each sample was calculated approximately 4.6 and 4.2 Gigabase (Gb) for BCa tumor and control urinary bladder tissue samples respectively. The clean read ratios were also calculated as 97 % for these each sample (**Appendix D**).

Two samples for case 1, three samples for case 2, three samples for case 3, and two samples for case 4 were used for the determination of up-regulated and down-regulated genes. The genes with zero expression in all samples were removed from the dataset. After the RNA sequencing datasets was normalized via DEseq normalization, the fold change of each gene existed in BCa tumor cases were calculated separately according to control urinary bladder sample. According to 25 % standard error (SE) showing the more variable expression values among samples, the FC values of genes larger than 25 % SE were filtered to obtain low variable expression values. Gene numbers were summarized in **Table 3.3**.

Table 3.3: Number of genes that up- and down-regulated after 25 % standard error filtration. FC, fold change; SE, standard error; Case 1, pTa-LG; Case 2, pTa-HG; Case 3, pT1-HG; and Case 4, pT2-HG or invasive.

Cases (SE % ≤ 25)	**Down-regulated genes (FC ≤ 0.5)**	**Up-regulated genes (FC ≥ 2.0)**	**Total genes**
Control - Case 1	2608	2330	4938
Control - Case 2	1814	2241	4055
Control - Case 3	1721	2223	3944
Control - Case 4	2234	2233	4467

Venn diagram was used to show a relationship among differentially expressed up and down-regulated gene numbers that are common in four BCa cases. FCs above 2.0 was considered up-regulated genes (**Figure 3.10**), and FCs below 0.5 was considered down-regulated genes (**Figure 3.11**). This Venn diagrams described the number of up-regulated genes was more than down-regulated ones commonly expressed in four cases. In addition to commonly up- and down-regulated gene numbers, each case contained various gene that discrete them from other cases.

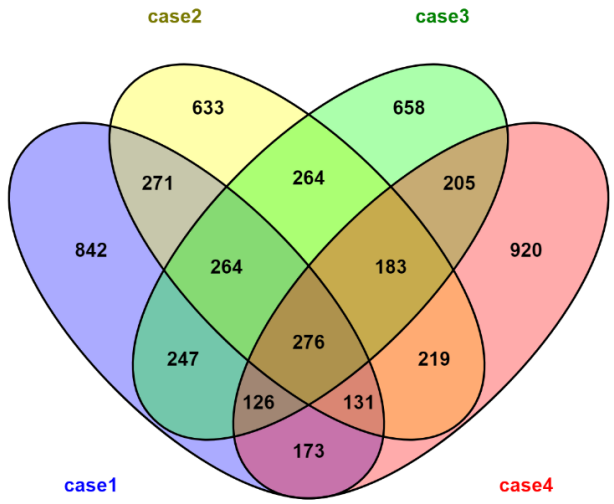

Figure 3.10: Venn diagram showing the up-regulated gene numbers between four BCa cases. Case 1, pTa-LG; Case 2, pTa-HG; Case 3, pT1-HG; and Case 4, pT2-HG or invasive.

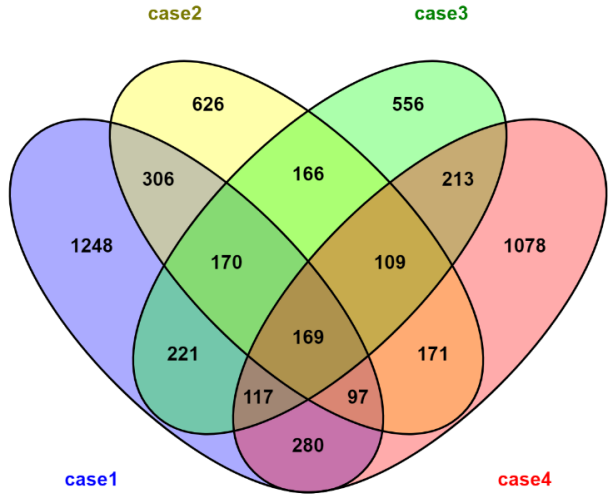

Figure 3.11: Venn diagram showing the down-regulated gene numbers between four BCa cases. Case 1, pTa-LG; Case 2, pTa-HG; Case 3, pT1-HG; and Case 4, pT2-HG or invasive.

The expression levels of up- and down-regulated genes in four BCa cases were also shown in **Figure 3.12**. Total of 395 differentially expressed genes have been determined in BCa tumors compared to non-cancerous urinary bladder tissue as a control. The expression level of these gene above 2.0 ($\log_2 2$) for up-regulated genes and below 0.5 ($-\log_2 2$) for down-regulated genes.

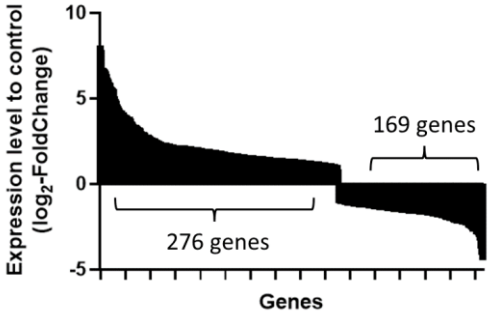

Figure 3.12: Expression levels of differentially up- and down-regulated genes between BCa cases, including, case 1, case 2, case 3, and case 4.

Heatmap of 395 commonly expressed genes have been constructed to show the differential expressed gene clusters between four cases (**Figure 3.13**). Which group genes existed in a case together based on the similarity of their gene expression pattern. In the heatmap each row represents a gene (up- or down- regulate) and each column represent a case of BCa. The colour and their intensity of each row or gene described the changes of gene expression according to normalization used in the heatmap (Pearson correlation and average linkage methods). According to the heatmap clustering, there were four main clustering group as shown in left dendrogram (**Figure 3.13**). Case 4 (pT2-HG) was distinguished from other cases. In addition, other case 1 (pTa-LG), case 2 (pTa-HG), and case 3 (pT1-HG) also described group genes clustered as different dendrogram clustering as shown in the (**Figure 3.13**).

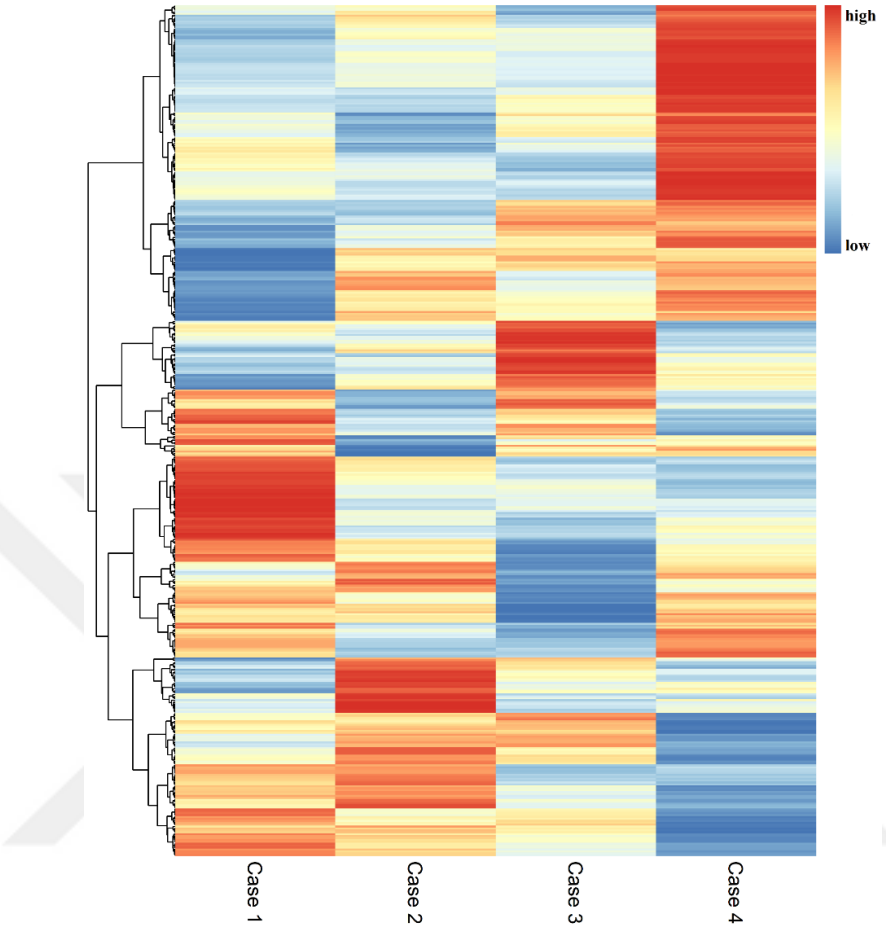

Figure 3.13: Cluster analysis of differentially expressed up- and down regulated genes in BCa tumors. Expression values for each gene (row) are normalized across all samples (columns) by Z-score. row clustering was applied. Case 1, pTa-LG; Case 2, pTa-HG; Case 3, pT1-HG; and Case 4, pT2-HG or invasive.

3.7 Determination of top 50 up-regulated genes

Fallowing the order of all differentially expressed genes in all cases, the 50 up-regulated genes with expression values were shown in **Figure 3.14**. Expression values of each same gene between cases have shown that variation of gene expression levels was closer to each other. In addition, to determine statistically significant up-regulated genes in BCa tumors, these top 50 up-regulated genes were analysed in The Cancer Genome Atlas (TCGA) and The Genotype-Tissue Expression (GTEx) datasets by GEPIA. Top 10 statistically significant up-regulated genes were determined (**Figure 3.15**).

These genes significantly expressed BCa tumors when compared to normal urinary bladder tissue with 99 % confidence interval.

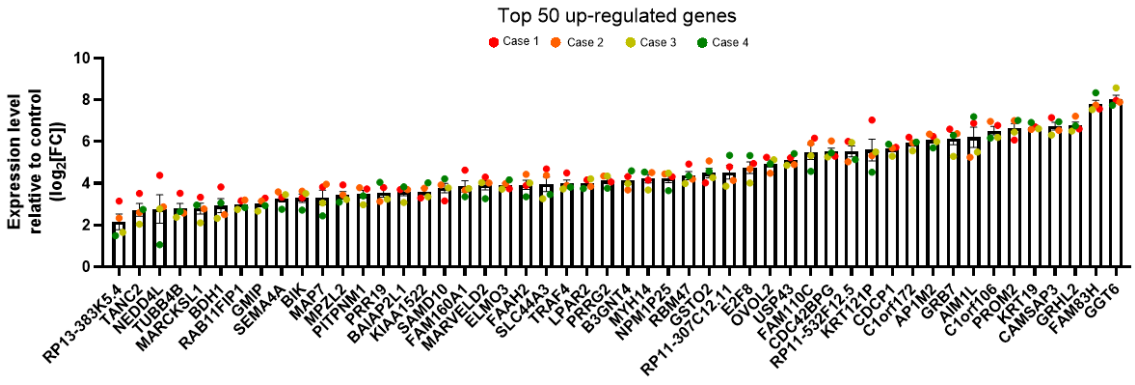

Figure 3.14: Top 50 differentially up-regulated common genes expressed in four BCa cases. Case 1, pTa-LG; Case 2, pTa-HG; Case 3, pT1-HG; and Case 4, pT2-HG or invasive. Each coloring spheres present gene expression value (as fold change) in relevant BCa case. Bars show standart deviation (± SEM).

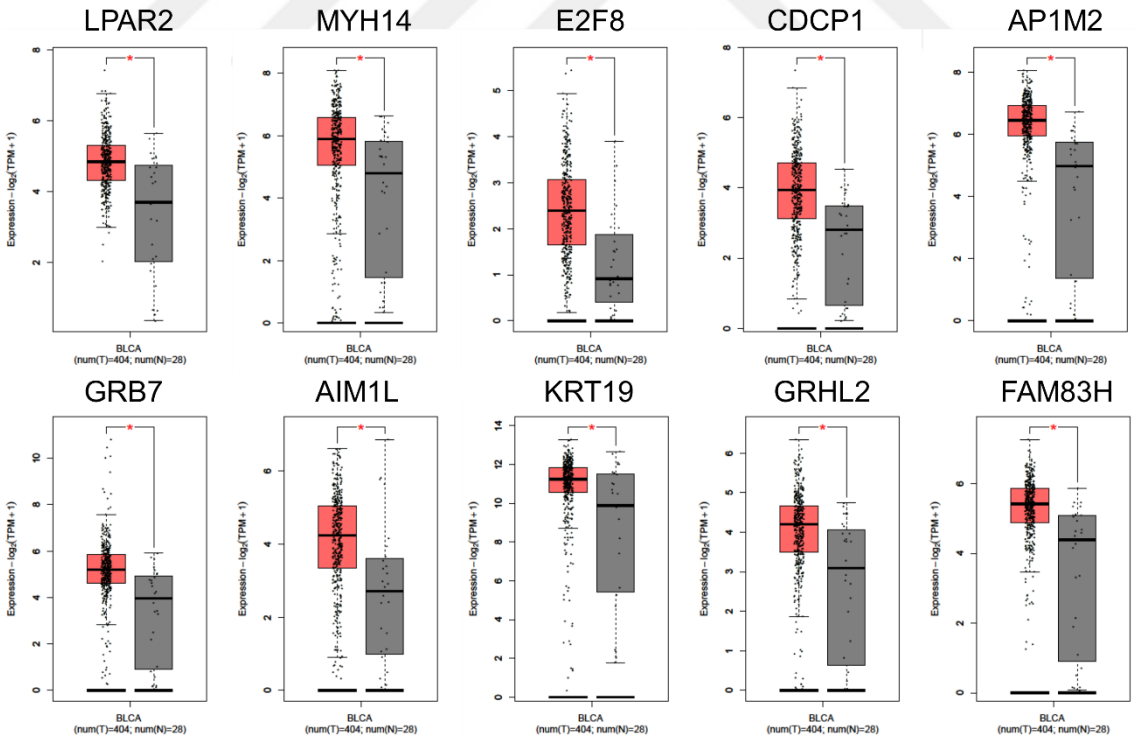

Figure 3.15: Tumor/normal differential expression analysis for top 10 up-regulated genes statistically. Bladder urothelial carcinoma (BLCA) dataset, cancer types of interest; TCGA normal + GTEx normal, control dataset. (p-value<0.01; Log_2FC cut off, 1)

The open name and potential function of these genes were also summarized in **Table 3.4**. Among these genes, exact open name and function of one gene (C1orf106) was not identified in the literature. Other genes mainly responsible for intracellular, membrane, and metabolic functions.

Table 3.4: Top 10 up-regulated genes with open name and potential cellular function (p-value<0.01; Log_2[FC] cut off, 1).

Gene	Open name	Function
LPAR2	Lysophosphatidic acid receptor 2	G-protein coupled receptors
MYH14	Myosin heavy chain 14	Cell shape
E2F8	E2F transcription factor 8	Cell cycle
CDCP1	CUB domain containing protein 1	cell adhesion
AP1M2	AP-1 complex subunit mu-2	Adaptor protein
GRB7	Growth factor receptor-bound protein 7	Adaptor protein
AIM1L	Beta/gamma crystallin domain-containing protein 2	Cytoplasmic protein
KRT19	Cytokeratin	Intermediate filament
GRHL2	Grainyhead-like genes	Transcription factor
FAM83H	Family with sequence similarity 83 member H	Cytoplasmic protein

3.8 Determination of top 50 down-regulated genes

The 50 down-regulated genes with expression values were shown in **Figure 3.16**. Expression values of each same gene between cases have shown that variation of gene expression levels was closer to each other. In addition, to determine statistically significant up-regulated genes in BCa tumors, these top 50 up-regulated genes were analysed in The Cancer Genome Atlas (TCGA) and The Genotype-Tissue Expression (GTEx) datasets by GEPIA. Top 10 statistically significant up-regulated genes were determined (**Figure 3.17**). These genes significantly expressed BCa tumors when compared to normal urinary bladder tissue with 99 % confidence interval.

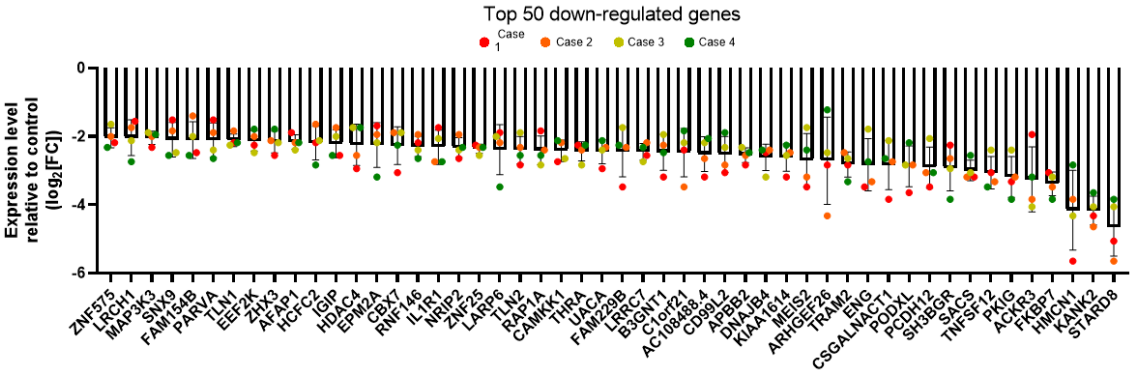

Figure 3.16: Top 50 differentially down-regulated common genes expressed in four BCa cases. Case 1, pTa-LG; Case 2, pTa-HG; Case 3, pT1-HG; and Case 4, pT2-HG or invasive. Each coloring spheres present gene expression value (as fold change) in relevant BCa case. Bars show standart deviation (±SD)

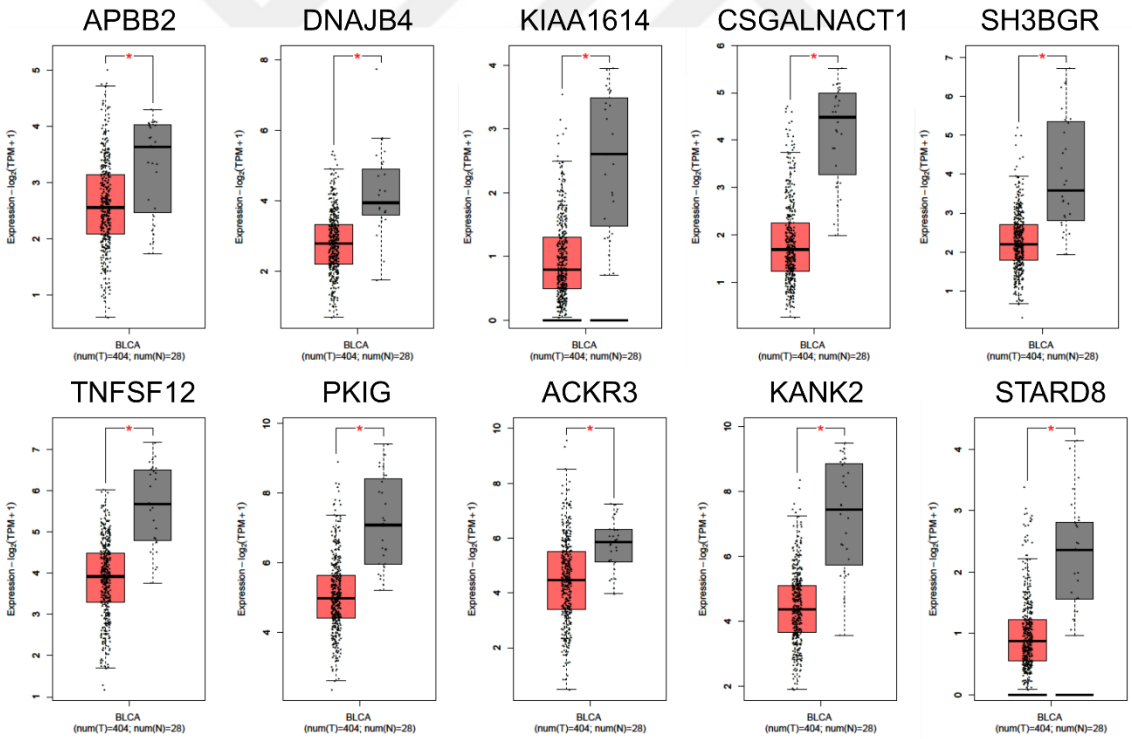

Figure 3.17: Tumor/normal differential expression analysis for top 10 up-regulated genes statistically. Bladder urothelial carcinoma (BLCA) dataset, cancer types of interest; TCGA normal + GTEx normal, control dataset. (p-value<0.01; Log_2FC cut off, 1)

The open name and potential function of these genes were also summarized in **Table 3.5**. These genes mainly responsible modulate cell adhesion, polarity and transcriptional regulatory functions.

Table 3.5: Top 10 down-regulated genes with open name and potential cellular function (p-value<0.01; Log_2FC cut off, 1).

Gene	Open name	Function
APBB2	Amyloid beta precursor protein binding family B member 2	muscle cell strength
DNAJB4	DnaJ heat shock protein family (Hsp40) member B4	Chaperone
KIAA1614	-	-
CSGALNACT1	Chondroitin sulfate N-acetylgalactosaminyltransferase 1	N-acetylglucosamine transferase
SH3BGR	SH3 domain binding glutamate rich protein	Migration / angiogenesis
TNFSF12	TNF superfamily member 12	TNF signaling
PKIG	CAMP-dependent protein kinase inhibitor gamma	protein kinase inhibitor
ACKR3	Atypical chemokine receptor 3	Chemokine signaling
KANK2	KN motif and ankyrin repeat domains 2	Transcriptional regulation
STARD8	StAR related lipid transfer domain containing 8	GTPase activation

3.9 Cell survival analysis of top 50 up- and down-regulated genes

Top 50 up- and down-regulated genes were assessed for their overall survival (OS) and disease / relapse free survival (DFS or RFS) analysis based on their gene expression. OVOL2 was only gene that was found to have statistically significant effects on DFS in BLCA dataset **(Figure 3.18**). The hazard ratio (HR) of the gene was calculated as 1.5 in BLCA dataset. When expression level of the gene was analyzed in BRCA of TCGA, OVOL2 expression was found to increased but not statistically significant.

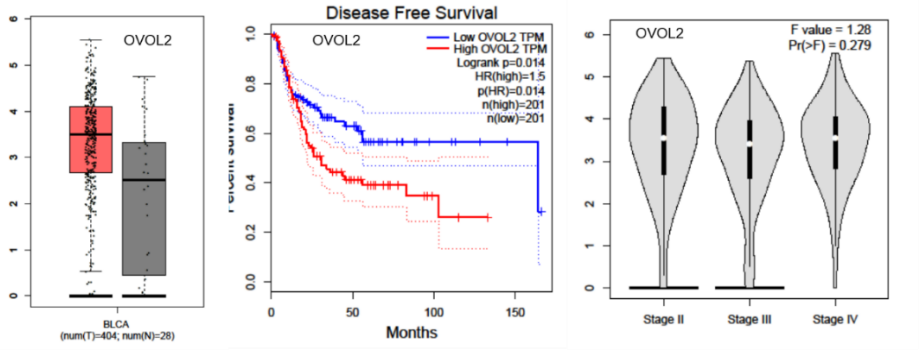

Figure 3.18: Gene expression and survival analysis of OVOL2 gene in BLCA cancer dataset. Group cut off was adjusted as median, 50%; confidence interval, 95%.

3.10 Analysis of gradient genes expressed in bladder cancer

Genes expressed in gradient across the tumor cases of the BCa were determined with 0.2 FC difference between case pairs. These gradients contain thirty-six genes and two Long noncoding RNA (lincRNA) that were also clustered and visualized by heatmap (**Figure 3.19**).

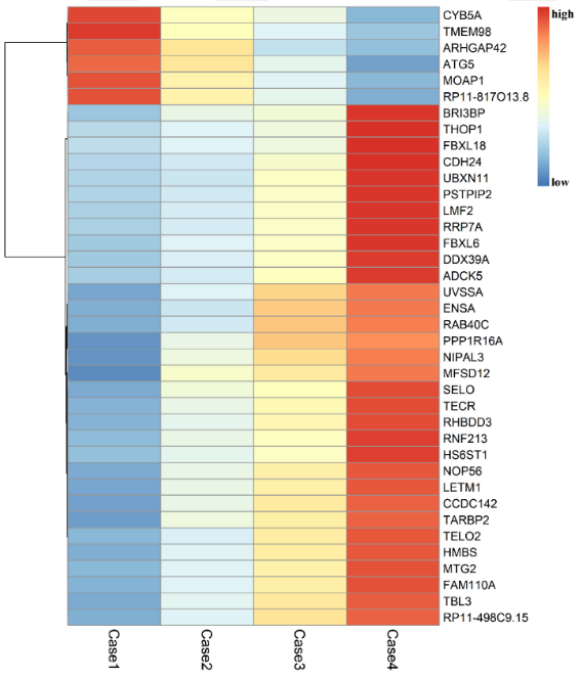

Figure 3.19: Cluster analysis of differentially expressed up- and down-regulated gradient genes in BCa tumors. Expression values for each gene (row) are normalized across all samples (columns) by Z-score. row clustering was applied. Case 1, pTa-LG; Case 2, pTa-HG; Case 3, pT1-HG; and Case 4, pT2-HG or invasive.

Among these gradients, genes that were increased at least two-fold and continued to be increasing of expression until case 4 (pT2-HG) were identified. In contrast, genes that were increased at least two-fold in case 1 (pTa-LG) but then the expressions conversely continued to decrease were also identified **(Figure 3.20)**. In addition, open name and potential biological functions of these genes were also summarized in **Table 3.6**. These functions composed of various biological regulation, including cell cycle, DNA repair, migration, mitochondrial ion pumps, heme synthesis, phosphatases, and RNA processing.

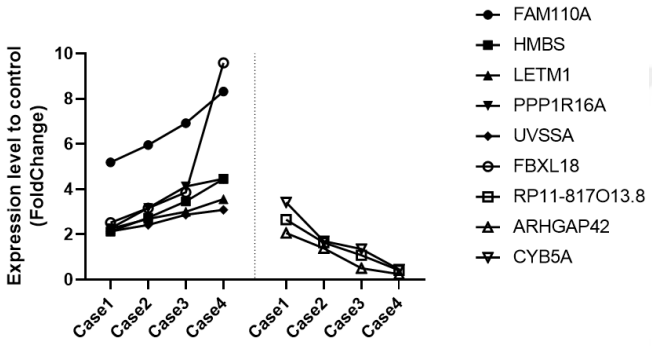

Figure 3.20: Gradient expressing genes between cases of BCa. Case 1, pTa-LG; Case 2, pTa-HG; Case 3, pT1-HG; and Case 4, pT2-HG or invasive.

Table 3.6: Gradient expressing genes between cases of BCa showing the open name and potential biological functions. ↑, up regulated gene; ↓, down-regulated gene.

Gene	Expression	Open name	Function in
FAM110A	↑	Family with sequence similarity 110 member A	Cell cycle
HMBS	↑	Hydroxymethylbilane synthase	Heme synthesis
LETM1	↑	Leucine zipper and EF-hand containing transmembrane protein 1	Mitochondrial proton/calcium antiporter
PPP1R16A	↑	Protein phosphatase 1 regulatory subunit 16A	Phosphatase
UVSSA	↑	UV stimulated scaffold protein A	DNA repair
FBXL18	↑	F-box and leucine rich repeat protein 18	Cell cycle
RP11-817O13.8	↓	lincRNA	RNA processing
ARHGAP42	↓	Rho GTPase activating protein 42	Migration
CYB5A	↓	Cytochrome b5 type A	Electron carrier

3.11 Determination of differently expressed urinary bladder enriched genes

Urinary bladder-enriched genes (188 genes) that have high expression levels compared to other tissue types were obtained from the HPA. These genes were investigated in the up-and down-regulated genes expressed in four cases of BCa, and the numbers of genes were summarized in **Table 3.7**. The determined genes were then compared to each other to identify genes that were case specific for BCa. The case specific up-regulated genes and down-regulated genes had been shown in **Figure 3.21** and **Figure 3.22**, respectively.

Table 3.7: Numbers of bladder enriched genes that up- and down-regulated in four cases of BCa. FC, fold change; SE, standard error; Case 1, pTa-LG; Case 2, pTa-HG; Case 3, pT1-HG; and Case 4, pT2-HG or invasive.

Cases (SE % ≤ 25)	Down-regulated genes (FC ≤ 0.5)	Up-regulated genes (FC ≥ 2.0)	Total genes
Control - Case 1	15	17	32
Control - Case 2	11	21	32
Control - Case 3	5	17	22
Control - Case 4	11	11	22

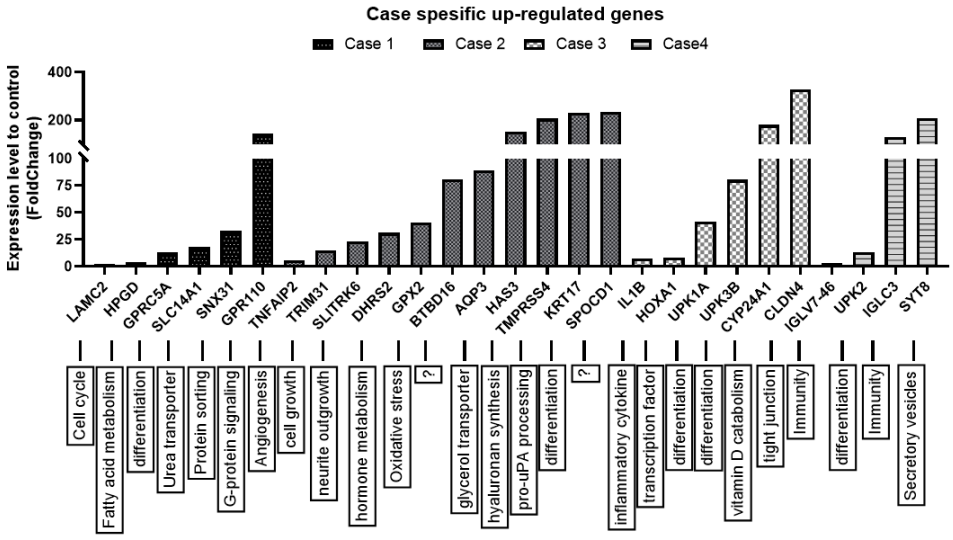

Figure 3.21: Expression levels of Case specific up-regulated genes expressed in specific cases. Case 1, pTa-LG; Case 2, pTa-HG; Case 3, pT1-HG; and Case 4, pT2-HG or invasive. Boxes contained potential biological function of relevant gene.

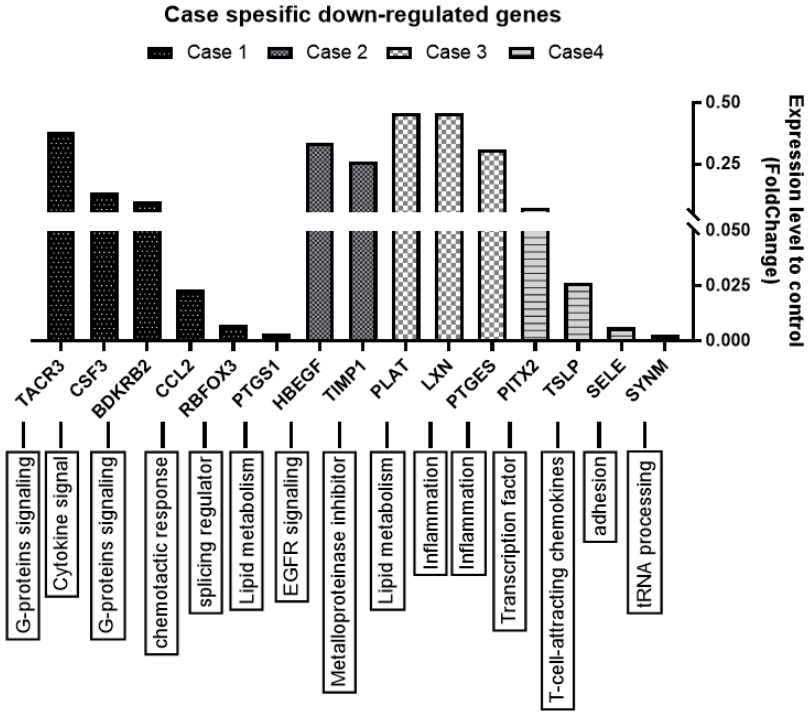

Figure 3.22: Expression levels of Case specific up-regulated genes expressed in specific cases. Case 1, pTa-LG; Case 2, pTa-HG; Case 3, pT1-HG; and Case 4, pT2-HG or invasive. Boxes contained potential biological function of relevant gene.

3.12 Determination of secreted and up-regulated genes

The number of secreted genes and their comparison among four cases were summarized by Venn diagram in **Figure 3.23**. Total of eleven genes were determined that have above 2 FC expression at least three cases when compared to control urinary bladder tissue.

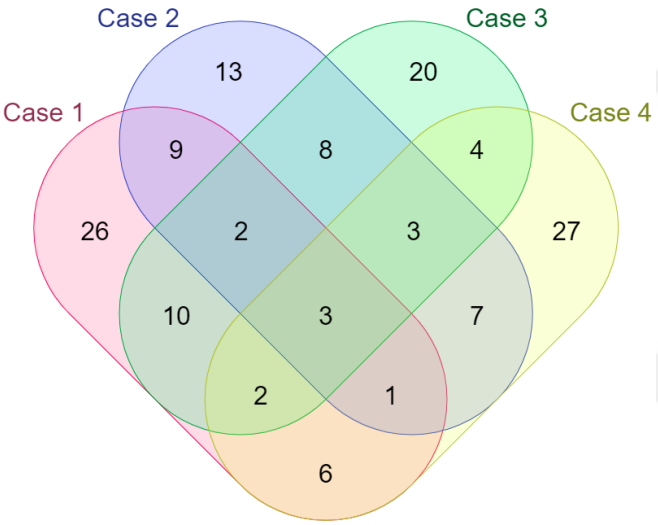

Figure 3.23: Venn diagram showing the number of secreted and up-regulated gene number. Case 1, pTa-LG; Case 2, pTa-HG; Case 3, pT1-HG; and Case 4, pT2-HG or invasive.

The gene expression values of these eleven secreted genes were also show in **Figure 3.24A**. The expression level of C4ORF48 and PRSS22 genes were also analyzed in BLCA dataset of TCGA have shown that these two genes significantly up-regulated in the BCa tumors compared to control urinary bladder tissue expression value **(Figure 3.24B)**. Potential functions of these gene were also shown in **Table 3.8**.

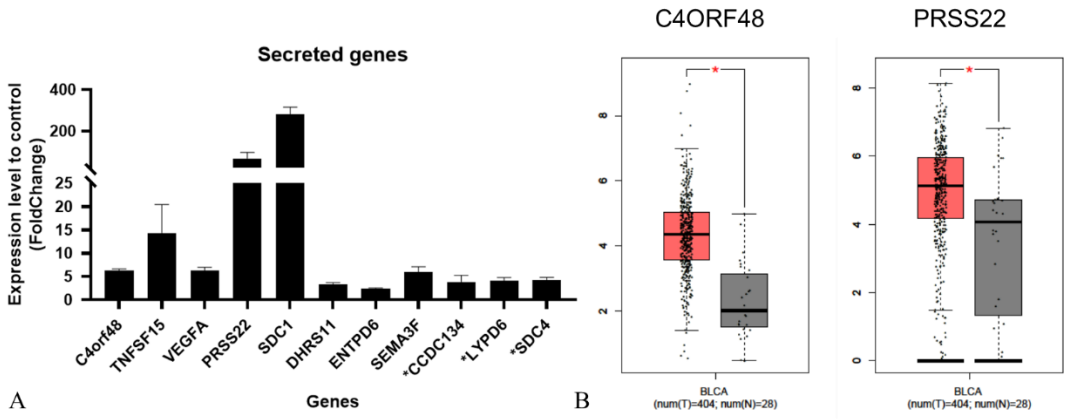

Figure 3.24: Expression level of secreted genes in BCa. A, secreted genes existed in at least 3 cases; B, TCGA gene expression analysis of two secreted genes (C4ORF48 and PRS22). BLCA, bladder urothelial carcinoma; T, tumor; N, normal; * genes expressed in four cases of BCa.

Table 3.8: The open name and potential function of secreted and up-regulated eleven genes. * genes expressed in four cases of BCa.

Gene	Open name	Function
C4ORF48	Chromosome 4 open reading frame 48	Unknown
TNFSF15	TNF superfamily member 15	Angiogenesis
VEGFA	Vascular endothelial growth factor A	Angiogenesis
PRSS22	Serine protease 22	Peptidase activity
SDC1	Syndecan 1	Cell-matrix interaction
DHRS11	Dehydrogenase/reductase 11	Oxidoreductase
ENTPD6	Ectonucleoside triphosphate diphosphohydrolase 6	Hydrolyse
SEMA3F	Semaphorin 3F	Cell motility
CCDC134*	Coiled-coil domain containing 134	Cytokine-like function
LYPD6*	LY6/PLAUR domain containing 6	nicotinic acetylcholine receptors regulator
SDC4*	Syndecan 4	Exosome biogenesis

3.13 Principle component analysis of secreted genes

Principal component analysis of these eleven secreted genes had been performed in TCGA tumor and control and GTEx control datasets, including whole blood, Bladder Urothelial Carcinoma (BLCA), Kidney renal papillary cell carcinoma (KIRP), and Lymphoid Neoplasm Diffuse Large B-cell Lymphoma (DLBC) **(Figure 3.25)**. Considering the expression levels of these eleven genes in these databases, the BLCA tumor group was separated from other tumor groups as well as BLCA normal group. In addition, expression level of these genes had been found to be differentiated from whole blood expressions.

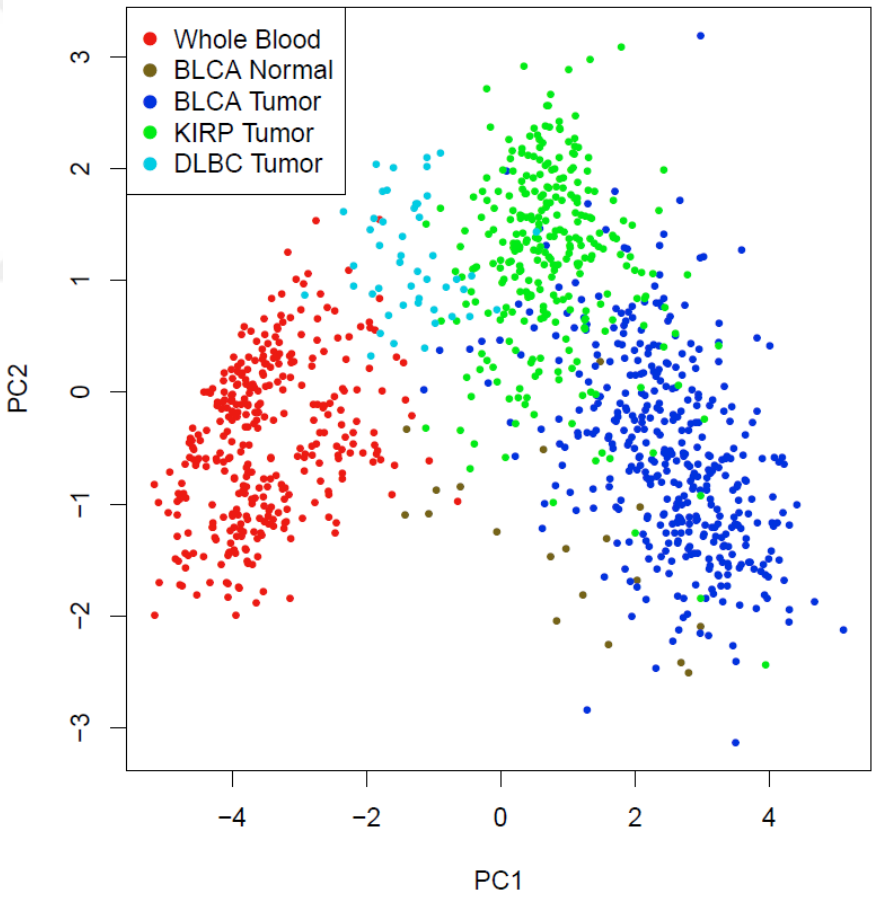

Figure 3.25: Principal component analysis plot showing the expression levels of eleven secreted genes. Analysis was performed by TCGA and GTEx datasets, including whole blood, Bladder Urothelial Carcinoma (BLCA), Kidney renal papillary cell carcinoma (KIRP), and Lymphoid Neoplasm Diffuse Large B-cell Lymphoma (DLBC).

3.14 Integration of image-features extracted from blood and urine samples with expression levels secreted genes

The gene expression levels of eleven secreted genes were co-visualized with features that extracted from whole blood and urine samples among patients. Features extracted from whole blood images integrated with the gene expression (with C4orf48 gene) were shown for tumor staging in **Figure 3.26**. Patients' features that have expression levels of these secreted eleven genes were found to be dispersed orientation among all patients. In addition, Features extracted from whole blood images integrated with the gene expression (with C4orf48 gene) were shown for tumor grading in **Figure 3.27**

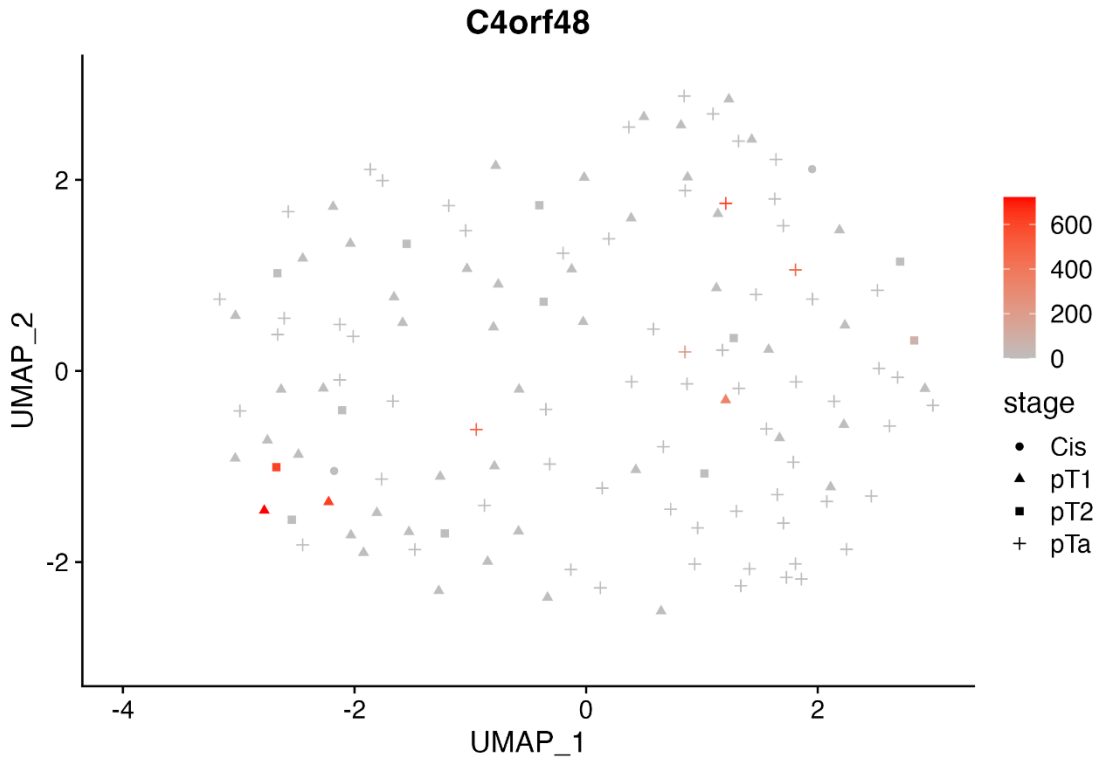

Figure 3.26: Gene expression distribution of C4orf48 gene on image features of whole blood. Patients that have gene expression in RNAseq were shown in red color according to expression level of related gene. Genes was also shown in upper side of each plot.

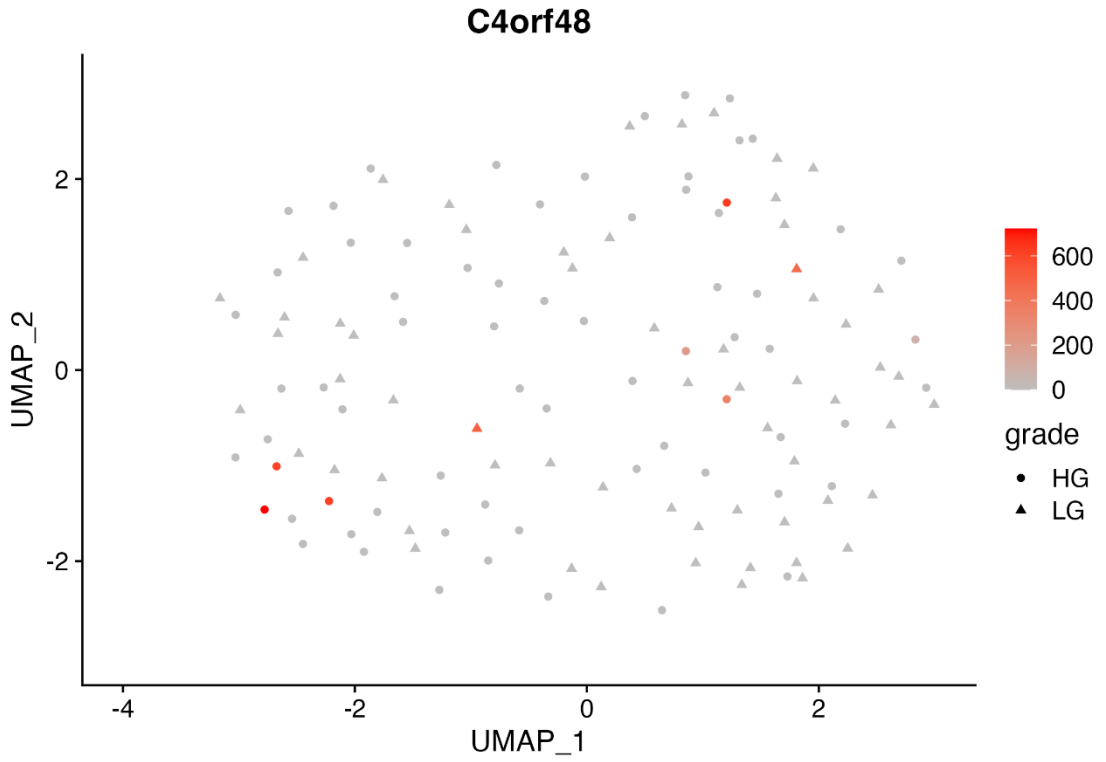

Figure 3.27: Gene expression distribution of C4orf48 gene on image features of whole blood. Patients that have gene expression in RNAseq were shown in red color according to expression level of related gene. Genes was also shown in upper side of each plot.

Patients' features were classified according to their staging or grading properties and that plots integrated with gene expression levels for eleven genes were shown in **Figure 3.28** and **Figure 3.29** respectively.

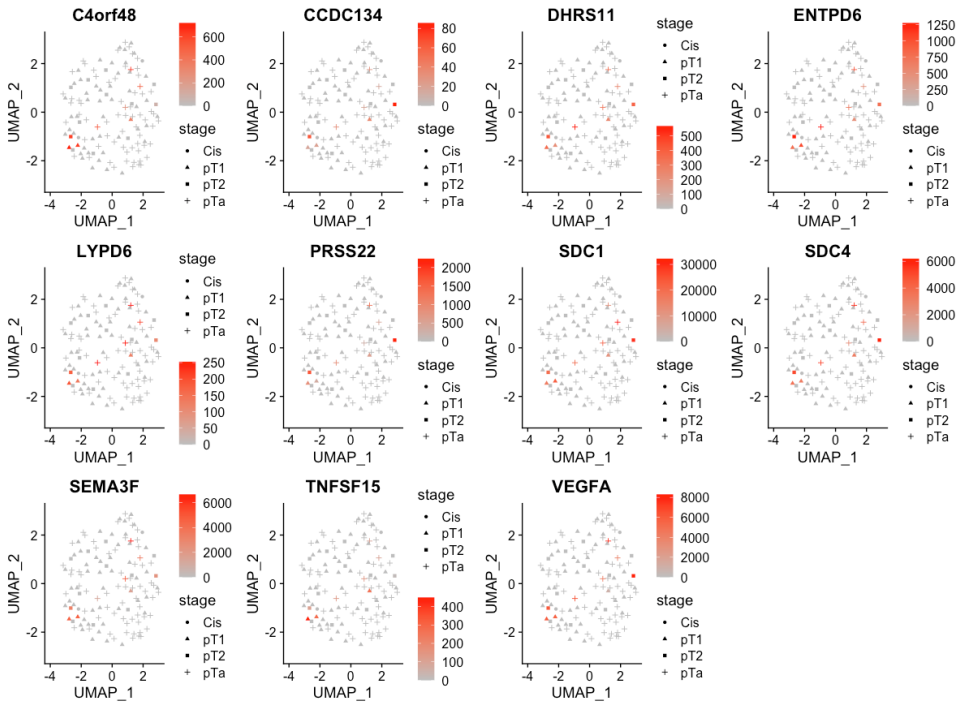

Figure 3.28: UMAP plots of features extracted from whole blood integrated with expression levels of eleven secreted genes and staging of patients.

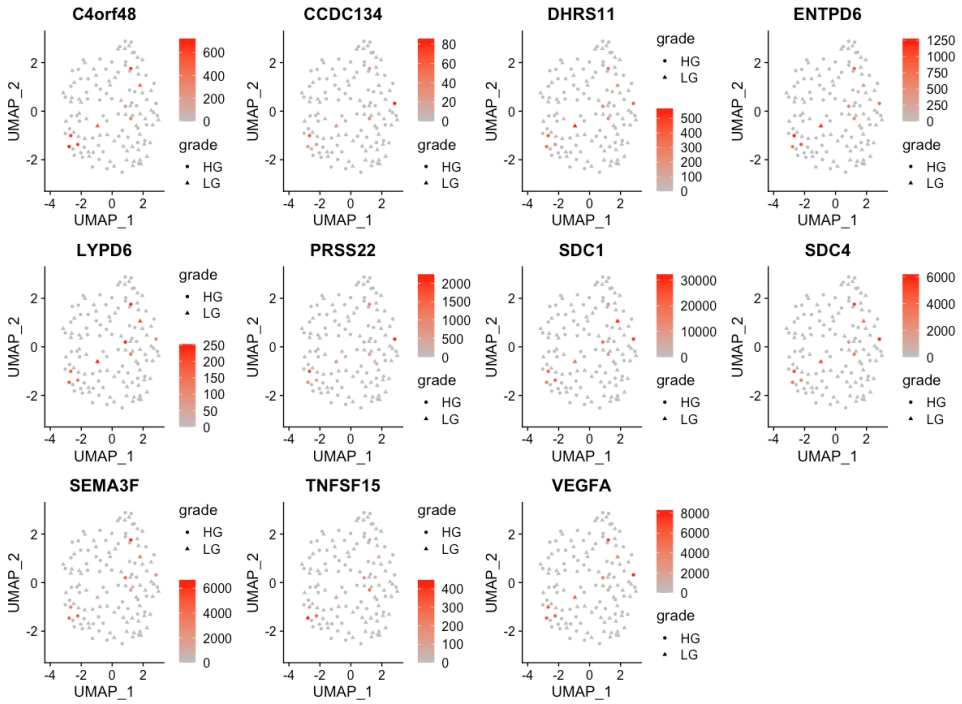

Figure 3.29: UMAP plots of features extracted from whole blood integrated with expression levels of eleven secreted genes and grading of patients.

Features extracted from urine-KCl images integrated with the gene expression (with C4orf48 gene) were shown for tumor staging (**Figure 3.30**) and grading (**Figure 3.31**).

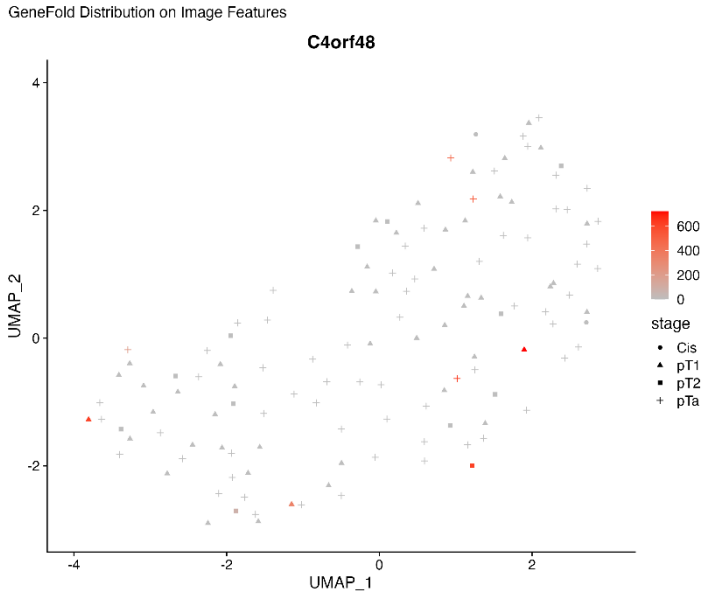

Figure 3.30: Gene expression distribution of C4orf48 gene on image features of urine-KCl. Patients that have gene expression in RNAseq were shown in red color according to expression level of related gene. Genes was also shown in upper side of each plot.

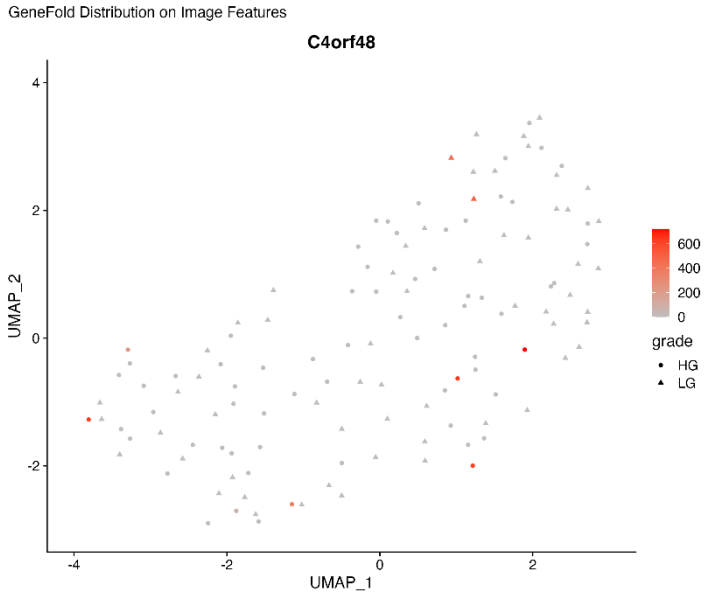

Figure 3.31: Gene expression distribution of C4orf48 gene on image features of urine-KCl. Patients that have gene expression in RNAseq were shown in red color according to expression level of related gene. Genes was also shown in upper side of each plot.

Plots integrated with gene expression levels for eleven genes were shown in **Figure 3.32** and **Figure 3.33** respectively.

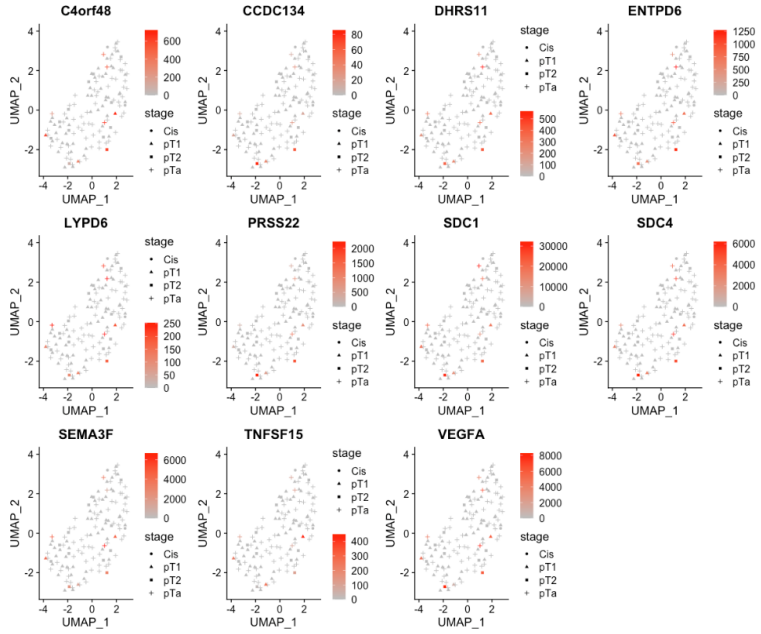

Figure 3.32: UMAP plots of features extracted from urine-KCl integrated with expression levels of eleven secreted genes and grading of patients.

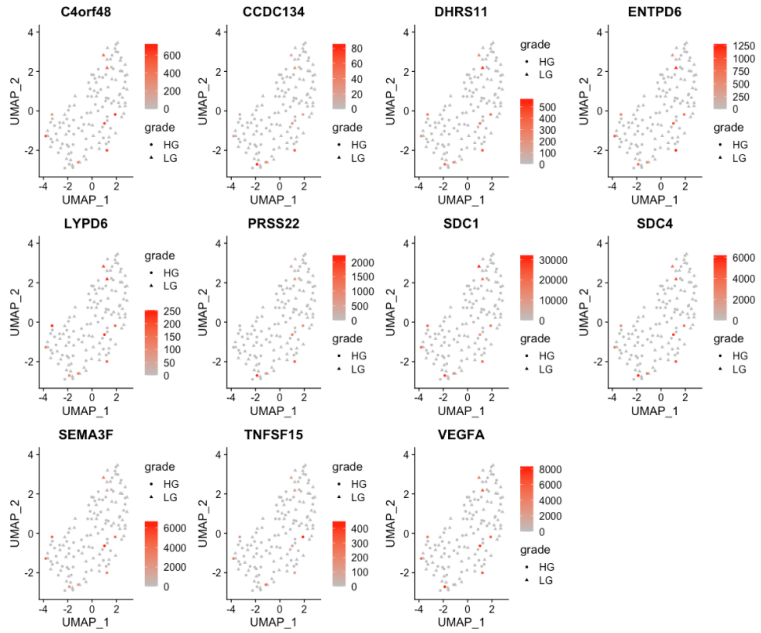

Figure 3.33: UMAP plots of features extracted from urine-KCl integrated with expression levels of eleven secreted genes and grading of patients.

Features extracted from urine-KCl+$MgCl_2$ images integrated with the gene expression (with C4orf48 gene) were shown for tumor staging (**Figure 3.34**) and grading (**Figure 35**).

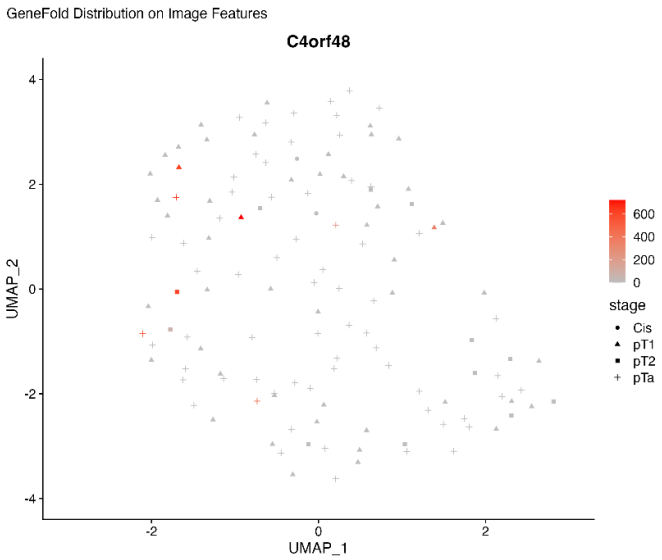

Figure 3.34: Gene expression distribution of C4orf48 gene on image features of urine-KCl+$MgCl_2$. Patients that have gene expression in RNAseq were shown in red color according to expression level of related gene. Genes was also shown in upper side of each plot.

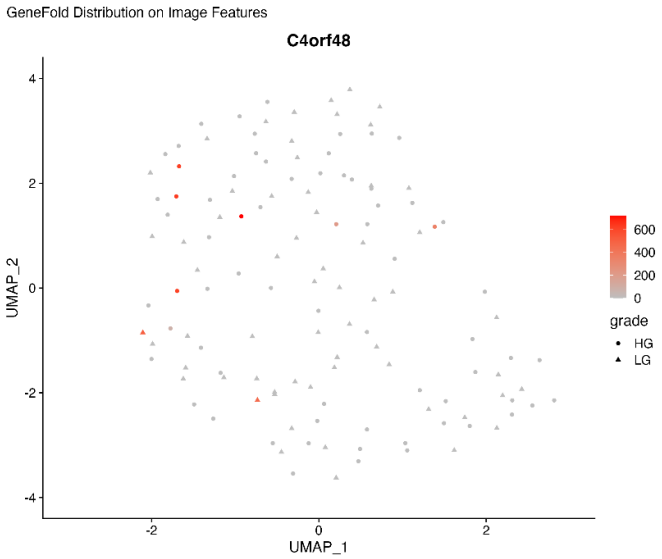

Figure 3.35: Gene expression distribution of C4orf48 gene on image features of urine-KCl+$MgCl_2$. Patients that have gene expression in RNAseq were shown in red color according to expression level of related gene. Genes was also shown in upper side of each plot.

Plots integrated with gene expression levels for eleven genes were shown in **Figure 3.36** and **Figure 3.37** respectively.

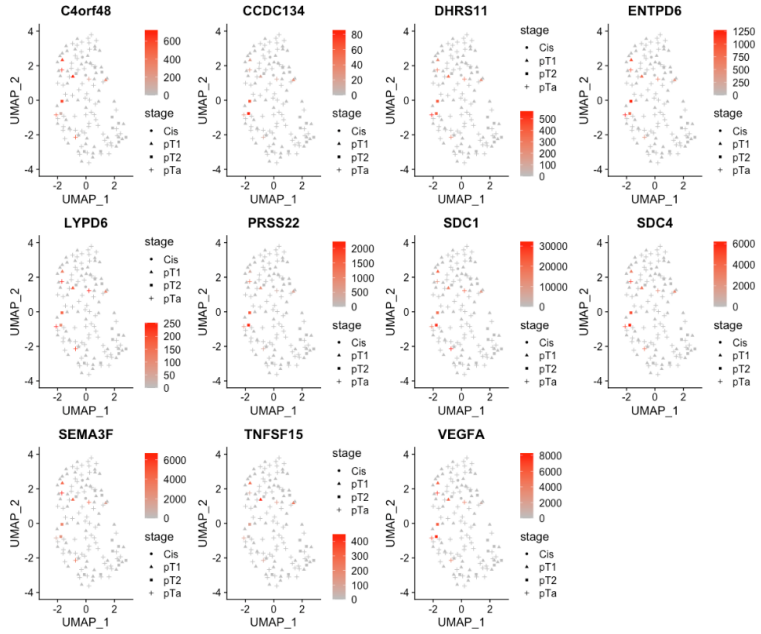

Figure 3.36: UMAP plots of features extracted from urine-KCl+$MgCl_2$ integrated with expression levels of eleven secreted genes and grading of patients.

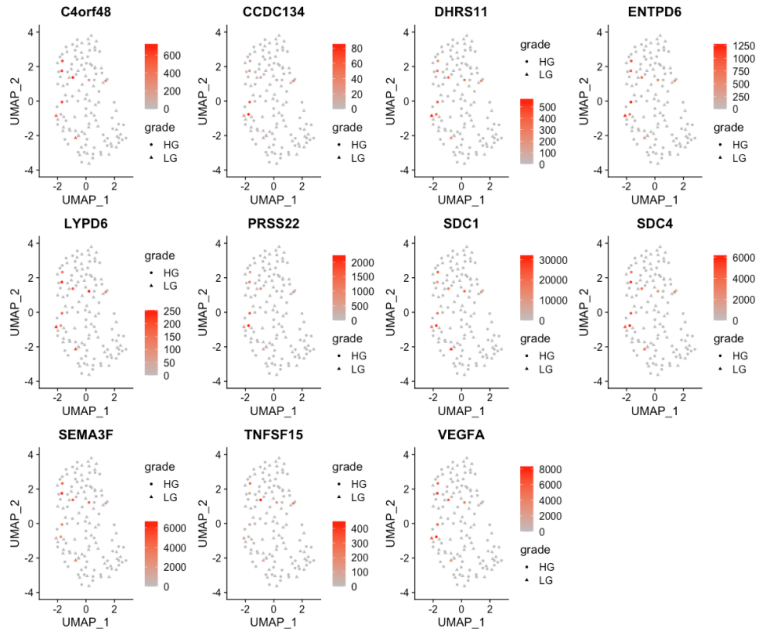

Figure 3.37: UMAP plots of features extracted from urine-KCl+$MgCl_2$ integrated with expression levels of eleven secreted genes and grading of patients.

Heatmap have been constructed to show the expression levels of eleven secreted genes among patients for tomor staging (**Figure 3.38**) and grading (**Figure 3.39**).

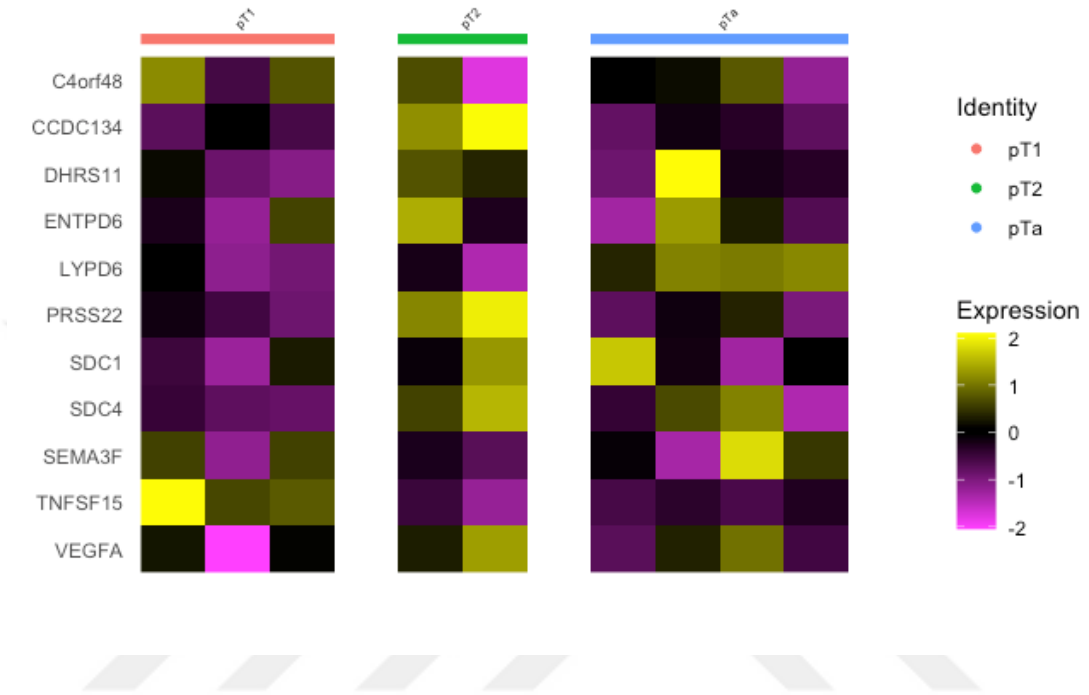

Figure 3.38: The heatmap showing expression level of eleven secreted genes according to patient staging.

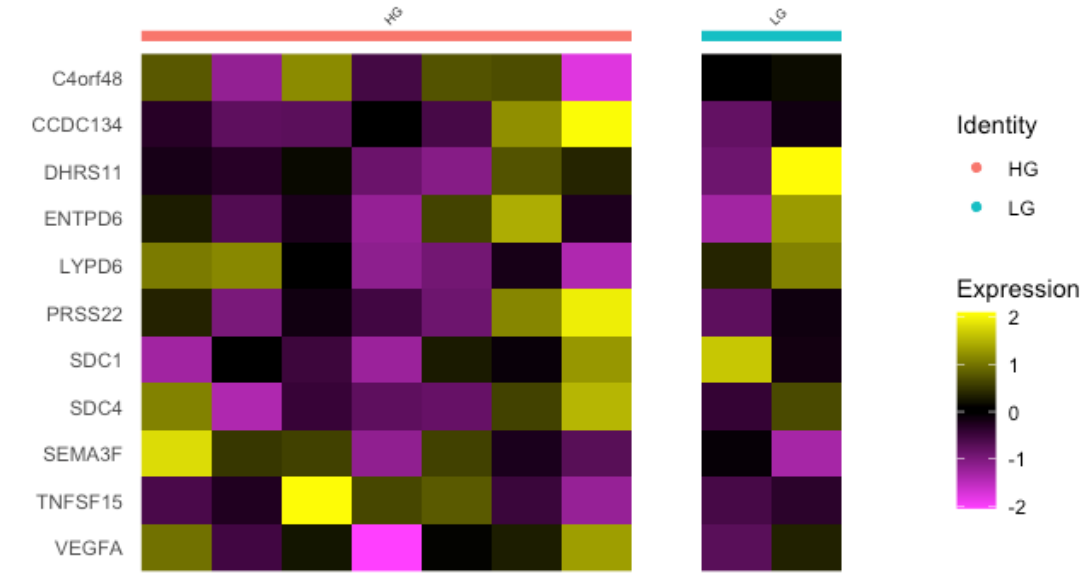

Figure 3.39: The heatmap showing expression level of eleven secreted genes according to patient grading.

3.15 Gene ontology and pathway enrichment analysis of differentially expressed gene in four cases of BCa

Gene ontology (GO) and pathway enrichment analysis was performed using all genes that have expression in RNA seq data after 25 % SE filtration process. Up-regulated and down-regulated genes were used to perform GO and pathway enrichment analysis results to show molecular function, biological function, cellular compartment, KEGG and Reactome were summarized with pie charts in **Figure 3.40**. The biological processes group contained the most detailed information among them.

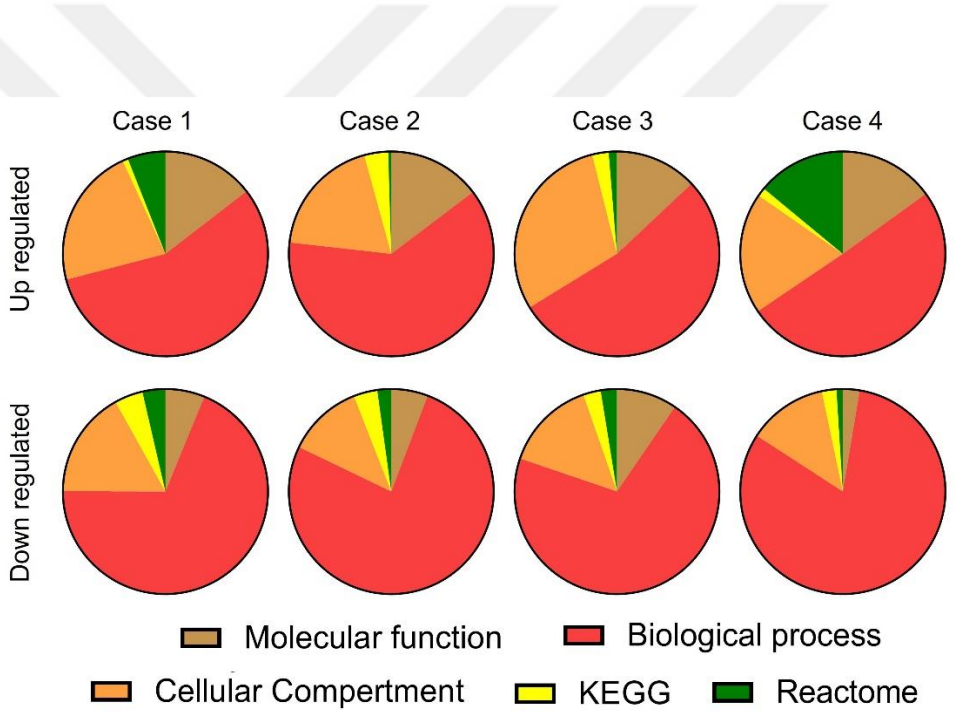

Figure 3.40: Gene ontology and pathway enrichment results of up- and down-regulated genes existed in four cases of BCa. Case 1, pTa-LG; Case 2, pTa-HG; Case 3, pT1-HG; and Case 4, pT2-HG or invasive. KEGG, Kyoto Encyclopedia of Genes and Genomes.

The enriched pathways considered biological process of these genes were summarized for case 1, case 2, case 3, and case 4 of BCa in **Table 3.9**. Up-regulated and down-regulated genes were also investigated separately in the analysis have shown that various biological pathways work in together in the same case simultaneously.

Table 3.9: Enriched pathways constructed by up- and down regulated genes expressed in four cases of BCa.

Up regulated		Down regulated
	Case 1 (pTa- LG)	
- Response to osmotic stress		- elastic fiber assembly
- Nitrogen compound transporter		- chondroitin sulfate proteoglycan metabolic process
- Organic substance transporter		- collagen biosynthetic process
- Regulation of Axonogenesis		- Leukocytes and T cell immunity
- Regulation of cell size		- Cell-cell adhesion
	Case 2 (pTa- HG)	
- Epithelial cell differentiation		- Cell adhesion mediated by integrin
- Stress activated MAPK cascade		- Negative regulation of secretory pathway
- Stem cell proliferation		- Lymphocyte regulation
- DNA replication		- Regulation of ERK cascade
	Case 3 (pT1- HG)	
- Mitochondrial gene expression		- Actin cytoskeleton organization
- RNA metabolism		- Adhesion
	Case 4 (pT2- HG)	
- Cell cycle		- Muscle regulation
- Protein *N*-linked glycosylation		- Epithelial to mesenchymal transition
- Cytoskeletal organization		- TGFβ signalling
- DNA repair		- Regulation of angiogenesis
- Intrinsic apoptotic pathway		- Cell migration

3.16 Variation analysis

Somatic variant analysis was performed for the BCa tumors and control urinary bladder tissue. Variants were detected for 203 genes that are composed of highly mutated in literature, up-regulated, gradient, and secreted gene from our RNAseq data. The genes found in somatic variant analysis had compared each other and control sample (**Figure 3.41**). Two genes (B4GALT3 and TINAGL1) were found to have same patients spesific variations in case 1 (pTa-LG) when compared to other cases and control urinary bladder tissue (**Table 3.10**). Among all cases and control tissue, three genes (CALR, ERBB2, and LAMA5) were found to show various somatic variants.

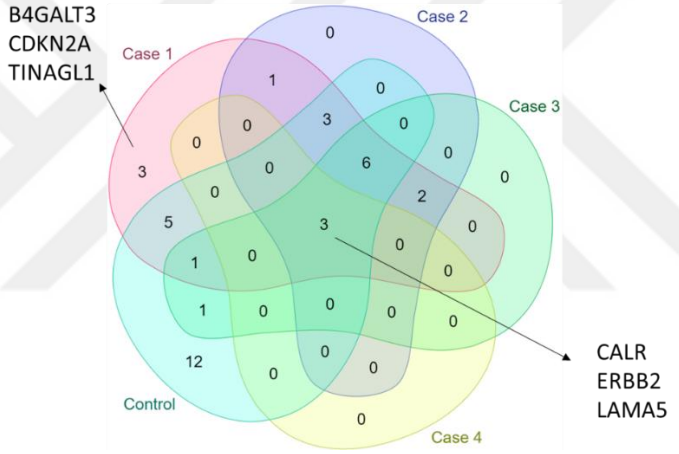

Figure 3.41: Venn diagram showing the gene number of somatic variants in BCa cases and control urinary bladder tissue. Case 1, pTa-LG; Case 2, pTa-HG; Case 3, pT1-HG; and Case 4, pT2-HG or invasive.

Table 3.10: Somatic variations of B4GALT3 and TINAGL1 genes in case 1. AD, allele depth; DP, depth of coverage.

Gene	Case/Patient	AD	DP	Start-End	Reference	Altered
B4GALT3	1/1	3	42	161143488	A	G
B4GALT3	1/2	4	193	161143488	A	G
TINAGL1	1/1	5	234	32050570	A	C
TINAGL1	1/2	4	603	32050570	A	C

3.17 In vitro analysis of two target genes expressed in BCa tumors

Molecular analysis was performed for two differentially expressed genes selected from the RNAseq data, including LGALS3 and SH3D21. One of the genes, LGALS3 was down-regulated, and SH3D21 was also up-regulated gene. The differentially expression levels of these two genes were also shown in **Table 3.11.**

Table 3.11: RNAseq expression fold changes for LGALS3 and SH3D21 genes. SE, standard error.

Gene name	Expression level to control (Fold change)			
	Case 1	Case 2	Case 3	Case 4
LGALS3	0.47 (SE, 8 %)	0.34 (SE, 38 %)	0.38 (SE, 37 %)	0.30 (SE, 29 %)
SH3D21	2.69 (SE, 20 %)	3.66 (SE, 18 %)	3.45 (SE, 25 %)	11.56 (SE, 25 %)

3.18 Establishment and assessment Gal-3 K.O of T24 bladder cell line

LGALS3 gene codded a protein, galectin 3 (Gal-3). The designed gRNAs were cloned to TLCV2 vector (**Appendix E**). Using with CRISPR system, Gal-3 protein expression was diminished completely from T24 bladder cancer cell line as shown in **Figure 3.42A**. In addition, MTT assay was performed to investigate cell viability/metabolic activity (**Figure 3.42B**), and T24 Gal-3 KO cells had no detrimental metabolic effect based on MTT assay.

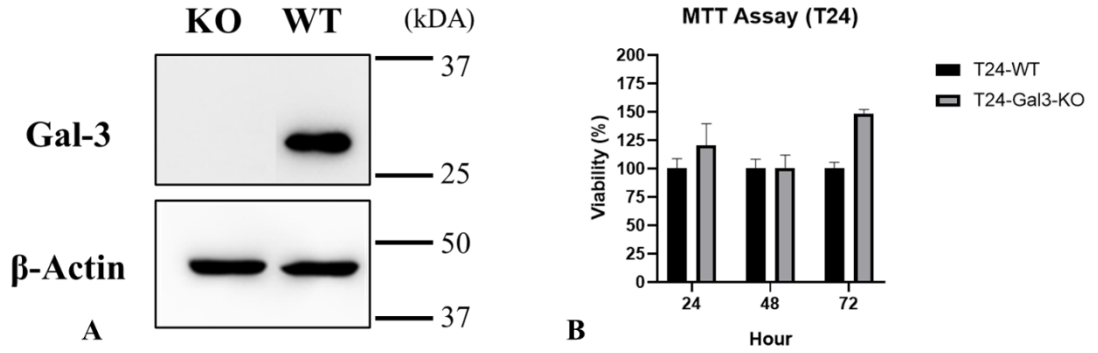

Figure 3.42: Immune staining of total protein extract and MTT assay for T24 WT and T24-Gal3 KO cell.

Gal-3 expression was also checked by immunofluorescence staining in WT and Gal-3 KO T24 cells (**Figure 3.43**). Gal-3 expression was not observed in T24 Gal-3 KO cells.

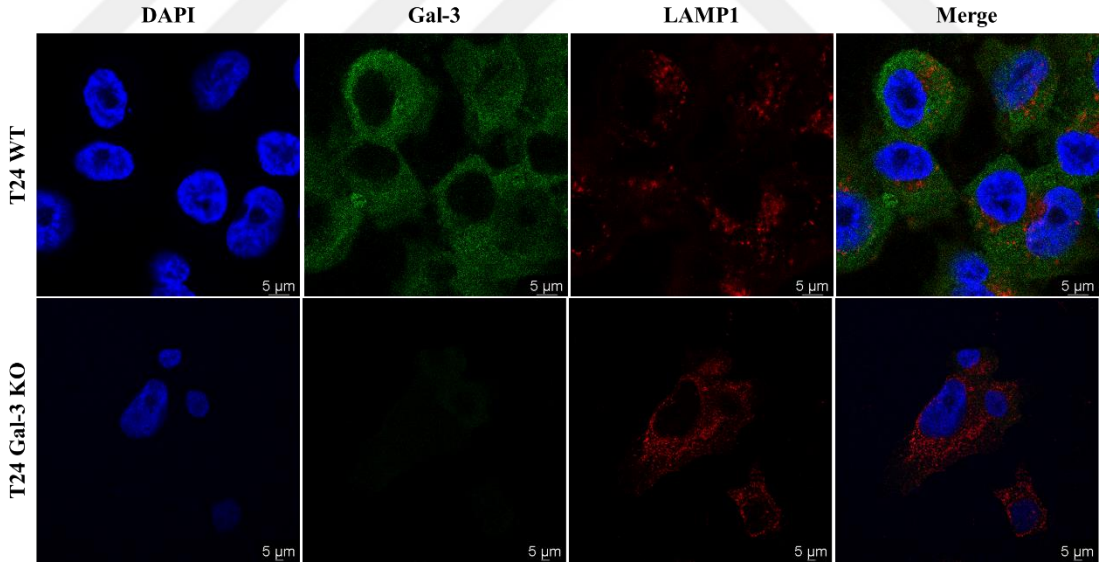

Figure 3.43: Immunofluorescence staining of Gal-3 and Lamp1 proteins in WT and Gal-3 KO T24 cells.

Scratch assay was performed to investigate the two-dimensional (2D) cell migration ability of wild type and Gal-3 KO T24 cell line (**Figure 3.44**). During the test, cell migration ability of the cells was investigated in four time points (0, 24, 48, and 72 hrs). Within the 24 hrs, T24 Gal-3 KO cells were seen to close the distance more between scratch edges compared to wild type control cell line (**Figure 3.45**).

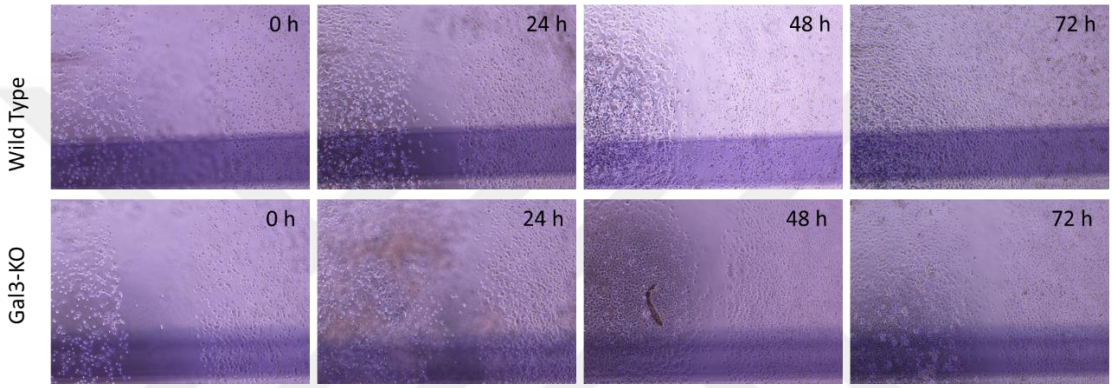

Figure 3.44: Scratch assay pictures of WT and Gal-3 KO T24 cell lines within four time points (0, 24, 48, and 72 hrs).

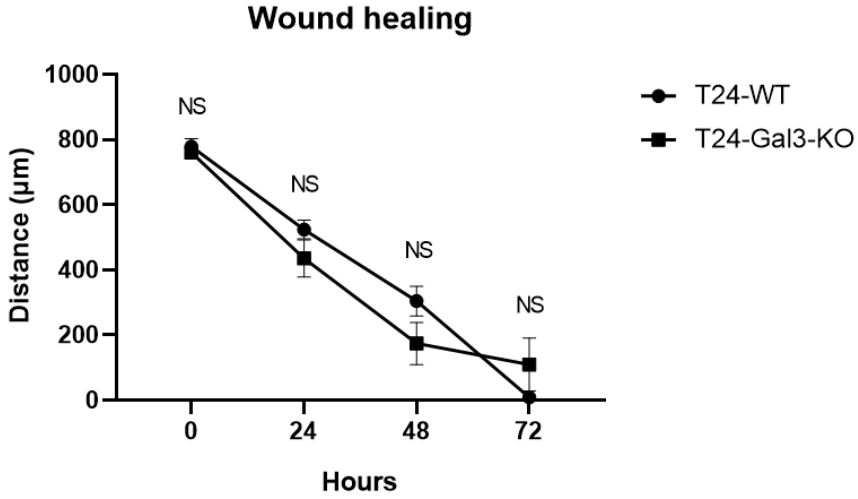

Figure 3.45: Comparison of 2D migration ability of WT and Gal-3 KO T24 cell lines within four time points according to Scratch assay pictures (0, 24, 48, and 72 hrs). Bars shows standard error mean (SEM).

Transwell migration assay was also performed to test to migratory response of T24 cells to different FBS concentrations (**Figure 3.46**). The ability of migration were found to not change between WT and Gal-3 KO conditions of T24 cells.

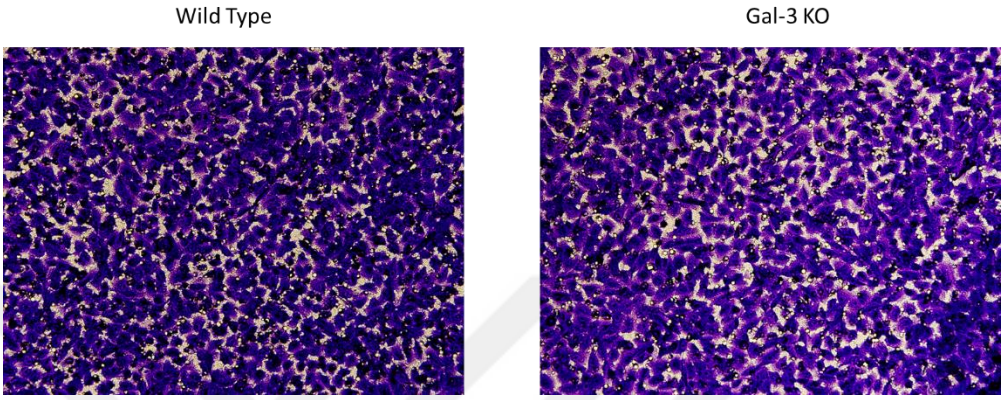

Figure 3.46: Transwell migration assay of WT and Gal-3 KO T24 cell lines.

As well as 2D cell culture experiments, WT and Gal-3 KO cells were assessed in Matrigel based 3D cell culture environment (**Figure 3.47**). In the experiment, cells were investigated at four time points (1, 4, 8, and 12 days), and diameter of intact spheroids were assessed. The diameter of the Gal-3 KO cells spheroids was found to increase in 3D cell culture compared to WT T24 cells (**Figure 3.48**).

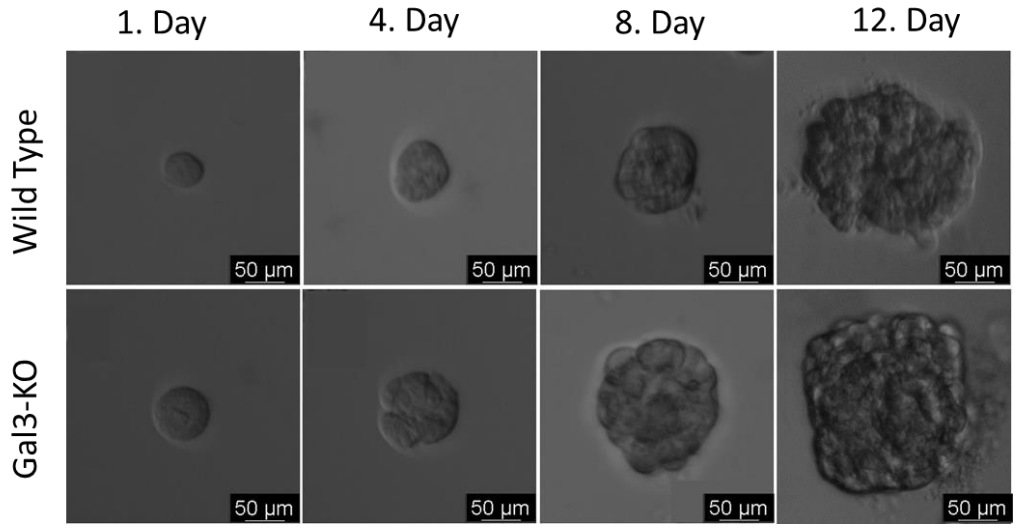

Figure 3.47: Spheroid structures of WT and Gal-3 KO T24 cells in 3D cell culture environment.

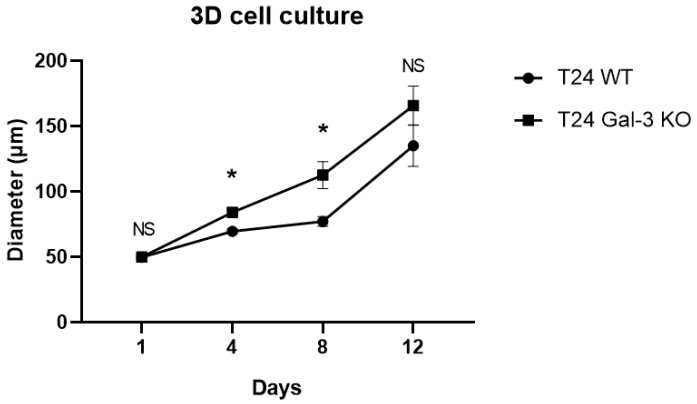

Figure 3.48: Spheroid diameters of WT and Gal-3 KO T24 cells in 3D cell culture environment. Bars shows standard error mean (SEM).

3.19 Determination of SH3D21 protein expression by qPCR

As an up-regulated gene, SH3D21 was assessed expression level qPCR analysis (**Figure 3.49**). The gene was found to have up-regulated gene in BCa tumors compared to control urinary bladder tissue. In addition, two bladder cell lines (RT4 and T24) were assessed for expression of the gene. SH3D21 was found to have more expression than tumors and controls in RT4 cells. In contrast, T24 cells expression of the gene was not found to be significant compared to control and tumor tissues.

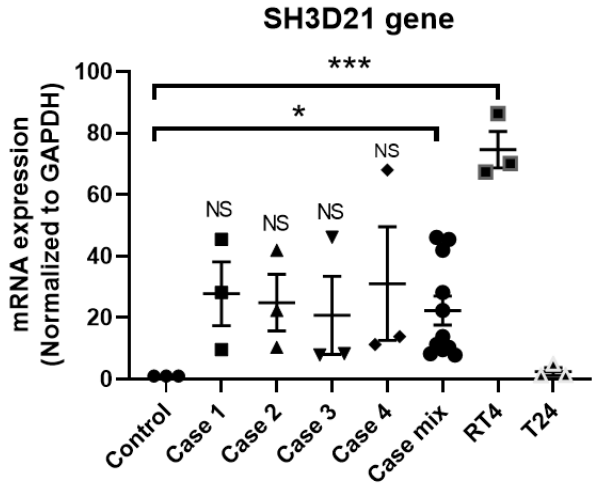

Figure 3.49: mRNA expression level of SH3D21 gene in bladder tumors and cell lines. Bars shows standard error mean (SEM).

CHAPTER 4
DISCUSSION

Bladder cancer, one of the most common urinary tract malignancies, necessitates costly and invasive diagnosis and treatment methods and strict follow-up throughout patient's lifetime. For instance, cystoscopy, a commonly used diagnostic method for BCa, is an effective but invasive approach that requires qualified professionals and facilities for accurate diagnosis of the cancer [106–109]. Indeed, false negatives and procedure-related complications are not uncommon [110–112]. On the other hand, there are no specific and reliable serum or urine markers for BCa, rendering large screens and field diagnosis difficult. Hence, practical, cost-effective and accurate diagnosis methods need to be developed.

AI-based applications are widely applied in diagnostic medicine, especially in radiology and pathology fields. Magnetic resonance imaging (MRI) and computerized tomography (CT) images, pictures of stained tissue slides or cytology analysis results, are among primary sources of data used in AI-based applications [82]. Images of serum or urine droplet patterns were not so far analyzed in the context of this cancer type. We had previously analyzed the effects of salt mixtures on droplet pattern formation of bovine serum albumin (BSA) solutions and discovered that mixtures induced formation of various complex patterns [98].

Several studies focus on the physical-chemical properties of evaporation and the consequences or effects on the formation of various droplet patterns and their reproducibility [113]. The behavior of droplet patterns is typical in pure liquids but has been found to be more complex in liquids containing multiple components [114]. Studies by adding salts or mimicking biological fluids using salt or isolated proteins (like BSA-salt solutions) tried to decipher how different patterns are forming and how specific they are [98]. In the more complex liquids (salvia or tears), evaporating liquids will turn into solid or gel, and these types of drops generally end up cracking and various pattern morphologies [115,116]. Further analysis using organic or original biological fluids such as

blood and urine in their crude forms or combination with other chemicals or solutions tried to test the formation of different patterns. The serum plasma pattern of dried blood plasma could help diagnose diseases [117]. The characteristic of blood plasma pattern is differentiated from healthy individuals to blood diseases such as Hepatitis B [118]. In addition, analysis of anemic patients' dried whole blood pattern results in a different pattern profile compared to healthy individuals [119]. The morphological features of dried blood serum drop from patients with cancer, including breast and lung cancer, have shown considerable differences in the dried pattern [120].

Convolutional neural networks (CNNs) are structurally designed to handle spatially correlated data such as images [89,121]. Since AI model also eliminates many errors derived from human handling such as visual and mental fatigue, stress and burn-out, their uses as an assistant tool may prove beneficial to increase correct diagnosis and follow-up [82]. Transfer learning is another widely used strategy to combat over fitting when the size of the data is smaller than desired [122]. Our proposed CNN architecture pretrained on ImageNet systematically apply across those blood and urine droplets, enabling comparisons to reveal shared spatial behaviours and underlying morphological features which precisely differentiate cancerous and control samples. According to ROC plot of the model, predictions of the BCa with whole blood, urine (KCl and KCl+$MgCl_2$) have shown up to 90 % AUC. This has shown that our proposed model trained with blood and urine images can be a candidate clinical assistant tool for BCa diagnosis.

Data augmentation is performed with many methods, including random sectioning, horizontal and vertical rotation, rotation around an arbitrary angle, and cutting [123]. The data augmentation is suitable for applying a transformation to an image that maps the points of the image to a different position or manipulates the image intensity values to produce augmented images, such as cystoscopy images [124]. Blood and urine images from patients have not directly reflected specific tumors. However, they are strictly related secretome and transcriptome of the tumor [125–127]. Thereby, image size of droplets (1360x1024 for blood and 4140x4096 for urine) without dividing into image patches to preserve whole pattern in images to boost the learning only with small pre-

processing process, including discarded background noise. Therefore, learning step was performed by 512 extracted features from full images sizes of the ResNet-18 model feeds into a fully connected layer with 512 neurons.

Whole blood and urine are composed of non-volatile components, including cells, macromolecules, inorganic salts, and suspended particles [128]. The percentage of these components may be changed according to various conditions, such as cancer [125–127]. Many physiological factors, metabolism, and hormones significantly affect the rheological properties of urine and whole blood (plasma and cells) [90,91]. The evaporation of blood and urine may lead to the self-organization of their components on the solid substrate to form characteristic patterns [129]. According to parameters considered in the clinic from our BCa patient and control cohorts, evaluated parameters are at normal levels in clinical thresholds. The competition between the adhesion of blood drops to glass substrate and evaporation induced shrinkage of the drops lead to the formation of cracks [91,130]. It has shown that dried drops pattern of whole blood from a sick and healthy individual is different [119]. The application of dried blood pattern for medical diagnosis were primarily carried out with plasma [131]. It has shown that dried human plasma pattern diagnoses metastatic carcinoma [117]. Various interesting patterns of dried drops of human plasma could be attributed to the disease. It is suggested that different cracking patterns could also reveal different pathological information. However, the use of whole blood pattern for medical diagnosis was rarely reported [132]. Besides whole blood samples, urine is an alternative biological liquid because it can describe any changes in the bladder. Our proposed AI-based analysis of whole blood and urine (KCl and KCl+$MgCl_2$) have shown that bladder cancer is precisely predicted in our cohort.

Our proposed whole blood and urine-based AI approach presents many advantages for BCa diagnosis in medicine. (1) Interpreting morphologies of dried human blood and urine could be an effective tool for evaluating BCa patients. The application of dried blood analysis especially performed with a serum of a patient, need a separation step for blood. However, our proposed whole blood directly can be used in the clinical diagnosis of BCa. (2) In addition, plasma and serum samples could be contaminated by hemolysis of

red blood cells (RBCs), but we included the RBCs in our analysis after hemolysis derived from freeze-thaw cycles. (3) In clinical cases, a tumor is reflected by cystoscopy investigation and pathological staining. To increase the potential ratio of diagnosis, our proposed AI-based diagnostic method can support this issue. (4) Because of the high recurrence ratio of bladder tumors, the follow-up of patients after any operation is an important issue. The gold criterion is also cystoscopy for patient follow-up. Thanks to our proposed AI-based perspective, routine blood and urine samples can quickly check patients' follow-up for BCa.

The discovery of novel biomarkers has been performed for high-performance molecular signatures of BCa by various research efforts in the last decades. A number of tests have been approved by the FDA, including NMP22, NMP22 BladderChek test, BTA TRAK, BTA Stat, Urovision, and Immunocyte etc. ([133]). These tests have used various type of markers derived from various sources, including protein biomarkers, genomic biomarkers, epigenetic biomarkers, transcriptomic, and combination of -omic biomarkers [134]. However, none of them are enough to facilitate early detection and risk stratification of BCa in routine clinical implementation [135]. Integration of these traditional assays with next generation sequencing, advancement of novel bioinformatic tools, and artificial intelligence based diagnostic techniques give rise to a promising achievement for cancer managements.

Protein based biomarkers are the most commonly used test signatures for detection of bladder cancer. Single or multiple protein targets can be used in various assays. The main assay platforms used in BCa diagnosis are immune assays, spectrometry, and ELISA ([134]). Among these protein targets, Orosomucoid 1 (ORM1) as an acute phase transport protein have been quantified with sensitivity of 92 %, and specificity of 94 % in urine [136]. Another markers HTRA1 as a serine protease is measured with a sensitivity of 93 % and specificity of 96 % [137]. In addition, many protein markers have been investigated by various platform in different test methods, including survivin (inhibitor of apoptosis), soluble Fas (activator of apoptosis), Amplified in Breas Cancer 1 (AIBC1, modulator of proliferation), Eukaryotic initiation factor 2 (EIF5A2), Nuclear matrix protein (NMP22),

Cytokeratin 8 or 18 (intermediate filament), hyaluronic acid (non-sulphated proteoglycan) [134,138]. In addition, multiple protein markers have been also used for BCa diagnosis such as, Interleukin 8 (IL-8), Matrix metallopeptidase 9 (MMP9), and vascular endothelial growth factor A (VEGFA) [139]. Many of these works underline the importance of nuclear, transport, and extracellular region related proteins for BCa diagnosis.

Several genomic biomarkers have also been investigated for the detection of BCa. Telomerase reverse transcriptase (TERT) is the most mutated gene with up to 70 % in BCa [140]. But there are some controversial results have been reported from other researchers [141]. TERT mutation is also associated with a 5-fold increase relative to recurrence ([141]). In addition, FGFR3 mutation is more common in low grade and significantly associated with shorter time to recurrence [142,143]. Other mutations (TP53, PIK3CA, and RAS) have been found limited performance. The diagnostic performance of FGFR3 and TERT with PI3K3CA, TP53, and RAS improved bladder cancer detection but only achieve a sensitivity of 73 % [144].

Heterogeneity is an important context in cancer diagnosis and follow up. It can be divided into two main categories, intratumoral (between tumor cells) and intertumoral (between patients) heterogeneity [145]. The most significant signs of intertumoral heterogeneity can be carried by biological liquids (blood and urine) between different patients. These complexities of the biological samples consist of various biological markers that released from cancer cells as well as patient based additional contents. Hence, tumor specific secreted molecules present in these biological fluids may contribute complexity for their contents. The advancement sequencing technology (such as RNAseq) also helps to understand intratumoral heterogeneity ([146,147]). Based on the high throughput transcription level differences of genes may describe the pathological or histological differences between tumors. As we can see pathway enrichment analysis, various different pathways can be activated or deactivated in different pathological conditions in BCa.

The use of AI-based technology in BCa diagnosis and treatments has potential to improve patients' quality of life by preventing unnecessary complication. In addition, integration of the AI-based diagnosis from blood and urine samples with tumor-secreted markers may be a significant way to diagnose BCa. One important point is the selection of these secreted markers from tumors. Another important point is that expression level of these markers planning to use as BCa marker for AI-based integrated diagnostic system should have differentiated expression level among diseases and normal blood. Thereby, integration of AI-based diagnosis with RNAseq expression values should be an alternative preoperative method to explain the diagnosis and follow up as well as heterogenicity of BCa.

Glycosylation changes and their recognition by glycan-specific lectins is valuable issue in cancer [148]. Translating basic research of glycan-lectin interactions into clinical management is a significant way of promising new perspectives in cancer research and personalized medicine. Considering the recent studies about glycosylation and lectins in BCa, the emerging data have shown the potential roles of galectins in improving new strategies for understanding BCa management [149]. Galectins have ability to support cancer cell survival by intracellular and extracellular mechanisms suggests this molecule is an important component of the tumor microenvironment that potentially could be targeted for therapy [150] .

Galectin-3 (gal-3) mediates cell-cell and cell-matrix interactions when expressed in the plasma membrane and extracellular matrix. Also, its nuclear expression regulates gene transcription and pre-mRNA processing [151]. Different expression patterns of the gal-3 have been suggested in many cancers, including ovarian, and prostate cancer; its role and function in BCa formation and maintenance have not been comprehensively investigated. Many types of cells and tissues express the protein, especially prominent in macrophages and epithelial cells, with various levels in development, tissue regeneration, and cancer [152]. The broad expression pattern of galectins contributes to tumorigenesis, proliferation, angiogenesis, metastasis, and immunosuppression in cancer [153].

Main pathways frequently dysregulated in bladder cancer are cell cycle regulation, PI3K signalling, and chromatin remodelling [58]. Potential roles of gal-3 in these biological pathways may highlight its importance for BCa behaviour [154]. In 3D cell culture environment, gal-3 KO T24 bladder cells creating more higher diameter spheroids may show the importance of gal-3 in bladder tumor behaviour compared to WT T24 cells.

Src homology 3 (SH3) domain containing proteins are involved in the regulation of various significant cellular process, including cell proliferation, migration, and cytoskeletal modification [155] . In addition, high affinity peptides that binds SH3 domain are used in drug development. SH3 domain of proteins carry important roles in substrate recognition, membrane localization and regulation of kinase activity [156]. SH3D21 is expressed in nucleus and plasma membrane of various human tissues [102]. According to gene coding variant analysis, two variants of SH3D21 have been found to be associated with resting metabolic rate [157]. In addition, a genome scale CRISPR/Cas9 screening have shown the SH3D21that is a sensitizer as a therapeutic target in pancreatic cancer cell line Panc1 [158]. Considering the RNAseq and qPCR results, SH3D21 is also highly expressed in BCa tumor samples as well as RT4 bladder cell line. However, the high expression has not been seen in another BCa cell line, T24. This may be explained the genetic background of these two cell lines, RT4 (CDKN2A and TSC1 mutations) and T24 (HRAS and TP53 mutations).

CHAPTER 5
CONCLUSION

Urothelial bladder cancer is one of the most common types of cancer among genitourinary tumors worldwide. Cancer's high recurrence and progression ratio deeply thread patients' life expectancies and qualities. Cystoscopy is the critical standard for BCa diagnosis and follow-up. However, it is an invasive approach that requires qualified professionals and facilities for accurate cancer diagnosis. Indeed, FDA approved various test kits have been known, but the usability of these kits has seen no practical in routine clinical tests. In addition, diagnostic accuracy of single biomarkers various considerably, and performance evaluation of these tests relies on the prevalence of underlying diseases of the patients.

Over the last decades, technological advances have facilitated the transformation of biological information into the clinical management for various diseases such as cancer. Artificial intelligence and RNA sequencing technologies are two significant developments in medical diagnose and follow-up of patients because of their usability and capability of explanation for an enormous amount of biological information. Despite various proteomics, genomics, and epigenomics markers that have been known for BCa, there is no exact biological marker with high sensitivity and specificity for diagnosis and follow-up of BCa. Hence, practical, cost-effective and accurate diagnosis methods need to be developed.

Blood and urine are the two main component of the body, and their content give rise to various amount of information about patient healthcare. Bladder tumors are strictly connected urine and blood and cause diagnostic opportunity for BCa. Hence, previously many tests based on urine markers had been try to develop to diagnose BCa. However, these are not reliable in clinical practice. In addition, tumor heterogenicity is other difficulty to diagnosis and follow-up of patients. Thereby, high sensitivity and reliable markers need to develop for BCa.

In the thesis, by analyzing droplet patterns of blood and urine samples from patients, we developed a deep learning- and artificial intelligence-assisted quick, cheap, and reliable diagnosis method. Droplet pattern analysis of evaporated blood or urine deposits was performed using patient and normal control samples. Our proposed AI-assistant solution (ResNet-18 pre-trained ImageNet) can be systematically applied across droplets, enabling comparisons to reveal shared spatial behaviors and underlying morphological patterns, which precisely differentiate patient-derived samples from controls with high accuracy. The AI-based approach that tried to integrate with RNAseq gene expression data for patients may be an alternative diagnostic potential between BCa tumor.

The four cases composed of different stage and grade tumors were also investigated for RNAseq expression profiles. Potential targets that may be used in bladder cancer diagnosis, recurrence, and follow up have been determined. In addition, candidate genes had been determined for BCa in different gene classes, including, gradient and secreted genes.

Two potential target genes have also been investigated for bladder cancer, LGALS3 and SH3D21 in molecular perspectives. LGALS3 (coded Gal-3 protein) was found to be a down regulated gene in RNAseq analysis. Indeed, Gal-KO bladder cell line T24 has shown the increased cell diameter when compared to Gal-WT T24 cells in cell culture system. In contrast, SH3D21 gene was found to have increased expression level in RNA sequencing analysis. Despite no information about the gene function, we also found that SH3D21 gene also over-expressed in bladder tumors samples by qPCR analysis.

BIBLIOGRAPHY

1. Betts, J. G. *et al. Anatomy & Physiology SENIOR CONTRIBUTING AUTHORS*.
2. Zhou, H., Guo, C. C. & Ro, J. Y. *Urinary Bladder Pathology*.
3. Bladder Cancer — Cancer Stat Facts. https://seer.cancer.gov/statfacts/html/urinb.html.
4. Jost, S. P., Goslingt, J. A. & Dixon, J. S. The morphology of normal human bladder urothelium. *J. Anat* **167**, 103–115 (1989).
5. Dixon, J. S. & Gosling, J. A. Histology and fine structure of the muscularis mucosae of the human urinary bladder. *J Anat* **136**, 265 (1983).
6. Ro, J. Y., Ayala, A. G. & El-Naggar, A. Muscularis mucosa of urinary bladder. Importance for staging and treatment. *American Journal of Surgical Pathology* **11**, 668–673 (1987).
7. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **71**, 209–249 (2021).
8. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2018. *CA Cancer J Clin* **68**, 7–30 (2018).
9. Antoni, S. *et al.* Bladder Cancer Incidence and Mortality: A Global Overview and Recent Trends. *Eur Urol* **71**, 96–108 (2017).
10. Rodríguez Faba, O. & Palou, J. Predictive factors for recurrence progression and cancer specific survival in high-risk bladder cancer. *Curr Opin Urol* **22**, 415–420 (2012).
11. Silverman, D. T., Koutros, S., Figueroa, J. D., Prokunina-Olsson, L. & Rothman, N. Bladder cancer. *Schottenfeld and Fraumeni Cancer Epidemiology and Prevention, Fourth Edition* 977–996 (2017) doi:10.1093/OSO/9780190238667.003.0052.
12. Teoh, J. Y. C. *et al.* Global Trends of Bladder Cancer Incidence and Mortality, and Their Associations with Tobacco Use and Gross Domestic Product Per Capita. *Eur Urol* **78**, 893–906 (2020).
13. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **71**, 209–249 (2021).
14. Tran, L., Xiao, J. F., Agarwal, N., Duex, J. E. & Theodorescu, D. Advances in bladder cancer biology and therapy. *Nature Reviews Cancer 2020 21:2* **21**, 104–121 (2020).
15. Berdik, C. & Ashour, M. Unlocking bladder cancer. *Nature 2017 551:7679* **551**, S34–S35 (2017).
16. Matsumoto, K. *et al.* Late Recurrence and Progression in Non-muscle–invasive Bladder Cancers After 5-year Tumor-free Periods. *Urology* **75**, 1385–1390 (2010).
17. Anastasiadis, A. & de Reijke, T. M. Best practice in the treatment of nonmuscle invasive bladder cancer. *http://dx.doi.org/10.1177/1756287211431976* **4**, 13–32 (2011).
18. Tabayoyong, W. B. *et al.* Systematic Review on the Utilization of Maintenance Intravesical Chemotherapy in the Management of Non-muscle-invasive Bladder Cancer. *Eur Urol Focus* **4**, 512–521 (2018).
19. Edge, S. B. & Compton, C. C. The american joint committee on cancer: The 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol* **17**, 1471–1474 (2010).

20. Ploeg, M. & Witjes, J. A. Bladder Cancer Diagnosis and Detection: Current Status. *Bladder Tumors:* 63–77 (2011) doi:10.1007/978-1-60761-928-4_4.

21. Montironi, R., Cheng, L., Scarpelli, M. & Lopez-Beltran, A. Pathology and Genetics: Tumours of the Urinary System and Male Genital System: Clinical Implications of the 4th Edition of the WHO Classification and Beyond. *Eur Urol* **70**, 120–123 (2016).

22. Humphrey, P. A., Moch, H., Cubilla, A. L., Ulbright, T. M. & Reuter, V. E. The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs—Part B: Prostate and Bladder Tumours. *Eur Urol* **70**, 106–119 (2016).

23. Kamat, A. M. *et al.* Bladder cancer. *The Lancet* **388**, 2796–2810 (2016).

24. Lenis, A. T., Lec, P. M. & Chamie, K. Bladder Cancer: A Review. *JAMA* **324**, 1980–1991 (2020).

25. Wallmeroth, A. *et al.* Patterns of Metastasis in Muscle-Invasive Bladder Cancer (pT2–4): An Autopsy Study on 367 Patients. *Urol Int* **62**, 69–75 (1999).

26. DeMarini, D. M. Genotoxicity of tobacco smoke and tobacco smoke condensate: a review. *Mutation Research/Reviews in Mutation Research* **567**, 447–474 (2004).

27. Castaño-Vinyals, G. *et al.* Bulky DNA Adduct Formation and Risk of Bladder Cancer. *Cancer Epidemiology, Biomarkers & Prevention* **16**, 2155–2159 (2007).

28. Veglia, F. *et al.* DNA adducts and cancer risk in prospective studies: a pooled analysis and a meta-analysis. *Carcinogenesis* **29**, 932–936 (2008).

29. Fadl-Elmula, I. Chromosomal changes in uroepithelial carcinomas. *Cell Chromosome* **4**, (2005).

30. Knowles, M. A. Molecular subtypes of bladder cancer: Jekyll and Hyde or chalk and cheese? *Carcinogenesis* **27**, 361–373 (2006).

31. Knowles, M. A. & Knowles, M. A. Molecular pathogenesis of bladder cancer. *International Journal of Clinical Oncology 2008 13:4* **13**, 287–297 (2008).

32. Hashmi, A. A. *et al.* Prognostic significance of p16 immunohistochemical expression in urothelial carcinoma. *Surgical and Experimental Pathology 2019 2:1* **2**, 1–8 (2019).

33. Veerakumarasivam, A. *et al.* High-Resolution Array-Based Comparative Genomic Hybridization of Bladder Cancers Identifies Mouse Double Minute 4 (MDM4) as an Amplification Target Exclusive of MDM2 and TP53. *Clinical Cancer Research* **14**, 2527–2534 (2008).

34. Zaharieva, B. *et al.* High-throughput tissue microarray analysis of CMYC amplificationin urinary bladder cancer. *Int J Cancer* **117**, 952–956 (2005).

35. Watters, A. D. *et al.* Genetic aberrations of c-myc and CCND1 in the development of invasive bladder cancer. *British Journal of Cancer 2002 87:6* **87**, 654–658 (2002).

36. Panani, A. D., Babanaraki, A., Malianga, E. & Roussos, C. Numerical Aberrations of Chromosomes 9 and 11 Detected by FISH in Greek Bladder Cancer Patients. *Anticancer Res* **24**, 3857–3862 (2004).

37. Lamy, A. *et al.* Molecular Profiling of Bladder Tumors Based on the Detection of FGFR3 and TP53 Mutations. *J Urol* **176**, 2686–2689 (2006).

38. Tomlinson, D. C., Baldo, O., Hamden, P. & Knowles, M. A. FGFR3 protein expression and its relationship to mutation status and prognostic variables in bladder cancer. *J Pathol* **213**, 91–98 (2007).

39. Lee, C. S. D., Yoon, C. Y. & Witjes, J. A. The past, present and future of cystoscopy: the fusion of cystoscopy and novel imaging technology. *BJU Int* **102**, 1228–1233 (2008).

40. Brown, F. M. URINE CYTOLOGY: Is it Still the Gold Standard for Screening? *Urologic Clinics of North America* **27**, 25–37 (2000).

41. Brausi, M. *et al.* Variability in the Recurrence Rate at First Follow-up Cystoscopy after TUR in Stage Ta T1 Transitional Cell Carcinoma of the Bladder: A Combined Analysis of Seven EORTC Studies. *Eur Urol* **41**, 523–531 (2002).

42. Vriesema, J. L. J., Poucki, M. H., Kiemeney, L. A. L. M. & Witjes, J. A. Patient opinion of urinary tests versus flexible urethrocystoscopy in follow-up examination for superficial bladder cancer: a utility analysis. *Urology* **56**, 793–797 (2000).

43. Sievert, K. D. *et al.* Economic aspects of bladder cancer: What are the benefits and costs? *World J Urol* **27**, 295–300 (2009).

44. Filbeck, T. *et al.* 5-aminolevulinic acid-induced fluorescence endoscopy applied at secondary transurethral resection after conventional resection of primary superficial bladder tumors. *Urology* **53**, 77–81 (1999).

45. Grimbergen, M. C. M., van Swol, C. F. P., Jonges, T. G. N., Boon, T. A. & van Moorselaar, R. J. A. Reduced Specificity of 5-ALA Induced Fluorescence in Photodynamic Diagnosis of Transitional Cell Carcinoma after Previous Intravesical Therapy. *Eur Urol* **44**, 51–56 (2003).

46. Hirata, M. *et al.* Evaluation of microvessels in colorectal tumors by narrow band imaging magnification. *Gastrointest Endosc* **66**, 945–952 (2007).

47. Tearney, G. J. *et al.* Optical Biopsy in Human Urologic Tissue Using Optical Coherence Tomography. *J Urol* **157**, 1915–1919 (1997).

48. Santos, L. *et al.* Ki-67 index enhances the prognostic accuracy of the urothelial superficial bladder carcinoma risk group classification. *Int J Cancer* **105**, 267–272 (2003).

49. Rhijn, B. V. van *et al.* The fibroblast growth factor receptor 3 (FGFR3) mutation is a strong indicator of superficial bladder cancer with low recurrence rate. *undefined* (2001).

50. Christoph, F. *et al.* Methylation of tumour suppressor genes APAF-1 and DAPK-1 and in vitro effects of demethylating agents in bladder and kidney cancer. *Br J Cancer* **95**, 1701 (2006).

51. The Association of Death-associated Protein Kinase Hypermethylation with Early Recurrence in Superficial Bladder Cancers1 | Cancer Research | American Association for Cancer Research. https://aacrjournals.org/cancerres/article/62/14/4048/509012/The-Association-of-Death-associated-Protein-Kinase.

52. Friedrich, M. G. *et al.* Prognostic relevance of methylation markers in patients with non-muscle invasive bladder carcinoma. *Eur J Cancer* **41**, 2769–2778 (2005).

53. Mhawech, P., Herrmann, F., Coassin, M., Guillou, L. & Iselin, C. E. Motility-related protein 1 (MRP-1/CD9) expression in urothelial bladder carcinoma and its relation to tumor recurrence and progression. *Cancer* **98**, 1649–1657 (2003).

54. Sitnikova, L. *et al.* IMP3 Predicts Aggressive Superficial Urothelial Carcinoma of the Bladder. *Clinical Cancer Research* **14**, 1701–1706 (2008).

55. Rodríguez-Alonso, A., Pita-Fernández, S., González-Carreró, J., Luis Nogueira-March, J. & Schmitz-Dräger, B. J. p53 and ki67 expression as prognostic factors for cancer-related survival in stage T1 transitional cell bladder carcinoma. *Eur Urol* **41**, 182–189 (2002).

56. Goddard, J. C., Sutton, C. D., Jones, J. L., O'Byrne, K. J. & Kockelbergh, R. C. Reduced Thrombospondin-1 at presentation predicts disease progression in superficial bladder cancer. *Eur Urol* **42**, 464–468 (2002).

57. Habuchi, T. Prognostic Markers for Bladder Cancer. *Bladder Tumors:* 139–163 (2011) doi:10.1007/978-1-60761-928-4_8.

58. Weinstein, J. N. *et al.* Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature 2014 507:7492* **507**, 315–322 (2014).

59. Rieger-Christ, K. M. *et al.* Identification of fibroblast growth factor receptor 3 mutations in urine sediment DNA samples complements cytology in bladder tumor detection. *Cancer* **98**, 737–744 (2003).

60. van Rhijn, B. W. G. *et al.* FGFR3 and P53 Characterize Alternative Genetic Pathways in the Pathogenesis of Urothelial Cell Carcinoma. *Cancer Res* **64**, 1911–1914 (2004).

61. Mitra, A. P., Datar, R. H. & Cote, R. J. Molecular pathways in invasive bladder cancer: New insights into mechanisms, progression, and target identification. *Journal of Clinical Oncology* **24**, 5552–5564 (2006).

62. FGFR3 and TP53 Gene Mutations Define Two Distinct Pathways in Urothelial Cell Carcinoma of the Bladder | Cancer Research | American Association for Cancer Research. https://aacrjournals.org/cancerres/article/63/23/8108/510791/FGFR3-and-TP53-Gene-Mutations-Define-Two-Distinct.

63. Mitra, A. P., Datar, R. H. & Cote, R. J. Molecular pathways in invasive bladder cancer: New insights into mechanisms, progression, and target identification. *Journal of Clinical Oncology* **24**, 5552–5564 (2006).

64. El-Deiry, W. S. *et al.* WAR, a Potential Mediator of ~53 Tumor Suppression. *Cell* **75**, 817–825 (1993).

65. Wu, X., Bayle, J. H., Olson, D. & Levine, A. J. The p53-mdm-2 autoregulatory feedback loop. *Genes Dev* **7**, 1126–1132 (1993).

66. Kawamoto, K. *et al.* p16INK4a and p14ARF methylation as a potential biomarker for human bladder cancer. *Biochem Biophys Res Commun* **339**, 790–796 (2006).

67. Chellappan, S. P., Hiebert, S., Mudryj, M., Horowitz, J. M. & Nevins, J. R. The E2F transcription factor is a cellular target for the RB protein. *Cell* **65**, 1053–1061 (1991).

68. Mirra, A. P., Lin, H., Datar, R. H. & Cote, R. J. Molecular Biology of Bladder Cancer: Prognostic and Clinical Implications. *Clin Genitourin Cancer* **5**, 67–77 (2006).

69. Chatterjee, S. J. *et al.* Hyperphosphorylation of pRb: a mechanism for RB tumour suppressor pathway inactivation in bladder cancer. *J Pathol* **203**, 762–770 (2004).

70. Jones, A. & Crew, J. Vascular endothelial growth factor and its correlation with superficial bladder cancer recurrence rates and stage progression. *Urologic Clinics of North America* **27**, 191–197 (2000).

71. Crew, J. P. *et al.* Vascular Endothelial Growth Factor Is a Predictor of Relapse and Stage Progression in Superficial Bladder Cancer'. *Cancer Res* **57**, 5281–5285 (1997).

72. Pepper, M. S., Ferrara, N., Orci, L. & Montesano, R. Vascular endothelial growth factor (VEGF) induces plasminogen activators and plasminogen activator inhibitor-1 in microvascular endothelial cells. *Biochem Biophys Res Commun* **181**, 902–906 (1991).

73. Brown, N. S., Jones, A., Fujiyama, C., Harris, A. L. & Bicknell, R. Thymidine Phosphorylase Induces Carcinoma Cell Oxidative Stress and Promotes Secretion of Angiogenic Factors 1.

74. Nutt, J. E., Mellon, J. K., Qureshi, K. & Lunec, J. Matrix metalloproteinase-1 is induced by epidermal growth factor in human bladder tumour cell lines and is detectable in urine of patients with bladder tumours. *British Journal of Cancer 1998 78:2* **78**, 215–220 (1998).

75. Dameron, K. M., Volpert, O. v., Tainsky, M. A. & Bouck, N. Control of Angiogenesis in Fibroblasts by p53 Regulation of Thrombospondin-1. *Science (1979)* **265**, 1582–1584 (1994).

76. Kumar, Y., Koul, A., Singla, R. & Ijaz, M. F. Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. *J Ambient Intell Humaniz Comput* **1**, 1 (2022).

77. Goldenberg, S. L., Nir, G. & Salcudean, S. E. A new era: artificial intelligence and machine learning in prostate cancer. *Nat Rev Urol* **16**, 391–403 (2019).

78. Sorace, J. *et al.* Integrating pathology and radiology disciplines: An emerging opportunity? *BMC Med* **10**, 1–6 (2012).

79. Xiao, C., Choi, E. & Sun, J. Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. *Journal of the American Medical Informatics Association* **25**, 1419–1428 (2018).

80. Vijayan, V., Connolly, J., Condell, J., McKelvey, N. & Gardiner, P. Review of Wearable Devices and Data Collection Considerations for Connected Health. *Sensors 2021, Vol. 21, Page 5589* **21**, 5589 (2021).

81. Karczewski, K. J. & Snyder, M. P. Integrative omics for health and disease. *Nat Rev Genet* **19**, 299 (2018).

82. Borhani, S., Borhani, R. & Kajdacsy-Balla, A. Artificial intelligence: A promising frontier in bladder cancer diagnosis and outcome prediction. *Crit Rev Oncol Hematol* **171**, (2022).

83. Bishop, C. M. Pattern Recogniton and Machine Learning. *Information Science and Statistics* 738 (2006).

84. Hastie, T., Friedman, J. & Tibshirani, R. The Elements of Statistical Learning. (2001) doi:10.1007/978-0-387-21606-5.

85. Lidströmer, N. & Ashrafian, H. Artificial Intelligence in Medicine. (2020) doi:10.1007/978-3-030-58080-3.

86. Poggio, T. & Liao, Q. Theory I: Deep networks and the curse of dimensionality. *Bull. Pol. Ac.: Tech* **66**, 2018.

87. Lecun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature 2015 521:7553* **521**, 436–444 (2015).

88. Bengio, Y., Goodfellow, I. & Courville, A. Deep Learning. (2015).

89. Bakator, M. & Radosav, D. Deep Learning and Medical Diagnosis: A Review of Literature. *Multimodal Technologies and Interaction 2018, Vol. 2, Page 47* **2**, 47 (2018).

90. Chen, R., Zhang, L., Zang, D. & Shen, W. Blood drop patterns: Formation and applications. *Adv Colloid Interface Sci* **231**, 1–14 (2016).

91. Sobac, B. & Brutin, D. Desiccation of a sessile drop of blood: Cracks, folds formation and delamination. *Colloids Surf A Physicochem Eng Asp* **448**, 34–44 (2014).

92. Marieb, E. N. & Keller, S. M. *Essentials of Human Anatomy & Physiology plus Pearson Mastering Anatomy & Physiology with Pearson eText, Global Edition*. *Essentials of <h>Human</h> <h>Anatomy</h> <h>& Physiology</h> plus Pearson Mastering <h>Anatomy</h> & <h>Physiology</h> with Pearson eText, Global Edition* (2017).

93. Cramer, P. Organization and regulation of gene transcription. *Nature* vol. 573 45–54 Preprint at https://doi.org/10.1038/s41586-019-1517-4 (2019).

94. Yan, X. *et al.* Four novel biomarkers for bladder cancer identified by weighted gene coexpression network analysis. *J Cell Physiol* **234**, 19073–19087 (2019).

95. Schiffer, E. *et al.* Prediction of Muscle-invasive Bladder Cancer Using Urinary Proteomics. *Diagnosis* doi:10.1158/1078-0432.CCR-09-0226.

96. Habuchi, T. Prognostic Markers for Bladder Cancer. *Bladder Tumors:* 139–163 (2011) doi:10.1007/978-1-60761-928-4_8.

97. McNeil, B. K., Ekwenna, O. O. & Getzenberg, R. H. Molecular Signatures of Bladder Cancer. *Bladder Tumors:* 91–119 (2011) doi:10.1007/978-1-60761-928-4_6.

98. Pathak, B., Christy, J., Sefiane, K. & Gozuacik, D. Complex pattern formation in solutions of protein and mixed salts using dehydrating sessile droplets. *Langmuir* **36**, 9728–9737 (2020).

99. Noorbakhsh, J. *et al.* Deep learning-based cross-classifications reveal conserved spatial behaviors within tumor histological images. *Nature Communications 2020 11:1* **11**, 1–14 (2020).

100. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Nature Precedings 2010* 1–1 (2010) doi:10.1038/npre.2010.4282.1.

101. Eisenberg, D. T. A., Kuzawa, C. W. & Hayes, M. G. Improving qPCR telomere length assays: Controlling for well position effects increases statistical power. *American Journal of Human Biology* **27**, 570–575 (2015).

102. Pontén, F., Jirström, K. & Uhlen, M. The Human Protein Atlas—a tool for pathology. *J Pathol* **216**, 387–393 (2008).

103. Chen, G. *et al.* Comprehensive Identification and Characterization of Human Secretome Based on Integrative Proteomic and Transcriptomic Data. *Front Cell Dev Biol* **7**, 299 (2019).

104. Tang, Z. *et al.* GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res* **45**, W98–W102 (2017).

105. Tang, Z., Kang, B., Li, C., Chen, T. & Zhang, Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res* **47**, W556–W560 (2019).

106. Chang, S. S. *et al.* Diagnosis and Treatment of Non-Muscle Invasive Bladder Cancer: AUA/SUO Guideline. *J Urol* **196**, 1021–1029 (2016).

107. Faiena, I., Rosser, C. J., Chamie, K. & Furuya, H. Diagnostic biomarkers in non-muscle invasive bladder cancer. *World Journal of Urology 2018 37:10* **37**, 2009–2016 (2018).

108. Biardeau, X., Lam, O., Ba, V., Campeau, L. & Corcos, J. Prospective evaluation of anxiety, pain, and embarrassment associated with cystoscopy and urodynamic testing in clinical practice. *Can Urol Assoc J* **11**, 104–110 (2017).

109. Oeyen, E. *et al.* Bladder Cancer Diagnosis and Follow-Up: The Current Status and Possible Role of Extracellular Vesicles. *Int J Mol Sci* **20**, 821 (2019).

110. Herr, H. W. & Donat, S. M. Quality control in transurethral resection of bladder tumours. *BJU Int* **102**, 1242–1246 (2008).

111. Raitanen, M. P. *et al.* Routine follow-up cystoscopy in detection of recurrence in patients being monitored for bladder cancer. *Ann Chir Gynaecol* **90**, 261–265 (2001).

112. Burke, D. M., Shackley, D. C. & O'Reilly, P. H. The community-based morbidity of flexible cystoscopy. *BJU Int* **89**, 347–349 (2002).

113. Cameron, J. M., Butler, H. J., Palmer, D. S. & Baker, M. J. Biofluid spectroscopic disease diagnostics: A review on the processes and spectral impact of drying. *J Biophotonics* **11**, e201700299 (2018).

114. Diddens, C. *et al.* Evaporating pure, binary and ternary droplets: thermal effects and axial symmetry breaking. *J Fluid Mech* **823**, 470–497 (2017).

115. Annarelli, C. C., Fornazero, J., Bert, J. & Colombani, J. Crack patterns in drying protein solution drops. *European Physical Journal E* **5**, 599–603 (2001).

116. Pearce, E. I. & Tomlinson, A. Spatial location studies on the chemical composition of human tear ferns. *Ophthalmic and Physiological Optics* **20**, 306–313 (2000).

117. Rapis, E. A change in the physical state of a nonequilibrium blood plasma protein film in patients with carcinoma. *Technical Physics 2002 47:4* **47**, 510–512 (2002).

118. Martusevich, A. K., Zimin, Y., Bochkareva, A. & Kimovich, A. MORPHOLOGY OF DRIED BLOOD SERUM SPECIMENS OF VIRAL HEPATITIS. *Hep Mon* vol. 7 207–210 Preprint at https://sid.ir/paper/306146/en (2007).

119. Brutin, D., Sobac, B., Loquet, B. & Sampol, J. Pattern formation in drying drops of blood. *J Fluid Mech* **667**, 85–95 (2011).

120. Yakhno, T. A. *et al.* The informative-capacity phenomenon of drying drops. *IEEE Engineering in Medicine and Biology Magazine* **24**, 96–104 (2005).

121. Goodfellow, I. *et al.* Generative adversarial networks. *Commun ACM* **63**, 139–144 (2020).

122. Weiss, K., Khoshgoftaar, T. M. & Wang, D. D. A survey of transfer learning. *J Big Data* **3**, 1–40 (2016).

123. Shorten, C. & Khoshgoftaar, T. M. A survey on Image Data Augmentation for Deep Learning. *J Big Data* **6**, 1–48 (2019).

124. Chlap, P. *et al.* A review of medical image data augmentation techniques for deep learning applications. *J Med Imaging Radiat Oncol* **65**, 545–563 (2021).

125. Dubois, J., Rueger, J., Haubold, B., Far, R. K. K. & Sczakiel, G. Transcriptome analyses of urine RNA reveal tumor markers for human bladder cancer: validated amplicons for RT-qPCR-based detection. *Oncotarget* **12**, 1011 (2021).

126. Smalley, D. M., Sheman, N. E., Nelson, K. & Theodorescu, D. Isolation and identification of potential urinary microparticle biomarkers of bladder cancer. *J Proteome Res* **7**, 2088–2096 (2008).

127. Osman, I. *et al.* Novel Blood Biomarkers of Human Urinary Bladder Cancer. *Clinical Cancer Research* **12**, 3374–3380 (2006).

128. OpenStax by Rice University. *The Cardiovascular System: Blood Vessels and Circulation*. *Anatomy & Phisiology* (OpenStax, 2016).

129. Sefiane, K. On the Formation of Regular Patterns from Drying Droplets and Their Potential Use for Bio-Medical Applications. *Journal of Bionic Engineering 2010 7:4* **7**, S82–S93 (2010).

130. Bou Zeid, W. & Brutin, D. Influence of relative humidity on spreading, pattern formation and adhesion of a drying drop of whole blood. *Colloids Surf A Physicochem Eng Asp* **430**, 1–7 (2013).

131. Buzoverya, M. E., Shcherbak, Y. P. & Shishpor, I. v. Quantitative estimation of the microstructural inhomogeneity of biological fluid facies. *Technical Physics 2014 59:10* **59**, 1550–1555 (2014).

132. Brutin, D., Sobac, B. & Nicloux, C. Influence of substrate nature on the evaporation of a sessile drop of blood. *J Heat Transfer* **134**, (2012).

133. Lokeshwar, S. D. *et al.* Molecular Oncology of Bladder Cancer from Inception to Modern Perspective. *Cancers 2022, Vol. 14, Page 2578* **14**, 2578 (2022).

134. Tan, W. S. *et al.* Novel urinary biomarkers for the detection of bladder cancer: A systematic review. *Cancer Treat Rev* **69**, 39–52 (2018).

135. Schmitz-Dräger, B. J. *et al.* Molecular Markers for Bladder Cancer Screening, Early Diagnosis, and Surveillance: The WHO/ICUD Consensus. *Urol Int* **94**, 1–24 (2015).

136. Li, F. *et al.* The increased excretion of urinary orosomucoid 1 as a useful biomarker for bladder cancer. *Am J Cancer Res* **6**, 331 (2016).

137. Lorenzi, T. *et al.* HtrA1 in human urothelial bladder cancer: A secreted protein and a potential novel biomarker. *Int J Cancer* **133**, 2650–2661 (2013).

138. Humayun-Zakaria, N., Ward, D. G., Arnold, R. & Bryan, R. T. Trends in urine biomarker discovery for urothelial bladder cancer: DNA, RNA, or protein? *Transl Androl Urol* **10**, 2787808–2782808 (2021).

139. Rosser, C. J., Dai, Y., Miyake, M., Zhang, G. & Goodison, S. Simultaneous multi-analyte urinary protein assay for bladder cancer detection. *BMC Biotechnol* **14**, 1–6 (2014).

140. Allory, Y. *et al.* Telomerase Reverse Transcriptase Promoter Mutations in Bladder Cancer: High Frequency Across Stages, Detection in Urine, and Lack of Association with Outcome. *Eur Urol* **65**, 360–366 (2014).

141. Descotes, F. *et al.* Non-invasive prediction of recurrence in bladder cancer by detecting somatic TERT promoter mutations in urine. *British Journal of Cancer 2017 117:4* **117**, 583–587 (2017).

142. Couffignal, C. *et al.* The Diagnostic and Prognostic Performance of Urinary FGFR3 Mutation Analysis in Bladder Cancer Surveillance: A Prospective Multicenter Study. *Urology* **86**, 1185–1191 (2015).

143. Critelli, R. *et al.* Detection of multiple mutations in urinary exfoliated cells from male bladder cancer patients at diagnosis and during follow-up. *Oncotarget* **7**, 67435 (2016).

144. Ward, D. G. *et al.* Multiplex PCR and Next Generation Sequencing for the Non-Invasive Detection of Bladder Cancer. (2016) doi:10.1371/journal.pone.0149756.

145. Dagogo-Jack, I. & Shaw, A. T. Tumour heterogeneity and resistance to cancer therapies. *Nature Reviews Clinical Oncology 2017 15:2* **15**, 81–94 (2017).

146. Fan, J. *et al.* Linking transcriptional and genetic tumor heterogeneity through allele analysis of single-cell RNA-seq data. *Genome Res* **28**, 1217–1227 (2018).

147. Contreras-Trujillo, H. *et al.* Deciphering intratumoral heterogeneity using integrated clonal tracking and single-cell transcriptome analyses. *Nature Communications 2021 12:1* **12**, 1–14 (2021).

148. Dube, D. H. & Bertozzi, C. R. Glycans in cancer and inflammation — potential for therapeutics and diagnostics. *Nature Reviews Drug Discovery 2005 4:6* **4**, 477–488 (2005).

149. Martínez-Bosch, N. *et al.* Galectins in prostate and bladder cancer: tumorigenic roles and clinical opportunities. *Nature Reviews Urology 2019 16:7* **16**, 433–445 (2019).

150. Ruvolo, P. P. Galectin 3 as a guardian of the tumor microenvironment. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1863**, 427–437 (2016).

151. Liu, F. T. & Rabinovich, G. A. Galectins as modulators of tumour progression. *Nature Reviews Cancer 2005 5:1* **5**, 29–41 (2005).

152. Varki, Ajit; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. & E. *Essentials of Glycobiology, 3rd edition*. *Cold Spring Harbor (NY)* (Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 2015).

153. Thijssen, V. L. Galectins in Endothelial Cell Biology and Angiogenesis: The Basics. *Biomolecules* **11**, (2021).

154. Fritsch, K. *et al.* Galectin-3 interacts with components of the nuclear ribonucleoprotein complex. *BMC Cancer* **16**, 1–10 (2016).

155. Kurochkina, N. & Guha, U. SH3 domains: modules of protein–protein interactions. *Biophysical Reviews 2012 5:1* **5**, 29–39 (2012).

156. Gmeiner, W. H. & Horita, D. A. Implications of SH3 domain structure and dynamics for protein regulation and drug design. *Cell Biochem Biophys* **35**, 127–140 (2001).

157. Hellwege, J. N. *et al.* Association of gene coding variation and resting metabolic rate in a multi-ethnic sample of children and adults. *BMC Obes* **4**, 1–11 (2017).

158. Masoudi, M., Seki, M., Yazdanparast, R., Yachie, N. & Aburatani, H. A genome-scale CRISPR/Cas9 knockout screening reveals SH3D21 as a sensitizer for gemcitabine. *Scientific Reports 2019 9:1* **9**, 1–9 (2019).

APPENDIX

Appendix A: List of control individuals considered in this study. "-", no record for control volunteers; C, control.

Control number	Gender	Age	Type	Origin of tumor	Grade	Stage	Invasiveness
C1	Male	63	-	-	-	-	-
C2	Male	74	-	-	-	-	-
C3	Male	66	-	-	-	-	-
C4	Female	78	-	-	-	-	-
C5	Female	62	-	-	-	-	-
C6	Male	67	-	-	-	-	-
C7	Female	70	-	-	-	-	-
C8	Male	32	-	-	-	-	-
C9	Male	49	-	-	-	-	-
C10	Male	72	-	-	-	-	-
C11	Male	30	-	-	-	-	-
C12	Male	45	-	-	-	-	-
C13	Male	52	-	-	-	-	-
C14	Male	36	-	-	-	-	-
C15	Male	51	-	-	-	-	-
C16	Male	33	-	-	-	-	-
C17	Male	46	-	-	-	-	-
C18	Male	41	-	-	-	-	-
C19	Male	39	-	-	-	-	-
C20	Female	54	-	-	-	-	-
C21	Female	54	-	-	-	-	-
C22	Female	73	-	-	-	-	-
C23	Female	28	-	-	-	-	-
C24	Male	50	-	-	-	-	-
C25	Female	27	-	-	-	-	-
C26	Male	28	-	-	-	-	-
C27	Male	61	-	-	-	-	-
C28	Female	52	-	-	-	-	-
C29	Female	53	-	-	-	-	-
C30	Female	72	-	-	-	-	-
C31	Male	47	-	-	-	-	-
C32	Male	20	-	-	-	-	-
C33	Female	44	-	-	-	-	-
C34	Female	79	-	-	-	-	-
C35	Female	39	-	-	-	-	-
C36	Female	71	-	-	-	-	-
C37	Male	58	-	-	-	-	-

C38	Male	46	-	-	-	-	-
C39	Female	40	-	-	-	-	-
C40	Male	65	-	-	-	-	-
C41	Female	32	-	-	-	-	-
C42	Male	38	-	-	-	-	-
C43	Female	39	-	-	-	-	-
C44	Female	83	-	-	-	-	-
C45	Male	60	-	-	-	-	-
C46	Male	78	-	-	-	-	-
C47	Female	37	-	-	-	-	-
C48	Male	61	-	-	-	-	-
C49	Female	56	-	-	-	-	-
C50	Female	40	-	-	-	-	-
C51	Female	31	-	-	-	-	-
C52	Female	47	-	-	-	-	-
C53	Male	54	-	-	-	-	-
C54	Male	35	-	-	-	-	-
C55	Male	43	-	-	-	-	-
C56	Male	42	-	-	-	-	-
C57	Female	73	-	-	-	-	-
C58	Female	56	-	-	-	-	-
C59	Female	74	-	-	-	-	-
C60	Female	62	-	-	-	-	-
C124P	Male	76	-	-	-	-	-
C135P	Male	62	-	-	-	-	-
C138P	Male	70	-	-	-	-	-
C141P	Male	61	-	-	-	-	-

Appendix B: List of cancer patients considered in this study. UCC, urothelial cell carcinoma; LG, low grade; HG, high grade; "-", information not known; P, patients; Cis, carcinoma in situ; NMBIC, non-muscle invasive bladder cancer; MBIC, muscle invasive bladder cancer.

Patient number	Gender	Age	Type	Origin of tumor	Grade	Stage	Invasiveness
P1	Male	72	UCC	Primary	HG	pT1	NMIBC
P2	Male	61	UCC	Primary	HG	pT1	NMIBC
P3	Male	59	UCC	Primary	LG	pTa	NMIBC
P4	Male	54	UCC	Primary	HG	pTa	NMIBC
P5	Male	72	UCC	Primary	LG	pTa	NMIBC
P6	Male	53	UCC	Recurrence	HG	pT1	NMIBC
P7	Female	82	UCC	Primary	HG	pT1	NMIBC
P9	Male	62	UCC	Primary	HG	pT2	MIBC
P11	Male	62	UCC	Primary	LG	pTa	NMIBC
P12	Male	80	UCC	Primary	LG	pTa	NMIBC
P13	Male	59	UCC	Primary	HG	pT2	MIBC
P14	Male	65	UCC	Primary	-	pT1	NMIBC
P15	Male	77	UCC	Primary	HG	pT1	NMIBC
P17	Female	61	UCC	Primary	HG	pT1	NMIBC
P18	Male	49	UCC	Primary	HG	pT2	MIBC
P19	Male	80	UCC	Recurrence	HG	pTa	NMIBC
P21	Male	69	UCC	Recurrence	LG	pTa	NMIBC
P22	Male	73	UCC	Recurrence	HG	pT1	NMIBC
P23	Male	62	UCC	Primary	LG	pTa	NMIBC
P24	Male	60	UCC	Primary	HG	pT1	NMIBC
P25	Male	69	UCC	Primary	HG	pTa	NMIBC
P26	Male	39	UCC	Primary	LG	pTa	NMIBC
P27	Female	66	UCC	Primary	HG	pT2	MIBC
P28	Female	66	UCC	Primary	HG	pT1	NMIBC
P29	Male	72	UCC	Primary	HG	pT1	NMIBC
P30	Male	23	UCC	Primary	LG	pTa	NMIBC
P31	Male	68	UCC	Primary	HG	pTa	NMIBC
P32	Male	70	UCC	Primary	LG	pTa	NMIBC
P33	Male	49	UCC	Recurrence	HG	pT1	NMIBC
P34	Female	73	UCC	Primary	HG	pT1	NMIBC
P35	Male	65	UCC	Primary	LG	pTa	NMIBC
P36	Male	70	UCC	Primary	HG	pT1	NMIBC
P37	Male	52	UCC	Primary	LG	pTa	NMIBC
P38	Male	86	UCC	Primary	HG	pT2	MIBC
P39	Male	74	UCC	Primary	HG	pT1	NMIBC
P40	Male	51	UCC	Recurrence	LG	pTa	NMIBC
P41	Male	50	UCC	Primary	HG	pTa	NMIBC
P42	Male	69	UCC	Primary	HG	pT1	NMIBC
P43	Male	74	UCC	Primary	HG	pT2	MIBC
P44	Male	59	UCC	Recurrence	HG	pT1	NMIBC
P45	Female	89	UCC	Primary	HG	pT1	NMIBC
P46	Male	67	UCC	Primary	LG	pTa	NMIBC
P47	Male	63	UCC	Primary	LG	pTa	NMIBC
P48	Male	51	UCC	Primary	HG	pT2	MIBC

P50	Male	76	UCC	Primary	LG	pTa	NMIBC
P52	Male	84	UCC	Primary	HG	pT1	NMIBC
P53	Male	46	UCC	Primary	HG	pTa	NMIBC
P54	Female	70	UCC	Primary	LG	pTa	NMIBC
P55	Male	48	UCC	Primary	LG	pT1	NMIBC
P56	Male	71	UCC	Primary	LG	pTa	NMIBC
P57	Male	69	UCC	Primary	LG	pTa	NMIBC
P59	Male	69	UCC	Primary	HG	pT1	NMIBC
P60	Male	71	UCC	Primary	HG	pT1	NMIBC
P61	Male	61	UCC	Primary	HG	pT1	NMIBC
P62	Male	68	UCC	Primary	LG	pT1	NMIBC
P63	Male	67	UCC	Primary	HG	pT2	MIBC
P64	Male	78	UCC	Primary	HG	pTa	NMIBC
P65	Male	82	UCC	Primary	LG	pT1	NMIBC
P66	Male	73	UCC	Primary	HG	pTa	NMIBC
P67	Male	69	UCC	Primary	HG	pTa	NMIBC
P68	Male	45	UCC	Primary	LG	pTa	NMIBC
P69	Male	72	UCC	Recurrence	HG	pT1	NMIBC
P70	Male	79	UCC	Recurrence	HG	pT1	NMIBC
P71	Female	65	UCC	Primary	LG	pTa	NMIBC
P73	Male	63	UCC	Recurrence	LG	pTa	NMIBC
P74	Male	66	UCC	Recurrence	LG	pTa	NMIBC
P75	Male	75	UCC	Primary	HG	pT1	NMIBC
P76	Male	64	UCC	Primary	HG	pT1	NMIBC
P78	Male	71	UCC	Primary	HG	pT1	NMIBC
P79	Male	68	UCC	Recurrence	LG	pTa	NMIBC
P80	Female	64	UCC	Primary	HG	pT1	NMIBC
P81	Male	36	UCC	Primary	LG	pTa	NMIBC
P82	Male	45	UCC	Primary	LG	pTa	NMIBC
P83	Male	68	UCC	Primary	HG	pT1	NMIBC
P84	Male	52	UCC	Recurrence	HG	pTa	NMIBC
P85	Female	83	UCC	Recurrence	HG	pT2	MIBC
P86	Female	62	UCC	Recurrence	LG	pTa	NMIBC
P87	Male	83	UCC	Recurrence	LG	pTa	NMIBC
P88	Male	58	UCC	Recurrence	LG	pTa	NMIBC
P89	Female	40	UCC	Recurrence	LG	pTa	NMIBC
P91	Male	59	UCC	Recurrence	LG	pTa	NMIBC
P92	Male	73	UCC	Primary	LG	pTa	NMIBC
P95	Female	55	UCC	Primary	HG	pT1	NMIBC
P96	Male	86	UCC	Primary	HG	pT1	NMIBC
P97	Male	79	UCC	Primary	HG	pT2	MIBC
P98	Male	61	UCC	Primary	LG	pTa	NMIBC
P99	Male	61	UCC	Primary	HG	pT1	NMIBC
P100	Male	49	UCC	Recurrence	HG	pT1	NMIBC
P101	Male	81	UCC	Primary	LG	pTa	NMIBC
P103	Male	67	UCC	Recurrence	LG	pTa	NMIBC
P104	Male	57	UCC	Recurrence	LG	pTa	NMIBC
P105	Female	68	UCC	Primary	HG	pT1	NMIBC
P106	Male	59	UCC	Primary	LG	pT1	NMIBC
P107	Male	73	UCC	Primary	LG	pTa	NMIBC
P108	Male	53	UCC	Primary	LG	pTa	NMIBC

P109	Female	54	UCC	Recurrence	LG	pTa	NMIBC
P110	Male	69	UCC	Primary	LG	pTa	NMIBC
P111	Male	62	UCC	Recurrence	LG	pTa	NMIBC
P112	Male	66	UCC	Primary	LG	pTa	NMIBC
P113	Male	37	UCC	Primary	HG	pT1	NMIBC
P114	Female	62	UCC	Primary	LG	pTa	NMIBC
P115	Male	68	UCC	Primary	HG	pTa	NMIBC
P117	Male	67	UCC	Primary	LG	pTa	NMIBC
P118	Male	74	UCC	Primary	HG	pT2	MIBC
P119	Male	63	UCC	Primary	HG	pTa	NMIBC
P120	Male	88	UCC	Recurrence	HG	pTa	NMIBC
P121	Female	73	UCC	Recurrence	LG	pTa	NMIBC
P122	Female	61	UCC	Primary	HG	pT1	NMIBC
P123	Female	63	UCC	Primary	LG	pTa	NMIBC
P125	Male	51	UCC	Primary	LG	pT1	NMIBC
P126	Female	72	UCC	Recurrence	HG	pT1	NMIBC
P127	Male	62	UCC	Primary	HG	pT1	NMIBC
P128	Male	60	UCC	Primary	LG	Cis	NMIBC
P129	Male	56	UCC	Primary	HG	pTa	NMIBC
P131	Male	66	UCC	Primary	HG	pT1	NMIBC
P132	Male	42	UCC	Recurrence	LG	pTa	NMIBC
P134	Male	62	UCC	Primary	HG	pT2	MIBC
P136	Male	68	UCC	Primary	HG	pT1	NMIBC
P137	Male	74	UCC	Recurrence	LG	pTa	NMIBC
P139	Male	54	UCC	Recurrence	LG	pTa	NMIBC
P140	Male	59	UCC	Recurrence	LG	pTa	NMIBC
P143	Male	63	UCC	Recurrence	LG	pTa	NMIBC
P144	Male	61	UCC	Primary	LG	pTa	NMIBC
P145	Male	73	UCC	Recurrence	LG	pT1	NMIBC
P146	Male	79	UCC	Primary	HG	pT1	NMIBC
P147	Male	67	UCC	Primary	HG	pT1	NMIBC
P148	Male	86	UCC	Primary	LG	Cis	NMIBC
P149	Male	50	UCC	Primary	LG	pTa	NMIBC
P150	Male	66	UCC	Recurrence	HG	pTa	NMIBC
P151	Male	64	UCC	Primary	LG	pT1	NMIBC

Appendix C: Statistical analysis of blood test parameters considered in clinics for control individuals and bladder cancer patients. * These values are out of clinical range.

	Number / Percent (%)		Mean	
Parameter (Unit)	(Control)	(Patient)	(Control)	(Patient)
Glucose (mg/dl)	40 (59.7)	63 (47.4)	98.7	115.1*
BUN (mg/dl)	42 (62.7)	90 (67.7)	12.6	17.9
Creatinine (mg/dl)	57 (85.1)	84 (63.2)	0.9	0.98
AST (U/L)	52 (77.6)	84 (63.2)	25.0	23.3
ALT (U/L)	52 (77.6)	84 (63.2)	30.0	22.9
Sodium (mEq/L)	56 (83.6)	92 (69.2)	140.9	139.7
Potassium (mEq/L)	56 (83.6)	87 (65.4)	4.4	4.5
Chloride (mEq/L)	30 (44.8)	34 (25.6)	103.6	103.1
Calcium (mg/dL)	16 (23.9)	29 (21.8)	9.5	9.6
Phosphorus (mg/dL)	4 (6.0)	14 (10.5)	3.2	3.6
Urea (mg/dL)	30 (44.8)	25 (18.8)	26.0	38.0
WBC $x10^3$	62 (92.5)	96 (72.7)	7.3	8.3
HGB (g/dL)	61 (91.0)	83 (62.4)	13.4	13.5
HCT (%)	61 (91.0)	76 (57.1)	39.6	39.6
PLT x 10^3 /μL	62 (92.5)	97 (72.9)	259.8	241.6
PT (sec)	49 (73.1)	65 (48.9)	10.9*	13.6*
aPTT (sec)	47 (70.1)	51 (38.3)	27.0*	28.6
INR	49 (73.1)	64 (48.1)	1.01	1.04

Appendix D: Clean reads quality metrics.

Sample	Pathology	Total raw reads (M)	Total clean reads (Gb)	Total clean bases (4.61)	Clean reads ration (%)
Sample1	pTa-LG	47.48	46.05	4.61	97.00
Sample2	pTa-LG	47.48	46.12	4.61	97.15
Sample3	pTa-HG	47.48	46.07	4.61	97.05
Sample4	pTa-HG	47.48	45.9	4.59	96.68
Sample5	pTa-HG	47.48	46.33	4.63	97.58
Sample6	pT1-HG	47.48	46.15	4.61	97.2
Sample7	pT1-HG	47.48	45.99	4.6	96.86
Sample8	pT1-HG	47.48	45.94	4.59	96.76
Sample10	pT2-HG	47.48	46.16	4.62	97.23
Sample11	pT2-HG	47.48	45.98	4.60	96.85
Sample12	Control	43.29	41.92	4.19	96.83

Appendix E: Sequencing of TLCV2 vector after cloning of gal-3 gRNA1 (A) and gRNA2 (B).