Molecular Characterisation of Bladder and Upper Tract Urothelial Carcinoma

Hugo Alexandre Aguado Robert

MSc by Research

University of York

Biology

December 2025

Abstract

Despite morphological similarities, bladder (BUC) and upper tract urothelial carcinomas (UTUC) exhibit distinct clinical behaviours and survival outcomes. This study investigated whether these disparities arise from the distinct embryological origins of their respective tissues. The transcriptomic profiles of normal and malignant urothelium were analysed using short-read Illumina and long-read Oxford Nanopore sequencing to identify gene expression and isoform differences. In normal tissues, 3,698 differentially expressed genes were identified. Ureteric urothelium retained a distinct HOX gene signature reflecting its mesodermal origin, whereas bladder urothelium exhibited elevated immune signalling and barrier function markers. In malignant tissues, RNA degradation limited comprehensive isoform analysis, however, seven genes exhibited significant differential expression. *AQP3* was upregulated in UTUC, whilst metallothioneins (*MT2A*, *MT1E*) were upregulated in BUC. Furthermore, Gene Set Enrichment Analysis indicated different enrichment in hypoxia and oxidative stress pathways between the tumour types. These findings demonstrate that adult urothelium retains positional identity markers and distinct molecular programmes. Consequently, the difference between UTUC and BUC may originate from the transcriptomic landscape of the tissue of origin rather than tumour specific mutations alone. This highlights the necessity for site specific therapeutic strategies rather than treating urothelial carcinoma as a uniform disease.

Declaration

I declare that this thesis is a presentation of original work, and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as references.

Table of contents

Chapter 1: Introduction

Urothelial carcinoma has over 10,000 new diagnoses annually in the UK and has large differences in patient outcomes that current risk stratification systems fail to adequately predict. Whilst bladder and upper tract urothelial carcinomas arise from morphologically identical tissue, they exhibit markedly different clinical behaviours, upper tract urothelial carcinomas present at more advanced stages at diagnosis and have poorer long term survival rates compared to bladder urothelial carcinomas (National Cancer Institute, 2023a; Nally *et al.*, 2024). These clinical differences suggest that fundamental molecular differences exist in these tissues, which remain poorly understood, representing a significant gap in developing targeted therapeutic strategies.

1.1 The Urinary Tract

The urinary tract plays a vital role in filtering out waste and excess fluid from the blood and removing them from the body as urine. The urinary tract can be divided into two parts, the upper urinary tract comprising the kidneys and the ureters, and the lower urinary tract, which consists of the bladder and the urethra, Figure 1.1 (Mahadevan, 2016).

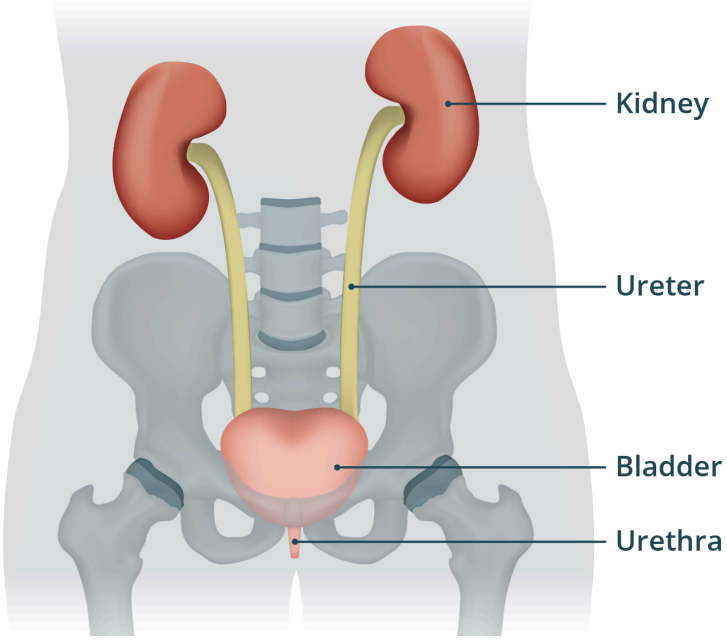

Figure 1.1: Anatomy of the urinary tract. A diagram of the urinary tract showing the kidneys, ureters, bladder and urethra. Sourced from (Healthdirect Australia, 2025).

The kidneys continuously filter blood, removing waste and extra water, thereby producing urine. The urine is subsequently transported to the bladder, through the ureters. The ureters are bilateral muscular tubular structures approximately 26 cm long with a 3 to 4 mm diameter that connect the kidneys to the bladder (Lescay *et al.*, 2025). They are capable of propelling

urine through peristaltic contractions of the muscles in the ureter wall, moving urine from the renal pelvis to the bladder (Zhu *et al.*, 2025).

The bladder serves multiple functions within the urinary tract, it must collect, store and expel urine. In order to perform these functions it must be both flexible and impermeable, to accommodate the changes in volume and prevent leakage (Andersson and Arner, 2004). The human bladder is also required to void urine periodically, during which its shape changes considerably, from oval shaped when full to flattened by the overlying bowel when empty (Damaser and Lehman, 1995). This differs from the mice, where they use urination as a form of social communication, which influences their voiding patterns (Arakawa *et al.*, 2008). The final structure of the urinary tract, the urethra, serves as a conduit through which the urine stored in the bladder is expelled from the body.

The structural architecture of the organs that comprise the urinary tract enables them to carry out their specialised functions. The bladder is composed of four distinct layers, each of which contribute to its function, Figure 1.2. The histological structure of the bladder wall layers, from luminal to serosal comprises; the urothelium, which functions as a barrier, the lamina propria, which supplies the bladder with blood and lymphatic systems, the muscularis propria, which consists of detrusor muscle and facilitates voiding, and the serosa and perivesical fat which provide support and mobility during voiding (Bolla *et al.*, 2025).

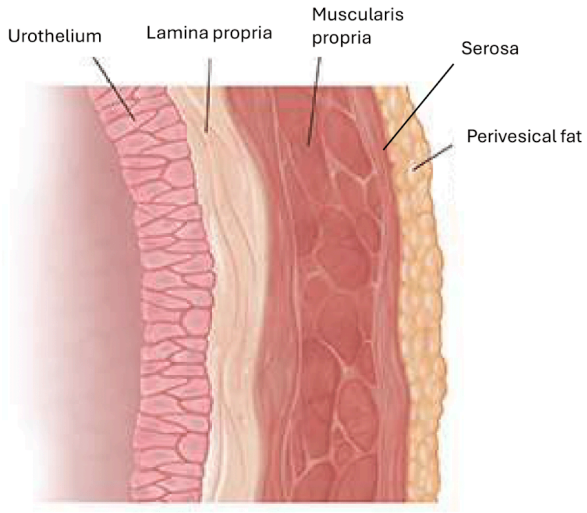

Figure 1.2: The layers of the bladder wall. A diagram of the bladder wall, showing its four distinct layers, the urothelium, lamina propria, muscularis propria, made up of detrusor muscle, and the serosa. The perivesical fat which surrounds the bladder is also shown. Adapted from (Saint Luke's, 2025).

The ureter wall is similarly characterised by a layered structure, and comprises various layers, specialised for its function in transporting urine. The innermost layers consist of the urothelium and the lamina propria which form the mucosa, this is characterised by folds which help prevent reflux of urine, and a muscle layer composed of two layers of smooth muscle, which perform the peristaltic contractions, Figure 1.3 (Moinuddin and Dhanda, 2015; Lescay *et al.*, 2025).

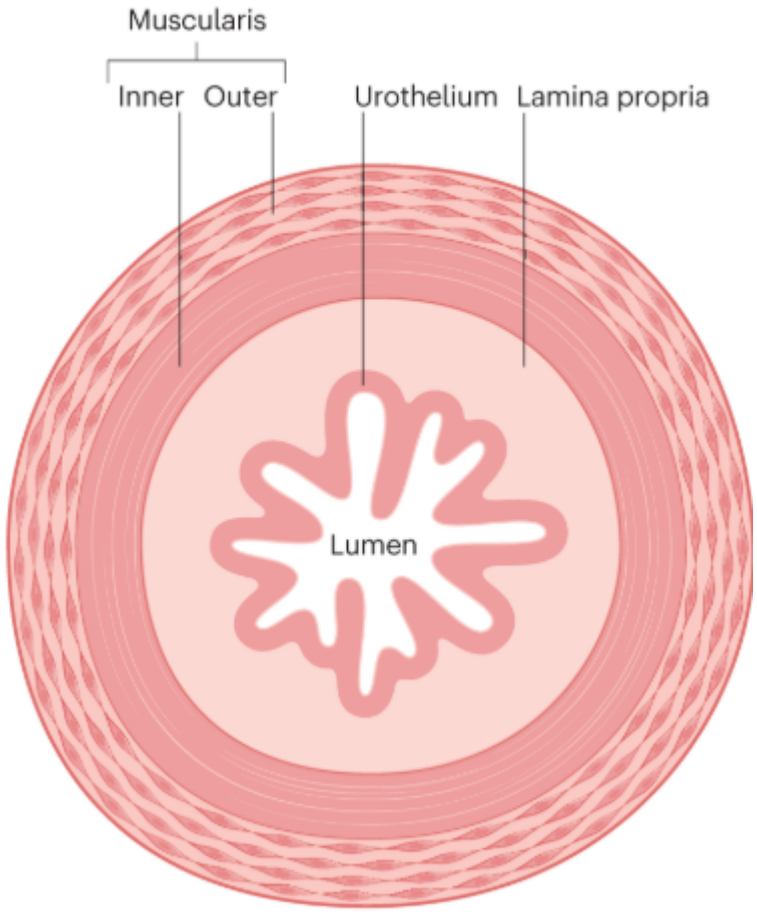

Figure 1.3: The layers of the ureter wall. A diagram showing the layers of the ureter wall including the urothelium, lamina propria and the muscularis propria, made up of smooth muscle where the inner is longitudinally arranged and outer is circularly arranged. Adapted from (O'Meara *et al.*, 2024).

Understanding how these structures form during development provides important insights into potential distinctions in their molecular identities, which may persist into adulthood.

1.1.1 Development of the urinary tract

The urothelium appears to be a singular tissue, like other epithelial linings such as those found in the gastrointestinal tract or the respiratory tract (Wan *et al.*, 2018). However, it is unique in that it originates from different parts of the embryo. The urothelium which lines the renal pelvis and ureters is derived from the mesoderm whereas the urothelium lining the bladder and proximal urethra originates from the endoderm, Figure 1.4 (Georgas *et al.*, 2015).

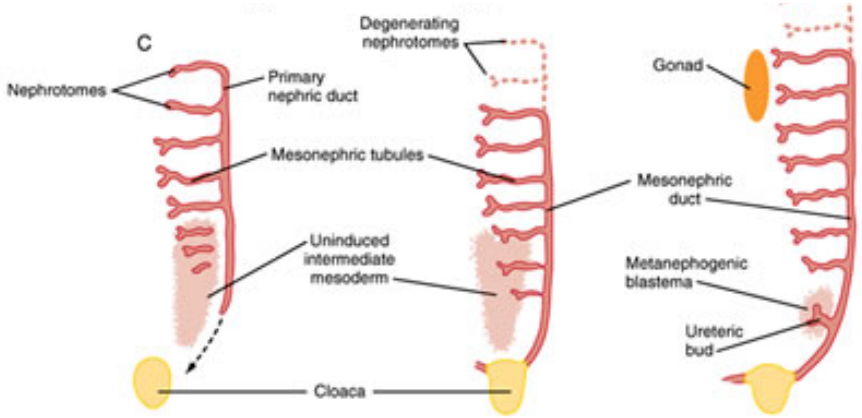

Figure 1.4: Embryological development of the urinary tract. A diagram showing the chronological development of the ureteric bud which gives rise to the upper urinary tract, the ureter and renal pelvis, which grows from the mesonephric duct. The bladder develops from the cloaca. Sourced from (Duke University School of Medicine, 2015).

More specifically, the ureters develop from the mesodermal ureteric bud, which can be described as an outgrowth of the Wolffian duct (Bohnenpoll and Kispert, 2014). The bladder develops from the endodermal urogenital sinus following the division of the embryonic cloaca (Rasouly and Lu, 2013). Through the process of epithelial-mesenchymal interactions, the simple bilayered epithelium differentiates into multilayered urothelium (Baskin *et al.*, 1996). These distinct embryological origins result in a mature urothelium which, whilst appearing uniform throughout the urinary tract, is required to meet the differing demands and functions of each organ.

The development of the urinary tract is controlled through various molecular pathways, disruption to these pathways leads to both congenital anomalies and adult malignancies. During embryogenesis the formation and differentiation of the urinary tract requires the coordination of various signalling pathways, including retinoic acid (RA) signalling, sonic hedgehog (SHH) signalling, TGF-beta, Wnt, fibroblast growth factor (FGF) and HOX transcription factors (Rasouly and Lu, 2013). These pathways crosstalk and ensure correct tissue development, cellular differentiation and organ morphogenesis (Aulehla and Pourquié, 2010).

Among these pathways RA signalling serves as a master regulator of cellular differentiation, maintaining progenitor population in the urinary tract whilst controlling the timing and extent of differentiation across urinary organs (Bohnenpoll, Weiss, *et al.*, 2017). RA signalling works in concert with SHH signalling, which coordinates epithelial-mesenchymal interactions and induces BMP4 expression, ensuring proper tissue architecture (Bohnenpoll, Wittern, *et al.*, 2017). This SHH-BMP4 interaction is crucial for urothelial stratification and barrier formation (Meuser *et al.*, 2022). FGF signalling drives branching morphogenesis of the ureters and maintains progenitor stemness in the nephrons (Barak *et al.*, 2012). The Wnt pathway controls epithelial to mesenchymal transitions during nephrogenesis (Park, Valerius and McMahon, 2007). HOX transcription factors ensure the appropriate development at each position in the urinary tract (Libretti and Aeddula, 2025).

The clinical significance of these signalling pathways extends beyond embryogenesis as they remain active in adult tissues and can become dysregulated in urological malignancies

(Kumar *et al.*, 2021). In urothelial carcinomas, loss of RA signalling contributes to the compromised barrier function and the dysregulation of cellular differentiation (Zupančič *et al.*, 2020). The loss of SHH accelerates tumour progression, whilst FGF signalling becomes hyperactivated, driving rapid and uncontrolled cell growth (Kim *et al.*, 2019; Garje *et al.*, 2020). Understanding the differences in these molecular programmes in the different organs of the urinary tract may be able to explain the differences in the urothelial carcinoma rates at these different sites.

1.1.2 The urothelium

The urinary tract presents one of the harshest chemical environments in the body, due to exposure to urine, which is composed of various metabolic waste products, toxins and concentrations of solutes that have the potential to damage tissues (Ogobuiro and Tuma, 2025). The urothelium is a specialised epithelial tissue which serves to meet this need, protecting underlying tissues by forming a complete and uninterrupted lining of the walls of the urinary passages (Winder *et al.*, 2014; Dalghi *et al.*, 2020). Not only must the urothelium be able to serve this protective function, but it must also be flexible enough to accommodate to changes in surface area resulting from the voiding and expansion of the bladder from urine (Dalghi *et al.*, 2021).

The urothelium is composed of three stratified cell layers, comprising basal, intermediate and superficial cells, the latter of which are sometimes termed umbrella cells, Figure 1.5 (Jaimes-Parra *et al.*, 2016). The superficial cells are responsible for the impermeability of the urinary tract, through the formation of urothelial plaques on the luminal surface and tight junctions, which offer transcellular resistance and limit the permeability to water, solutes, and toxins (Hu *et al.*, 2000; Acharya *et al.*, 2004). The intermediate cells are hypothesised to confer the urothelium its fast acting regenerative properties, quickly replenishing damaged superficial cells (Wang *et al.*, 2018). The cells responsible for the urothelium's long-term regeneration remain debated. Whilst modelling suggests that basal cells may be the origin of this trait, experimental evidence indicates that basal and suprabasal cell populations are able to regenerate the entire urothelial structure through cellular plasticity rather than through stem cell hierarchies, with no urothelial stem cell being identified to date (Wezel, Pearson and Southgate, 2014; Siegel *et al.*, 2025).

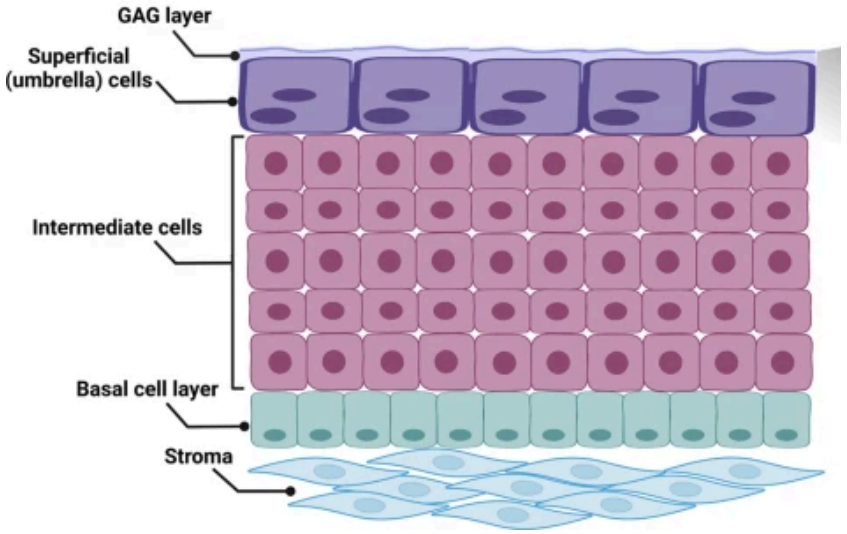

Figure 1.5: The different types of cells making up the urothelium. A figure showing the layers of the urothelium, showing the superficial cells, intermediate cell layers, basal cell layer and the underlying stroma. Adapted from (Jafari and Rohn, 2022).

Under homeostatic conditions the urothelium is quiescent, which means that they are in cell cycle arrest, with one of the slowest turnovers of ~200 days (Cooper, 1972). This reduces the risk of mutations that could arise due to high proliferation, which is especially important in such a harsh environment. Upon injury however, the cells are capable of completely restoring the injured tissue within 72 hours, thereby restoring its protective role (Paraskevopoulou, Papafotiou and Klinakis, 2016). Given these unique properties, researchers have developed various model systems in which to study the urothelium.

1.1.3 Models of the urothelium and urothelial carcinoma

Understanding the molecular basis of the urothelium has been limited by the inherent drawbacks of the available experimental models, which include cell culture systems, rodent models and fresh human tissue.

Cell culture models are widely used in studying the urothelium, with transformed cell lines such as T24, UM-UC-3 and RT4, as they retain key oncogenic mutations observed in human tumours (Earl *et al.*, 2015). However, two dimensional systems are limited by their inability to recreate the biological complexity and cellular interactions found in vivo, which is particularly relevant when comparing such similar tissues (Zhang and Atala, 2013; Hustler *et al.*, 2018; Tse *et al.*, 2022). The UROtsa cell line is able to form stratified layers resembling in-situ urothelium, however it does not accurately represent a fully differentiated urothelium as it lacks uroplakins (Rossi *et al.*, 2001; Eblin, Bredfeldt and Gandolfi, 2008). Whilst early finite lifespan normal-human urothelial (NHU) cells from surgical specimens, which can stratify and form a barrier, struggled to achieve complete differentiation as they lack uroplakins and certain expression characteristics of superficial cells, later advancements enabled these cells to develop terminal differentiation markers like uroplakins (Cross *et al.*, 2005; Varley and Southgate, 2008; Dalghi *et al.*, 2020; Jafari and Rohn, 2022). These models exhibit the multi-layered stratification and high electrical resistance characteristic of the human

urothelium in-vivo (Baker, Shabir and Southgate, 2014; Dalghi *et al.*, 2020; Jafari and Rohn, 2022).

Whilst three-dimensional organoid models provide more physiologically relevant organisation, their closed spheroid shape fails to recapitulate tubular geometry and mechanical forces characteristic of the urinary tract (Whyard *et al.*, 2020). Ultimately neither 2D stratified cultures nor 3D approaches are able to capture the complete biological complexity as they both lack systems such as mechanical stretch, blood vessels, immune infiltration and organ interactions, which may influence the molecular differences between bladder and ureter urothelium (Wei *et al.*, 2022). Non-spheroid 3D models, such as microfluidic organ-on-a-chip systems, are capable of combining a multi-layered urothelium with continuous fluid flow and mechanical stretch to more accurately simulate the urinary tract (Hou *et al.*, 2023; Paduthol *et al.*, 2026).

Though three dimensional models provide more physiologically relevant cellular organisation than monolayer cultures, rodent models provide the opportunity for the study of urothelial carcinoma in in-vivo conditions (John and Said, 2017; Gills *et al.*, 2018; Mullenders *et al.*, 2019). Despite morphologically similar neoplasms in rodent models to those of humans, their translational relevance remains uncertain (Oyasu, 1995; John and Said, 2017). This is due to significant species differences found in this tissue, in humans uroplakins are only expressed by superficial cells, whereas mice express uroplakins in other cell layers, including the basal cells, impacting the way the urothelium differentiates and consequently extrapolating findings from these models into human systems is not always possible (Cohen and Lawson, 1995; Zhang *et al.*, 1999).

Despite the respective advantages of cell culture and rodent models, they present limitations in studying urothelial biology that cannot be overlooked when the aims of research are to understand the human urothelium. Chemically induced rodent models which generate tumours which are similar to humans histologically and genetically, but are unable to fully model immune interactions, heterogeneity and luminal subtype (Williams, Lee and Theodorescu, 2008). Furthermore, discrepancies exist in gene expression between these rodent models and humans (Stavrou and Ross, 2015). Genetically engineered rodent models, in which researchers clone oncogenes or delete tumour suppressor genes, typically investigate only single-gene alterations which do not capture the polygenic nature of urothelial carcinomas (Zhang et al., 1999; Park et al., 2021). Transplant rodent models, where cancer cells are implanted into an animal, require immunocompromised hosts and may not be clinically relevant (Jäger et al., 2015; Y.-R. Liu et al., 2019).

Fresh human samples can be considered the most true to in-vivo human biology, as they preserve tissue architecture and its complex microenvironment (Vitek *et al.*, 2022). This enables the study of the genetic profile of the urothelium and urothelial carcinoma, without culture-induced mutations that would otherwise be observed in the cell culture approaches (Chen *et al.*, 2020; Zeber-Lubecka *et al.*, 2022). However, practical challenges associated with fresh tissue have limited their use. Ethical approval processes result in delays, whilst the time-constrained processing places burdens on clinical teams and causes sample degradation. These logistical constraints typically restrict studies to small sample sizes.

Advances in sequencing technologies have provided opportunities to characterise the genetic and transcriptomic profiles of these model systems. Primary human tissue samples offer distinct advantages over the other systems for studying these landscapes, by preserving molecular signatures without culture-induced mutations and avoiding confounding effects from species differences which could obscure human tissue specific differences.

1.2 Urothelial Carcinoma

Urothelial carcinoma (UC) is the cancer of the urothelium, the most prominent of all bladder cancers, accounting for approximately 90% of all cases (Leslie, Soon-Sutton and Aeddula, 2025). In comparison with other cancers, bladder cancer is the 10th most common cancer in the UK, with approximately 10,000 people per year being diagnosed and a corresponding annual mortality of 5,000 deaths (UK National Screening Committee, 2020). Whilst most people can be readily treated with surgery, the five-year recurrence for bladder cancer is approximately 70% (Chamie *et al.*, 2013; Ma *et al.*, 2024).

Although bladder cancer is commonly understood to refer to urothelial carcinoma, urothelial carcinoma occurs not only in the bladder but also throughout the urinary tract. As Table 1.1 shows, whilst bladder cancer cases are the most common form with 8752 new diagnoses in 2020, upper tract disease accounted for 1447 cases, ureter and renal pelvis counts combined. Despite the anatomical spread of urothelial carcinoma most research has focused primarily on bladder cancer, consequently leading to a paucity in the literature focused on upper tract disease in comparison. This has led to the anatomical division of urothelial carcinomas in bladder urothelial carcinomas (BUC) and upper tract urothelial carcinomas (UTUC) (Saginala *et al.*, 2020). However, when clinicians refer to differing types of urothelial carcinoma they typically refer to invasion depth rather than anatomical location, distinguishing between muscle-invasive and non-muscle invasive disease. As explored in the sections below, each of these classifications have important implications in understanding urothelial carcinoma progression and outcome.

Cancer site	Incidence counts (2020)			Mortality counts (2020)		
	Both sexes	Male	Female	Both sexes	Male	Female
Malignant neoplasms of the bladder	8752	6391	2361	4754	3299	1455
Malignant neoplasms of the ureter	687	442	245	221	122	99
Malignant neoplasms of renal pelvis	760	452	308	32	16	16
Malignant neoplasms of other and unspecified urinary organs	209	182	27	387	243	144

Table 1.1: Cancer incidence and mortality counts by urinary tract cancer type in England, 2020. The table shows the number of new cancer diagnoses (incidence) and deaths (mortality) for various cancer sites of the urinary system, with combined data for both sexes and also by sex. Data taken from the NHS England Cancer Data Hub using ICD-10 codes: bladder (C67), ureter (C66), renal pelvis (C65), and other and unspecified urinary organs (C68). Source: (National Disease Registration Service, 2024).

Classifying urothelial carcinomas clinically is typically done using tumour-node-metastasis (TNM) staging, classifying tumours on the basis of the depth of invasion (T), lymph node involvement (N), and distant metastasis (M). The T stage progresses from Ta (noninvasive papillary) and Tis (carcinoma in situ) to T1 (lamina propria invasion), these are considered the non-muscle invasive disease, T2 (muscle invasion), T3 (perivesical, peripelvic or periureteric tissue), and T4 (adjacent organ invasion), the latter of which are considered muscle invasive disease (Cheng *et al.*, 2009; Honda *et al.*, 2019). Node involvement staging ranges from N0 (no nodes) to N3 (common iliac nodes), whilst M staging goes from no metastasis (M0), distant lymph nodes (M1a), and non-lymph node distant metastases (M1b) (Quarles *et al.*, 2021). Overall staging combines these stages to help inform clinical decision making, Figure 1.6 (Paner *et al.*, 2018).

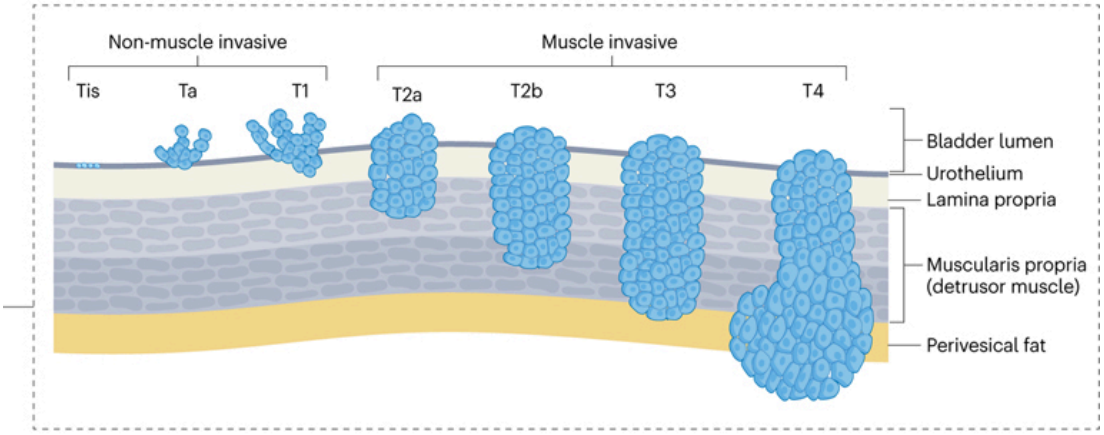

Figure 1.6: Tumour stages of urothelial carcinoma. A figure showing the stages of urothelial carcinoma, with non-muscle invasive disease, stages Tis, Ta and T1, and muscle invasive disease, T2a, T2b, T3 and T4. Adapted from (Dyrskjøt *et al.*, 2023).

1.2.1 Bladder urothelial carcinoma (BUC)

Approximately 90-95% of urothelial carcinomas are bladder urothelial carcinomas (BUC), with an incidence rate of approximately 10-20/100,000 people per year (Richters, Aben and Kiemeney, 2020). BUC has a high male predominance, 4:1 male to female ratio, and patients present at an average of 73 years (Hemelt *et al.*, 2009; Downes, 2022; National Cancer Institute, 2023b). BUC is traditionally classified by invasion depth, non-muscle invasive bladder cancer (NMBIC) which comprises stages Tis, Ta and T1 and muscle invasive bladder cancer (MIBC) which comprises stages T2, T3 and T4. Whilst traditional clinical staging guides treatment for bladder cancer, the heterogeneous nature of BUC outcomes has suggested that there may be differentiating factors that would allow for a more accurate treatment leading to better outcomes.

This hypothesis drove multiple research groups to develop molecular classification systems based on gene expression profiles, as later described in Section 1.3. For MIBC, six major molecular classification methods were created from Baylor, University of North Carolina, MD Anderson Cancer Center, the Cancer Genome Atlas (TCGA), Cartes d'Identité des Tumeurs Curie (CIT), and Lund, each identifying different numbers of molecular subtypes (Choi *et al.*, 2014; Damrauer *et al.*, 2014; Rebouissou *et al.*, 2014; Robertson *et al.*, 2017; Marzouka *et al.*, 2018; Mo *et al.*, 2018). The variability between these classification methods was addressed by Kamoun et al. 2020, developing a consensus classification system which reconciled the major classification approaches into six MIBC subtypes, luminal papillary (LumP), luminal non-specified (LumNS), luminal unstable (LumU), stroma-rich, basal/squamous (Ba/Sq) and neuroendocrine-like (NE) (Kamoun *et al.*, 2020). Efforts towards molecular subtyping of NMIBC were led primarily by the UROMOL consortium, with its initial work classifying NMIBC into three transcriptomic classes, which was later refined and expanded into four classes, 1, 2a, 2b and 3 (Hedegaard *et al.*, 2016; Lindskrog *et al.*, 2021).

Each subtype has distinct molecular, histological and disease progression profiles, though there are overarching similarities throughout BUC tumours. These similarities include *TERT*

promoter mutations that occur in 80% of patients, though mutations in *TP53* (48%), *FGFR3* (40%) and *PIK3CA* (22%) are also common (Allory *et al.*, 2014; Robertson *et al.*, 2017; van Rhijn *et al.*, 2020; Li, Liu, *et al.*, 2023). However, their prevalence can differ widely by clinical stage, with *FGFR3* and *PIK3CA* mutations being more prevalent in NMIBC, whereas *TP53* mutations are more common in MIBC (Wang *et al.*, 2022; Komura *et al.*, 2023). The molecular subtypes in both MIBC and NMIBC align along the luminal-basal axis, whereby luminal tumours expression signatures are similar to the intermediate and superficial urothelial layers and the basal tumours more similar to the basal urothelial layer (Hedegaard *et al.*, 2016; Kamoun *et al.*, 2020). Immune infiltration also differs between molecular subtypes, with Ba/Sq and Stroma-rich MIBC tumours, and NMIBC 2b class tumours, showing higher levels of tumour-infiltrating immune cells, potentially contributing to recurrence and survival rates (Lindskrog *et al.*, 2021; Benítez *et al.*, 2023). The stroma-rich MIBC subtype is particularly controversial, with it being debated whether it represents true tumour biology or sampling contamination from surgical resectioning procedures, though recent research supports its biological validity (Koll *et al.*, 2023; Contreras-Sanz *et al.*, 2025).

The Urothelial carcinoma: Stratified Treatment and Oncological outcomes (GUSTO) trial is currently investigating whether molecular subtype guided neoadjuvant therapy treatment improves MIBC patient outcomes, though no evidence currently exists to show that molecular subtypes provide clinical utility beyond the currently established parameters for treatment decisions (Kamoun *et al.*, 2020; Griffin *et al.*, 2024). Furthermore, the International Society of Urological Pathology concludes that current data are ‘Insufficient to support routine clinical use’ (Warrick *et al.*, 2024). This reveals that perhaps the clustering of groups by bulk tissue gene expression is flawed at the conceptual level, as increasingly sophisticated methodologies have been developed and used to subtype groups. The superiority of traditional clinical classifications of these tumours indicates that the clustering classification systems may be capturing histological features recognised by histopathologists rather than novel biological insights (Morera *et al.*, 2020). These failures to capture the biological insights at the gene expression level may indicate that more data would improve these classifications, such as alternative splicing patterns, see Sections 1.3 and 1.4, potentially providing more meaningful biological and clinical distinctions between urothelial carcinomas.

1.2.2 Upper tract urothelial carcinoma (UTUC)

Whilst BUC represents the majority of urothelial carcinomas, upper tract tumours present as their own clinical entity, with their own distinct molecular landscape, despite treatment approaches typically being adapted from BUC. Approximately 5-10% of urothelial carcinomas are upper tract urothelial carcinomas (UTUC), with an incidence rate of around 1-2/100,000 people per year (Benamran *et al.*, 2020). The male to female ratio of UTUC is 2:1, lower than the observed 4:1 ratio for BUC (Shariat *et al.*, 2011).

Previous studies have identified five distinct molecular subtypes of UTUC based on genetic alterations:*TP53*/*MDM2*-mutated, *RAS*-mutated, *FGFR3*-mutated, hypermutated, and triple-negative, with each subtype exhibiting distinct clinicopathological features (Fujii *et al.*, 2021; Sugawara, Koyama and Inamura, 2022). Lynch syndrome, seems to match the hypermutated UTUC subtype, caused by mutations in DNA mismatch repair (Fujii *et al.*, 2021). The molecular subtypes in UTUC seem to share a slight correspondence with the

MIBC subtypes, with the *FGFR3*-mutated subtype in UTUC overlapping with the MIBC LumP subtype and the *TP53*-mutated subtype corresponding to the more aggressive MIBC subtypes such as LumU and Ba/Sq (Fujii *et al.*, 2021; Meeks *et al.*, 2021). However, these differences in subtyping classification suggest that UTUC treatment may be improved by site specific therapeutic approaches rather than using direct extrapolation from BUC classifications.

1.2.3 Comparison of UTUC and BUC

A key difference between BUC and UTUC is their clinical presentation and aggressiveness. BUC is more commonly diagnosed at non-invasive stages, with 25% presenting at invasive stages, whereas UTUC is more commonly diagnosed at invasive stages, ~60%, due to the thinness of the ureteric wall (Witjes *et al.*, 2021; Grabe-Heyne *et al.*, 2023; Lefort *et al.*, 2023; Rouprêt *et al.*, 2023). This large difference makes UTUC inherently more aggressive than BUC, with the incidence of metastatic UTUC at the initial time of diagnosis ~15% compared to BUC's ~5% (Stecca, Abdeljalil and Sridhar, 2021; Hu *et al.*, 2022). However, this difference in invasiveness may reflect diagnostic challenges rather than inherent biological aggressiveness. UTUC can be difficult to diagnose in its early stages and the patient may have no symptoms, potentially leading to a detection bias of the disease in its later stages (Rouprêt *et al.*, 2018; Lin *et al.*, 2023). The survival outcomes seem to reflect this difference in stage at presentation, with BUC having a higher overall five-year survival rate at 71% compared to UTUC's 57%, and also in muscle invasive and non-muscle invasive stages (45% versus 36% and 90% versus 70% respectively) (van Doeveren *et al.*, 2021; Cancer Research UK, 2022; National Cancer Institute, 2023a; Ślusarczyk *et al.*, 2023).

Whilst both forms of cancer share common risk factors such as smoking, occupational chemical exposure and aristolochic acid (AA) exposure, they also have distinct profiles (Decarli *et al.*, 1985; Clavel *et al.*, 1994; Li *et al.*, 2020; Zhao, Wang and Liang, 2022; Xie *et al.*, 2024). Lynch syndrome, for example, has an up to 75-fold increased risk for UTUC compared with a 7.5-fold increased risk for bladder cancer (Skeldon *et al.*, 2013; Fujii *et al.*, 2021; Nassour *et al.*, 2023). The mutational signatures also differ between the cancers, with Apolipoprotein B mRNA editing catalytic polypeptide-like (APOBEC) being the most common signature in both cancers, but with differing rates, 94% and 84% respectively (Glaser *et al.*, 2018; He *et al.*, 2022).

Large genomic profiling studies have identified various differences in mutational frequencies between UTUC and BUC. The Necchi et al. study found that *TERT* promoter mutations were more frequent in BUC (68%) than in UTUC (47%), *TP53* mutations were also more common in BUC (58%) than UTUC (49%) whilst *CDKN2A* was more frequent in UTUC (40%) than in BUC (35%) (Necchi *et al.*, 2021). Other notable differences between UTUC and BUC were in *FGFR3* (21% versus 14%) and *HRAS* (7% versus 3%) (Lefort *et al.*, 2023).

However, these genomic comparisons must be interpreted with caution due to methodological limitations. The issue is that the inherent difference in stage distribution at presentation between anatomical sites creates a confounding factor which may obscure true biological differences when accounting for it. This stage bias is exemplified by *FGFR3* mutations, which are a characteristic of the UROMOL Class 1 subtype and are associated with 44% of NMIBC but only 15% of MIBC (Komura *et al.*, 2023). The reported higher

FGFR3 frequency in UTUC (21% versus 14%) appears paradoxical given UTUC's predominantly invasive presentation (Lefort *et al.*, 2023). These findings suggest that stage stratified analysis is required to determine if these are stage dependent enrichment patterns or whether they represent true biological differences. These clinical and genomic differences translate into distinct patterns of disease behaviour, including in how, when and where these cancers recur after treatment.

1.2.4 Recurrence

Recurrences can occur throughout the urinary tract, regardless of the original tumour location. Several mechanisms have been proposed to explain this phenomenon, with two main theories emerging as the most significant. The first is field cancerisation, where areas near a tissue which has previously had a tumour are found to harbour precancerous changes and genetic alterations, increasing the risk of developing additional or recurrent tumours (Strandgaard *et al.*, 2024). A second possible explanation for recurrence is the incomplete clearance of the tumour during surgical resection. Studies have demonstrated that incomplete resection is likely a contributing factor to early recurrences, where approximately 20% of patients are found to have tumours in their first follow-up cystoscopic evaluation (Jeong *et al.*, 2022; Teoh *et al.*, 2022). These two mechanisms may operate on different time scales, with incomplete resection typically manifesting as early recurrence, whereas field cancerisation effects tend to produce later recurrences (Bryan *et al.*, 2010).

Recurrence patterns differ between UTUC and BUC. The five year recurrence rate for UTUC is 2 to 13%, whilst the risk of developing BUC after UTUC is higher, at 29% (Ishioka *et al.*, 2015; Yamashita *et al.*, 2024). In comparison the risk of UTUC after BUC is 2-6%, lower than BUC after UTUC, explained by the ureteral meatus' anti-reflux system which may prevent the dissemination of cancer cells from the bladder (Sanderson *et al.*, 2007; Lefort *et al.*, 2023). For BUC specifically the overall 5 year rate of recurrence for NMIBC is 62%, compared to 45% for MIBC (Jobczyk *et al.*, 2020; Squires *et al.*, 2025).

These differences in recurrence patterns in BUC and UTUC have led to the development of site-specific treatment strategies, moving away from the traditional one-way approach for urothelial carcinomas.

1.2.5 Standard of care and treatment

Historically UBC and UTUC had the same clinical recommendation for treatment, until 2011, where UTUC was long treated as a common carcinoma at a rare location where treatment decisions were made by adopting the UBC clinical recommendations (Rouprêt *et al.*, 2015; Szarvas *et al.*, 2016).

The current recommendations for bladder cancer from the European Association of Urology for diagnosis and treatment are summarised in the paragraphs below (Babjuk *et al.*, 2019; Masson-Lecomte *et al.*, 2025; van der Heijden *et al.*, 2025): Cystoscopies are used as the primary diagnostic tool, this is where a cystoscope, a thin tube with a camera, is inserted into the urethra and bladder to monitor for tumours. A transurethral resection of bladder tumour (TURBT) may also be performed, this is where all visible lesions are removed. Magnetic resonance imaging (MRI) can be used to differentiate between T1 and T2 tumours, thereby

being able to differentiate between MIBC and NMIBC. Computed tomography (CT) urography is used for staging in patients with confirmed MIBC, where the chest, pelvis and abdomen are scanned. For non-metastatic MIBC radical cystectomy, where the bladder is removed, can be performed. Neoadjuvant chemotherapy can be performed before surgery to shrink tumours. Post surgery nivolumab, a checkpoint inhibitor, can be used to improve patient outcomes (Bajorin *et al.*, 2021). For metastatic BUC the first line of treatment should be the use of enfortumab vedotin (EV) and the checkpoint inhibitor pembrolizumab (Powles *et al.*, 2024). If patients are ineligible for this, traditional platinum-based chemotherapy should be used followed by immunotherapy which can include the checkpoint inhibitor avelumab (Powles *et al.*, 2020). Post treatment patients should be monitored for recurrence.

For NMIBC treatment is stratified into different at-risk groups, with low, intermediate and high groups, which depend on tumour size, tumour amount and prior occurrence rates. Low risk patients should receive one course of chemotherapy immediately. Patients in the intermediate risk group should receive one year of bacillus Calmette-Guerin (BCG) intravesical immunotherapy or various courses of chemotherapy for one year. Patients in the high risk group should receive BCG immunotherapy for one to three years. BCG is a live vaccine used to stimulate the immune system, and is often used after TURBT therapy. In this high risk group an immediate radical cystectomy may be considered, this is where the whole bladder and nearby lymph nodes are removed, and is recommended in BCG non-responders. For MIBC patients neoadjuvant cisplatin-based chemotherapy is offered, followed by a radical cystectomy.

UTUC diagnosis and treatment has to deal with several challenges in comparison with BUC. Tumour tissue obstruction can occur, in 20 to 32% of cases. CT urography should be used due to its high sensitivity of 92%. A ureteroscopy is used if necessary to confirm the diagnosis in ambiguous cases, and can include taking biopsies of suspicious lesions. For the low risk UTUC, kidney sparing treatments are offered as the primary treatment option, with endoscopic management of tumours and a distal ureterectomy, where part of the ureter is removed. For high risk and recurrent low risk UTUC, a radical nephroureterectomy is performed, where the kidney, ureter and part of the bladder are removed. Platinum-based adjuvant chemotherapy should also be offered. Similar to MIBC recommendations, the first line of treatment recommended is the use of EV and pembrolizumab (Powles *et al.*, 2024). Avelumab maintenance therapy is recommended to patients with stable disease after platinum based chemotherapy.

However, despite these treatment recommendations, challenges remain in achieving optimal patient outcomes. BCG therapy, whilst effective, faces ~40% failure rates (Zlotta, Fleshner and Jewett, 2009; Maroof, Paramore and Ali, 2024). Similarly, cisplatin-based therapies show heterogeneous patient responses (Benkhadra *et al.*, 2022). This exemplifies that current risk stratification systems which inform therapies offered to patients, which rely on clinicopathological observations, such as tumour grade and stage, do not explain the differences in treatment outcomes.

Given that morphologically similar tumours can have different treatment responses and survival rates, molecular differences likely exist which we are not able to explain on the gene level alone. A deeper understanding of transcript isoforms and alternative splicing is

therefore needed to better understand what specific transcript variants are driving tissue specific cancer behaviours.

1.3 Gene Expression Profiling

Tissues have different molecular profiles, due to differences in their function (Wu *et al.*, 2024). This is especially useful for the investigation of tissues with similar morphological characteristics, as they may have molecular differences.

Molecular profiling enables clinicians to better understand and differentiate between types of cancers, allowing for more personalised and therefore effective treatments than could be achieved previously (Malone *et al.*, 2020; Scimeca *et al.*, 2024). This has been done in both UTUC and BUC, with various subtypes identified with different pathological presentations and potential therapies, as described in Sections 1.2.1 and 1.2.2 (Sugawara, Koyama and Inamura, 2022; Koll *et al.*, 2024). Yet clinical utility of molecular subtyping varies across studies and no significant differences in survival have been found between subtypes (Hardy *et al.*, 2022). These limitations may reflect the focus on gene level expression data, which overlooks the complexities introduced by alternative splicing which could provide more clinically relevant information on molecular characteristics using transcript level expression data.

1.3.1 High throughput technologies

RNA sequencing (RNA-seq) allows for a genome-wide view of gene expression by sequencing all expressed RNA molecules at a certain point in time, including messenger RNA (mRNA), non-coding RNAs and transcript isoforms (Mortazavi *et al.*, 2008; Wang, Gerstein and Snyder, 2009). RNA-seq involves extracting RNA from a biological sample, converting it into complementary DNA (cDNA) through reverse transcriptase and sequencing this cDNA to characterise and quantify the transcriptome (Nagalakshmi, Waern and Snyder, 2010). RNA-seq allows for the identification of novel transcripts, splice variants and the quantification of transcripts (Halperin *et al.*, 2021). However, the choice of sequencing technology changes the type of biological information that can be obtained from transcriptomic data. Advances in sequencing technologies have made it possible to do molecular profiling at larger scales and reduced costs.

The technology used for sequencing impacts the information that can be obtained, including in RNA-seq (Sharon *et al.*, 2013). Prior to the development of RNA-seq technologies, microarrays were used for gene expression profiling and have been used in several established molecular classification systems, including in the early versions of the Lund and CIT cohorts, though these have now incorporated RNA-seq data too (Sjödahl *et al.*, 2012; Biton *et al.*, 2014; Kamoun *et al.*, 2020; Cotillas *et al.*, 2024). Short read technologies, such as Illumina sequencing, offer high accuracy and typically generate read lengths of 75 to 300 base pairs (bp), this makes them useful for transcript quantification and differential expression analysis across different conditions (Meirelles *et al.*, 2024).

Long read technologies, such as PacBio and Oxford Nanopore (ONT), have a lower accuracy than short read, though this is improving considerably, with accuracies now reaching over 99%, and read lengths from 300 to over 100,000 bp (Espinosa *et al.*, 2024).

These longer lengths of reads enable structural variants and splice junctions to be captured in one read, which mean that they are particularly useful in capturing novel splice variants and transcript isoforms which would be harder to resolve or missed entirely in short read (Ahsan *et al.*, 2023). This ability of long-read sequencing to capture the full length of transcripts makes them well suited for studying alternative splicing.

A comprehensive genomic analysis of bladder cancers demonstrated the clinical significance of alternative splicing in urothelial carcinoma, which was able to associate these events with survival (He *et al.*, 2018). However, the He et al. study analysed only bladder urothelial carcinomas, and did not examine urothelial carcinomas at other sites like upper tract urothelial carcinomas, and used short read RNA sequencing data to infer splicing events, which may miss novel isoforms and complex splicing patterns only detectable through long read approaches.

1.4 Alternative Splicing

Alternative splicing of protein-coding messenger RNAs is an essential regulatory mechanism in eukaryotic gene expression, which helps to diversify protein functions (Birzele, Csaba and Zimmer, 2008). Alternative splicing also occurs on non-coding RNAs (Sekulovski *et al.*, 2021; Marasco and Kornblihtt, 2023). Around 95% of human genes undergo alternative splicing and do so through seven different ways, exon skipping (also known as cassette exons), mutually exclusive exons, alternative donor (5′ splice) sites, alternative acceptor (3′ splice) sites, intron retention, alternative promoters and alternative polyadenylation sites, Figure 1.7 (Blencowe, 2006; Pan *et al.*, 2008).

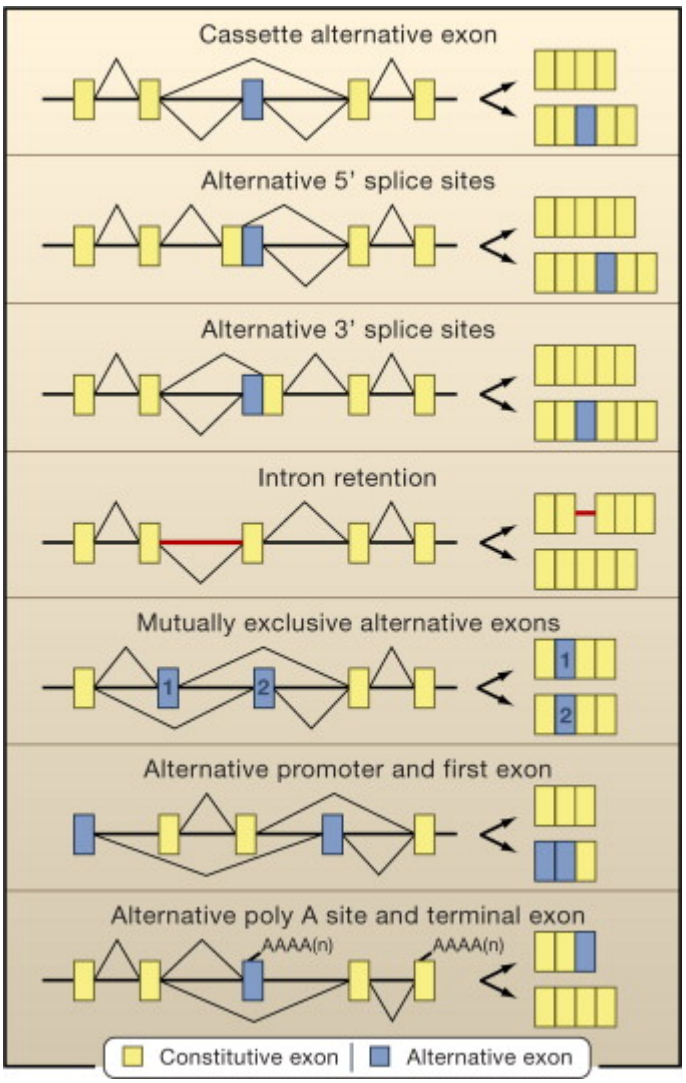

Figure 1.7: The seven different types of splicing in metazoans. A diagram showing the seven types of splicing event which can occur in metazoans. The depicted splicing events are cassette alternative exon skipping or inclusion, alternative 5' or 3' splice sites, intron retention, mutually exclusive exons, alternative promoters, and alternative polyadenylation (polyA) sites. Yellow boxes show constitutive exons and blue boxes show alternative exons. Sourced from (Blencowe, 2006).

Alternative splicing dysregulation is a hallmark for cancer, characterised by the production of splicing isoforms that disrupt proliferative and apoptotic processes (Climente-González *et al.*, 2017). These splicing alterations drive cancer progression by directly impacting cell cycle control, invasion and metastatic potential (Grelet *et al.*, 2017; Liu *et al.*, 2018). Furthermore, splicing-factor alterations have different profiles in different cancers (Urbanski, Leclair and Anczuków, 2018). Advances in long-read RNA sequencing have been able to show novel splice variants of oncogenes in cancers, such as in renal cell carcinoma (Lee *et al.*, 2024).

Despite these advances in splicing characterisation, studies have focused primarily on individual anatomical sites, leaving tissue specific alternative splicing differences between the upper tract and the bladder largely unexplored. Whilst short-read RNA sequencing has identified differentially expressed genes in bladder cancer, comprehensive isoform analysis

using long-read technologies has not been used to identify and understand differences between UTUC and BUC, despite their distinct clinical behaviours.

Previous transcriptomic studies have primarily focused on gene expression differences, many using short read technologies which cannot fully capture alternative splicing events. This represents a current gap in the molecular understanding of urothelial carcinomas and their heterogeneity, as alternative splicing is an important mechanism which influences development and cancer progression.

1.5 Project Rationale

Despite the morphological similarities and shared classification as urothelium, the tissue in the bladder and ureter give rise to carcinomas with different clinical behaviours, including differences in incidence rates, disease aggressiveness and patient outcomes (Lefort *et al.*, 2023). These findings suggest that our understanding of the urothelium in these tissues, and treating them as functionally equivalent, may be incomplete.

The differences in these tissues are likely due to various factors, including functional, mechanical and diagnostic differences. Whilst the urothelium lines both the bladder and the upper urinary tract, UTUC and BUC behave like two distinct diseases (Szarvas *et al.*, 2016). The physical structure of the upper tract compared to the bladder alters how tumours grow, invade and are treated. The bladder has a thick and multi-layered smooth muscle wall, the detrusor, as explained in Section 1.1, whereas the ureter and renal pelvis have a thin muscularis layer. Therefore, UTUC is able to invade local tissues and metastasise much faster than BUC (O'Meara *et al.*, 2024).

The physical location of the organs creates a disparity in how accurately diseases can be diagnosed and staged. UTUC is often under-staged and under-diagnosed compared to BUC due to a combination of reasons, including anatomical inaccessibility and instrument limitations (Mori *et al.*, 2022; Baard *et al.*, 2024). In bladder cancer, surgeons can pass a rigid instrument directly through the urethra and use a wire loop to cut deep into the tumour to the muscle layer underneath. Whether the cancer has invaded that muscle is critical for staging. The upper urinary tract is less accessible (Kolawa, D'Souza and Tulpule, 2023). The current standard for sampling UTUC is a ureteroscopic biopsy, where a flexible camera is threaded through the urethra and up into the ureter. The instruments that fit inside are much smaller and pinch off shallow surface samples as going deeper risks puncturing the ureter wall, which can cause serious complications including the spreading of cancer cells (Huynh *et al.*, 2025).

How the organs handle urine and toxins changes the tumour landscape. The longer exposure to urine in the bladder favours BUC, whilst the concentration of specific filtered toxins in the renal pelvis favours UTUC (Chen *et al.*, 2012; Jubber *et al.*, 2023). The bladder's function is storage, meaning that carcinogens can sit in the bladder for hours, exposing the lower tract (Gaffney *et al.*, 2023; Pan *et al.*, 2025). The upper tract's function is transit and therefore has less exposure to urine and the toxins it contains (Giudici *et al.*, 2021).

The distinct embryological origins of these tissues, as described in Section 1.1.1, may persist into adulthood (Staack *et al.*, 2005). Whilst the urothelium in these tissues appears morphologically similar in histological examination, molecular analysis reveals distinct immunohistochemical differences between them, reinforcing that their respective cancers are distinct (Wiessner *et al.*, 2022). It is apparent that UTUC presents more aggressively, with higher rates of muscle invasion at diagnosis and poorer survival outcomes than BUC (Fiuk and Schwartz, 2016). It seems implausible that such differences arise from molecularly identical tissues.

Previous studies have identified some transcriptomic differences between UTUC and BUC, yet they have not fully explained the differences in clinical behaviour (Sfakianos *et al.*, 2021; Li *et al.*, 2024). Crucially, these studies have relied on short-read sequencing technologies which are able to capture gene expression levels well, but miss the complexity of transcript isoforms. Alternative splicing, whereby a gene is able to produce multiple protein variants, remains an underexplored regulatory mechanism that could explain the molecular basis of tissue specific differences. Notably, retinoic acid signalling, as described in Section 1.1.1 as a master regulator of development in the urinary tract, also directly regulates the splicing machinery (Wang, Soprano and Soprano, 2015).

Given that alternative splicing drives both tissue differentiation in development and cancer progression, and that long-read sequencing enables comprehensive isoform characterisation, it is hypothesised that differences in alternative splicing between the bladder and ureter urothelia contribute to their distinct clinical behaviours of their respective carcinomas.

1.6 Aims

To test this hypothesis, this study sets out to analyse RNA sequencing data from normal and cancerous tissue samples from both bladder and ureter tissue. The specific objectives of this research are to determine the molecular basis for the distinct clinical behaviours observed between these two morphologically similar tissues.

This investigation has three primary aims:

i) Characterise the transcriptomic differences between the normal bladder and ureter urothelium, in terms of their alternative splicing patterns and gene expression profiles.

ii) Investigate the key differences between bladder and ureter urothelial carcinomas, with respect to their molecular signatures.

iii) Assess the retention of tissue specific identities in malignancy by comparing normal and cancerous samples, and evaluate whether developmental and transcriptomic signatures play a role in site specific tumour biology.

This study aims to explore whether the contribution of alternative splicing and transcriptomic variation underlies the clinical differences between bladder and ureter urothelial carcinomas. Clarification of the role of transcript usage and alternative splicing, including its influence on

cellular differentiation and tumour progression pathways in urothelial carcinomas, may contribute to more informed clinical decisions leading to improved patient outcomes.

Chapter 2: Methods

A comprehensive bioinformatics pipeline was developed to address three research aims investigating transcriptomic differences in urothelial carcinoma, Figure 2.1. The analysis used parallel processing of short-read and long-read RNA-sequencing data through modular subworkflows, consisting of five core analytical stages, quality control and preprocessing, read alignment and transcript quantification, differential gene expression analysis, biological pathway analysis, and isoform switch analysis. For Aim 2, the cancer comparison, an additional molecular subtyping analysis was performed specifically for bladder cancer samples using consensus classifier methods. The pipeline was implemented in Nextflow (version 23.10.0) to ensure reproducibility, scalability, and parallel execution across computational resources (Di Tommaso *et al.*, 2017). This project was undertaken on the Viking Cluster, which is a high performance computing facility provided by the University of York.

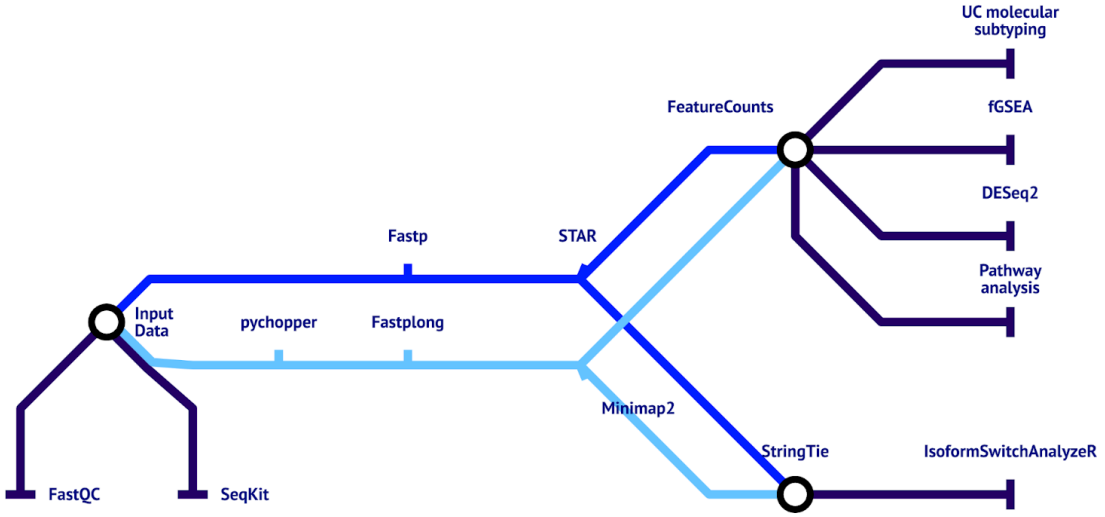

Figure 2.1: Schematic overview of the bioinformatics pipeline using short-read and long-read input data. The workflow is designed as a parallel processing network. This graph illustrates the different analytical paths for short-read (Illumina, royal blue track) and long-read (Oxford Nanopore, light blue track) RNA-sequencing data. The purple tracks represent modules which both types of sequencing go through. Raw sequencing reads enter the pipeline and undergo statistical tests using FastQC and SeqKit. Short reads are processed using Fastp, whilst long reads undergo full length cDNA identification via Pychopper and filtering via Fastplong. Reads are aligned to the GRCh38 human reference genome using STAR (short-read) and Minimap2 (long-read). Post-alignment, the workflow diverges into two quantification streams. Gene level counts are generated using featureCounts, whilst transcript level abundances are assembled and quantified using StringTie to capture isoform diversity. The quantified data feed into four different analyses, gene level counts are normalised and modelled using DESeq2 to identify significant changes between conditions. Gene counts also undergo Gene Set Enrichment Analysis (GSEA) and specific urothelial signalling pathway scoring. Transcript level data from StringTie are processed by IsoformSwitchAnalyzeR to detect functional isoform switching. A bladder

cancer specific branch classifies samples into consensus molecular subtypes on gene counts. The entire pipeline is implemented in Nextflow (v23.10.0) to facilitate reproducible and parallel execution.

2.1 Ethical Approval and Patient Consent

Ethics approval was obtained from the University of York Biology Ethics Committee (BEC reference AM202410) for the project titled "Reuse and reapplication of existing human cancer sequencing datasets". Human ureter specimens were collected under NHS Research Ethics Committee approval Leeds (East) REC 99/095, for the anonymous use of excess tissue following renal transplant surgery. Bladder samples, originating from rigid biopsies collected during investigations of uterine prolapse, were accessed from URoBank (NHS REC approved research tissue bank 16/YH/0396) containing anonymous donor specimens collected with informed patient consent. Local institutional research approval for these activities was granted by the Department of Biology Ethics Committee.

Relevant urothelial carcinoma samples had been identified within the Leeds Multidisciplinary Research Tissue Bank (LMRTB; University of Leeds, UK; REC 25/YH/0022; IRAS 352968). Fresh frozen tumour and adjacent normal samples had been collected through the Leeds Teaching Hospitals Trust with full informed consent after ethical review, and stored with pseudo-anonymised project IDs at the LMRTB with persistent consent for use in research, including sequencing. Tissues were transferred to the York Tissue Bank under HTA regulations and after favourable ethical review by The University of York Biology Ethics Committee (BEC-AM202504).

2.2 Data Acquisition and Sample Information

This study used RNA-sequencing data from two distinct cohorts: normal urothelial tissue samples (n=6) and urothelial carcinoma samples (n=7). Both short-read (Illumina) and long-read (Oxford Nanopore Technologies) sequencing platforms were used for transcriptomic analysis.

Normal tissue samples were obtained from six patients undergoing surgical procedures at York Teaching Hospital NHS Foundation Trust and St James's University Hospital, Leeds. The cohort comprised three bladder specimens from female patients aged 54-77 years undergoing cystoscopy, and three ureter specimens from living kidney transplant donors. All samples were processed within 24-48 hours of retrieval. Urothelial epithelium was enzymatically dissociated from underlying stromal tissue using established protocols, and total RNA was extracted using TRIzol reagent (Southgate *et al.*, 1994). RNA integrity numbers ranged from 7.1 to 9.8 (median: 9.4), indicating high quality RNA suitable for sequencing.

Tumour tissue samples were obtained from seven patients who underwent surgical resection for urothelial carcinoma, made up of three bladder cancer specimens and four upper tract specimens (three renal pelvis, one ureter). All samples were high-grade (G2-G3) tumours spanning pathological stages Ta to T4. Fresh frozen tissues, obtained from the Leeds Multidisciplinary Research Tissue Bank, were cryosectioned and the RNA extracted using

standard TRIzol protocols. RNA quality assessment revealed substantial degradation with RIN values ranging from 4.4 to 6.8 (median: 4.9). Despite falling below optimal thresholds, samples with RIN ≥4.0 were selected for ONT long-read sequencing using standard sample protocols.

2.3 Reference Genome and Annotation Files

The human reference genome GRCh38 (primary assembly) was obtained from ENSEMBL. Gene annotations were derived from GENCODE version 47 (gencode.v47.chr_patch_hapl_scaff.annotation.gtf). Gene set definitions for pathway analysis were obtained from the Molecular Signatures Database (MSigDB), Hallmark gene sets (v2023.2) and Reactome pathways (v2024.1). Functional annotation resources for isoform analysis included CPAT hexamer frequency tables and logistic regression models for human coding potential assessment, and the Pfam-A HMM database for protein domain prediction.

2.4 Short-read Analysis

2.4.1 Short-read preprocessing and quality control

Short-read data underwent quality assessment, adapter trimming, and filtering to prepare sequences for alignment. FastQC (v0.12.1) and SeqKit (v2.9.0) evaluated raw and processed data quality, whilst Fastp (v0.24.0) performed adapter trimming with automatic adapter detection (--detect_adapter_for_pe), quality filtering (minimum Phred score Q20, -q 20), polyX tail trimming (--trim_poly_x), and minimum read length filtering (50 bp, --length_required 50) (Andrews, 2010; Chen *et al.*, 2018; Shen, Sipos and Zhao, 2024). Input files were raw paired-end FASTQ files and the output included trimmed FASTQ files and quality control reports. This approach was used for computational efficiency and ability to process paired-end data whilst maintaining read pairing integrity (Chen *et al.*, 2018).

2.4.2 Short-read alignment and quantification

Trimmed short reads were aligned to the reference genome and quantified at the gene and transcript levels. STAR (v2.7.11b) performed splice-aware alignment with optimised splice junction detection (--sjdbOverhang 150), featureCounts (Subread v2.0.6) quantified gene level expression in paired-end mode (-p, -t exon, -g gene_name), and StringTie (v2.2.1) performed transcript level quantification through a three-stage workflow, individual assembly (-G -e -B), merging (--merge), and re-quantification with prepDE.py matrix generation (Dobin *et al.*, 2013; Liao, Smyth and Shi, 2014; Pertea *et al.*, 2015). Input files were the trimmed FASTQ files and GRCh38 and GENCODE v47 annotation reference. The output files included sorted BAM files and gene and transcript count matrices. STAR was used for splice junction accuracy, featureCounts for computational efficiency and DESeq2 compatibility, and StringTie for reference-guided assembly balancing isoform sensitivity with false discovery control (Dobin *et al.*, 2013; Liao, Smyth and Shi, 2014, 2019; Pertea *et al.*, 2015).

2.4.3 Short-read differential expression analysis

Differential expression analysis identified genes with statistically significant expression differences between bladder and ureter tissues. DESeq2 (v1.42.0) performed negative binomial generalised linear modelling with low-count filtering (minimum 10 reads), variance-stabilising transformation for PCA, and Benjamini-Hochberg adjusted *p*-values (Love, Huber and Anders, 2014). EnhancedVolcano (v1.24.0) generated graphs and fgsea (v1.28.0) performed gene set enrichment analysis (GSEA) against Hallmark MSigDB and Reactome databases (Blighe, Rana and Lewis, 2023). Input files were the featureCounts gene level count matrices. The output included differential expression results, volcano plots, PCA plots with eigenvector analysis, and GSEA results. This integrated approach was selected because DESeq2 appropriately models count data with small sample sizes, PCA reveals major sources of variation, and GSEA detects coordinated pathway level changes beyond individual gene lists (Love, Huber and Anders, 2014).

2.4.4 Short-read biological pathway expression analysis

Biological pathway analysis quantified expression of genes involved in urothelial biology, developmental patterning, and key signalling cascades. R scripts (R v4.3.1) with ggplot2 (v3.4.2), dplyr (v1.1.2), and tidyr (v1.3.0) calculated mean expression across curated gene sets including HOX genes, luminal/basal markers, urothelial differentiation markers, and signalling pathways (Sonic Hedgehog, FGF, Wnt/beta-catenin, TGF-beta, interferon response, immune cell markers, proliferation markers) (Wickham, 2016; Wickham *et al.*, 2025; Wickham, Vaughan and Girlich, 2025). Input files were TPM normalised count matrices and the output files included expression plots and pathway activity scores. This approach was used to provide comprehensive functional characterisation of transcriptomic differences beyond individual genes, using curated gene sets representing biological processes relevant to urothelial tissue identity and cancer development. Statistical comparisons of pathway activity scores were performed using Welch's two sample t-test (Fagerland, Sandvik and Mowinckel, 2011; Ullah *et al.*, 2019). Whilst non-parametric tests are often applied to transcriptomic data such as the Wilcoxon rank-sum test, they lack the ability to achieve a significance threshold of $p < 0.05$ at the cohort sizes used in this study ($n=3$ and $n=4$) (de Winter, 2013). The use of parametric testing was therefore justified as pathway activity scores represent the mean expression of large gene sets which approximate a normal distribution (Mar, 2019). This approach maintains the necessary statistical power to detect pathway level differences even if the underlying individual gene counts are skewed.

2.4.5 Short-read isoform switch analysis

Isoform switch analysis identified genes where the dominant transcript isoform changes between conditions and predicted functional consequences. IsoformSwitchAnalyzeR (v2.6.0) performed DEXSeq-based differential transcript usage testing (FDR < 0.05, minimum 10% isoform fraction change), ORF prediction (analyzeORF, minimum 100 nucleotides), coding potential assessment (CPAT v3.0.4, cutoff 0.725), protein domain annotation (Pfam via HMMER v3.3.2, E-value < 0.001), and functional consequence prediction evaluating intron retention, coding potential changes, ORF similarity (Jaccard < 0.9), nonsense-mediated decay sensitivity, and domain gains/losses (Wang *et al.*, 2013; Vitting-Seerup and Sandelin, 2019; Paysan-Lafosse *et al.*, 2025). The input files were the StringTie transcript level TPM

counts, merged GTF annotation, and GRCh38 reference genome. The output was made up of the significant isoform switches with predicted functional consequences. This approach was used because isoform switches can alter protein function through domain changes, coding potential alterations, or post-transcriptional regulation, and IsoformSwitchAnalyzeR was specifically designed to predict biological impact beyond detecting switches (Vitting-Seerup and Sandelin, 2019).

2.5 Long-read Analysis

2.5.1 Long-read preprocessing

Long-read data underwent quality assessment, full length read identification, and adapter trimming to prepare sequences for isoform analysis. FastQC (v0.12.1) and SeqKit (v2.9.0) evaluated raw and processed data quality, Pychopper (v2.7.10) identified full length cDNA reads based on primer presence and orientation and Fastplong (v0.2.2) performed quality filtering (minimum Q10, 40% threshold) and polyX tail trimming (Andrews, 2010; Chen, 2024; epi2me-labs, 2024; Shen, Sipos and Zhao, 2024). The input files were the raw Oxford Nanopore FASTQ files and the output included processed full length reads passing quality filters and QC reports. This approach was used because accurate isoform analysis requires high confidence full length transcripts and the use of Pychopper and Fastplong addresses the need for full length transcripts and high quality data (Chen, 2024; epi2me-labs, 2024).

2.5.2 Long-read alignment and quantification

Processed long-reads were aligned to the reference genome and quantified at the gene and transcript levels. Minimap2 (FLAIR v2.0.0) performed splice-aware alignment with cDNA-specific parameters (-ax splice -uf --secondary=no), featureCounts (Subread v2.0.6) quantified gene level expression in long-read mode (-L --largestOverlap -s 0), and StringTie (v2.2.1) performed transcript level quantification in long-read mode (-L) through a three-stage workflow, individual assembly (-G), merging, and re-quantification (-e -B) with prepDE.py3 matrix generation (Liao, Smyth and Shi, 2014; Pertea *et al.*, 2015; Li, 2018). The input files were the quality filtered full length reads and GRCh38 and GENCODE v47 annotation references. The output included sorted BAM files and gene/transcript count matrices. This approach was used because long-read RNA-seq allows us to look at full length transcripts and combining featureCounts for robust gene level quantification with StringTie's transcript assembly capabilities allows for both gene level reliability for differential expression and isoform resolution precision for detecting alternative splicing (Ament *et al.*, 2025).

2.5.3 Long-read differential expression analysis

Differential expression analysis followed the methodology described in Section 2.4.3, using DESeq2 (v1.42.0), EnhancedVolcano (v1.24.0), and fgsea (v1.28.0) with identical parameters (Love, Huber and Anders, 2014; Blighe, Rana and Lewis, 2023). The input comprised featureCounts gene level count matrices from long-read quantification and the output included differential expression results, volcano plots, PCA plots with eigenvector analysis, and GSEA results.

2.5.4 Long-read biological pathway expression analysis

Biological pathway analysis followed the approach described in Section 2.4.4, using identical gene sets and analytical methods applied to long-read derived TPM normalised count matrices.

2.5.5 Long-read isoform switch analysis

Isoform switch analysis identified genes where the dominant transcript isoform changes between conditions and predicted functional consequences, following the methodology described in Section 2.4.5. IsoformSwitchAnalyzeR (v2.6.0) was applied with identical parameters and thresholds to long-read derived transcript quantifications (Wang *et al.*, 2013; Vitting-Seerup and Sandelin, 2019; Paysan-Lafosse *et al.*, 2025). The input files were the StringTie transcript level TPM counts from long-read quantification, the merged GTF annotation, and the GRCh38 reference genome. The output included significant isoform switches with predicted functional consequences.

2.5.6 Urothelial carcinoma molecular subtyping

Molecular subtyping classified bladder cancer samples into consensus molecular subtypes. The consensusMIBC classifier assigned samples using the getConsensusClass() function with log2 transformed TPM normalised expression data (Kamoun *et al.*, 2020). It categorised samples into six subtype classifications, Luminal Papillary (LumP), Luminal Unstable (LumU), Luminal Non-Specified (LumNS), Basal/Squamous (Ba/Sq), Stroma-rich, and Neuroendocrine-like (NE-like), as well as separation levels quantifying the confidence of classification. The input consisted of the TPM normalised counts from bladder cancer samples. The output included subtype assignments and separation levels. This approach was selected because consensus classification provides robust subtype assignments by integrating multiple established systems, whilst separation levels enable the identification of ambiguous classifications.

2.6 Code Availability

All bioinformatics analyses were implemented as a modular Nextflow pipeline (version 23.10.0) using DSL2 syntax. The complete pipeline source code, including all subworkflows, process modules, and configuration files, is available at https://github.com/hugoaguadeau/urothelial_pipeline.

Chapter 3: Transcriptomic Characterisation of Normal Bladder and Ureter Urothelium

3.1 Introduction

This chapter addresses the first research aim, to characterise the molecular differences between normal bladder and ureter urothelium at both gene and transcript levels. Whilst these tissues appear morphologically similar and share a common urothelial identity, as discussed in Section 1.1.1, their distinct embryological origins may result in molecular differences maintained into adulthood. Understanding these baseline differences is necessary to contextualise the distinct clinical behaviours observed in bladder urothelial carcinoma (BUC) and upper tract urothelial carcinoma (UTUC).

To achieve this aim short-read Illumina and long-read Oxford Nanopore Technologies (ONT) sequencing were performed on six normal urothelial samples, three bladder samples and three ureter samples. Multiple analytical approaches were used to characterise tissue specific differences, using differential expression analysis, principal component analysis, gene set enrichment analysis (GSEA), transcript level isoform switch analysis, and targeted pathway analyses. This analysis revealed substantial transcriptomic divergence between these anatomically adjacent tissues, with 3,698 differentially expressed genes and pathway level differences showing that there is transcriptomic divergence between morphologically similar tissues.

3.2 Sample Characteristics

Normal urothelial tissue samples were obtained from six patients undergoing procedures at two centres, the York Teaching Hospital NHS Foundation Trust in York and Saint James's University Hospital in Leeds. The cohort was composed of three bladder specimens and three ureter specimens, Table 3.1. All samples were collected from patients without malignancy and were obtained as surgical surplus tissue.

Bladder samples were collected from female patients aged 54-77 years undergoing investigative cystoscopy following uterine prolapse, where the uterus slips down into or protrudes out of the vagina. Ureter samples were obtained from kidney transplant donors, two female patients aged 43 and 49 years and one male patient whose age was not recorded.

All the samples were processed within 24 to 48 hours of retrieval. Bladder specimens were stripped at 37 degrees celsius for approximately one and a half hours prior to P0 sequencing (Cross *et al.*, 2005). Ureter samples were stripped at either 37 degrees celsius for four hours (n = 2) or 4 degrees celsius overnight (n = 1).

Sample ID	Tissue	Age	Sex	Operation Type	Clinical Context	RIN
Y2631	Bladder	67	F	Rigid Cystoscopy	Uterine Prolapse	9.1
Y2632	Bladder	54	F	Rigid Cystoscopy	Uterine Prolapse	9.8
Y2796	Bladder	77	F	Rigid Cystoscopy	Uterine Prolapse	7.1
Y2391	Ureter	49	F	Kidney Transplant	Kidney Donor	9.3
Y2392	Ureter	43	F	Kidney Transplant	Kidney Donor	9.5
Y2396	Ureter	N/A	M	Kidney Transplant	Kidney Donor	9.5

Table 3.1: Clinical and processing characteristics of normal urothelial tissue samples. Clinical and demographic data for six normal urothelial samples. Bladder samples (n= 3) were from female patients undergoing cystoscopy. Ureter samples (n= 3) were from kidney transplant donors. Sample Y2796 tested positive for bacterial contamination via lysogeny broth and had a reduced RNA integrity compared to other samples. RIN, RNA integrity number. F, female. M, male. N/A, not available.

Urothelium from bladder and ureteric samples was dissociated from the underlying stroma (Southgate *et al.*, 1994). Urothelial cells were collected directly into TRIzol (Invitrogen Europe Ltd.) and RNA was extracted as per the manufacturer's instructions. RNA concentration and integrity were assessed using NanoDrop, Qubit and TapeStation before submission for Illumina short read (Oxford Genomics Centre, UK) or Oxford Nanopore Technologies PromethION (Bioscience Technology Facility, The University of York, UK) long read sequencing.

3.3 Short-read Analysis

3.3.1 Short-read sequencing quality and alignment

Short-read Illumina sequencing of six normal urothelial samples yielded 186.4 million paired-end reads across 40.74 Gb, with comparable sequencing depth between tissues (ureter: 34.0 million reads, bladder: 28.1 million reads, $p = 0.16$, Welch’s two sample t-test, n=3 bladder, n=3 ureter), Table 3.2. Mean Q30 scores exceeded 90% across all samples, with characteristic forward-to-reverse read quality degradation (R1: 94.6%, R2: 91.9%). All samples passed core FastQC metrics, with warnings for per-base sequence content and duplication levels.

Sample	Tissue	Read Pairs (million)	Total Bases (Gb)	Read Length (bp)	Mean Q30 (%)	Mean GC (%)
Y2391	Ureter	36.5	5.48	75	93.9	49.0
Y2392	Ureter	31.7	4.76	75	94.0	48.5
Y2396	Ureter	33.7	5.06	75	93.7	49.0
Ureter mean	**-**	**34.0**	**5.10**	**75**	**93.9**	**48.8**
Y2631	Bladder	25.8	7.79	151	92.0	51.0
Y2632	Bladder	24.9	7.52	151	91.7	50.2
Y2796	Bladder	33.8	10.13	150	94.0	51.2
Bladder mean	**-**	**28.1**	**8.50**	**151**	**92.6**	**50.7**
Overall	**-**	**186.4**	**40.74**	**-**	**93.2**	**49.8**

Table 3.2: Summary of raw Illumina short-read sequencing data quality metrics for normal bladder and ureter urothelial samples. Raw sequencing statistics generated by SeqKit (version 2.9.0) analysis of paired-end Illumina sequencing data from six normal urothelial tissue samples (three ureter, three bladder). Read pairs indicates the number of paired-end reads sequenced (in millions). Total bases represents the total number of bases sequenced in gigabases (Gb). Read length shows the sequencing read length in base pairs (bp). Mean Q30 represents the mean percentage of bases with Phred quality scores ≥30 (corresponding to 99.9% base calling accuracy) averaged across R1 and R2 reads. Mean GC indicates the mean guanine-cytosine content percentage averaged across R1 and R2 reads. Group means are provided for each tissue type (ureter n=3, bladder n=3), with overall totals shown for the cohort (n=6). Ureter samples were sequenced using 75 bp paired-end reads, whilst bladder samples were sequenced using 150-151 bp paired-end reads.

Adapter trimming and quality filtering using fastp (version 0.24.0) retained 98.4% of reads (183.4 million) whilst improving mean Q30 scores from 93.2% to 93.8%. GC content analysis showed that there were tissue specific differences, with bladder samples having significantly higher GC content (50.7% ± 0.5%) compared to ureter samples (48.8% ± 0.3%, p = 0.008, Welch's two sample t-test, n=3 bladder, n=3 ureter).

Alignment to GRCh38 using STAR (version 2.7.11b) achieved high mapping rates across all samples, Table 3.3. Overall, 94.8% of reads mapped uniquely to single genomic loci (range: 93.6-96.5%), whilst 4.3% multi-mapped to 2-20 loci, with only 0.9% unmapped. Total mapping rates exceeded 98% for both tissues. Gene level quantification using featureCounts (version 2.0.6) with GENCODE v47 annotation yielded mean assignment rate of 65.8%, with bladder samples achieving slightly higher rates (66.7%) than ureter samples (64.7%). Unassigned reads included multi-mapping (13.4%), intergenic/intronic regions (11.4%), overlapping features (8.6%), and unmapped reads (0.8%).

Sample	**Tissue**	**Input Reads (Million)**	**Uniquely mapped (%)**	**Multimapping (%)**	**Total mapped (%)**	**Unmapped (%)**
Y2391	Ureter	35.7	93.6	5.8	99.5	0.5
Y2392	Ureter	31.1	93.8	5.7	99.4	0.5
Y2396	Ureter	33.0	94.8	4.7	99.4	0.5
Ureter mean	**-**	**33.3**	**94.1**	**5.4**	**99.4**	**0.5**
Y2631	Bladder	25.4	94.9	3.5	98.4	1.6
Y2632	Bladder	24.6	95.3	3.6	98.8	1.1
Y2796	Bladder	33.6	96.5	2.8	99.3	0.7
Bladder mean	**-**	**27.9**	**95.6**	**3.3**	**98.8**	**1.1**
Overall	**-**	**183.4**	**94.8**	**4.3**	**99.1**	**0.9**

Table 3.3: STAR genome alignment statistics for processed short-reads from normal bladder and ureter urothelial samples. Alignment statistics from STAR (version 2.7.11b)

mapping to GRCh38 reference genome for six normal urothelial samples (ureter: Y2391, Y2392, Y2396, bladder: Y2631, Y2632, Y2796). Input reads indicates quality filtered paired-end reads in millions (from fastp output). Uniquely mapped shows percentage aligning to single genomic locus. Multimapping represents reads aligning to 2-20 loci. Total mapped is the sum of unique and multi-mapped reads. Unmapped indicates alignment failures. All percentages are relative to input reads. Group means shown for each tissue type.

3.3.2 Principal component analysis

Principal component analysis (PCA) was performed on variance-stabilised transformed count data to assess sample relationships and identify major sources of transcriptomic variation. The first two principal components explained 60.44% of total variance, Figure 3.1A. PC1 (36.61% variance) clearly separated samples by tissue type, with bladder and ureter samples forming distinct, non-overlapping clusters. PC2 (23.83% variance) reflected intra-group variation, potentially capturing individual biological differences or technical factors, with bladder sample Y2796 separated along this PC. Samples did not cluster according to RIN scores, Figure 3.1B.

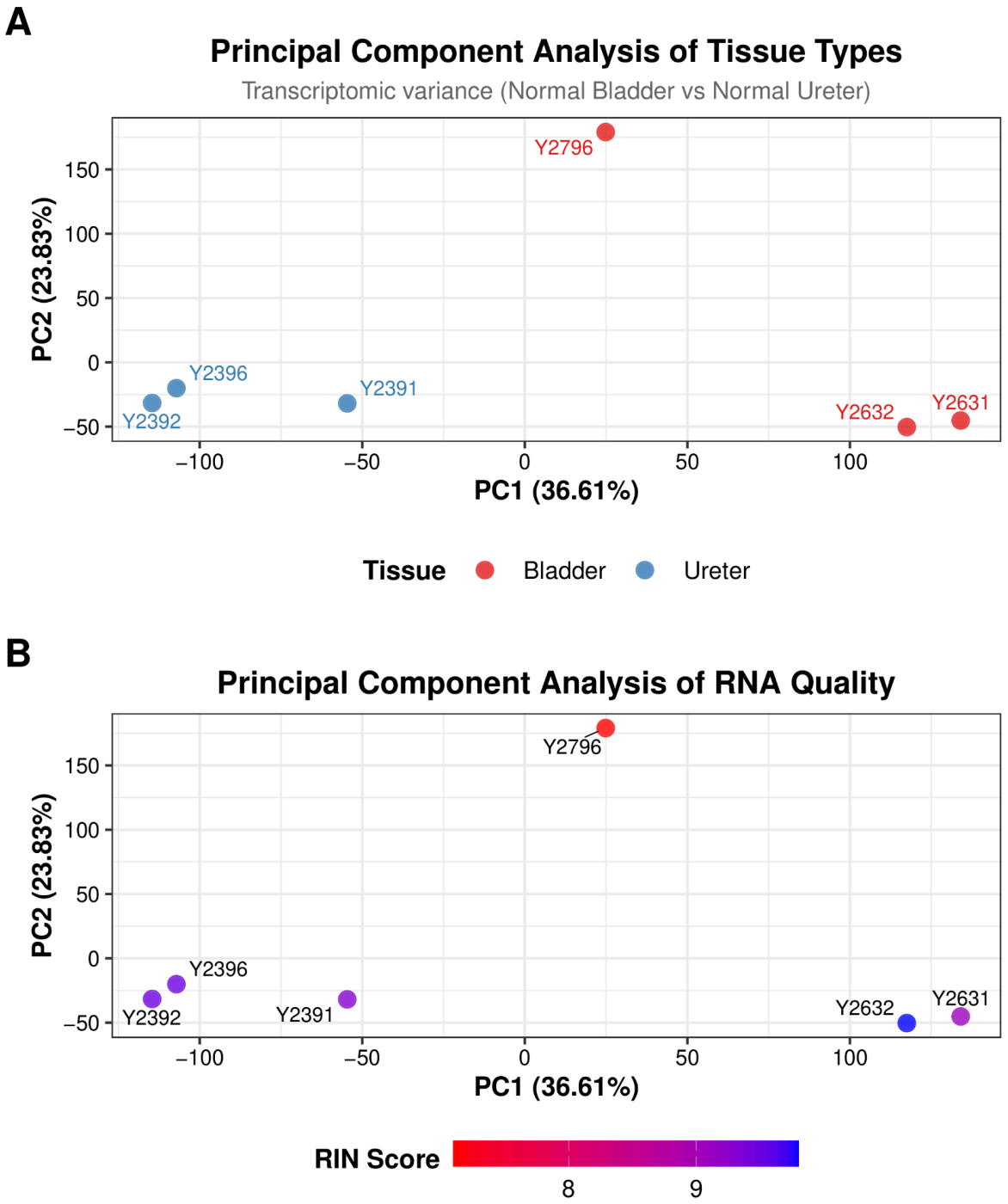

Figure 3.1: Principal component analysis of gene expression data. (A) Principal component analysis showing tissue specific clustering. Bladder samples (red, n=3) and ureter samples (blue, n=3) form distinct, non-overlapping clusters along PC1 (36.61% of variance), whilst PC2 (23.83% of variance) reflects intra-group variation. Sample identifiers are displayed. (B) The same principal component analysis with the samples coloured by RNA integrity number (RIN) score. RIN values range from 7.1 (red) to 9.8 (blue). Samples do not cluster according to RIN. Together, PC1 and PC2 explain 60.44% of total variance.

Gene loading analysis showed distributed contributions of many genes rather than dominance by individual genes. Top PC1 contributors had uniform loading magnitudes (mean: 0.0091). Contributors included *TMEM150C*, *ACER3*, *POM121* and *PRNP*. PC2 contributors (mean loading: 0.0113) included cell cycle regulators (*ORC1*, *ORC6*, *RAD51*, *CCNB1*, *E2F1*) and cytoskeletal genes (*CSRP1*, *ARHGDIB*).

3.3.3 Differential expression

Differential gene expression analysis was performed comparing normal bladder urothelium (n=3) against normal ureter urothelium (n=3) using DESeq2 (v1.42.0). Genes were defined as significantly differentially expressed if they had a false discovery rate (FDR) adjusted *p*-value < 0.05 following Benjamini-Hochberg correction. A total of 4,632 genes (14.1% of input) were excluded by independent filtering due to low counts.

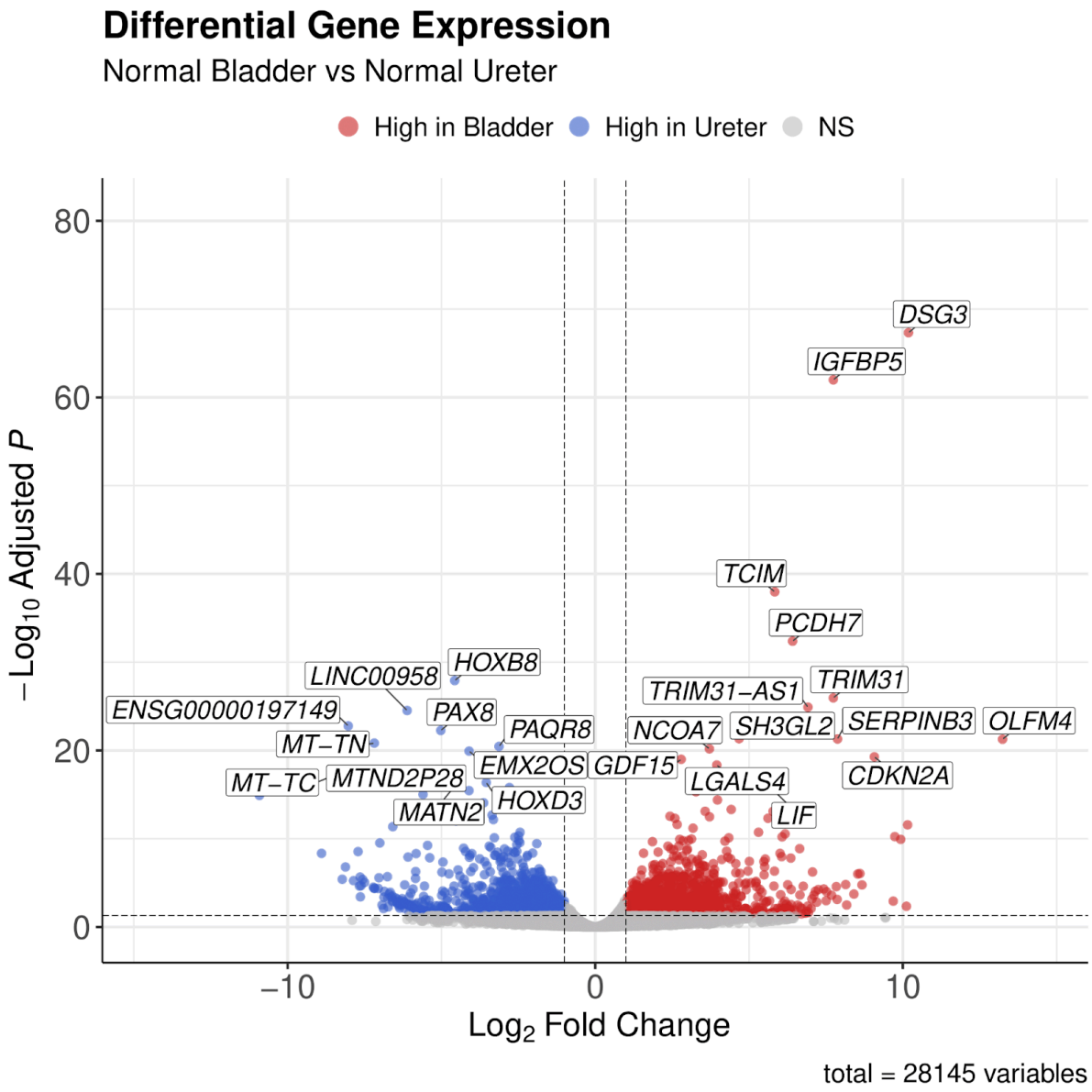

Figure 3.2: Differential gene expression between normal bladder and ureter urothelium. A volcano plot displaying log2 fold change versus adjusted *p*-value for differential gene expression analysis between normal bladder urothelium (n=3) and normal

ureter urothelium (n=3). Each point represents a single gene from 28,145 genes tested. Positive log2 fold changes indicate higher expression in bladder urothelium and negative values indicate higher expression in ureter urothelium. Red points are genes significantly upregulated in bladder, blue points are genes significantly upregulated in ureter and grey points represent non-significant changes. Dashed lines mark significance thresholds (vertical: ±1.0 log2FC and horizontal: adjusted $p = 0.05$). The top 25 most significantly differentially expressed genes are labelled. A total of 3,698 genes (13.1%) were significantly differentially expressed, with 1,981 upregulated in bladder and 1,717 upregulated in ureter.

Of the 28,145 genes tested, 3,698 (13.1%) exhibited significant differential expression, with 1,981 genes upregulated in the bladder and 1,717 genes upregulated in the ureter, Figure 3.2. Upregulated genes in the bladder had a mean log2 fold change (log2FC) of 2.31, whilst genes upregulated in the ureter had a mean log2FC of -2.48.

The most significantly upregulated gene in the bladder was *OLFM4* (log2FC = +13.24), with other upregulated genes such as *DSG3* (log2FC = +10.18, adjusted $p = 4.7 \times 10^{-68}$), *IGFBP5* (log2FC = +7.74), *SERPINB3* (log2FC = +7.88), *SERPINB4* (log2FC = +6.83) and PCDH7 (log2FC = +6.41). The most significantly upregulated gene in the ureter was *SIM1* (log2FC = -10.92), with HOX transcription factors also including *HOXB2* (log2FC = -4.62)*, HOXB8* (log2FC = -4.57), *HOXB9* (log2FC = -3.63), *PAX8* (log2FC = -5.02), mitochondrial transcripts *MT-TC* (log2FC -8.54) and *MT-TN* (log2FC -7.18) as well as *MATN2* (log2FC = -4.43).

3.3.4 Biological pathway analysis

To investigate functional differences between bladder and ureter urothelium, targeted pathway analyses examined curated gene sets representing developmental patterning, differentiation markers, signaling pathways, and immune responses.

3.3.4.1 HOX gene expression

HOX expression was analysed across 39 genes in four chromosomal clusters to determine whether developmental programmes persist in adult tissues.

Overall HOX pathway activity was significantly higher in ureter (13.52 ± 4.40 TPM) versus bladder (4.43 ± 2.70 TPM, $p = 0.049$, Welch's two sample t-test, n=3 bladder, n=3 ureter), Figure 3.3A. The HOXB cluster had the largest difference and had significantly higher expression in the ureter (ureter: 30.77 TPM, bladder: 5.49 TPM, $p = 0.036$, Welch's two sample t-test, n=3 bladder, n=3 ureter), with HOXD also elevated in ureter (13.03 vs 2.79 TPM), Figure 3.3B. HOXA expression was similar between tissues (bladder: 8.34 TPM, ureter: 9.07 TPM).

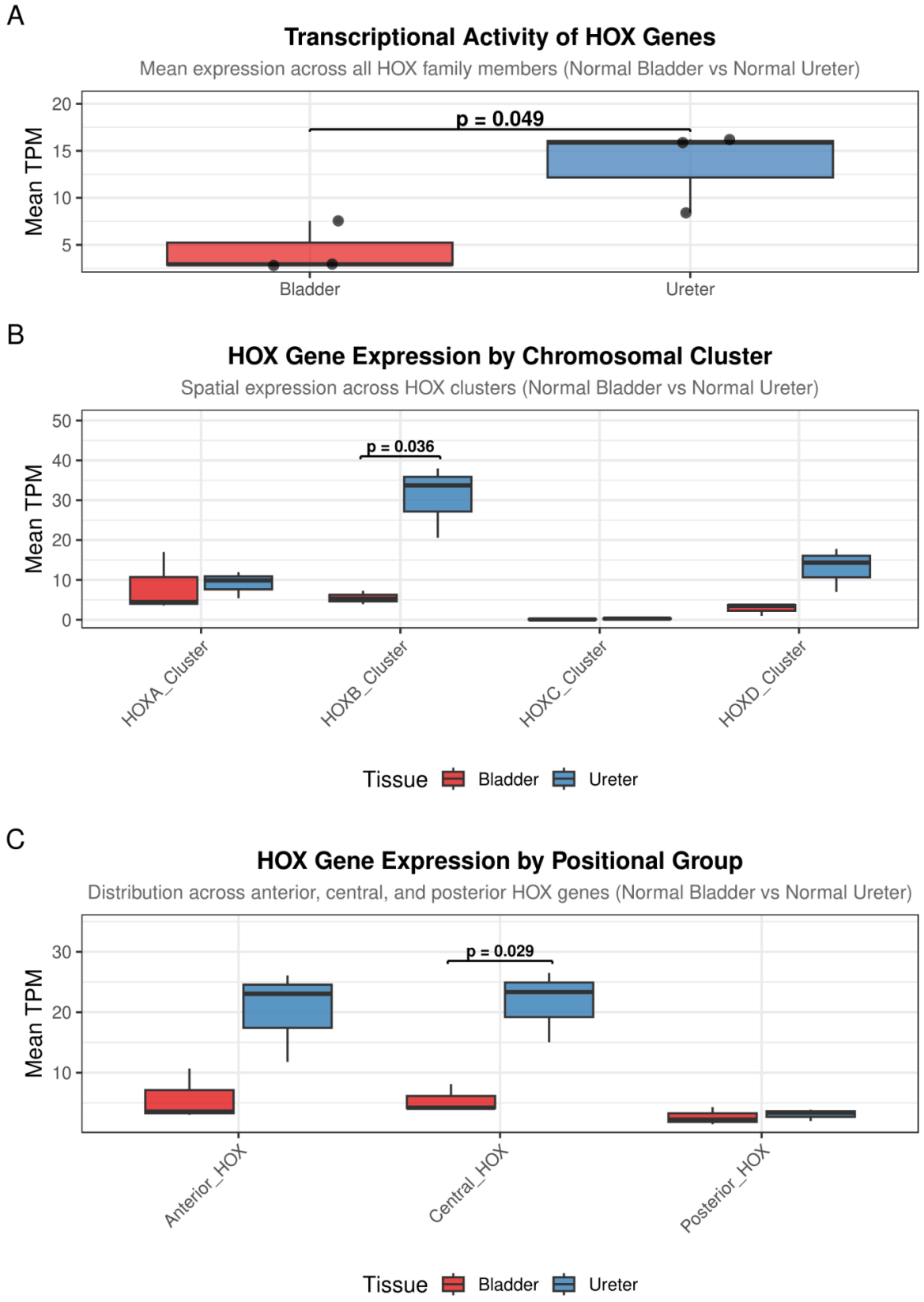

Figure 3.3: HOX gene expression between normal bladder and ureter urothelium. Box plots showing mean TPM of overall HOX expression, HOX expression by cluster and HOX expression by positional group in bladder (red, n=3) and ureter (blue, n=3) urothelium with

individual samples overlaid as points. Statistical significance was assessed using Welch's two-sample t-test, brackets with *p*-values indicate significant differences between groups. **(**A) A boxplot showing overall HOX pathway activity score (mean expression across 39 detected genes) with individual samples overlaid as points, bladder (red, n=3) and ureter (blue, n=3) samples. The ureter displays significantly higher pathway activity than bladder (B) A boxplot showing mean HOX gene expression (TPM) across chromosomal clusters in bladder and ureter. The HOXB is significantly more expressed in the ureter than the bladder samples. (C) A box plot showing mean HOX gene expression grouped by anterior-posterior position. Central HOX genes had significantly higher expression in the ureter compared to the bladder samples.

Positional analysis showed significantly elevated central (ureter: 21.63 TPM, bladder: 5.46 TPM, $p = 0.029$, Welch's two sample t-test, n=3 bladder, n=3 ureter) and elevated anterior HOX genes (ureter: 20.31 TPM, bladder: 5.76 TPM), whilst posterior genes had similar expression (ureter: 3.08 TPM, bladder: 2.68 TPM), Figure 3.3C.

3.3.4.2 Sonic hedgehog signalling pathway

SHH pathway activity was analysed across 36 genes to determine whether differential expression persists in adult bladder and ureter tissues.

Expression of SHH pathway associated genes was significantly higher in bladder urothelium (56.23 ± 7.68 TPM) versus ureter urothelium (22.65 ± 3.69 TPM, $p = 0.007$, Welch's two sample t-test, n=3 bladder, n=3 ureter), Figure 3.4A. This difference was primarily driven by elevated expression of the SHH ligand itself which was 3 fold higher in bladder (122 vs 41 TPM), along with pathway associated genes including *VEGFA* and *ETS2.* However, the canonical downstream target GLI1 showed minimal expression (~0.1 TPM) in both tissues.

3.3.4.3 Fibroblast growth factor signalling pathway

The expression of 55 pathway genes was analysed to determine whether differential activity exists between adult tissues.

FGF pathway activity was significantly higher in bladder (96.44 ± 9.61 TPM) versus ureter (73.64 ± 9.13 TPM, $p = 0.041$, Welch's two sample t-test, n=3 bladder, n=3 ureter), Figure 3.4B. Individual bladder samples ranged from 88.65 to 107.17 TPM, whilst ureter samples ranged from 63.32 to 80.70 TPM.

3.3.4.4 Wnt beta-catenin signalling pathway

Expression of 42 pathway genes was analysed to determine whether differential activity exists between adult tissues.

Wnt pathway activity was higher in bladder (36.53 ± 8.00 TPM) versus ureter (20.27 ± 2.27 TPM), though this did not achieve statistical significance ($p = 0.063$, Welch's two sample t-test, n=3 bladder, n=3 ureter), Figure 3.4C. Individual bladder samples ranged from 29.05 to 44.97 TPM, whilst ureter samples ranged from 17.66 to 21.80 TPM.

3.3.4.5 TGF-beta signalling pathway

The expression of 54 pathway genes was analysed to determine whether differential activity exists between adult tissues.

TGF-beta pathway activity was significantly higher in bladder (156.12 ± 8.91 TPM) versus ureter (123.22 ± 5.54 TPM, p = 0.009, Welch's two sample t-test, n=3 bladder, n=3 ureter), Figure 3.4D. Individual bladder samples ranged from 150.89 to 166.41 TPM, whilst ureter samples ranged from 116.94 to 127.41 TPM, showing tissue separation with low intra-group variability.

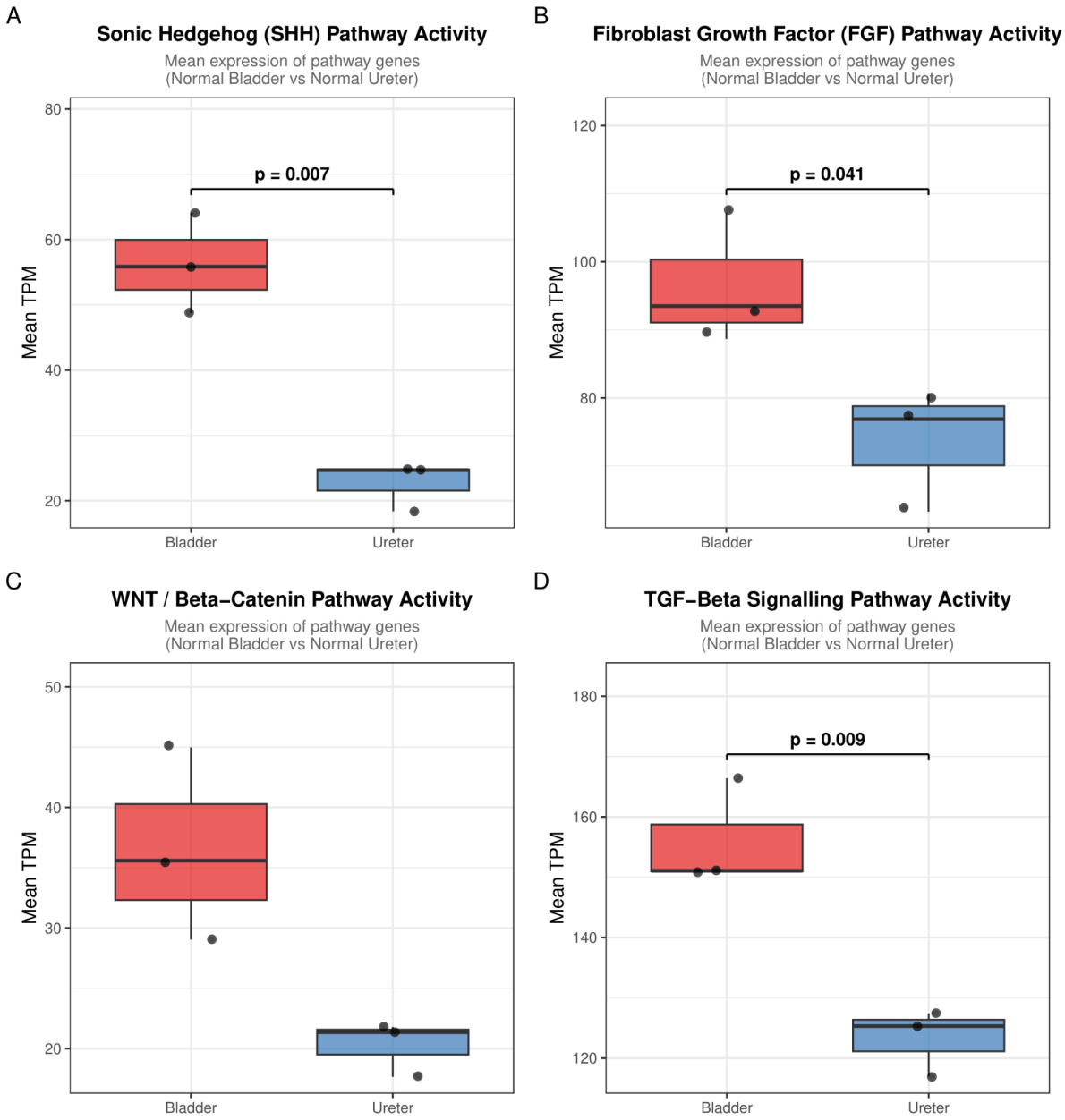

Figure 3.4: Signalling pathway expression between normal bladder and ureter urothelium. Box plots showing overall pathway activity scores (mean TPM) in bladder (red, n=3) and ureter (blue, n=3) urothelium with individual samples overlaid as points. Statistical

significance was assessed using Welch's two-sample t-test, brackets with *p*-values indicate significant differences between groups. (A) SHH pathway activity score (mean TPM across 36 pathway genes), bladder urothelium had significantly higher activity than ureter. (B) FGF pathway activity score (mean TPM across 55 pathway genes), bladder had significantly higher activity compared to ureter samples. (C) Wnt/beta-catenin pathway activity score (mean TPM across 42 pathway genes), bladder urothelium showed higher activity than ureter, though this did not reach significance. (D) TGF-beta pathway activity score (mean TPM across 54 pathway genes), bladder had significantly higher activity compared to ureter samples.

3.3.4.6 Interferon response

Overall interferon pathway activity was elevated in bladder (16.92 ± 5.62 TPM) versus ureter (7.70 ± 1.31 TPM), though this did not achieve statistical significance ($p = 0.098$, Welch's two sample t-test, n=3 bladder, n=3 ureter), Figure 3.5A. Individual bladder samples ranged from 11.50 to 22.72 TPM, whilst ureter samples ranged from 6.73 to 9.20 TPM, showing tissue separation.

The type II interferon had the largest difference, with bladder expression (49.98 ± 67.99 TPM) substantially exceeding ureter (0.91 ± 1.22 TPM). However, a lot of heterogeneity in expression was observed, with one bladder sample showing 127.52 TPM whilst others showed 0.60 and 21.82 TPM. Type II receptors were significantly elevated in bladder compared to ureter ((173.76 ± 32.76 TPM vs 84.53 ± 18.41 TPM, $p = 0.024$, Welch's two sample t-test, n=3 bladder, n=3 ureter). Type III receptors were also elevated in bladder compared to ureter tissue (bladder: 16.48 ± 6.61 TPM, ureter: 9.11 ± 1.20 TPM). Type I receptors had similar expression between tissues (bladder: 14.22 ± 6.38 TPM, ureter: 11.59 ± 1.03 TPM), Figure 3.5C.

3.3.4.7 Immune cell marker expression

To characterise immune cell populations, the expression of 14 cell type specific markers was analysed across nine immune cell categories (naive CD4+ T cells, memory CD4+ T cells, CD8+ T cells, B cells, NK cells, CD14+ monocytes, FCGR3A+ monocytes, dendritic cells, and platelets) to determine whether differential immune cell infiltration exists between adult bladder and ureter tissues. Platelets were chosen to be included as immune markers due to their involvement in innate and adaptive immune responses and impact on the EMT transition in tumours and the tumour microenvironment (Chen *et al.*, 2023; Nicolai, Pekayvaz and Massberg, 2024)

Overall immune marker expression was significantly higher in bladder (34.10 ± 2.59 TPM) versus ureter (22.94 ± 4.03 TPM, $p = 0.021$, Welch's two sample t-test, n=3 bladder, n=3 ureter), Figure 3.5B. Individual bladder samples ranged from 31.72 to 36.85 TPM, whilst ureter samples ranged from 20.55 to 27.60 TPM, showing clear tissue separation. Naive CD4+ T cell markers had the largest difference (bladder: 9.97 TPM, ureter: 2.00 TPM), followed by NK cell markers (bladder: 2.37 TPM, ureter: 0.81 TPM) and CD8+ T cell markers (bladder: 5.68 TPM, ureter: 2.59 TPM). Memory CD4+ T cells, B cells and CD14+ monocytes also showed elevated expression in bladder tissue, whilst dendritic cells and FCGR3A+ monocyte markers had similar expression between tissues, Figure 3.5D.

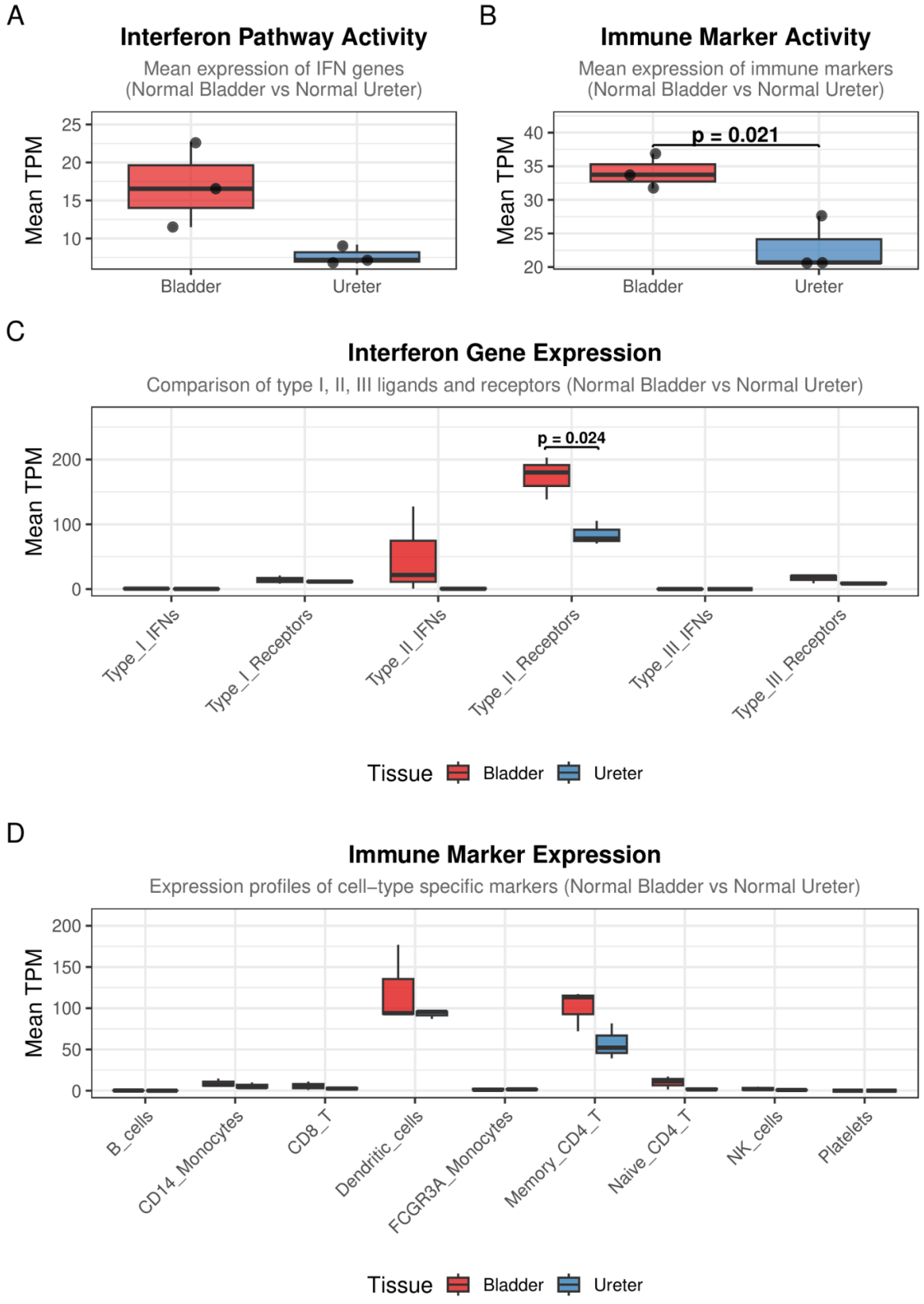

Figure 3.5: Interferon and immune marker gene expression between normal bladder and ureter urothelium. Box plots showing overall activity scores (mean TPM) in bladder (red, n=3) and ureter (blue, n=3) urothelium with individual samples overlaid as points.

Statistical significance was assessed using Welch's two-sample t-test, brackets with *p*-values indicate significant differences between groups. (A) Overall interferon marker gene activity score (28 genes), bladder urothelium showed higher activity than ureter but this did not reach statistical significance. (B) Overall immune marker gene activity score (14 genes), bladder urothelium showed significantly higher activity than ureter. (C) Mean gene expression (TPM) across IFN type I, II and III and their respective receptors in bladder and ureter. Type II IFN receptors showed significantly higher expression in bladder compared to ureter. (D) Mean gene expression (TPM) across immune cell types in bladder and ureter.

3.3.4.8 Cell cycle and proliferation analysis

The expression of 56 genes across six cell cycle categories (G1/S transition, S phase, G2/M transition, M phase, proliferation markers, and checkpoint control) was analysed to determine whether differential proliferative activity exists between adult bladder and ureter tissues.

Overall proliferation pathway activity was significantly higher in bladder (46.13 ± 8.31 TPM) versus ureter (21.00 ± 3.25 TPM, $p = 0.023$, Welch's two sample t-test, n=3 bladder, n=3 ureter), Figure 3.6A. Individual bladder samples ranged from 38.45 to 54.94 TPM, whilst ureter samples ranged from 18.22 to 24.58 TPM. G1/S transition genes showed the largest difference (bladder: 117.58 ± 12.44 TPM, ureter: 43.94 ± 13.00 TPM, $p = 0.002$, Welch's two sample t-test, n=3 bladder, n=3 ureter), followed by G2/M transition genes (bladder: 18.78 ± 1.71 TPM, ureter: 11.94 ± 2.81 TPM, $p = 0.031$, Welch's two sample t-test, n=3 bladder, n=3 ureter) and checkpoint control genes (bladder: 16.62 ± 0.83 TPM, ureter: 11.72 ± 2.72 TPM). S phase genes showed similar expression between tissues (bladder: 22.06 ± 19.66 TPM, ureter: 18.43 ± 4.84 TPM), though with higher bladder variability, Figure 3.6C.

3.3.4.9 Luminal and basal marker expression

Expression of 52 differentiation markers (20 luminal, 20 basal, 12 EMT) was analysed. Overall marker activity was higher in bladder (403.98 ± 46.09 TPM) versus ureter (241.98 ± 52.18 TPM, $p = 0.016$, Welch's two sample t-test, n=3 bladder, n=3 ureter). All categories had elevated bladder expression, luminal markers (556.89 vs 367.98 TPM, $p = 0.041$, Welch's two sample t-test, n=3 bladder, n=3 ureter), basal markers (424.12 vs 226.82 TPM) and EMT markers (117.23 vs 55.96 TPM), Figure 3.6B.

Luminal-basal ratios were similar between tissues (bladder: 0.42, ureter: 0.78, $p = 0.452$, Welch's two sample t-test, n=3 bladder, n=3 ureter).

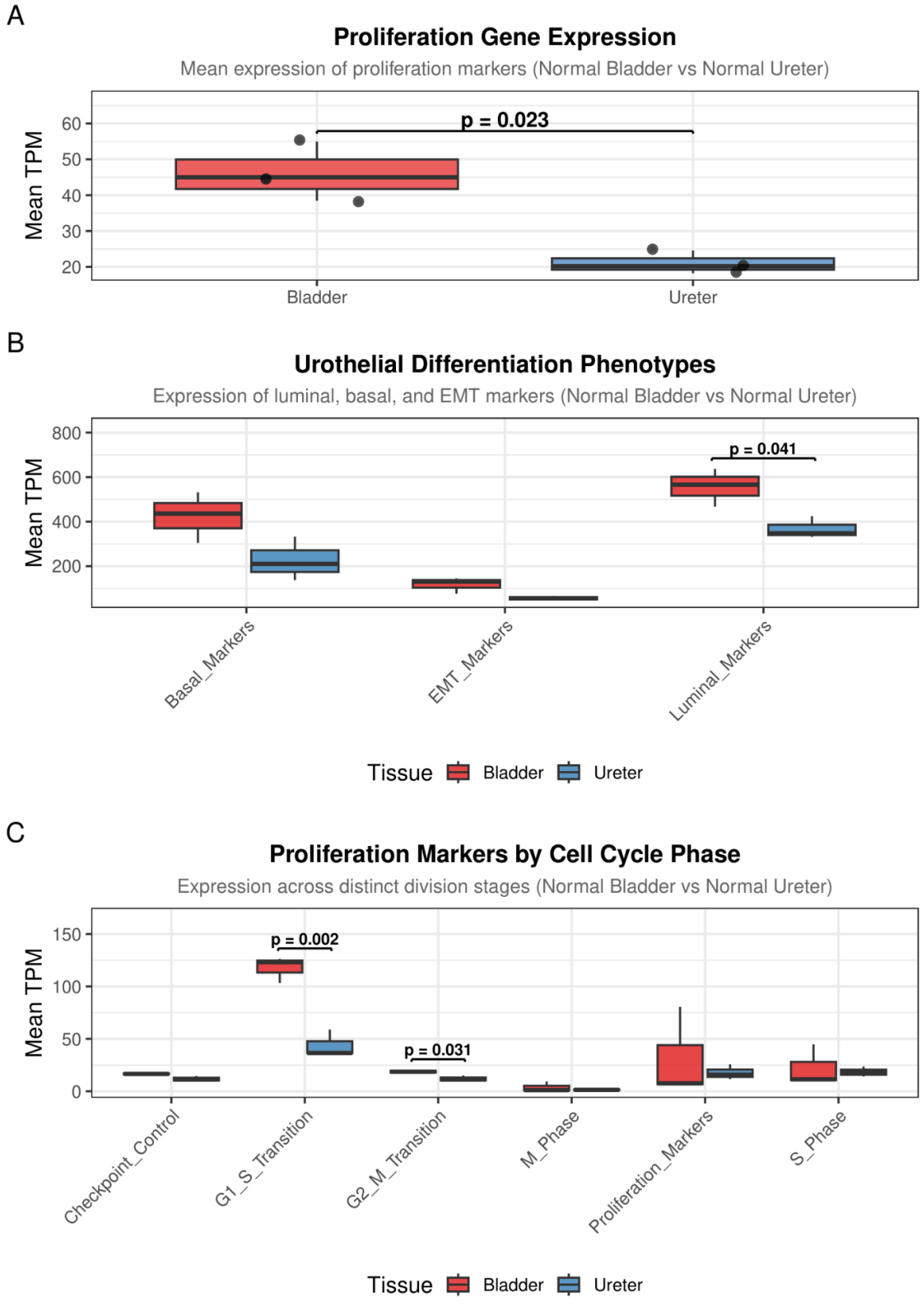

Figure 3.6: Proliferation, urothelial marker and cell cycle gene expression between normal bladder and ureter urothelium. Box plots showing mean gene expression (TPM) in bladder (red, n=3) and ureter (blue, n=3) urothelium, with individual samples overlaid as

points. Statistical significance was assessed using Welch's two-sample t-test, brackets with *p*-values indicate significant differences between groups. (A) Overall cell cycle and proliferation marker gene activity score (56 genes), bladder urothelium showed significantly higher activity than ureter. (B) Mean expression (TPM) of urothelial markers by category (luminal, basal and EMT), luminal markers showed significantly higher expression in bladder compared to ureter urothelium. (C) Mean gene expression (TPM) across cell cycle phases and checkpoints in bladder and ureter, G1/S transition and G2/M transition genes were significantly more expressed in bladder samples than ureter.

3.3.5 Gene set enrichment analysis

Gene Set Enrichment Analysis (GSEA) was performed to identify biological pathways and processes enriched in the differential expression data. The analysis used two complementary gene set collections, the Hallmark gene sets and Reactome pathways. The ranked gene list, comprising 28,145 genes, was derived from differential expression analysis between bladder and ureter urothelial samples.

3.3.5.1 Hallmark gene sets

Of the 50 gene sets tested using an fgsea exact permutation test (n=3 bladder, n=3 ureter), 35 achieved statistical significance at the FDR-corrected threshold (adjusted $p < 0.05$), whilst 38 had nominal significance (p-value < 0.05).

The pathway having the strongest enrichment was TNF alpha signalling via NF-kappa B (NES = 2.55, adjusted $p = 1.86 \times 10^{-16}$), followed closely by inflammatory response (NES = 2.42, adjusted $p = 3.33 \times 10^{-11}$) and allograft rejection (NES = 2.23, adjusted $p = 1.49 \times 10^{-7}$). Further analysis revealed significant enrichment of mTORC1 signalling (NES = 2.14, adjusted $p = 3.32 \times 10^{-7}$) and interferon gamma response (NES = 2.05, adjusted $p = 5.79 \times 10^{-6}$). Additional pathways include mitotic spindle organisation (NES = 2.03, adjusted $p = 9.88 \times 10^{-6}$), p53 pathway activation (NES = 1.99, adjusted $p = 4.90 \times 10^{-5}$), and G2M checkpoint regulation (NES = 1.93, adjusted $p = 2.34 \times 10^{-4}$).

3.3.5.2 Reactome pathways

The Reactome pathway analysis complemented and extended the Hallmark findings. Of the 1,017 pathways tested using an fgsea exact permutation test (n=3 bladder, n=3 ureter), 52 achieved FDR significance (adjusted $p < 0.05$), whilst 357 demonstrated nominal significance (p-value < 0.05).

The most significant pathway identified was signalling by interleukins (NES = 2.14, adjusted $p = 5.78 \times 10^{-9}$). This was followed by signalling by receptor tyrosine kinases (NES = 1.92, adjusted $p = 1.07 \times 10^{-5}$) and neutrophil degranulation (NES = 1.93, adjusted $p = 2.43 \times 10^{-5}$). Interferon signalling emerged as another key enriched pathway (NES = 1.97, adjusted $p = 2.28 \times 10^{-4}$). Programmed cell death pathways were also significantly enriched (NES = 2.06, adjusted $p = 4.06 \times 10^{-4}$).

Only one pathway, formation of the nephric duct (NES = -1.91, adjusted $p = 0.018$), achieved significant enrichment in ureter tissue.

3.3.6 Transcript level quantification

Transcript level quantification using StringTie assessed isoform diversity and alternative splicing patterns between tissues. A mean of 34,592 genes were detected (40.21% of annotated genes), with higher detection in ureter (36,777 genes) versus bladder (32,408 genes). However, bladder samples had higher median expression per gene (28-33 reads) compared to ureter (16 reads). Bladder samples had more high expression genes (39.8% with >100 counts) than ureter (34.7%), whilst ureter had more low-expression genes (44.9% with 1-10 counts) compared to bladder (37.2%).

At the transcript level, 129,480 transcripts were detected overall. Ureter samples had higher transcript detection (134,725) versus bladder (124,234), but bladder maintained slightly higher transcripts per gene (3.83 vs 3.66). Median transcript expression was higher in bladder (11-13 reads) than ureter (7-8 reads).

3.3.7 Isoform switch analysis

Isoform switch analysis identified genes exhibiting differential isoform usage between bladder and ureter urothelial tissue using the IsoformSwitchAnalyzeR package. The analysis detected 612 genes containing 809 isoforms involved in 899 switching events with predicted functional consequences, representing approximately 5.4% of expressed genes.

The gene demonstrating the strongest evidence for isoform switching was *GSTP1* (q-value = 2.75×10^{-97}). In the top switches there was a substantial enrichment of ribosomal proteins, including *RPL29* (q-value = 4.88×10^{-68}), *RPL10A* (q-value = 6.88×10^{-29}), *RPS20* (q-value = 7.54×10^{-29}), *RPL18* (q-value = 5.85×10^{-22}), and *RPS15* (q-value = 7.25×10^{-23}). Other notable switches include *DGKA*, *GADD45B* and *OAS3*.

There was an enrichment of intron retention in bladder tissue, in which 265 genes showed increased retention in bladder compared to 154 in ureter (63.2%, q-value = 2.50×10^{-13}). Alternative transcription start sites (ATSS) represented the most prevalent mechanism, with 587 genes showing increased ATSS usage in bladder versus 510 in ureter (53.5%, q-value = 1.01×10^{-5}). Other significant patterns included a bias towards exon inclusion in bladder (q-value = 3.17×10^{-3}) and loss of alternative 3' splice sites (q-value = 3.84×10^{-4}).

Isoform switch analysis identified 83 switching events resulting in complete open reading frame (ORF) loss in bladder, compared to 38 showing gain (31.4% gain, q-value = 6.58×10^{-5}), whilst 262 events produced shorter ORFs versus 160 producing longer ORFs (37.9% longer, q-value = 1.96×10^{-6}). 79 transcripts switched from coding to non-coding in bladder compared to 34 switching from non-coding to coding (30.1% coding gain, q-value = 4.62×10^{-5}). Finally, 90 switching events resulted in NMD sensitive isoforms in bladder versus 49 producing NMD insensitive isoforms (35.3% insensitive, q-value = 6.38×10^{-4}).

3.4 Long-read

3.4.1 Long-read sequencing quality and alignment

Long-read Oxford Nanopore sequencing of six urothelial samples generated 145.5 million raw reads with considerable variation in depth (17.5-48.8 million reads per sample), Table 3.4. Ureter sample Y2391 yielded substantially higher depth (48.8 million reads, 65.0 Gb) compared to remaining samples (17.5-21.5 million reads, 10.6-14.5 Gb). Raw data quality metrics showed Q20 percentages of 43.3-54.8% and Q30 percentages of 10.2-18.9%, with mean Phred scores of 9.86-11.27. Maximum read lengths ranged from 663 kb to 1,978 kb. All samples failed FastQC adapter content checks prior to preprocessing.

Sample	**Tissue**	**Total Reads**	**Total bases (Gb)**	**Mean length (bp)**	**Highest read length (bp)**	**GC content (%)**	**Q20 (%)**	**Q30 (%)**	**Mean Phred Score**
Y2391	Ureter	48,781,450	65.0	1,333.5	1,978,106	45.8	43.3	10.2	9.86
Y2392	Ureter	17,546,898	10.6	606.3	876,725	44.0	54.4	18.7	11.15
Y2396	Ureter	17,780,264	12.0	673.8	698,977	44.0	54.2	18.9	11.12
Y2631	Bladder	21,016,074	11.6	550.9	796,326	45.0	54.5	18.5	11.20
Y2632	Bladder	18,947,514	12.6	666.9	991,239	45.3	54.2	18.5	11.14
Y2796	Bladder	21,477,418	14.5	674.5	663,028	45.2	54.8	18.6	11.27

Table 3.4: Raw sequencing statistics and quality metrics from Oxford Nanopore long-read sequencing of normal bladder and ureter urothelial samples. Raw sequencing statistics from SeqKit (version 2.9.0) analysis of ONT cDNA RNA sequencing for six normal urothelial samples (ureter: Y2391, Y2392, Y2396, bladder: Y2631, Y2632,

Y2796). Total reads and bases indicate raw sequencing output. Mean length and highest read length show read size distributions in base pairs. GC content expressed as a percentage. Q20 and Q30 represent the percentage of bases with Phred scores ≥20 (99% accuracy) and ≥30 (99.9% accuracy), respectively. Mean Phred score is average quality across all bases.

Two stage preprocessing using Pychopper (version 2.7.10) for full length cDNA identification and Fastplong (version 0.2.2) for adapter trimming and quality filtering retained 8.6% of raw reads as high quality full length transcripts (12.5 million reads, 4.9 Gb total), Table 3.5. Pychopper identified full length reads in only 8-11% of input (mean: 9.6%), with the majority unusable due to absent primer sequences. Fastplong filtering retained 89.1% of full length reads (range: 87.2-94.9%), removing reads with low quality scores or excessive shortening after adapter trimming. Quality improvements were consistent across samples, with Q30 percentages increasing by mean 2.0% and mean Phred scores improving from 10.96 to 11.92. Mean read length decreased from 751 bp to 375.2 bp post-processing, with maximum read lengths decreasing from 663-1,978 kb (raw) to 8-11 kb (processed).

Sample	Tissue	Total Reads	Pass Reads	Pass rate (%)	Primers found	Primer (%)	Rescued reads	Full length reads	Full length reads (%)
Y2391	Ureter	48,781,450	44,906,397	92.06	4,251,653	8.72	198,050	4,449,703	9.12
Y2392	Ureter	17,546,898	16,333,308	93.08	1,518,889	8.66	68,962	1,587,851	9.05
Y2396	Ureter	17,780,264	16,544,392	93.05	1,639,058	9.22	69,491	1,708,549	9.61
Y2631	Bladder	21,016,074	19,466,553	92.63	2,107,816	10.03	72,444	2,180,260	10.37
Y2632	Bladder	18,947,514	17,855,389	94.24	1,908,532	10.07	79,635	1,988,167	10.49
Y2796	Bladder	21,477,418	20,418,960	95.07	2,010,814	9.36	80,384	2,091,198	9.74
Mean	-	24,258,270	22,587,500	93.36	2,239,460	9.34	94,828	2,334,288	9.73

Table 3.5: Pychopper full length cDNA identification statistics for Oxford Nanopore long-read sequencing data from normal bladder and ureter urothelial samples. Full length transcript identification from Pychopper (version 2.7.10) analysis of ONT sequencing data for six normal urothelial samples (ureter: Y2391, Y2392, Y2396; bladder: Y2631, Y2632, Y2796). Total reads indicates raw input. Pass reads shows reads passing quality filters, pass rate is percentage passing. Primers found indicates reads with both 5' and 3' primers detected, primer (%) shows the percentage. Rescued reads recovered by rescue algorithm. Full length reads combines primers found and rescued reads, full length reads (%) shows the percentage of total classified as full length.

Alignment to GRCh38 using Minimap2 (version 2.24) achieved mean mapping rate of 78.6% (range: 74.9-87.3%), with primary alignments accounting for 99.3-99.9% of mapped reads and minimal chimeric alignments (0.5-0.7%), Table 3.6. Overall, 10.3 million reads (81.9% of 12.6 million processed) successfully aligned. Gene level quantification using featureCounts

(version 2.0.6) with GENCODE v47 annotation yielded mean assignment rate of 56.7% (range: 53.0-65.2%) with 7.5 million reads successfully assigned to genes. Unassigned reads comprised unmapped reads (21.4%), reads in unannotated regions (14.4%), and ambiguous multi-gene overlaps (7.6%).

Sample	Tissue	Total Reads	Mapped Reads	Mapping Rate (%)	Unmapped Reads
Y2391	Ureter	5,987,602	5,228,671	87.32	758,931
Y2392	Ureter	1,175,691	910,321	77.43	265,370
Y2396	Ureter	1,250,392	979,353	78.32	271,039
Y2631	Bladder	1,251,509	951,570	76.03	299,939
Y2632	Bladder	1,407,666	1,053,569	74.85	354,097
Y2796	Bladder	1,481,308	1,153,796	77.89	327,512
Mean	-	2,092,361	1,712,880	78.64	379,481

Table 3.6: Minimap2 reference genome alignment statistics for processed long-read cDNA sequences from normal bladder and ureter urothelial samples. Alignment statistics from Minimap2 (version 2.24) splice-aware mapping (-ax splice -uf --secondary=no) to GRCh38 reference genome for six normal urothelial samples (ureter: Y2391, Y2392, Y2396, bladder: Y2631, Y2632, Y2796). Total reads indicates quality filtered, full length cDNA input from Pychopper and Fastplong processing. Mapped reads shows successful alignments. Mapping rate is the percentage of total reads mapped. Unmapped reads represents alignment failures. Mean values shown across all samples.

3.4.2 Principal component analysis

Principal component analysis (PCA) was performed on variance-stabilised gene expression data to assess sample relationships and identify major sources of transcriptomic variation.

The first two principal components explained 54.29% of total variance, PC1 (28.84% variance) and PC2 (25.45% variance). PC1 clusters by tissue type, Figure 3.7.

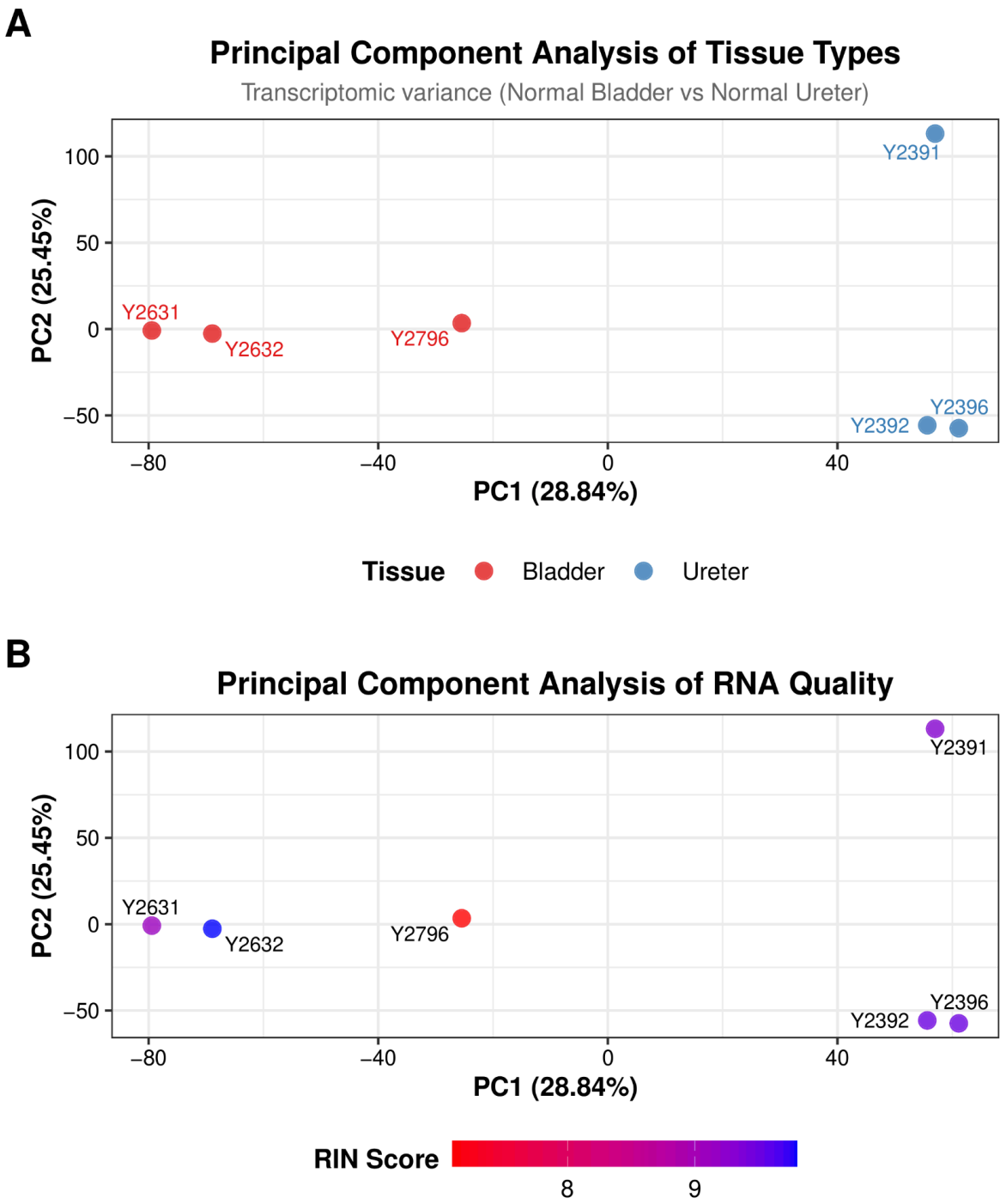

Figure 3.7: Principal component analysis of gene expression data. (A) Principal component analysis showing tissue specific clustering. Bladder samples (red, n=3) and ureter samples (blue, n=3) form distinct, non-overlapping clusters along PC1 (28.84% of variance), whilst PC2 (25.45% of variance) reflects intra-group variation. Sample identifiers are displayed. (B) The same principal component analysis with the samples coloured by RNA integrity number (RIN) score. RIN values range from 7.1 (red) to 9.8 (blue). Samples do not cluster according to RIN. Together, PC1 and PC2 explain 54.29% of total variance.

Gene loading analysis showed distributed contributions of many genes rather than dominance by individual genes. Top PC1 contributors had uniform loading magnitudes

(mean: 0.0150). Contributors included *LINC02518*, *NKTR*, *TTC39C* and *SLC35F6*, which have diverse functions including transcriptional regulation, immune function, cell cycle regulation and mitochondrial transport.

PC2 contributors (mean loading: 0.0160) included cell cycle regulators (*CDK2AP2*), metabolic enzymes (*PECR*), cytoskeletal genes (*INF2*) and signalling molecules (*PTPRK*).

3.4.3 Differential expression analysis

Differential gene expression analysis was performed comparing normal bladder urothelium (n=3) against normal ureter urothelium (n=3) using DESeq2 (v1.42.0). Genes were defined as significantly differentially expressed if they had a false discovery rate (FDR) adjusted *p*-value < 0.05 following Benjamini-Hochberg correction. A total of 4,439 genes (29.4% of input) were excluded due to NA values in adjusted *p*-values.

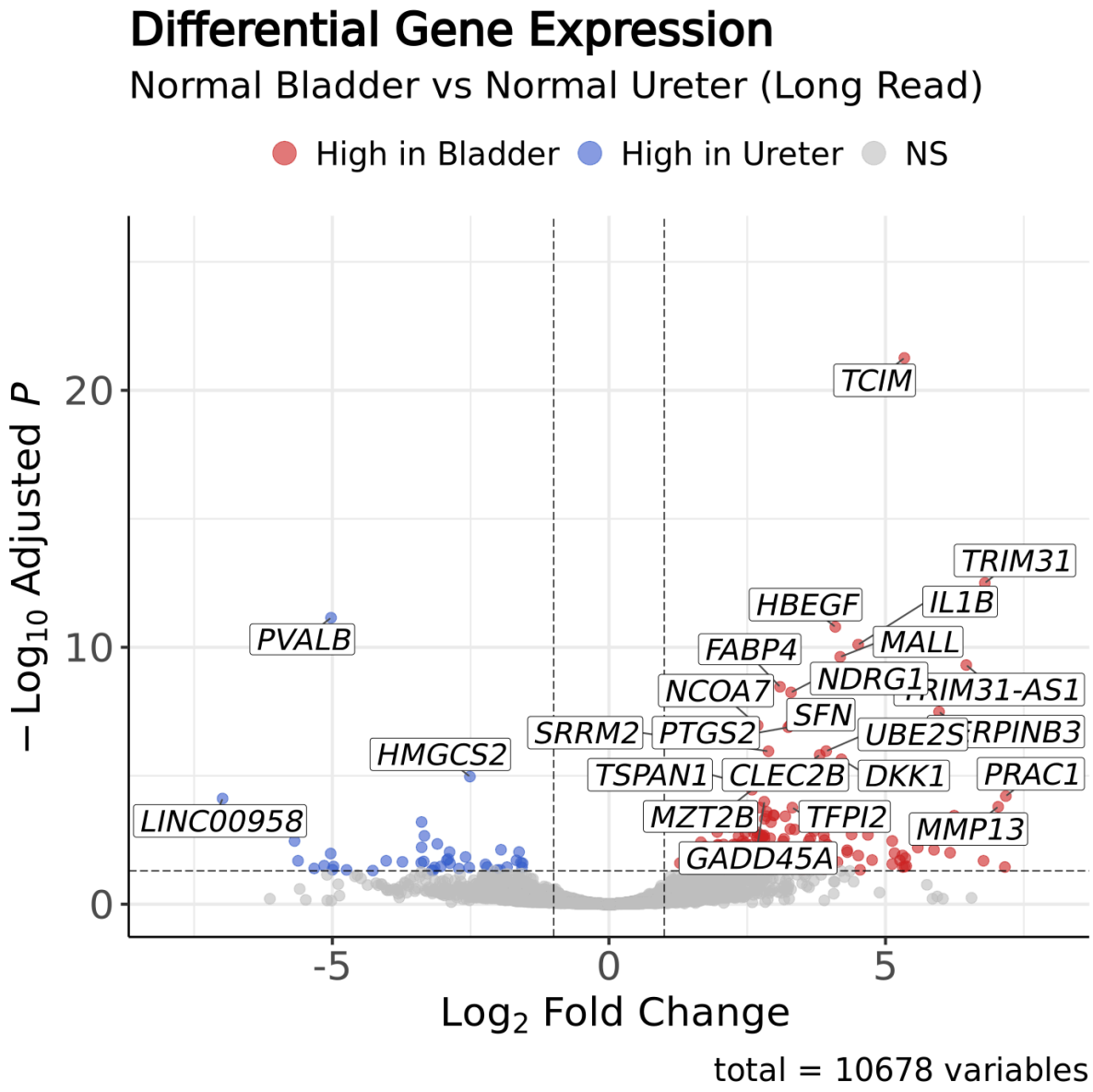

Figure 3.8: Differential gene expression between normal bladder and ureter urothelium. A volcano plot displaying log2 fold change versus adjusted *p*-value for

differential gene expression analysis between normal bladder urothelium (n=3) and normal ureter urothelium (n=3). Each point represents a single gene from 10,678 genes tested. Positive log2 fold changes indicate higher expression in bladder urothelium and negative values indicate higher expression in ureter urothelium. Red points are genes significantly upregulated in bladder, blue points are genes significantly upregulated in ureter and grey points represent non-significant changes. Dashed lines mark significance thresholds (vertical: ±1.0 log2FC and horizontal: adjusted $p = 0.05$). The top 25 most significantly differentially expressed genes are labelled. A total of 239 genes (2.24%) were significantly differentially expressed, with 197 upregulated in bladder and 42 upregulated in ureter.

Of the 10,678 genes tested, 239 (2.24%) had significant differential expression, Figure 3.8. The distribution of differentially expressed genes (DEGs) favoured bladder upregulation, with 197 genes upregulated (82.4% of DEGs) and 42 genes downregulated (17.6% of DEGs) relative to the ureter. Upregulated genes in the bladder had a mean log2 fold change (log2FC) of 3.00, whilst genes upregulated in the ureter (downregulated in bladder) had a mean log2FC of -3.34.

The most upregulated gene in the bladder was *TRIM31* (log2FC = 6.80), with other upregulated genes such as *SERPINB3* (log2FC = 5.97), *TCIM* (log2FC = 5.34), *IL1B* (log2FC = 4.51), *MALL* (log2FC = 4.18), *HBEGF* (log2FC = 4.09) and *SFN* (log2FC = 3.27). The most significantly upregulated genes in ureter were *PVALB* (log2FC = -5.02) and *HMGCS2* (log2FC = -2.51).

3.4.4 Biological pathway and gene expression analyses

Targeted biological pathway analyses were performed to characterise functional differences between bladder and ureter urothelium. Gene expression was normalised to transcripts per million (TPM) to account for both sequencing depth and gene length, allowing the direct comparison of expression levels across samples and genes.

3.4.4.1 HOX gene expression

HOX expression was analysed across 39 genes in four chromosomal clusters to determine whether developmental programmes persist in adult tissues.

Overall HOX pathway activity was higher in ureter (3.73 ± 1.51 TPM) versus bladder (1.23 ± 0.97 TPM, $p = 0.085$, Welch's two sample t-test, n=3 bladder, n=3 ureter), Figure 3.9A. The HOXB cluster had the largest difference (ureter: 8.17 TPM, bladder: 2.07 TPM), with HOXD also elevated in ureter (2.26 vs 0.45 TPM), Figure 3.9B. HOXA expression was similar between tissues (bladder: 2.12 TPM, ureter: 3.94 TPM). HOXC showed negligible expression in both tissues.

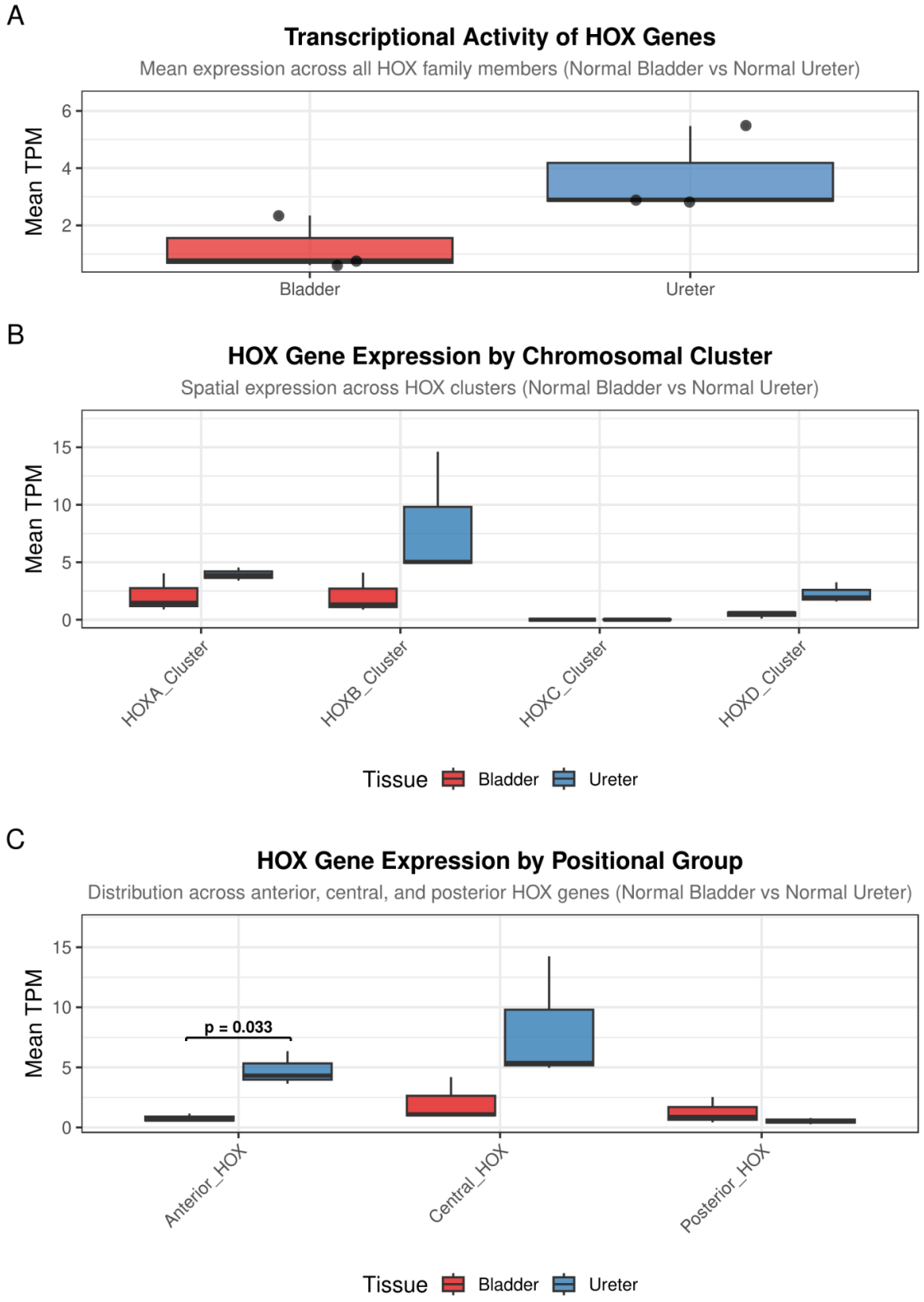

Figure 3.9: HOX gene expression between normal bladder and ureter urothelium. Box plots showing mean TPM of overall HOX expression, HOX expression by cluster and HOX expression by positional group in bladder (red, n=3) and ureter (blue, n=3) urothelium with

individual samples overlaid as points. Statistical significance was assessed using Welch's two-sample t-test, brackets with *p*-values indicate significant differences between groups. **(**A) A boxplot showing overall HOX pathway activity score (mean expression across 39 detected genes) with individual samples overlaid as points, bladder (red, n=3) and ureter (blue, n=3) samples. (B) A boxplot showing mean HOX gene expression (TPM) across chromosomal clusters in bladder and ureter. HOXB and HOXD clusters are higher in the ureter than the bladder samples. (C) A box plot showing mean HOX gene expression grouped by anterior-posterior position. Anterior and central HOX genes had increased expression in the ureter compared to the bladder samples.

Positional analysis showed significantly elevated anterior HOX genes in the ureter compared to the bladder (ureter: 4.77 TPM, bladder: 0.77 TPM, $p = 0.033$, Welch's two sample t-test, n=3 bladder, n=3 ureter), Figure 3.9C. Central HOX genes were also elevated (ureter: 8.19 TPM, bladder: 2.06 TPM) in ureter tissue, whilst posterior genes were higher in bladder (ureter: 0.53 TPM, bladder: 1.25 TPM).

3.4.4.2 Sonic hedgehog signalling pathway

Pathway activity was analysed across 36 genes to determine whether differential expression persists in adult bladder and ureter tissues.

Expression of SHH pathway associated genes was higher in bladder urothelium (6.70 ± 2.18 TPM) versus ureter urothelium (2.65 ± 1.10 TPM, $p = 0.065$, Welch's two sample t-test, n=3 bladder, n=3 ureter), Figure 3.10A.

This difference was primarily driven by elevated expression of the SHH ligand itself, which was higher in bladder (35.6 TPM) versus ureter (10.7 TPM), along with pathway associated genes including *VEGFA* (bladder: 20.4 TPM, ureter: 7.8 TPM) and *ETS2* (bladder: 130.5 TPM, ureter: 36.7 TPM). However, the canonical downstream target *GLI1* showed minimal expression (~0.04 TPM) in both tissues.

3.4.4.3 Fibroblast growth factor signalling pathway

The expression of 55 pathway genes was analysed to determine whether differential activity exists between adult tissues.

FGF pathway activity showed similar expression between bladder (22.56 ± 5.01 TPM) and ureter (26.72 ± 6.97 TPM, $p = 0.453$, Welch's two sample t-test, n=3 bladder, n=3 ureter), Figure 3.10B. Individual bladder samples ranged from 18.99 to 28.28 TPM, whilst ureter samples ranged from 19.28 to 33.10 TPM.

3.4.4.4 Wnt beta-catenin signalling pathway

Expression of 42 pathway genes was analysed to determine whether differential activity exists between adult tissues.

Wnt pathway activity was higher in bladder (7.37 ± 3.18 TPM) versus ureter (3.43 ± 1.46 TPM), though this did not achieve statistical significance ($p = 0.152$, Welch's two sample t-test, n=3 bladder, n=3 ureter), Figure 3.10C. Individual bladder samples ranged from 5.47 to 11.04 TPM, whilst ureter samples ranged from 2.28 to 5.08 TPM.

3.4.4.5 TGF-beta signalling pathway

The expression of 54 pathway genes was analysed to determine whether differential activity exists between adult tissues.

TGF-beta pathway activity showed similar expression between bladder (33.95 ± 3.44 TPM) and ureter (38.15 ± 14.59 TPM, p = 0.671, Welch's two sample t-test, n=3 bladder, n=3 ureter), Figure 3.10D. Individual bladder samples ranged from 31.18 to 37.81 TPM with low variability, whilst ureter samples showed higher variability ranging from 29.22 to 54.99 TPM.

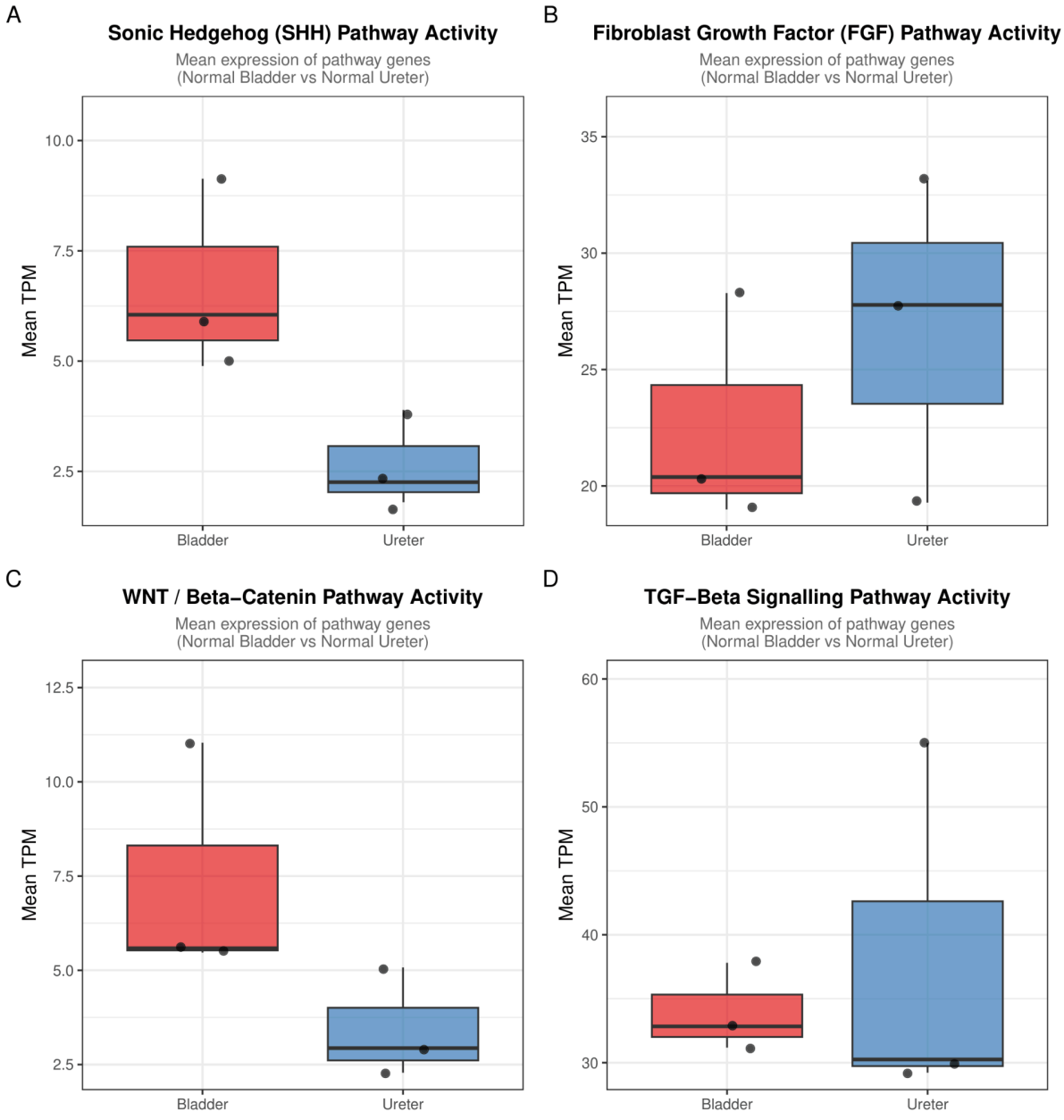

Figure 3.10: Signalling pathway expression between normal bladder and ureter urothelium. Box plots showing overall pathway activity scores (mean TPM) in bladder (red, n=3) and ureter (blue, n=3) urothelium with individual samples overlaid as points. Statistical significance was assessed using Welch's two-sample t-test, brackets with p-values indicate significant differences between groups. (A) SHH pathway activity score (mean TPM across

36 pathway genes), bladder urothelium showed higher activity than ureter though this did not reach significance. (B) FGF pathway activity score (mean TPM across 55 pathway genes), bladder and ureter urothelium showed similar activity. (C) Wnt/beta-catenin pathway activity score (mean TPM across 42 pathway genes), bladder urothelium showed higher activity than ureter, though this did not reach significance. (D) TGF-beta pathway activity score (mean TPM across 54 pathway genes), bladder and ureter urothelium showed similar activity.

3.4.4.6 Interferon response

The expression of 28 interferon and interferon receptor genes was analysed to determine whether differential activity exists between adult bladder and ureter tissues.

Overall interferon pathway activity was elevated in bladder (5.49 ± 2.56 TPM) versus ureter (3.52 ± 3.28 TPM), though this did not achieve statistical significance (p = 0.460, Welch's two sample t-test, n=3 bladder, n=3 ureter), Figure 3.11A. Individual bladder samples ranged from 3.40 to 8.35 TPM, whilst ureter samples ranged from 0.87 to 7.30 TPM.

Type II interferons had the largest difference, with bladder expression (44.40 ± 60.34 TPM) substantially exceeding ureter (0.81 ± 1.40 TPM). However, considerable heterogeneity was observed, with one bladder sample showing 113.49 TPM whilst others showed 2.07 and 17.64 TPM. Type II receptors showed similar expression between tissues (bladder: 47.50 ± 6.46 TPM, ureter: 41.89 ± 42.37 TPM), as did Type III receptors (bladder: 1.71 ± 0.60 TPM, ureter: 1.49 ± 0.77 TPM) and Type I receptors (bladder: 3.57 ± 1.26 TPM, ureter: 4.69 ± 2.47 TPM), Figure 3.11C.

3.4.4.7 Immune cell marker expression

To characterise immune cell populations, the expression of 14 cell type-specific markers was analysed across nine immune cell categories (naive CD4+ T cells, memory CD4+ T cells, CD8+ T cells, B cells, NK cells, CD14+ monocytes, FCGR3A+ monocytes, dendritic cells, and platelets) to determine whether differential immune cell infiltration exists between adult bladder and ureter tissues.

Overall immune marker expression was significantly higher in bladder (101.57 ± 7.62 TPM) versus ureter (47.70 ± 17.30 TPM, p = 0.022, Welch's two sample t-test, n=3 bladder, n=3 ureter), Figure 3.11B. Individual bladder samples ranged from 92.78 to 106.44 TPM, whilst ureter samples ranged from 27.85 to 59.56 TPM, showing clear tissue separation. Memory CD4+ T cell markers had the largest difference (bladder: 504.10 TPM, ureter: 226.97 TPM), followed by dendritic cell markers (bladder: 197.56 TPM, ureter: 101.13 TPM) and naive CD4+ T cell markers (bladder: 2.62 TPM, ureter: 0.90 TPM). NK cells and CD14+ monocytes also showed elevated expression in bladder tissue, whilst CD8+ T cells and FCGR3A+ monocyte markers had similar expression between tissues, Figure 3.11D.

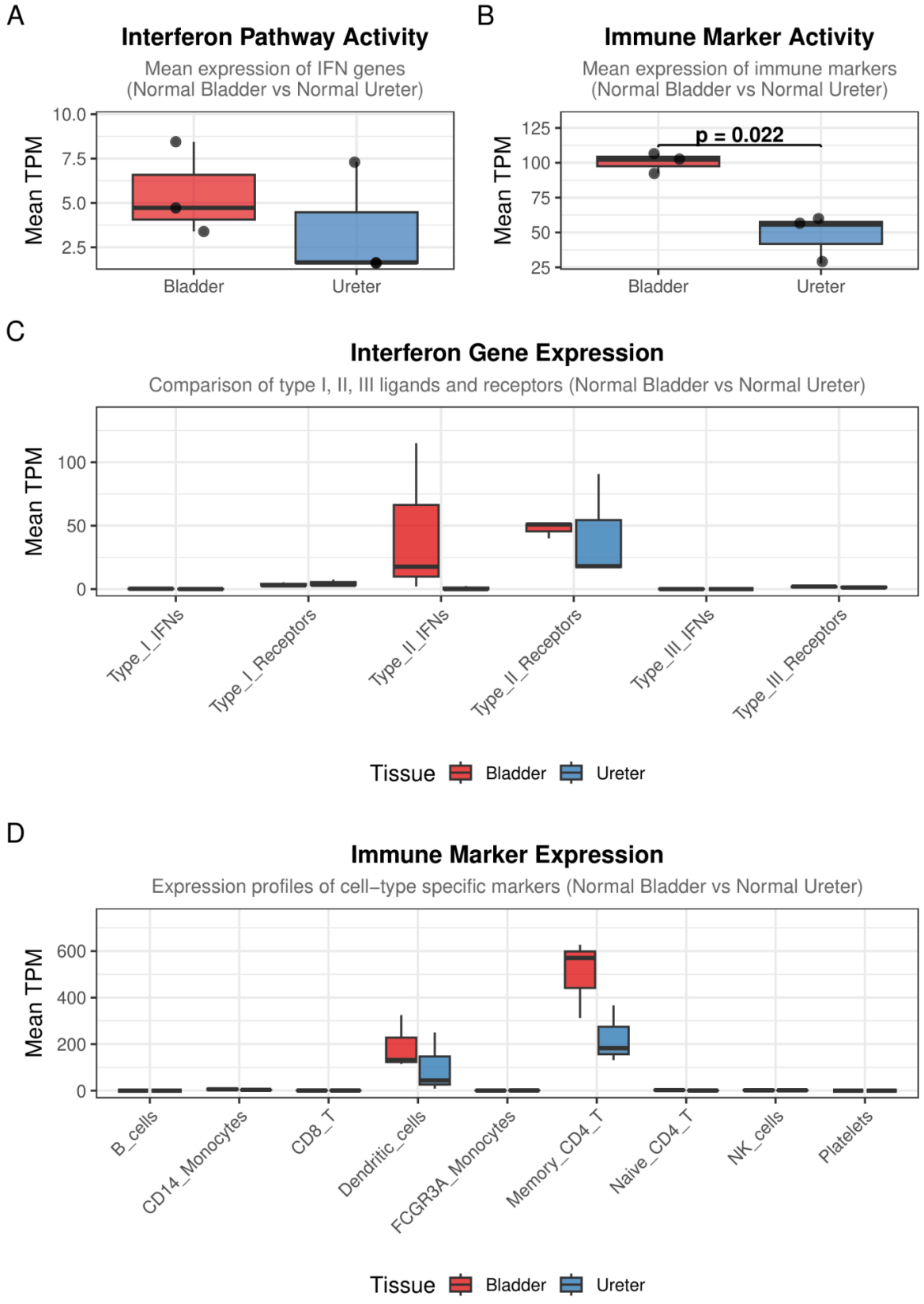

Figure 3.11: Interferon and immune marker gene expression between normal bladder and ureter urothelium. Box plots showing overall activity scores (mean TPM) in bladder (red, n=3) and ureter (blue, n=3) urothelium with individual samples overlaid as points.

Statistical significance was assessed using Welch's two-sample t-test, brackets with *p*-values indicate significant differences between groups. (A) Overall interferon marker gene activity score (mean TPM across 28 genes), bladder urothelium showed higher activity than ureter though this did not reach significance. (B) Overall immune marker gene activity score (mean TPM across 14 genes), bladder urothelium showed significantly higher activity than ureter. (C) Mean gene expression (TPM) across IFN type I, II and III and their respective receptors in bladder and ureter. (D) Mean gene expression (TPM) across immune cell types in bladder and ureter.

3.4.4.8 Cell cycle and proliferation

The expression of 56 genes across six cell cycle categories (G1/S transition, S phase, G2/M transition, M phase, proliferation markers, and checkpoint control) was analysed to determine whether differential proliferative activity exists between adult bladder and ureter tissues.

Overall proliferation pathway activity was higher in bladder (10.98 ± 3.41 TPM) versus ureter (6.04 ± 2.73 TPM), though this did not achieve statistical significance ($p = 0.125$, Welch's two sample t-test, n=3 bladder, n=3 ureter), Figure 3.12A. Individual bladder samples ranged from 7.90 to 14.64 TPM, whilst ureter samples ranged from 4.46 to 9.13 TPM. Checkpoint control genes were significantly higher in bladder than in ureter (bladder: 10.26 ± 0.95 TPM, ureter: 5.56 ± 1.29 TPM, $p = 0.01$, Welch's two sample t-test, n=3 bladder, n=3 ureter). Bladder had higher expression for both G1/S transition genes (bladder: 23.63 ± 4.67 TPM, ureter: 11.63 ± 8.01 TPM) and G2/M transition genes (bladder: 7.18 ± 7.36 TPM, ureter: 2.74 ± 0.48 TPM). S phase genes showed similar expression between tissues (bladder: 6.22 ± 6.59 TPM, ureter: 5.78 ± 1.38 TPM), though with higher bladder variability, Figure 3.12C.

3.4.4.9 Luminal and basal marker expression

Expression of 52 differentiation markers (20 luminal, 20 basal, 12 EMT) was analysed. Overall marker activity was significantly higher in bladder (211.52 ± 25.66 TPM) versus ureter (122.21 ± 32.48 TPM, $p = 0.022$, Welch's two sample t-test, n=3 bladder, n=3 ureter), Figure 3.12B. All categories had elevated bladder expression, with luminal markers (478.29 vs 257.96 TPM, $p = 0.016$, Welch's two sample t-test, n=3 bladder, n=3 ureter) and EMT markers (24.26 vs 16.93 TPM, $p = 0.026$, Welch's two sample t-test, n=3 bladder, n=3 ureter) having significantly higher expression in the bladder. Basal markers were also higher in the bladder but did not reach significance (48.99 vs 45.80 TPM).

Luminal-basal (L/B) ratios were similar between tissues (bladder: 3.36, ureter: 2.53, $p = 0.212$, Welch's two sample t-test, n=3 bladder, n=3 ureter).

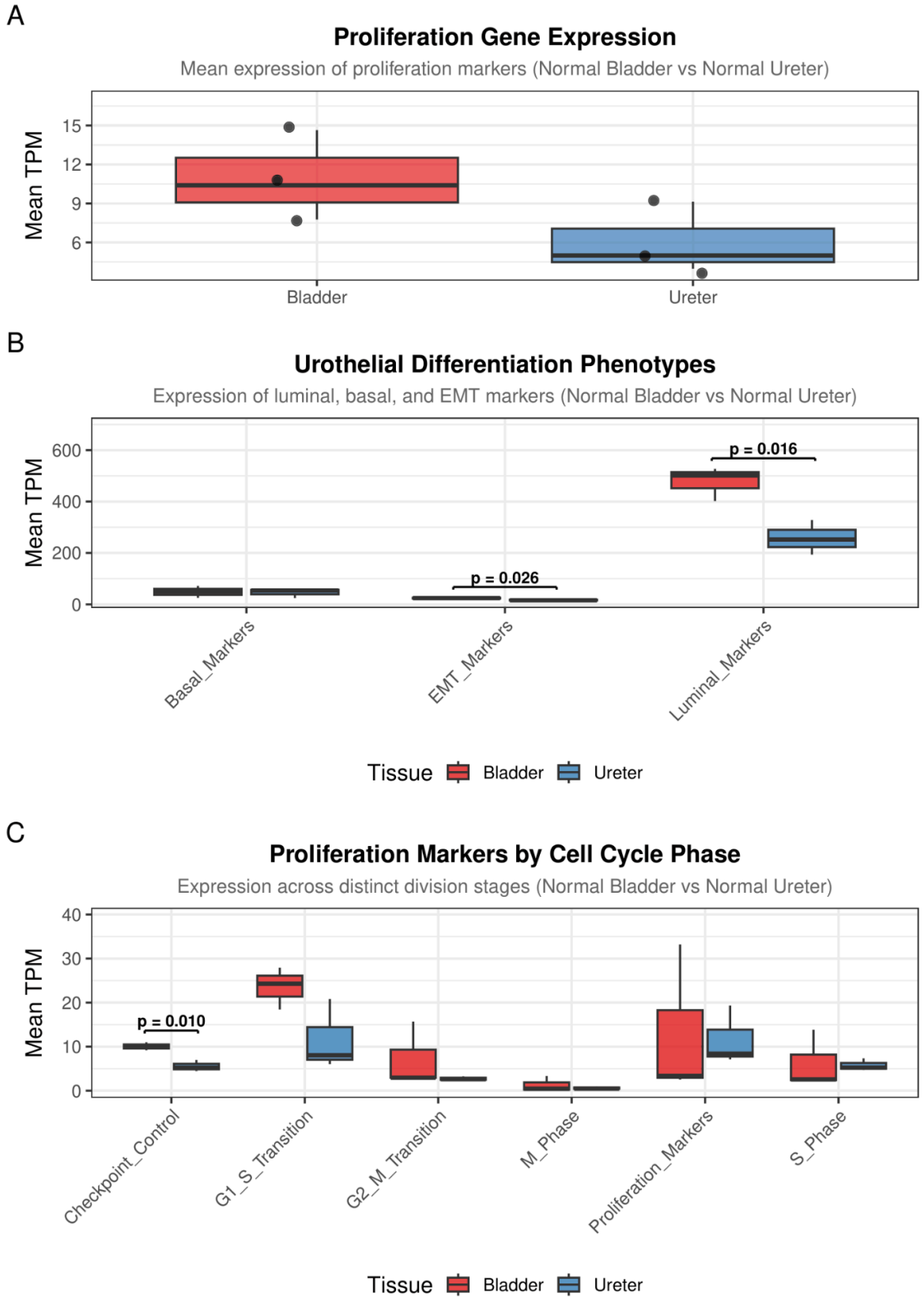

Figure 3.12: Cell cycle, proliferation and urothelial marker gene expression between normal bladder and ureter urothelium. Box plots showing mean gene expression (TPM) in bladder (red, n=3) and ureter (blue, n=3) urothelium, with individual samples overlaid as

points. Statistical significance was assessed using Welch's two-sample t-test, brackets with *p*-values indicate significant differences between groups. (A) Overall cell cycle and proliferation marker gene activity score (mean TPM across 56 genes), bladder urothelium showed higher activity than ureter though this did not reach significance. (B) Mean expression (TPM) of urothelial markers by category (luminal, basal, and EMT) in bladder and ureter, luminal and EMT markers had significantly higher expression in bladder compared to ureter urothelium. (C) Mean gene expression (TPM) across cell cycle phases and checkpoints in bladder and ureter, checkpoint control genes were significantly upregulated in the bladder.

3.4.5 Gene set enrichment analysis

Gene Set Enrichment Analysis (GSEA) was performed to identify biological pathways and processes enriched in normal bladder urothelial tissue compared to ureter urothelium using long-read RNA sequencing data. The analysis used two complementary gene set collections, the Hallmark gene sets and Reactome pathways. The ranked gene list, comprising 10,678 genes, was derived from differential expression analysis comparing bladder samples (n=3) to ureter samples (n=3), with ureter serving as the reference condition.

3.4.5.1 Hallmark gene sets

The GSEA analysis of Hallmark gene sets showed that there was enrichment of inflammatory and immune related pathways in bladder tissue. Of the 47 gene sets tested using an fgsea exact permutation test (n=3 bladder, n=3 ureter), 17 achieved statistical significance at the FDR-corrected threshold (adjusted $p < 0.05$), whilst 22 achieved nominal significance (p-value < 0.05). All significant pathways showed positive enrichment (NES > 0), indicating consistent upregulation in bladder relative to ureter tissue.

Inflammatory signalling pathways predominated in bladder urothelium. The pathway with the strongest enrichment was TNF alpha signalling via NF-Kappa B (NES = 2.03, adjusted $p = 1.41 \times 10^{-10}$), followed closely by inflammatory response (NES = 1.97, adjusted $p = 2.17 \times 10^{-6}$). There was also significant enrichment of IL-2/STAT5 signalling (NES = 1.84, adjusted $p = 5.62 \times 10^{-4}$) and interferon gamma response (NES = 1.84, adjusted $p = 5.62 \times 10^{-4}$).

Cell cycle related pathways were also significantly enriched, including G2M checkpoint (NES = 1.60, adjusted $p = 2.35 \times 10^{-2}$) as well as the apoptosis pathway (NES = 1.76, adjusted $p = 3.63 \times 10^{-3}$) and allograft rejection (NES = 1.71, adjusted $p = 1.08 \times 10^{-2}$).

3.4.5.2 Reactome pathways

The Reactome pathway analysis complemented and extended the Hallmark findings. Of the 811 pathways tested using an fgsea exact permutation test (n=3 bladder, n=3 ureter), 7 achieved FDR significance (adjusted $p < 0.05$), whilst 127 had nominal significance (p-value < 0.05). Consistent with the Hallmark results, 6 of the 7 FDR-significant pathways showed positive enrichment in bladder tissue, with only one pathway showing enrichment in the ureter urothelium.

The most significant pathway identified was cytokine signalling in the immune system (NES = 1.81, adjusted $p = 6.34 \times 10^{-9}$). This was followed by signalling by interleukins (NES = 1.82, adjusted $p = 1.44 \times 10^{-5}$) and interleukin-4/13 signalling (NES = 1.89, adjusted $p = 3.59 \times 10^{-3}$).

Interferon signalling was another key enriched pathway (NES = 1.70, adjusted $p = 1.28 \times 10^{-2}$), corroborating the interferon gamma response observed in the Hallmark analysis. Additional immune related pathways demonstrating significance include interleukin-1 family signalling (NES = 1.75, adjusted $p = 2.66 \times 10^{-2}$) and interleukin-10 signalling (NES = 1.82, adjusted $p = 3.46 \times 10^{-2}$).

Only one pathway had significant enrichment in ureter tissue, transcriptional regulation by MECP2 (NES = -2.01, adjusted $p = 1.63 \times 10^{-2}$).

3.4.6 Transcript quantification and expression analysis

Gene level quantification was performed using featureCounts (version 2.0.6) with the GENCODE v47 annotation. Assignment rates ranged from 53.0% (Y2396) to 65.2% (Y2391), with mean 56.7%, Table 3.7. Sample Y2391 had the highest amount of reads assigned (3.9 million), whilst remaining samples showed similar rates (53.0-57.1%). Overall, 7.5 million reads (59.9% of 12.6 million processed) were assigned to genes.

Sample	Tissue	Total Reads	Assigned reads	Assigned (%)	Unmapped reads	Unmapped (%)	Ambiguous reads	Ambiguous (%)
Y2391	Ureter	5,987,602	3,902,720	65.18	758,931	12.68	961,833	16.06
Y2392	Ureter	1,175,691	640,021	54.44	265,370	22.57	85,252	7.25
Y2396	Ureter	1,250,392	663,085	53.03	271,039	21.68	72,707	5.81
Y2631	Bladder	1,251,509	709,095	56.66	299,939	23.97	70,477	5.63
Y2632	Bladder	1,407,666	803,347	57.07	354,097	25.15	65,080	4.62
Y2796	Bladder	1,481,308	800,176	54.02	327,512	22.1	88,073	5.95
Mean	-	2,092,361	1,253,074	56.73	379,481	21.36	223,904	7.55

Table 3.7: Gene assignment statistics from featureCounts quantification of long-read cDNA sequences for normal bladder and ureter urothelial samples. Gene assignment statistics from featureCounts (version 2.0.6) using GENCODE v47 annotation and long-read parameters (-L, --largestOverlap) for six normal urothelial samples (ureter: Y2391, Y2392, Y2396, bladder: Y2631, Y2632, Y2796). Total reads includes both mapped and unmapped alignments from Minimap2. Assigned reads shows successful gene assignments, assigned (%) indicates the percentage. Unmapped reads represents alignment failures, unmapped (%) shows the percentage. Ambiguous reads indicates reads overlapping multiple genes, ambiguous (%) shows the percentage. All percentages relative to total reads. Mean values shown across all samples.

Unassigned reads comprised three categories, unmapped reads (12.7-25.2%, mean: 21.4%), reads with no overlapping features (6.1-19.5%, mean: 14.4%), representing successful alignments to unannotated or intronic regions, and ambiguous reads (4.6-16.1%, mean: 7.6%), overlapping multiple genes. Sample Y2391 had the lowest unannotated region mapping (6.1%) but highest ambiguity (16.1%).

Ureter samples had a mean assignment rate 57.6 ± 6.8%, slightly higher than bladder (55.9 ± 1.6%). Unannotated region mapping (ureter: 13.8 ± 6.8%, bladder: 15.0 ± 2.5%) and

ambiguity rates (ureter: 9.7 ± 5.6%, bladder: 5.4 ± 0.7%) were similar between tissues. The relationship between mapping and assignment rates showed 67.7-76.3% of mapped reads (mean: 72.1%) successfully assigned to genes, indicating approximately 29.3% of aligned reads map outside annotated gene boundaries.

FeatureCounts reported zero multi-mapping reads, ensuring quantified expression reflects uniquely mapping reads only. Negligible overlapping feature counts (0-2 reads per sample) were detected across the samples.

Overall, 7.5 million reads were successfully assigned across samples, with consistent quantification efficiency between tissues.

3.4.7 Isoform switch analysis

Isoform switch analysis was performed to identify genes exhibiting differential isoform usage between bladder and ureter urothelial tissue using the IsoformSwitchAnalyzeR package. The analysis examined 10,993 isoforms from 3,743 genes across three bladder and three ureter samples. No statistically significant isoform switches were detected in the long-read dataset using standard filtering criteria (dIF cutoff = 0.1, q-value < 0.05).

3.5 Discussion

3.5.1 Technical validation and platform concordance

The processed datasets met recommended quality thresholds for RNA-seq analysis. With mean Q30 scores exceeding 90% and total mapping rates above 98% for short read data, these results are consistent with benchmarks for high quality libraries (Dobin *et al.*, 2013; Hu *et al.*, 2024; Wang *et al.*, 2024). A possible explanation for the somewhat higher featureCounts assignment rates in bladder samples might be their longer read lengths, as longer reads span more of a transcript and are therefore more easily assigned to annotated features with confidence. FastQC warnings for per-base sequence content and elevated duplication levels were not considered indicative of poor data quality. Rather than indicating technical errors, sequence content bias is a known artefact of non-random hexamer priming, and high duplication rates are generally associated with highly expressed transcripts rather than technical error (Hansen, Brenner and Dudoit, 2010; Tan *et al.*, 2019). Regarding the long read data, initial FastQC warnings for adapter content were fully anticipated as nanopore cDNA libraries inherently retain sequencing adapters in their raw output that are subsequently removed during preprocessing rather than indicating poor initial data quality.

The low full length read recovery by Pychopper (8-11% of input) in the long read data, is in line with the known challenges of cDNA based nanopore sequencing where RNA degradation, incomplete reverse transcription or adapter loss during library preparation can prevent primer detection at both transcript ends (Verwilt, Mestdagh and Vandesompele, 2023). The mean mapping rate of 78.6% for long read data is considerably lower than the >98% achieved with short read data. This is an expected result given the distinct error profiles of nanopore sequencing and the more stringent full length read selection applied during preprocessing. The gene level assignment rate of 56.7% is similarly lower than the

short read equivalent (65.8%). It should be noted that sample Y2391 presented a substantial depth discrepancy. This technical imbalance drove a slightly higher mean assignment rate and ambiguity in the ureter cohort. However, when this outlier is excluded, assignment rates were comparable across tissues suggesting quantification efficiency is not inherently biased by tissue type. This discrepancy could be attributed to the reduced sequencing depth following full length filtering rather than alignment failure. Additionally, the observation that a substantial proportion of long reads (14.4%) mapped to unannotated regions is consistent with the capacity of long read sequencing to capture active transcription from novel loci or intergenic regions absent from GENCODE annotations. Furthermore, the complete absence of multi-mapping reads ensures that the quantified expression values robustly reflect uniquely mapping transcripts, thereby strengthening confidence in downstream analyses.

Across both platforms, a consistent ~2 percentage point elevation in GC content was observed in bladder samples relative to ureter, persisting across preprocessing steps and sequencing types. This cross platform concordance strongly suggests that the GC content difference is driven by a genuine transcriptomic distinction between tissue types rather than a technical artefact of either sequencing chemistry. Similarly, the clear separation of normal bladder and ureter samples along PC1 in both short read and long read PCA confirms that tissue identity is the dominant driver of transcriptomic variation across both datasets. The distributed PC1 gene loading magnitudes in both analyses indicate that tissue specific transcriptomic differences reflect coordinated changes across many genes rather than dramatic shifts in a small number. Biologically, these distributed contributors encompass system level differences rather than pathway specific changes, spanning membrane transport, ion binding, and nuclear pore components in the short read data, and transcriptional regulation, immune function, and mitochondrial transport in the long read data. The absence of RIN score clustering in either dataset indicates that RNA quality did not substantially drive transcriptomic variation. On the other hand, the separation of sample Y2796 along PC2 may reflect genuine biological variation in proliferative state. A possible explanation for this might be its potential association with bacterial contamination as it tested positive in a lysogeny broth test, whilst its low RIN score may also have contributed to its divergent clustering (Gallego Romero *et al.*, 2014). PC2 contributors were enriched for cell cycle regulators and cytoskeletal genes across both platforms. Although they presented non-overlapping gene sets, these results are likely to be related to differences in proliferative state or cellular stress across samples (Eymin *et al.*, 2001; Nogueira *et al.*, 2010; Kara *et al.*, 2015; Wang *et al.*, 2020; Dai *et al.*, 2023; Luo *et al.*, 2023; Ueda *et al.*, 2023; Yu *et al.*, 2023; Kurniawan *et al.*, 2024; Pellarin *et al.*, 2025; Zhao *et al.*, 2025). It is important to bear in mind that discrepancies between platforms were occasionally observed. In such cases the short read results were prioritised due to their superior sequencing depth. The agreement between platforms for the major biological themes, including HOX signatures, immune markers, proliferation patterns and GC content differences validates these differences and the conclusions drawn from them.

3.5.2 Retention of embryological identity

A recurrent theme in the transcriptomic data was an embryological signature distinguishing the two tissues. These findings are consistent with earlier histological and culture based evidence demonstrating that the urothelium comprises developmentally distinct lineages (Liang *et al.*, 2005; Riedel *et al.*, 2005). Previous studies have shown that the upper tract

differs from the bladder in terms of embryonic origin, uroplakin content and keratin expression. Because these differences persist after serial culture under identical conditions, it can be concluded that they are intrinsic to the cells rather than imposed by the local microenvironment.

This intrinsic molecular distinctiveness is heavily driven by HOX pathway activity. In the short read data the ureter had significantly higher activity in the HOXB and HOXD clusters and enrichment for anterior-central positional genes. In contrast, the long read data had higher posterior HOX activity in the bladder. The enrichment of these HOX transcription factors in the ureter is consistent with their established roles in anterior-posterior axis patterning (Inamura *et al.*, 2007; Hayashida *et al.*, 2010; Salmanidis *et al.*, 2013). Taken together, these results provide converging evidence that the adult urothelium preserves its embryological positional identity, specifically the mesodermal origin of the upper tract versus the endodermal origin of the bladder.

This developmental signature extends beyond only HOX genes. The specific enrichment of *PAX8* in the ureter aligns with its role in nephric duct development from which the upper tract derives, in contrast to the bladder which arises from the urogenital sinus (Tacha, Zhou and Cheng, 2011). The strong ureteric enrichment of *SIM1*, whilst classically characterised as a hypothalamic transcription factor, reflects its pleiotropic role in embryological differentiation extending to the nephric lineage (Brunskill *et al.*, 2008; Duplan *et al.*, 2009; Liu *et al.*, 2019). Additionally in the long read data, the ureter's isolated enrichment for transcriptional regulation by *MECP2*, a repressor involved in gene silencing, suggests that epigenetic regulatory mechanisms may maintain this upper tract tissue identity (Cheng and Qiu, 2014).

The finding that most supports these claims is the single ureter enriched pathway identified by GSEA, the formation of the nephric duct, the exact developmental structure from which the ureter originates (Nagalakshmi and Yu, 2015). This is supported by recent single cell transcriptomic profiling of the developing mammalian urinary tract. During embryonic development the specification of nephric duct and ureteric bud progenitors is characterised by a core gene signature encompassing *PAX8* alongside the HOX A/B/C clusters (Sanchez-Ferras *et al.*, 2021). The detection of this exact transcriptional signature in the adult ureteric urothelium in the present study suggests that these tissues do not converge into a uniform urothelium but that the adult upper tract urothelium preserves the positional identity.

3.5.3 Functional and physiological specialisation

The transcriptomes reflect the ongoing functional requirements of each tissue. Specifically, the bladder urothelium is adapted for barrier protection and repeated cycles of physical distension, whereas the ureter serves continuously as a peristaltic conduit. These differing mechanical and chemical environments act as drivers of transcriptomic variation. Due to these differences, the differential expression analysis identified distinct sets of upregulated genes for each anatomical site.

The bladder enrichment of desmosomal and protocadherin adhesion molecules alongside *IGFBP5* in the bladder suggests a robust barrier architecture suited to withstanding the requirement for urine storage and physical distension (Zhou *et al.*, 2017; Viehweger *et al.*,

2022; Rodríguez-Rojas *et al.*, 2024). Upregulation of the serine peptidase inhibitors *SERPINB3* and *SERPINB4* may further reflect a heightened defence against the concentrated solutes present in stored urine (Ho *et al.*, 2012; Harish and Uppuluri, 2018; Chuang *et al.*, 2023). In the long read data the pronounced expression of transcription regulators, such as *TCIM*, alongside genes driving inflammatory and growth factor signalling including *TRIM31*, *HBEGF* and *IL1B*, highlight a primed stress response consistent with maintaining the urothelial barrier (Chua *et al.*, 2000; Kim *et al.*, 2009; Lopez-Castejon and Brough, 2011; Tolino, Block and Klarlund, 2011; Guo *et al.*, 2018). Furthermore, the bladder's reliance on robust structural defence is supported by the enrichment of specific barrier protection and cell cycle markers such as *SERPINB3*, *MALL*, and *SFN (Magyar et al., 1997; Sano et al., 2004; Ho et al., 2012)*. The significantly higher luminal marker expression in the bladder is consistent with the robust urothelial barrier required for urine storage, whilst the elevated EMT marker expression may support a greater regenerative capacity and cellular plasticity necessary for morphological changes during these repeated distension cycles (Zhang and Weinberg, 2018; Jafari and Rohn, 2022). Importantly, the similar luminal to basal ratios between tissues suggest that despite absolute expression differences, both tissues maintain comparable relative differentiation proportions across the urothelium. The dramatically elevated *OLFM4* expression in the bladder, a marker of intestinal stem cells and an anti-apoptotic factor, may reflect a unique regenerative capacity inherent to the bladder urothelium (Liu *et al.*, 2012).

A likely explanation for the elevated SHH ligand expression in bladder urothelium is that it represents ongoing paracrine epithelial to mesenchymal signalling. This is consistent with developmental patterns where urothelial SHH coordinates stromal responses and suggests that higher baseline signalling is required to maintain adult bladder urothelial stratification and barrier function (Shin *et al.*, 2014). The near absent *GLI1* expression in both tissues indicates that canonical autocrine SHH signalling is quiescent in adult urothelium across both sequencing platforms. The increased expression of *VEGFA* and *ETS2* may reflect the bladder's mechanical stretch as outlined above, as these genes respond to mechanical stimuli and regulate epithelial responses, though they are not SHH specific (Nguyen *et al.*, 2000; Mackenzie and Ruhrberg, 2012). The lower pathway activity in the ureter may reflect the different functional requirements of this tissue, which serves primarily as a conduit and therefore may require less active ongoing SHH signalling to maintain barrier integrity and regulate epithelial remodelling in response to distension (Weiss *et al.*, 2013). Recent evidence suggests a more complex picture regarding SHH activity in the ureter. In a comprehensive single cell and spatial study of the human ureter, Fink et al. (2022) showed that a discrete population of SHH high basal progenitor cells drives organoid formation and urothelial differentiation (Fink *et al.*, 2022). This difference could be due to methodological differences between the studies, likely because a bulk pathway score aggregates average ligand abundance across the entire tissue which may dilute the signal of a rare cell group. The absence of *GLI1* expression observed is consistent with the findings of Fink et al. (2022) who showed that SHH acts in a paracrine, epithelial to stromal way rather than through canonical autocrine signalling.

The significantly elevated and consistent TGF-beta activity in the bladder, characterised by low intra-group variability in both tissues in the short read data suggests that this reflects a genuine tissue level difference rather than sampling noise. The demands of repeated bladder contraction and distension may require greater extracellular matrix maintenance and

tissue remodelling. Given that TGF-beta regulates collagen production and matrix deposition, which are crucial for maintaining structural integrity during filling and voiding cycles, this finding is consistent with its known roles (Streuli *et al.*, 1993; Chang *et al.*, 1998; Aitken and Bägli, 2009). Whilst the long read sequencing showed similar expression across tissues, the higher depth short read data suggests that the elevated TGF-beta signalling in bladder tissue represents a genuine biological difference.

Whilst statistically significant, the relatively modest magnitude of the FGF pathway difference suggests that both tissues maintain broadly similar baseline FGF pathway activity, and the similar expression observed in long read sequencing supports this interpretation. This baseline similarity is highly clinically relevant as it implies that the highly elevated FGF signalling frequently reported in upper tract urothelial carcinoma (UTUC) represents a cancer specific alteration rather than a pre-existing tissue difference (Sfakianos *et al.*, 2021). The trend towards elevated Wnt activity in the bladder must be interpreted with caution because of the small sample size and relatively high intra-group variability, both of which limit statistical power. Nevertheless, the concordant direction of effect across both sequencing platforms supports a biological difference that, if validated in larger cohorts, could reflect ongoing regulation of epithelial proliferation and differentiation (Shin *et al.*, 2011; Wiessner *et al.*, 2022).

The ureteric urothelium also has a transcriptional profile which reflects its role. The ureteric enrichment of *MATN2* is likely to reflect the distinct extracellular matrix requirements of a peristaltic conduit (Luo *et al.*, 2017). Whilst in the long read data the significant enrichment of calcium binding proteins like *PVALB* likely supports the continuous regulation of muscle contraction necessary for ureteric function (Wang, Martindale and Metzger, 2009). Elevated expression of mitochondrial enzymes involved in ketone body synthesis, such as *HMGCS2*, potentially reflect the distinct metabolic requirements of ureter urothelium adjacent to continuous peristaltic smooth muscle versus the distensible bladder (Asif *et al.*, 2022). More broadly, the upregulation of mitochondrial transcripts in the ureter across both platforms may indicate greater oxidative metabolic demands, potentially linked to the energy requirements of adjacent smooth muscle peristalsis or intrinsic metabolic differences within the ureteric lineage (Duan *et al.*, 2024).

The elevated proliferative gene expression in the bladder observed across both sequencing types, may be related to structural and functional differences between these tissues. The bladder urothelium comprises five to seven cell layers when relaxed, compared to two to three layers in the ureter, potentially requiring more active baseline proliferative machinery to maintain this thicker epithelium (Bolla *et al.*, 2025). Additionally, the bladder's repeated distension and contraction cycles subject the urothelium to mechanical stresses absent in the ureter, which may necessitate a more primed proliferative state for rapid injury response (Li, Hu, *et al.*, 2023; Liu *et al.*, 2024). The significant upregulation of the G1/S and G2/M transition genes in the short read and checkpoint control genes in the long read, rather than a uniform increase across all phases suggests that the bladder urothelium is maintained in a state of greater proliferative readiness rather than active cycling, which is consistent with its normally slow turnover under homeostatic conditions and rapid regeneration capabilities (Kullmann *et al.*, 2017).

3.5.4 Tissue specific immune landscapes

There was a strong inflammatory and immune related transcriptional signature in normal bladder urothelium relative to ureter. TNF-alpha/NF-kappa B signalling, the inflammatory response pathway and allograft rejection were among the top Hallmark enrichments in short read GSEA, suggesting that bladder urothelium has more pronounced activation of inflammatory signalling compared to ureter tissue. This is supported by Reactome analysis, where signalling by interleukins was the most significantly enriched pathway alongside neutrophil degranulation and interferon signalling, which is consistent with enhanced immune surveillance in bladder urothelium. In the long read GSEA, the upregulation of pathways including TNF-alpha, IL-2/STAT5 and the interferon gamma response similarly highlighted a highly active baseline immune state (Mahmud, Manlove and Farrar, 2013; Ivashkiv, 2018). This is further supported by the enrichment of allograft rejection and apoptosis pathways suggesting a tissue microenvironment primed for injury response and cellular turnover (LaRosa, Rahman and Turka, 2007).

The significance of the interferon gamma response in Hallmark analysis and interferon signalling in Reactome analysis across both sequencing platforms provides converging evidence for a role of interferon mediated immunity in the bladder. The elevated interferon receptor expression, particularly Type II receptors in the higher depth short read data, compared to the highly variable ligand expression, suggests that bladder tissue may maintain heightened immune surveillance capacity compared to the upper tract, this was consistent across both sequencing types (Castro *et al.*, 2018). The elevated immune marker expression in the bladder, with a predominance of T cell and NK cell markers is consistent with maintaining antimicrobial defences and was observed consistently across both sequencing platforms, as was the overall elevated immune cell infiltration compared to the ureter (Balin *et al.*, 2018; Sun *et al.*, 2023). Together these findings suggest a higher immune surveillance in the bladder compared to the ureter. A possible explanation for this may be an adaptation to greater microbial exposure from stored urine, potentially priming bladder tissue for stronger interferon mediated responses (Ingersoll and Albert, 2013; Hayes and Abraham, 2016). Furthermore, the enrichment of mTORC1 signalling and programmed cell death pathways alongside inflammatory pathways suggests the activation of cellular stress responses and apoptotic machinery in bladder tissue. The enrichment of cell cycle related pathways including mitotic spindle organisation, p53 pathway activation and G2M checkpoint regulation alongside the inflammatory phenotype may reflect the higher proliferative turnover of bladder urothelium which is subject to greater mechanical and chemical stress than ureter epithelium. The high variability in some immune markers, particularly NK cells and CD8+ T cells in the short read data and dendritic cells and NK cells in the long read data indicates heterogeneous immune infiltration across samples.

3.5.5 Post-transcriptional regulation and isoform switching

Although ureter tissue expresses a larger number of distinct genes and transcripts overall, bladder tissue shows higher expression per gene and maintains similar or slightly elevated transcript diversity per expressed gene. This suggests that bladder urothelium may focus transcriptional resources on a more restricted set of genes but maintains comparable isoform complexity for those genes that are expressed.

The isoform switch analysis reveals distinct post-transcriptional regulatory landscapes between bladder and ureter urothelium. The most significantly switched gene, *GSTP1*, encodes a detoxification enzyme that protects against xenobiotics and its differential isoform usage is particularly relevant given the bladder's prolonged exposure to concentrated urinary metabolites compared to the ureter (Gudur *et al.*, 2024). Prior studies using knockout models have noted the importance of GSTP in protecting the bladder epithelium against toxic urinary metabolites (Conklin *et al.*, 2009). The enrichment of ribosomal protein isoform switches, including *RPL29*, *RPL10A*, *RPS20*, *RPL18* and *RPS15*, may reflect tissue specific regulation of protein synthesis through ribosome heterogeneity suggesting that the two tissues differ not only in what they transcribe but in how they control translation (Genuth and Barna, 2018). There is an emerging concept of the 'specialised ribosome' with recent comprehensive atlases confirming that ribosomal proteins are differentially expressed across various tissues and developmental stages to regulate mRNA translation (Brunchault *et al.*, 2025). The specific isoform switches identified in the data, notably *RPL10A* and *RPL29*, are examples of this. Previous studies have shown that the incorporation of *RPL10A* confers differential translational selectivity to the ribosome (Shi *et al.*, 2017). Other notable switches in *DGKA*, *GADD45B* and *OAS3*, covering lipid signalling and immunity, stress response and interferon response respectively, are consistent with the broader inflammatory and immune transcriptional signature identified by GSEA.

The bladder's consistent bias towards isoforms with reduced protein coding capacity, evidenced by ORF loss, coding to non-coding switching, intron retention and NMD sensitive isoform production, suggests that bladder urothelium uses post-transcriptional mechanisms to selectively reduce protein output from specific genes rather than relying solely on transcriptional control. Retained introns frequently introduce premature termination codons leading to transcript degradation via nonsense mediated decay enabling the rapid reduction of protein levels without transcriptional silencing (Porras-Tobias, Caldera and Castro-Piedras, 2025). This regulatory strategy may reflect the distinct physiological demands of the bladder, which must accommodate repeated cycles of distension and contraction during urine storage, requiring dynamic and rapid protein level adjustments.

The prominence of isoform level differences in both detoxification and translational machinery is particularly relevant for bladder cancer biology, as both have established roles in urothelial carcinogenesis. No statistically significant isoform switches were detected in the long read dataset. This discrepancy could be attributed to insufficient sequencing depth rather than a biological absence of isoform regulation given that the long read dataset achieved only around 10% classification as full length by Pychopper, yielding very low reads assigned to genes per sample. It is important to bear in mind that IsoformSwitchAnalyzeR documentation explicitly states that at least three independent biological replicates per condition are needed for good statistical performance (Vitting-Seerup and Sandelin, 2019). With the minimum n=3 employed here the statistical power is substantially reduced, as isoform level tests must account for both sample biological variability and per gene isoform proportion variability. Therefore, the short read isoform findings stand as the basis for these conclusions as the long read data lacked sufficient depth to confirm or refute them.

Overall, these findings reveal that the transcriptomes of bladder and ureter urothelium differ across three levels of regulation, which genes are expressed, at what level they are expressed and their transcript output post-transcriptionally. The substantial transcriptomic

divergence between these tissues encompasses developmental patterning, barrier function, immune surveillance and post-transcriptional regulation. This supports that the tissues maintain distinct molecular identities. These tissue specific regulatory differences may contribute to the distinct molecular profiles and clinical behaviours of bladder and upper tract urothelial carcinomas and could influence differential disease susceptibility and therapeutic responses.

Chapter 4: Transcriptomic Comparison of Bladder and Upper Tract Urothelial Carcinomas

4.1 Introduction

This chapter presents results of the second research aim, characterising molecular differences between bladder urothelial carcinoma (BUC) and upper tract urothelial carcinoma (UTUC) at gene and transcript levels. As described in Section 1.2.3, whilst bladder and ureter urothelium appear morphologically similar, they give rise to carcinomas with distinct clinical behaviours. Chapter 3 demonstrated that these morphologically similar tissues maintain distinct molecular programmes, with 3,698 differentially expressed genes and coordinated differences in developmental, immune, and differentiation pathways. These baseline molecular differences between normal bladder and ureter urothelium provide context for understanding whether the distinct clinical behaviours of their respective carcinomas arise from malignant transformation of tissues with inherently different molecular identities. UTUC presents more aggressively, with ~60% of cases showing muscle invasion at diagnosis compared to 25% for BUC, and lower five-year survival rates (57% vs 71%) (van Doeveren *et al.*, 2021; Witjes *et al.*, 2021; Lefort *et al.*, 2023; National Cancer Institute, 2023a). Understanding molecular differences in tumours is necessary to explain distinct clinical behaviours and potentially identify site specific therapeutic targets.

To achieve this aim long-read Oxford Nanopore Technologies (ONT) sequencing was performed on seven tumour samples, three BUC and four UTUC. Multiple analytical approaches characterised tissue specific differences in these malignant samples, including gene level differential expression analysis, principal component analysis to identify overall transcriptomic similarity and genes contributing to tumour separation by location, and targeted biological analyses examining known pathways within urothelial carcinomas. This approach enabled comprehensive characterisation of the molecular landscape distinguishing BUC from UTUC. Having identified substantial transcriptomic differences between normal tissues in Chapter 3, this chapter investigates whether similar molecular distinctions persist in malignant samples and whether these explain the more aggressive clinical behaviour of UTUC.

4.2 Description of Samples

Tumour tissue samples were obtained from seven patients who underwent surgical procedures for urothelial carcinoma. The cohort was composed of three bladder cancer specimens and four upper tract cancer specimens, Table 4.1.

Sample ID	Tissue	Age	Sex	Grade and Stage	RIN
1752T	Renal Pelvis	74	M	G3 T4	4.4
2179T	Renal Pelvis	80	F	G3 T2	4.9
2237T	Renal Pelvis	79	M	G3 T3	5.1
504T	Bladder	63	F	G3 T2	6.2
989T	Bladder	71	M	G3 Ta	6.8
1829T	Ureter	67	M	G2 T2	4.9
1525T	Bladder	76	F	G3 T3	4.6

Table 4.1: Clinical and processing characteristics of urothelial carcinoma tissue samples. Clinical and demographic data for seven urothelial carcinoma samples. Bladder samples (n= 3) were from one male and two female patients. Upper tract samples (n=4, 3 renal pelvis, 1 ureter) from 3 male and one female patient. Grade (G) indicates histological differentiation (G2: moderately differentiated, G3: poorly differentiated). Tumor stage (T) indicates the depth of invasion: Ta (non-invasive papillary), T2 (muscle-invasive), T3 (invasion beyond organ wall), T4 (invasion into adjacent tissue). RIN, RNA integrity number. F, female. M, male.

The bladder cancer samples comprised three high-grade tumours, two muscle invasive and one non-muscle invasive. Upper tract cancer samples were made up of three renal pelvis tumours and one ureteric tumour, which were all muscle invasive.

Following RNA extraction, the long-read sequencing libraries were prepared using the Oxford Nanopore Technologies cDNA sequencing protocol and sequenced on the PromethION machine.

Tumour samples were all Optimal Cutting Temperature (OCT)-embedded, and cryosectioned in York before standard TRIzol RNA extraction following well-established protocols, with a subsequent sodium acetate cleanup. RNA QC using the Agilent Technologies TapeStation revealed that RNA in the majority (26/39) of samples was highly degraded with RIN values <4.0, with no samples having a RIN value greater than 7.

Whilst all samples failed the standard RIN value threshold for ONT, samples with RIN greater than 4 were taken forward for sequencing using both the standard and degraded sample cDNA library preparation kits.

Samples were taken forward in the knowledge it was likely the yield would be poor and that the proportion of full length transcripts would be low. However, the possibility of gaining relevant data was considered sufficient justification for continuing with this aim of this thesis.

4.3 Results

4.3.1 Raw sequencing data quality assessment

Long-read ONT sequencing of seven tumour samples yielded 14.5 million reads across 5.6 billion bases, with considerable variation between samples, Table 4.2. Bladder samples ranged from 0.89 to 7.79 million reads (mean: 3.23 million), whilst upper tract samples ranged from 1.13 to 1.32 million reads (mean: 1.22 million). This 2.01 million read difference was not statistically significant (p = 0.47, Welch's two-sample t-test, n=3 bladder cancer, n=4 upper tract cancer), though one bladder sample (1525T) yielded substantially higher counts.

Sample	**Tissue**	**Read Count (million)**	**Total Bases (Gb)**	**Mean Length (bp)**	**Max Length (bp)**	**Mean Q30 (%)**	**Mean GC (%)**
1752T	Renal pelvis	1.16	0.37	314.8	399,580	80.4	43.1
1829T	Ureter	1.24	0.29	235.9	426,118	76.9	42
2179T	Renal pelvis	1.13	0.24	213.7	307,685	76	42
2237T	Renal pelvis	1.32	0.39	296.7	156,666	79.6	44.2
Upper tract mean	**-**	**1.22**	**0.32**	**265.3**	**322,512**	**78.2**	**42.8**
1525T	Bladder	7.79	3.7	474.3	514,625	75.1	45
504T	Bladder	0.89	0.31	350.2	381,089	79	45.7
989T	Bladder	1	0.3	303	158,301	78.8	43.9
Bladder mean	**-**	**3.23**	**1.44**	**375.8**	**351,338**	**77.6**	**44.9**
Overall	**-**	**2.08**	**0.8**	**312.7**	**334,866**	**78**	**43.7**

Table 4.2: Summary of raw Oxford Nanopore sequencing metrics for urothelial carcinoma tumour samples. Raw sequencing data were generated from seven urothelial carcinoma samples comprising four upper urinary tract specimens (renal pelvis and ureter) and three bladder specimens. Read count indicates the total number of sequencing reads obtained per sample. Total bases represents the cumulative sequence length in gigabases (Gb). Mean length denotes the average read length in base pairs (bp). Max length indicates the longest single read detected in each sample. Mean Q30 refers to the percentage of bases with a quality score of ≥30, corresponding to a base call accuracy of ≥99.9%. Mean GC represents the guanine-cytosine content expressed as a percentage of total bases. Group means were calculated for upper tract (renal pelvis and ureter combined), bladder, and overall cohort metrics.

RNA integrity numbers (RIN) ranged between 4.4 to 6.8, and had a mean of 5.3. The mean read length was 312.7 bp (range: 213.7-474.3 bp).

All samples passed core FastQC metrics, with Q30 percentages of 75-80% (mean: 78.0%) and mean Phred scores of 18.7. However, adapter content warnings indicated the need for filtering. GC content analysis revealed tissue specific differences, with bladder samples having higher mean GC content (44.9% ± 0.9%) than upper tract samples (42.8% ± 1.1%, $p = 0.04$, Welch's two-sample t-test, n=3 bladder cancer, n=4 upper tract cancer). Sample 1525T had substantially higher read counts (7.79 million), the longest mean read length (474.3 bp) and the lowest Q30 percentage (75.1%).

4.3.2 Read preprocessing and filtering

Pychopper (version 2.7.10) classification identified full length cDNA reads based on primer presence. Across all samples, 14.5 million raw reads were processed, with only 4.3-10.9% containing identifiable primer sequences at both ends, Table 4.3. The majority (88.3-95.6%) were designated unusable due to absent primers. An additional 0.1-0.8% of reads were rescued through Pychopper's recovery algorithm. Bladder samples averaged 10.7% full length identification, whilst upper tract samples had an average of 5.8%, though this difference was driven by sample 1525T (11.7%). When 1525T was excluded, bladder samples showed 6.8% full length identification.

Sample	Tissue	Input Reads (million)	Primers Found (%)	Rescued reads (%)	Unusable reads (%)	Full Length reads (%)
1525T	Bladder	7.79	10.9	0.8	88.3	11.7
504T	Bladder	0.89	6.6	0.2	93.2	6.8
989T	Bladder	1.00	6.6	0.2	93.2	6.8
Bladder mean	**-**	**3.23**	**10.1**	**0.6**	**89.3**	**10.7**
1829T	Ureter	1.24	5.0	0.1	94.9	5.1
2179T	Renal pelvis	1.13	4.3	0.1	95.6	4.4
2237T	Renal pelvis	1.32	6.1	0.1	93.8	6.2
1752T	Renal pelvis	1.16	7.2	0.2	92.6	7.4
Upper tract mean	**-**	**1.21**	**5.7**	**0.1**	**94.2**	**5.8**
Overall mean	**-**	**2.08**	**8.5**	**0.6**	**90.9**	**9.1**

Table 4.3: Pychopper classification statistics for full length cDNA identification in urothelial carcinoma samples. Pychopper (version 2.7.10) was used to identify and classify full length cDNA reads from Oxford Nanopore sequencing data. Input reads represents the total number of reads (in millions) submitted to Pychopper for processing. Primers found denotes the percentage of reads in which both 5' and 3' primer sequences which were successfully detected, indicating high confidence, full length transcripts. Rescued reads refers to the percentage of reads that were recovered through the algorithm's rescue mechanism, which attempts to identify full length transcripts from reads with partially degraded or missing primer sequences. Unusable reads indicates the percentage of reads that could not be classified as full length due to absent or incomplete primer sequences. Full length reads represents the combined percentage of primers found and rescued reads, reflecting the total proportion of reads classified as full length cDNA molecules. Group means were calculated for bladder specimens (n=3), upper urinary tract specimens (n=4, comprising ureter and renal pelvis), and the overall cohort (n=7).

Subsequent Fastplong (version 0.2.2) filtering for quality and length retained 89.9% of full length reads from Pychopper whilst improving quality metrics, Table 4.4. A total of 133,649 reads were removed, primarily due to excessive shortening after adapter trimming (mean: 14,874 reads per sample) or low quality scores (mean: 4,221 reads per sample). Retention rates ranged from 87.1% (1752T) to 99.1% (989T), with similar retention between tissue types (bladder: 95.4%, upper tract: 90.7%). Base quality improved consistently, with Q30 percentages increasing by mean 3.9 percentage points (range: 2.9-5.0). Sample 1525T had the largest improvement (Q30: 78.3% to 83.3%).

Sample	Tissue	Input Reads	Output Reads	Retention (%)	Filtered (%)	Low Quality	Too Short	Q30 Before (%)	Q30 After (%)
1525T	Bladder	912,188	805,940	88.4	11.6	26,795	79,460	78.3	83.3
504T	Bladder	60,512	59,796	98.8	1.2	329	387	83.8	86.7
989T	Bladder	67,779	67,178	99.1	0.9	431	179	82.6	85.6
Bladder mean	**-**	**346,826**	**310,971**	**95.4**	**4.6**	**9,185**	**26,675**	**81.6**	**85.2**
1829T	Ureter	63,626	56,209	88.3	11.7	505	6,912	80.9	85.2
2179T	Renal pelvis	49,611	45,509	91.7	8.3	433	3,669	80	84.9
2237T	Renal pelvis	81,210	77,725	95.7	4.3	477	3,008	83.1	86.5
1752T	Renal pelvis	86,226	75,146	87.1	12.9	579	10,501	83	86.4
Upper tract mean	**-**	**70,168**	**63,647**	**90.7**	**9.3**	**498**	**6,023**	**81.8**	**85.8**
Overall mean	**-**	**188,736**	**169,643**	**92.8**	**7.2**	**4,221**	**14,874**	**81.7**	**85.5**

Table 4.4: Quality filtering and adapter trimming statistics from Fastplong processing of full length cDNA reads for urothelial carcinoma tumour samples. Fastplong (version 0.2.2) was used to perform quality based filtering and adapter trimming on full length cDNA reads identified by Pychopper. Input reads represents the number of full length reads submitted to Fastplong for processing. Output reads indicates the number of reads retained after quality filtering and trimming. Retention denotes the percentage of input reads that passed quality control thresholds and were retained for downstream analysis. Filtered represents the percentage of input reads that failed quality control criteria and were excluded from further analysis. Low quality indicates the number of reads discarded due to insufficient base quality (defined as >40% of bases with Phred quality score <10). Too short denotes the number of reads removed due to insufficient length following adapter and polyX tail trimming.

Q30 before and Q30 after represent the percentage of bases with Phred quality scores ≥30 (corresponding to 99.9% base call accuracy) prior to and following quality filtering, respectively. Group means were calculated for bladder specimens (n=3), upper urinary tract specimens (n=4, comprising ureter and renal pelvis), and the overall cohort (n=7).

The combined preprocessing pipeline retained a total of 1.19 million reads for downstream analysis, representing 8.2% of initial raw reads. This substantial attrition is due to Pychopper identifying 1.32 million full length transcripts and subsequent Fastplong filtering reducing this to the final 1.19 million high quality reads. The mean read length was of 181 bp in the processed data. Each sample contained 45,509 to 805,940 processed reads (mean: 169,643 reads per sample).

4.3.3 Alignment to the reference genome

Processed reads were aligned to GRCh38 using Minimap2 (version 2.28). Across all samples 1.19 million processed reads were aligned of which 830,334 were successfully mapped, Table 4.5. The mean mapping rate was 63.3%, with considerable variation ranging from 50.1% (2179T) to 78.3% (504T). Bladder samples had higher mean mapping rates (67.7 ± 14.4%) compared to upper tract samples (60.0 ± 8.9%), though this was not statistically significant. Primary alignments accounted for 99.8% of mapped reads, whilst supplementary alignments (chimeric reads) comprised only 0.2% indicating unambiguous read mapping.

Sample	Tissue	Total Reads	Mapped Reads	Mapping Rate (%)	Unmapped Reads	Primary Alignments	Supplementary alignments
1525T	Bladder	805,940	592,425	73.5	213,515	592,425	787
504T	Bladder	59,796	46,791	78.3	13,005	46,791	44
989T	Bladder	67,178	34,419	51.3	32,759	34,419	66
Bladder mean	**-**	**310,971**	**224,545**	**67.7**	**86,426**	**224,545**	**299**
1829T	Ureter	56,209	30,949	55.1	25,260	30,949	58
2179T	Renal pelvis	45,509	22,782	50.1	22,727	22,782	24
2237T	Renal pelvis	77,725	53,835	69.3	23,890	53,835	36
1752T	Renal pelvis	75,146	49,133	65.4	26,013	49,133	71
Upper tract mean	**-**	**63,647**	**39,175**	**60**	**24,473**	**39,175**	**47**

Overall mean	**-**	**169,643**	**118,619**	**63.3**	**51,024**	**118,619**	**155**

Table 4.5: Reference genome alignment statistics from Minimap2 for quality filtered long reads of urothelial carcinoma tumour samples. Alignment statistics from Minimap2 splice-aware mapping to GRCh38 reference genome. Total reads indicates quality filtered input. Mapped reads and mapping rate show successful alignments. Unmapped reads represents alignment failures. Primary alignments denotes best single alignment per read, supplementary indicates split alignments for chimeric reads. Group means calculated for bladder (n=3), upper tract (n=4), and overall cohort (n=7).

Three samples (2179T, 1829T and 989T) had mapping rates below 56%. These samples also had the lowest sequencing depth (45,509-67,178 reads). Sample 1525T contained 592,425 mapped reads, representing 71.3% of all mapped reads across the dataset. The supplementary alignment frequency was low at 0.2%.

4.3.4 Gene level quantification

Gene level quantification using featureCounts (version 2.0.6) with GENCODE v47 annotation assigned 757,440 reads (63.8% of processed reads) to annotated genes, Table 4.6. Mean assignment rate was 58.0%, with sample variation ranging from 44.7% (2179T) to 73.4% (504T). Bladder samples achieved higher mean assignment rates (62.6 ± 13.3%) compared to upper tract samples (54.6 ± 9%).

Sample	Tissue	Total Reads	Assigned	Assignment Rate (%)	Unmapped	No Features	Ambiguous
1525T	Bladder	805,940	538,433	66.7	213,515	10,489	44,290
504T	Bladder	59,796	43,878	73.4	13,005	445	2,512
989T	Bladder	67,178	32,011	47.6	32,759	666	1,808
Bladder mean	**-**	**310,971**	**204,774**	**62.6**	**86,426**	**3,867**	**16,203**
1829T	Ureter	56,209	27,907	49.6	25,260	1,088	2,012
2179T	Renal pelvis	45,509	20,322	44.7	22,727	885	1,599
2237T	Renal pelvis	77,725	49,761	64	23,890	1,212	2,898
1752T	Renal pelvis	75,146	45,128	60	26,013	1,442	2,634
Upper tract mean	**-**	**63,647**	**35,780**	**54.6**	**24,473**	**1,157**	**2,286**
Overall mean	**-**	**169,643**	**108,206**	**58**	**51,024**	**2,318**	**8,250**

Table 4.6: Gene level quantification statistics from featureCounts analysis of aligned long reads of urothelial carcinoma tumour samples. Gene level quantification from featureCounts using GENCODE v47 annotation and long-read parameters (-L, --largestOverlap). Total reads includes both mapped and unmapped alignments from Minimap2. Assigned shows reads successfully assigned to genes, assignment rate indicates percentage. Unmapped represents alignment failures. No features denotes mapped reads in unannotated regions. Ambiguous indicates reads overlapping multiple genes. Group means calculated for bladder (n=3), upper tract (n=4), and overall cohort (n=7).

Unassigned reads were made up of 30.1% unmapped, 1.4% in unannotated regions (no features), and 4.9% overlapping multiple genes (ambiguous). The low proportion of "no

features" reads (mean 1.5%, range 0.7-1.9%) was consistent across samples. Ambiguity rates ranged from 2.7% (989T) to 5.5% (1525T), with no significant tissue specific pattern.

Three samples (2179T, 1829T, 989T) had assignment rates below 50%, corresponding to the same samples with lowest mapping rates. Sample 1525T contained 538,433 assigned reads, representing 71.1% of all assigned reads across the dataset. Excluding this outlier, bladder samples averaged 37,945 assigned reads, comparable to the upper tract samples, 35,780 reads.

4.3.5 Differential expression analysis

Differential gene expression analysis was performed comparing bladder cancer (n=3) against upper tract urothelial carcinoma (n=4) using DESeq2 (v1.42.0). Genes were defined as significantly differentially expressed if they had a false discovery rate (FDR) adjusted *p*-value < 0.05 following Benjamini-Hochberg correction. A total of 5,196 genes (91.4% of input) were excluded due to NA values in adjusted *p*-values following DESeq2's independent filtering procedure. Of the 490 genes tested, only 7 (1.43%) had significant differential expression, Figure 4.1.

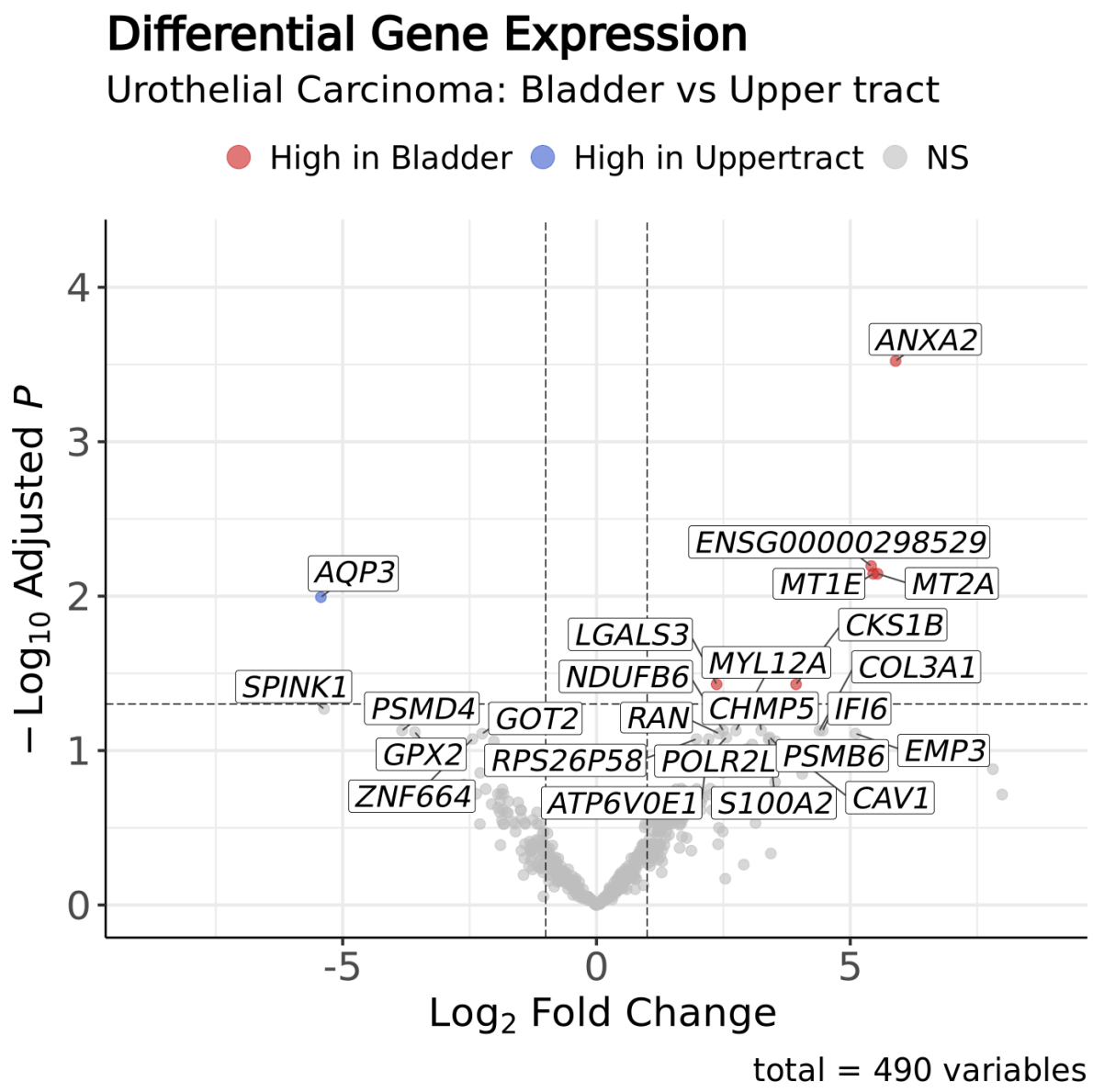

Figure 4.1: Differential gene expression between bladder cancer and upper tract urothelial carcinoma. A volcano plot showing log2 fold change versus adjusted *p*-value for differential gene expression analysis between bladder cancer (n=3) and upper tract urothelial carcinoma (n=4). Each point represents a single gene from 490 genes tested. Positive log2 fold changes indicate higher expression in bladder cancer and negative values indicate higher expression in upper tract cancer. Red points are genes significantly upregulated in bladder, blue points are genes significantly upregulated in upper tract and grey points represent non-significant changes. Dashed lines mark significance thresholds (vertical: ±1.0 log2FC and horizontal: adjusted *p*-value = 0.05). The top 25 most significantly differentially expressed genes are labelled. A total of 7 genes (1.43%) were significantly differentially expressed, with 6 upregulated in bladder cancer and 1 upregulated in upper tract cancer.

The distribution of differentially expressed genes (DEGs) strongly favoured bladder cancer upregulation, with 6 genes upregulated (85.7% of DEGs) and 1 gene downregulated (14.3% of DEGs) relative to upper tract cancer. Upregulated genes in bladder cancer had a mean

log2 fold change (log2FC) of 4.77, whilst the single gene upregulated in upper tract cancer (downregulated in bladder) had a log2FC of -5.43.

The most significantly upregulated gene in bladder cancer was *ANXA2* with a log2 fold change of 5.89 (adjusted $p = 3.00 \times 10^{-4}$). Bladder cancer also showed upregulation of metallothionein family members *MT2A* and *MT1E*, with log2 fold changes of 5.54 and 5.45 respectively (both adjusted $p = 7.16 \times 10^{-3}$).

The single gene downregulated in bladder cancer (upregulated in upper tract cancer) was *AQP3* (Aquaporin 3) with a log2 fold change of -5.43 (adjusted $p = 1.02 \times 10^{-2}$).

4.3.6 Principal component analysis

Principal component analysis (PCA) was performed on variance-stabilised count data to assess overall transcriptomic similarity between samples and to identify the sources of variation in the dataset. The first principal component (PC1) explained 26.5% of total variance, whilst the second component (PC2) explained 17.81%, Figure 4.2. Together, the first two principal components accounted for 44.31% of total transcriptomic variance.

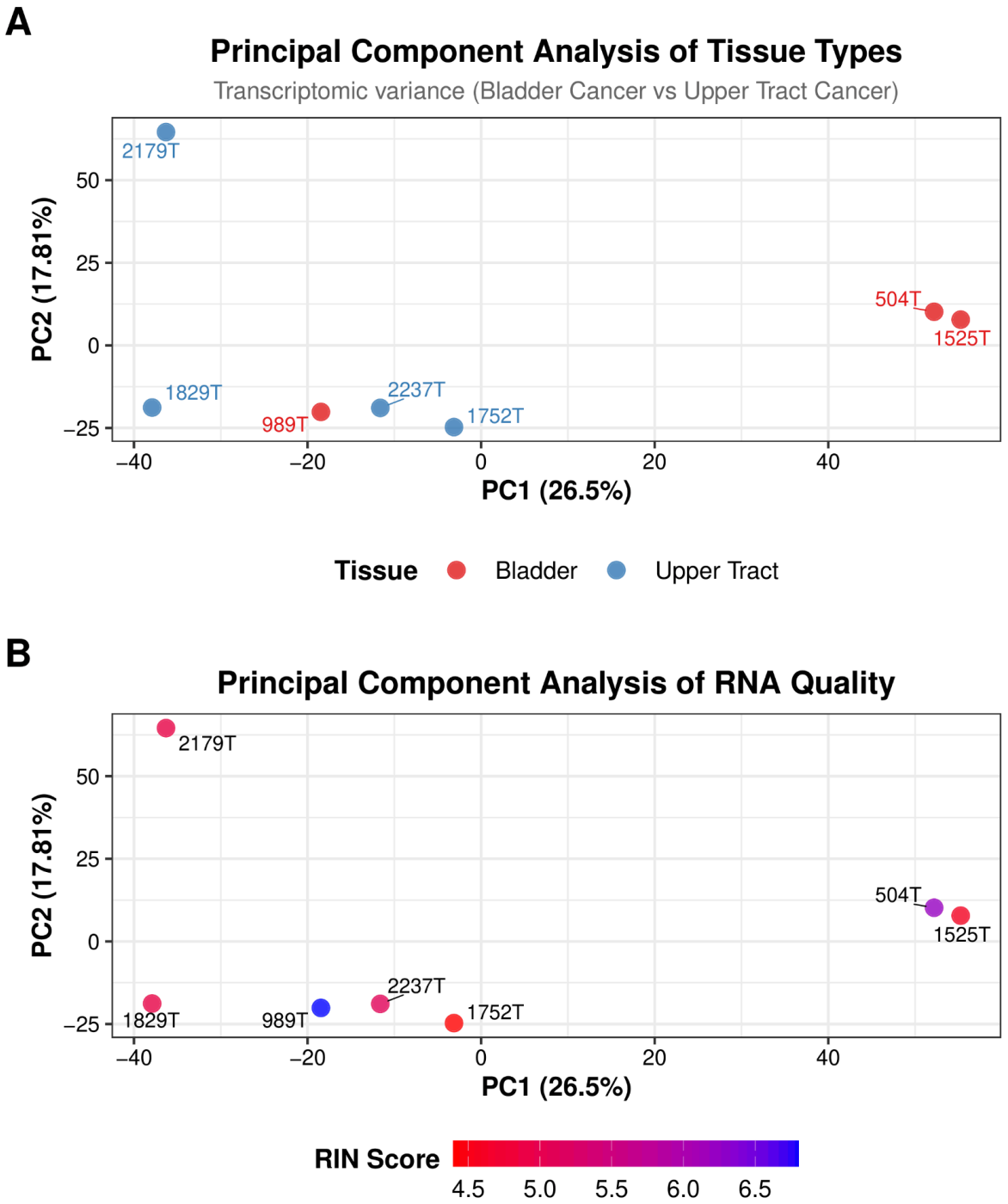

Figure 4.2: Principal component analysis of gene expression data. (A) Principal component analysis showing tissue specific clustering. Bladder samples (red, n=3) and upper tract samples (blue, n=4). Sample identifiers are displayed. (B) The same principal component analysis with the samples coloured by RNA integrity number (RIN) score. RIN values range from 4.4 (red) to 6.8 (blue). Samples do not cluster according to RIN. Together, PC1 and PC2 explain 44.31% of total variance.

The PCA plot showed no clear separation between bladder and upper tract samples along either PC1 or PC2. Samples from both tumour types were mixed in the PCA dimensions, with no distinct clustering by tissue type.

To identify which genes contribute most to the variation captured by PC1 and PC2, eigenvector analysis was performed on the PCA loadings. The top 20 genes contributing to PC1 included genes which were involved in multiple cellular processes, including *CHMP4A* (endosomal sorting), *MYL12B* (cytoskeletal regulation), *HSD17B10* (mitochondrial metabolism), and *GSTP1* (detoxification). PC2 had various genes which contributed to it including *ING3* (cell cycle and apoptosis), *SEC23A* (vesicle trafficking), *CD14* and *CD3D* (immune markers), and *ALDH2* (aldehyde metabolism).

4.3.7 Biological pathway and gene expression analyses

Targeted biological pathway analyses were performed to characterise functional differences between bladder and upper tract urothelial carcinomas. Gene expression was normalised to transcripts per million (TPM) to account for both sequencing depth and gene length, allowing the direct comparison of expression levels across samples and genes.

4.3.7.1 HOX gene expression

HOX expression was analysed across the same 39 genes in bladder tumours (n=3) and upper tract tumours (n=4). Overall HOX pathway activity was higher in upper tract tumours (2.58 ± 2.42 TPM) compared to bladder tumours (0.87 ± 0.11 TPM), though this difference did not reach statistical significance (p = 0.251, Welch's two-sample t-test, n=3 bladder cancer, n=4 upper tract cancer), Figure 4.3A.

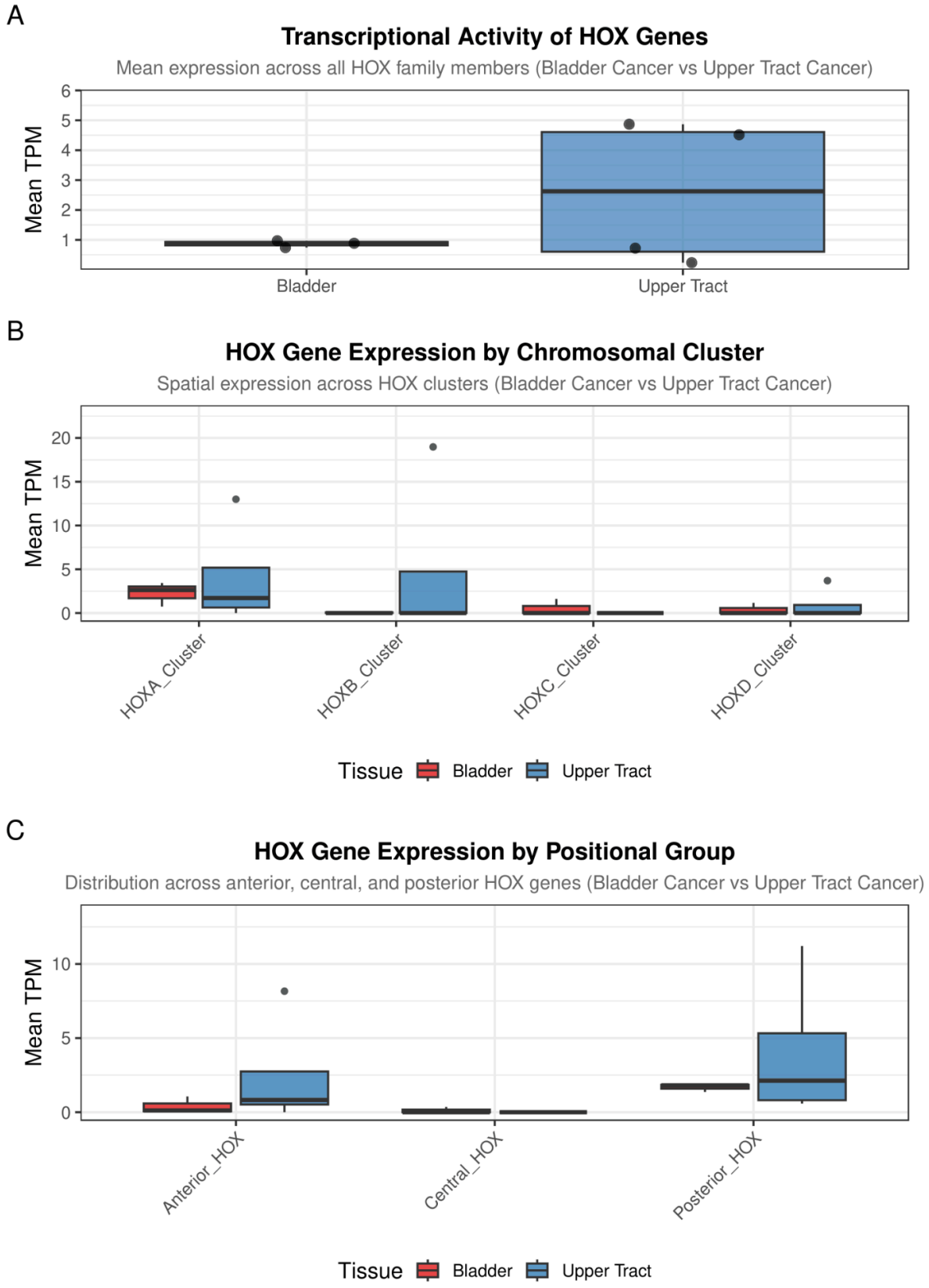

Figure 4.3: HOX gene expression in bladder and upper tract urothelial carcinomas. Box plots showing mean TPM of overall HOX expression, HOX expression by cluster and

HOX expression by positional group in bladder (red, n=3) and ureter (blue, n=4) urothelium with individual samples overlaid as points. Statistical significance was assessed using Welch's two-sample t-test, brackets with *p*-values indicate significant differences between groups. (A) Boxplot showing overall HOX pathway activity score (mean expression across 39 detected genes) in bladder tumours (red, n=3) and upper tract tumours (blue, n=4). Upper tract tumours display higher pathway activity than bladder tumours, though with considerable variability. (B) Boxplot showing mean HOX gene expression (TPM) across chromosomal clusters. HOXB and HOXD clusters are elevated in upper tract tumours, whilst HOXC expression is restricted to bladder specimens. (C) Boxplot showing mean HOX gene expression grouped by anterior-posterior position. Anterior and posterior HOX genes show increased expression in upper tract tumours, whilst central genes are bladder specific.

The HOXB cluster showed the biggest difference with upper tract tumours having higher expression (4.74 ± 9.49 TPM) compared to bladder tumours (0.05 ± 0.09 TPM), Figure 4.3B. Similarly, HOXD cluster expression was higher in upper tract specimens (0.92 ± 1.83 TPM) versus bladder (0.39 ± 0.67 TPM). In contrast, HOXC expression was detected exclusively in bladder tumours (0.54 ± 0.93 TPM) with complete absence in upper tract samples (0.00 ± 0.00 TPM).

Positional grouping demonstrated that anterior HOX genes were more abundant in upper tract tumours (2.44 ± 3.78 TPM) than bladder tumours (0.39 ± 0.58 TPM), Figure 4.3C. Central HOX genes had bladder specific expression (0.12 ± 0.20 TPM) with no detectable expression in upper tract samples. Posterior HOX genes were elevated in upper tract tumours (4.01 ± 4.95 TPM) compared to bladder (1.69 ± 0.28 TPM).

Some specific samples had much higher HOX expression, upper tract sample 1829T had elevated *HOXA1* (121.2 TPM), *HOXA9* (20.3 TPM), and *HOXD9* (33.0 TPM), whilst sample 2179T had increased *HOXB9* expression (179.4 TPM).

4.3.7.2 Sonic hedgehog signalling pathway

The expression of 36 SHH pathway-associated genes was analysed across bladder tumours and upper tract tumours.

SHH pathway activity was similar between bladder tumours (2.98 ± 1.80 TPM) and upper tract tumours (3.59 ± 0.73 TPM, $p = 0.626$, Welch's two-sample t-test, n=3 bladder cancer, n=4 upper tract cancer), Figure 4.4A.

The reduction in pathway activity was primarily driven by the diminished expression of key pathway components. The SHH ligand itself was undetectable in all tumour samples, contrasting with expression in normal bladder (35.6 TPM) and ureter (10.7 TPM). *VEGFA* expression was reduced and variable expression in both tumour types (bladder: 3.71 ± 6.42 TPM, upper tract: 26.90 ± 27.05 TPM), whilst *ETS2* had moderate expression in both sites (bladder: 28.05 ± 21.03 TPM, upper tract: 30.30 ± 11.29 TPM). The *GLI1* gene was undetectable in all samples. However, certain pathway associated genes showed selective upregulation, notably *TLE1* (bladder: 38.00 ± 46.94 TPM, upper tract: 29.30 ± 13.16 TPM) and *ADGRG1* (bladder: 10.04 ± 0.95 TPM, upper tract: 14.71 ± 16.19 TPM).

4.3.7.3 Fibroblast growth factor signalling pathway

The expression of 55 FGF pathway genes was analysed across bladder tumours (n=3) and upper tract tumours (n=4).

FGF pathway activity was significantly elevated in upper tract tumours (17.38 ± 3.39 TPM) compared to bladder tumours (7.38 ± 2.36 TPM, $p = 0.007$, Welch's two-sample t-test, n=3 bladder cancer, n=4 upper tract cancer), Figure 4.4B. Individual upper tract samples ranged from 13.08 to 21.36 TPM, whilst bladder tumours ranged from 5.07 to 9.79 TPM.

The elevated pathway activity in upper tract specimens was primarily driven by *FGFR3*, which had higher expression in upper tract tumours (255.31 ± 152.61 TPM) compared to bladder (39.26 ± 48.45 TPM).

4.3.7.4 Wnt beta-catenin signalling pathway

The expression of 42 Wnt pathway genes was analysed across bladder tumours (n=3) and upper tract tumours (n=4).

Wnt pathway activity was slightly higher in upper tract tumours (3.57 ± 1.96 TPM) compared to bladder tumours (2.62 ± 1.52 TPM), though this was not a significant difference ($p = 0.503$, Welch's two-sample t-test, n=3 bladder cancer, n=4 upper tract cancer), Figure 4.4C. Individual upper tract samples ranged from 1.78 to 6.35 TPM, whilst bladder tumours ranged from 1.58 to 4.37 TPM.

4.3.7.5 TGF-beta signalling pathway

The expression of 54 TGF-beta pathway genes was analysed across bladder tumours (n=3) and upper tract tumours (n=4).

TGF-beta pathway activity showed comparable expression between bladder tumours (20.41 ± 10.54 TPM) and upper tract tumours (23.56 ± 4.59 TPM, $p = 0.666$, Welch's two-sample t-test, n=3 bladder cancer, n=4 upper tract cancer), Figure 4.4D.

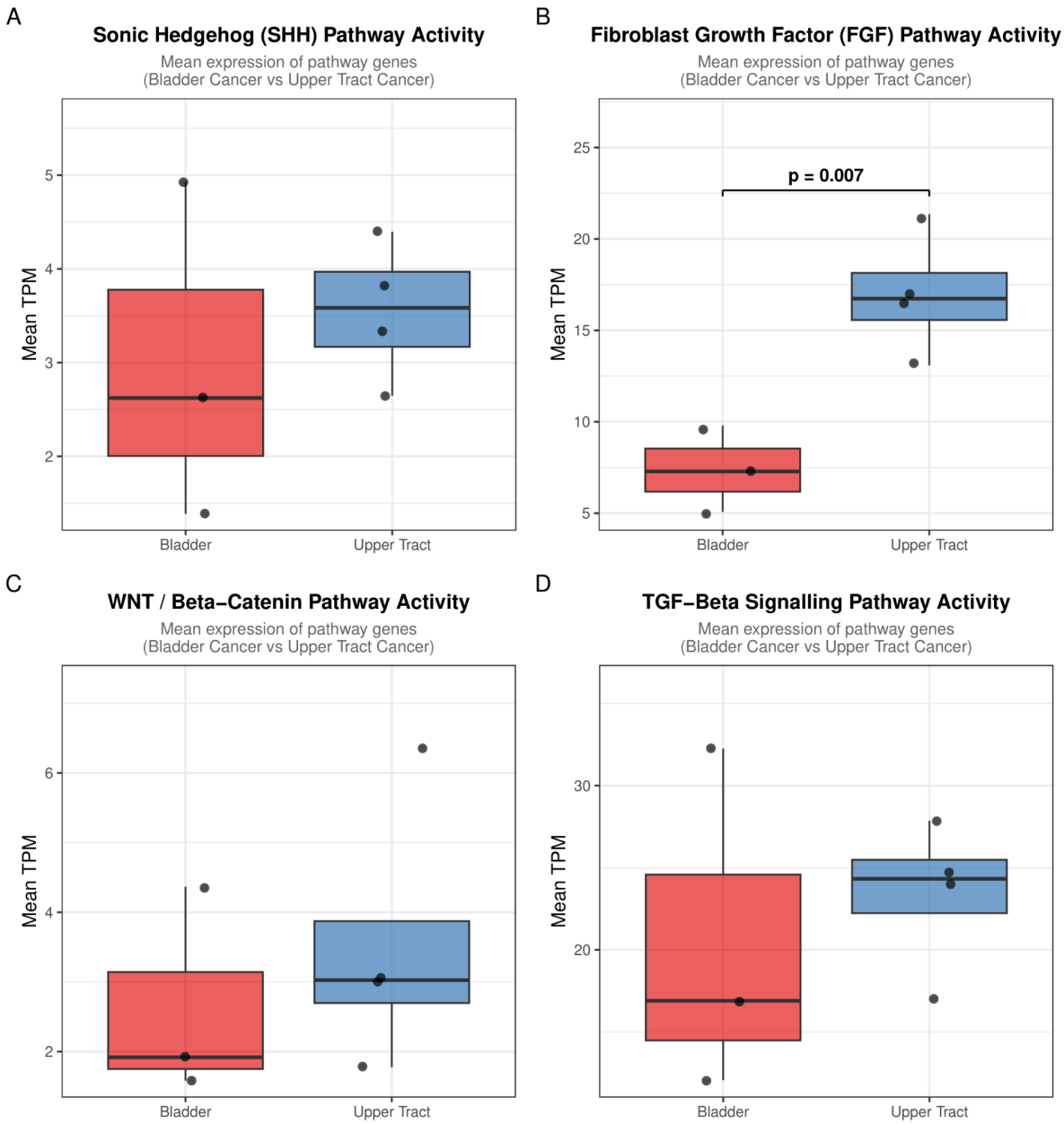

Figure 4.4: Signalling pathway expression in bladder and upper tract urothelial carcinomas. Box plots showing overall pathway activity scores (mean TPM) in bladder tumours (red, n=3) and upper tract tumours (blue, n=4) with individual samples overlaid as points. Statistical significance was assessed using Welch's two-sample t-test, brackets with *p*-values indicate significant differences between groups. (A) SHH pathway activity score (mean TPM across 36 pathway genes), activity was similar between tumour types and substantially reduced compared to normal bladder tissue. (B) FGF pathway activity score (mean TPM across 55 pathway genes), upper tract tumours showed significantly higher pathway activity than bladder tumours, driven primarily by elevated *FGFR3* expression. (C) Wnt/beta-catenin pathway activity score (mean TPM across 42 pathway genes), upper tract tumours showed slightly higher activity than bladder tumours though neither reached the expression levels seen in normal bladder tissue. (D) TGF-beta pathway activity score (mean TPM across 54 pathway genes), activity was comparable between tumour types and reduced compared to normal bladder tissue.

4.3.7.6 Interferon response

The expression of 28 interferon and interferon receptor genes was analysed across bladder tumours (n=3) and upper tract tumours (n=4).

Overall interferon pathway activity was elevated in upper tract tumours (1.64 ± 1.18 TPM) compared to bladder tumours (0.77 ± 0.51 TPM), though this did not achieve statistical significance (p = 0.252, Welch's two-sample t-test, n=3 bladder cancer, n=4 upper tract cancer), Figure 4.5A.

Type I interferons had minimal expression in bladder tumours (0.05 ± 0.08 TPM) but were slightly higher in upper tract samples (0.61 ± 1.23 TPM). Type II interferons (IFN-gamma) were completely absent in all tumour samples, contrasting with normal bladder expression (44.40 TPM). Type III interferons remained undetectable in both tumour types. Type II receptors demonstrated the highest expression amongst all categories (bladder: 9.68 ± 5.26 TPM, upper tract: 14.24 ± 11.36 TPM). Type I receptors had higher expression in upper tract tumours (2.88 ± 2.86 TPM) versus bladder (0.45 ± 0.77 TPM), whilst Type III receptors remained low in both groups, Figure 4.5C.

4.3.7.7 Immune cell marker expression

The expression of 14 cell type-specific markers across nine immune cell categories was analysed in bladder tumours (n=3) and upper tract tumours (n=4).

Overall immune marker expression showed considerable variability between tumour types, with no statistically significant differences observed for any individual markers, Figure 4.5B. CD14+ monocyte markers demonstrated variable expression, whilst *CD14* had varying levels of expression (bladder: 28.95 ± 31.13 TPM, upper tract: 35.21 ± 70.41 TPM), Figure 4.5D. Memory CD4+ T cell marker *S100A4* had elevated expression in upper tract specimens (543.20 ± 672.55 TPM) compared to bladder (63.29 ± 64.22 TPM), driven primarily by exceptionally high expression in sample 2237T (1,653.01 TPM).

Dendritic cell marker *CST3* had similar expression between tumour types (bladder: 60.72 ± 22.34 TPM, upper tract: 81.34 ± 50.65 TPM). NK cell markers had varying levels of expression. FCGR3A+ monocyte markers had variable expression across samples (MS4A7 bladder: 10.71 ± 11.69 TPM, upper tract: 7.99 ± 8.61 TPM). CD8+ T cell marker *CD8A*, B cell marker *MS4A1*, and platelet marker *PPBP* remained at negligible levels across all samples.

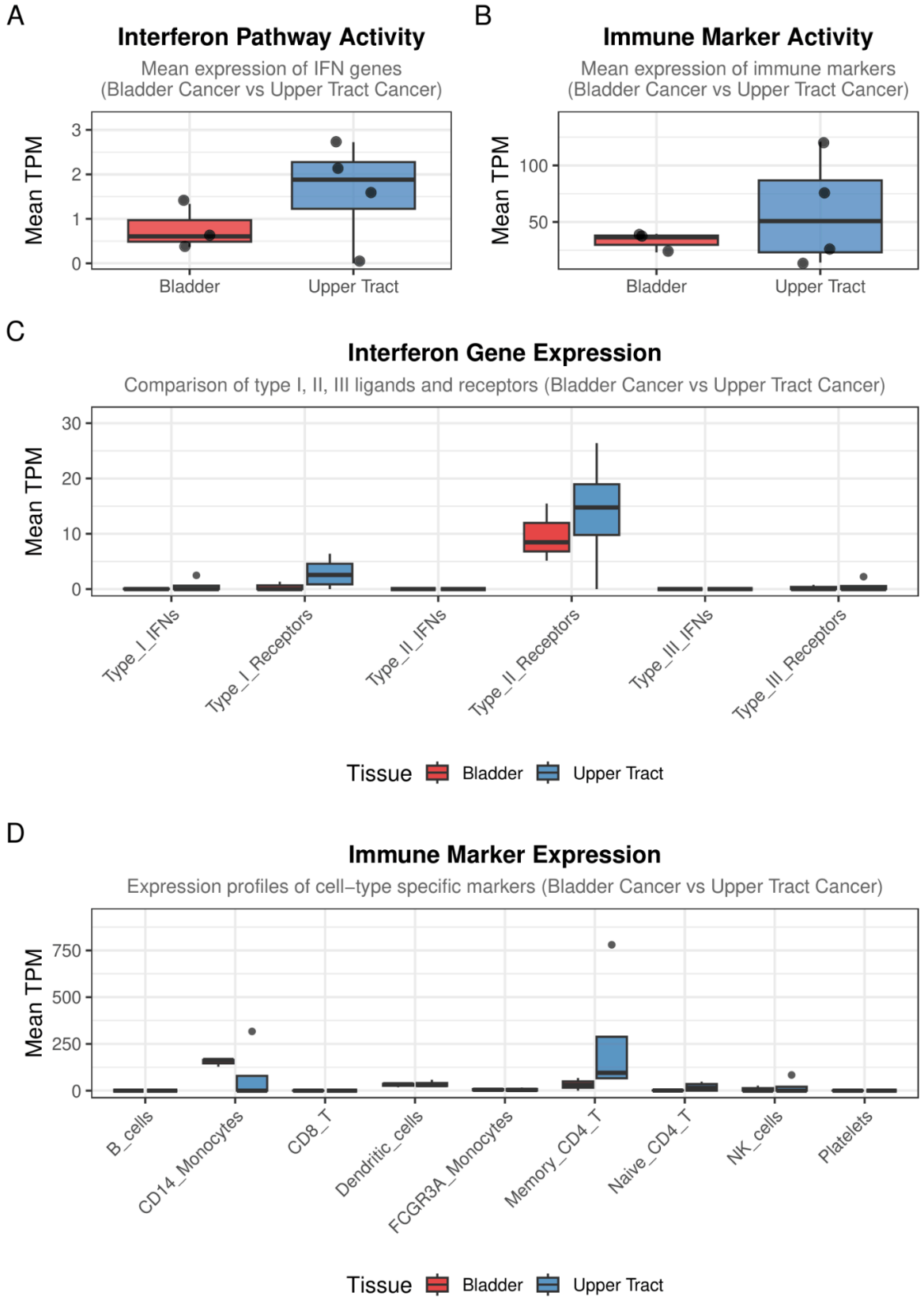

Figure 4.5: Interferon and immune marker gene expression in bladder and upper tract urothelial carcinomas. Box plots showing overall activity scores (mean TPM) in bladder tumours (red, n=3) and upper tract tumours (blue, n=4), with individual samples overlaid as

points. Statistical significance was assessed using Welch's two-sample t-test, brackets with *p*-values indicate significant differences between groups. (A) Overall interferon marker gene activity score (mean TPM across 28 genes). (B) Overall immune marker gene activity score (mean TPM across 14 genes). (C) Mean gene expression (TPM) across IFN types I, II and III and their respective receptors. All tumour types showed reduced expression compared to normal tissues. (D) Mean gene expression (TPM) across immune cell types. Immune marker expression was highly variable across samples with substantial inter-tumour heterogeneity and reduced overall expression compared to normal bladder tissue.

4.3.7.8 Cell cycle and proliferation analysis

The expression of 56 genes across six cell cycle categories was analysed in bladder tumours (n=3) and upper tract tumours (n=4).

Overall proliferation pathway activity was higher in bladder tumours (11.43 ± 5.92 TPM) compared to upper tract tumours (7.55 ± 4.85 TPM), though this did not achieve statistical significance (p = 0.408, Welch's two-sample t-test, n=3 bladder cancer, n=4 upper tract cancer), Figure 4.6A.

G2/M transition genes had the largest increase in expression in bladder tumours (25.84 ± 18.74 TPM) compared to upper tract specimens (10.13 ± 10.17 TPM). S phase genes had increased expression in bladder (13.23 ± 7.79 TPM versus upper tract: 9.61 ± 5.60 TPM), G1/S transition genes had comparable expression (bladder: 8.15 ± 4.84 TPM, upper tract: 6.57 ± 3.45 TPM), Figure 4.6C.

Proliferation markers had higher expression in upper tract tumours (18.25 ± 11.36 TPM) compared to bladder (11.97 ± 11.66 TPM), contrasting with cell cycle phase gene patterns. Checkpoint control genes were higher in upper tract specimens (5.63 ± 3.21 TPM versus bladder: 2.93 ± 1.75 TPM).

4.3.7.9 Luminal and basal marker expression

The expression of 52 differentiation markers was analysed across bladder tumours (n=3) and upper tract tumours (n=4).

Overall marker activity was higher in bladder tumours (88.97 ± 52.35 TPM) versus upper tract tumours (73.61 ± 32.12 TPM, p = 0.683, Welch's two-sample t-test, n=3 bladder cancer, n=4 upper tract cancer), Figure 4.6B.

Marker category analysis revealed divergent patterns. Luminal markers were nearly equivalent between sites (bladder: 154.94 TPM, upper tract: 153.61 TPM). However, basal markers were much higher in bladder tumours (59.04 vs 9.41 TPM), whilst EMT markers were higher in upper tract specimens (41.92 vs 26.42 TPM).

Luminal-basal (L/B) ratios differed between tumour types (bladder: 1.44 ± 5.36, upper tract: 3.93 ± 1.18, p = 0.508, Welch's two-sample t-test, n=3 bladder cancer, n=4 upper tract cancer). Despite this there was considerable inter-tumour heterogeneity with sample 504T having basal predominant phenotype (L/B: -3.96) and sample 989T having luminal predominant phenotype (L/B: 6.77).

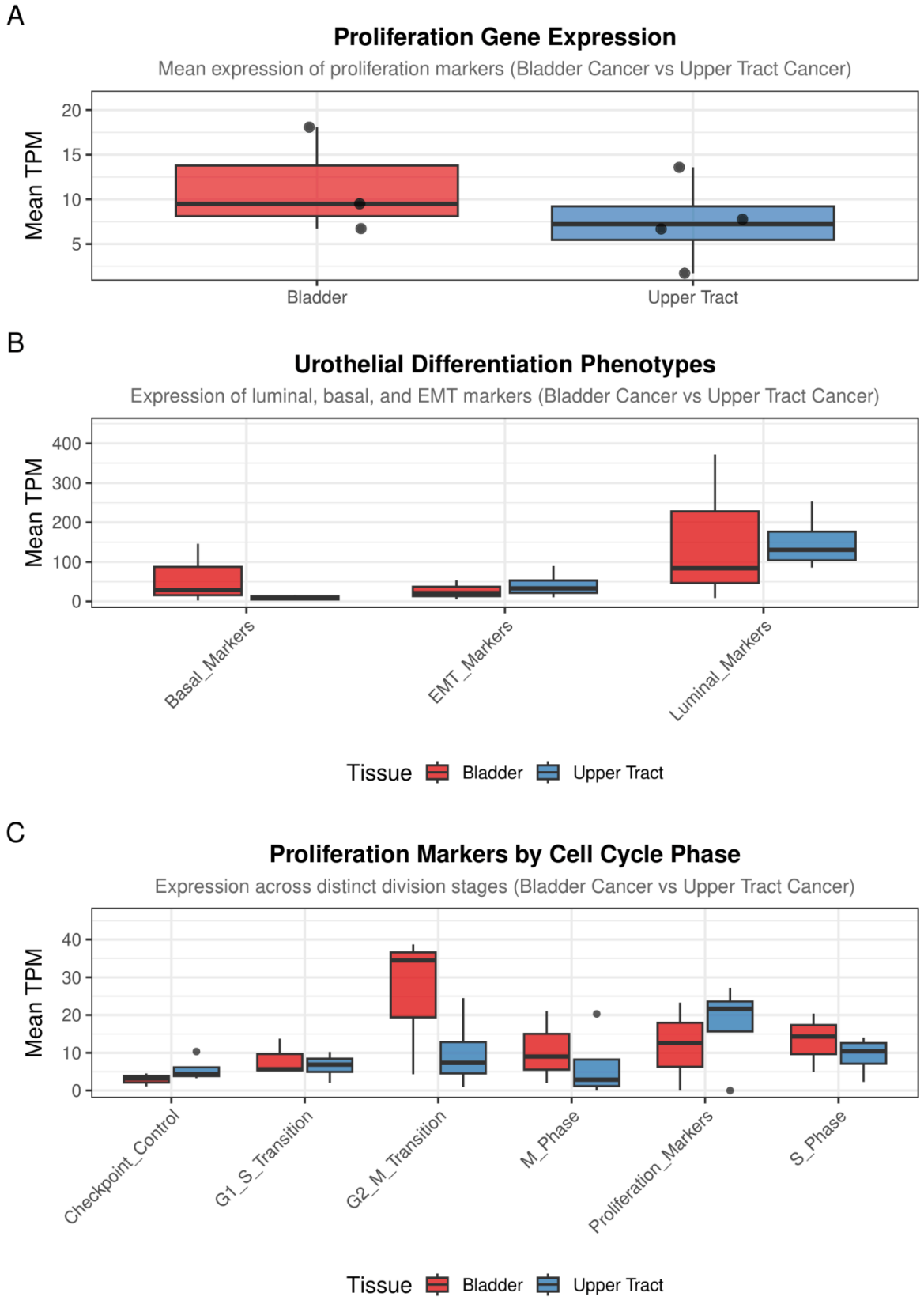

Figure 4.6: Cell cycle, proliferation and urothelial marker gene expression in bladder and upper tract urothelial carcinomas. Box plots showing mean gene expression (TPM) in bladder tumours (red, n=3) and upper tract tumours (blue, n=4), with individual samples

overlaid as points. Statistical significance was assessed using Welch's two-sample t-test, brackets with *p*-values indicate significant differences between groups. (A) Overall cell cycle and proliferation marker gene activity score (mean TPM across 56 genes). (B) Mean expression (TPM) of urothelial differentiation markers by category (luminal, basal and EMT) in bladder and upper tract tumours. Basal markers showed elevated expression in bladder specimens, whilst EMT markers were higher in upper tract tumours and luminal markers were comparable between sites, though none achieved significance. (C) Mean gene expression (TPM) across cell cycle phases and checkpoints.

4.3.8 Gene set enrichment analysis

Gene Set Enrichment Analysis (GSEA) was performed to identify biological pathways and processes enriched in bladder tumours compared to upper tract tumours using long-read RNA sequencing data. The analysis used two complementary gene set collections, the Hallmark gene sets and Reactome pathways. The ranked gene list, comprising 490 genes, was derived from differential expression analysis comparing bladder tumour samples (n=3) to upper tract tumour samples (n=4), with upper tract serving as the reference condition.

4.3.8.1 Hallmark gene sets

The GSEA analysis of Hallmark gene sets revealed enrichment of metabolic and stress response pathways in bladder tumours. Of the 6 gene sets tested using an fgsea exact permutation test (n=3 bladder cancer, n=4 upper tract cancer) 2 achieved statistical significance at the FDR-corrected threshold (adjusted $p < 0.05$), whilst 3 achieved nominal significance (p-value < 0.05). All significant pathways showed positive enrichment (NES > 0), indicating consistent upregulation in bladder tumours relative to upper tract specimens.

The pathway with the strongest enrichment was hypoxia (NES = 1.73, adjusted $p = 3.48 \times 10^{-2}$). This was followed by oxidative phosphorylation (NES = 1.50, adjusted $p = 3.67 \times 10^{-2}$).

4.3.8.2 Reactome pathways

The Reactome pathway analysis complemented the Hallmark findings, though with reduced statistical power due to the smaller gene set. Of the 122 pathways tested using an fgsea exact permutation test, n=3 bladder cancer, n=4 upper tract cancer, none achieved FDR significance (adjusted $p < 0.05$), whilst 12 achieved nominal significance (p-value < 0.05). All nominally significant pathways had positive enrichment in bladder tumours.

The most significant pathway identified was cytokine signalling in the immune system (NES = 1.63, p-value = 2.68×10^{-3}, adjusted $p = 0.137$). This was followed by RNA polymerase II transcription (NES = 1.63, p-value = 2.86×10^{-3}, adjusted $p = 0.137$) and HIV life cycle (NES = 1.73, p-value = 3.38×10^{-3}, adjusted $p = 0.137$).

4.3.9 StringTie quantification

Stringtie was used to quantify gene and transcript expression from the aligned long-read data. The total number of transcripts detected ranged from 671 to 4,326, with a mean of 1,336 transcripts per sample. The mean number of transcripts per gene was approximately 1.03 across all samples, indicating limited detection of alternative isoforms. Sample 1525T

had the highest transcript detection, with 4,326 transcripts and a mean of 1.09 transcripts per gene, whilst the remaining samples displayed means between 1.01 and 1.04 transcripts per gene.

4.3.10 Isoform switch analysis

Isoform switch analysis was performed to identify genes exhibiting differential isoform usage between bladder and upper tract urothelial carcinoma using the IsoformSwitchAnalyzeR package. The analysis examined 715 isoforms from 298 genes across three bladder and four upper tract tumour samples.

No statistically significant isoform switches were detected in the cancer long-read dataset using the standard filtering criteria (dIF cutoff = 0.1, q-value < 0.05).

4.3.11 Molecular subtype classification

Consensus molecular subtyping using the MIBC classifier identified two subtypes across the seven tumour samples: Basal/Squamous (Ba/Sq, n=2, 28.6%) and Luminal Papillary (LumP, n=5, 71.4%). Classification confidence was assessed through separation level scores, with overall mean separation of 0.582 (median: 0.629). Ba/Sq tumours had high confidence classifications with a mean separation of 0.905 (range: 0.884-0.926), whilst LumP tumours had lower confidence (mean: 0.453, range: 0.230-0.645). One sample (989T) had a particularly low separation score (0.230).

Subtype distribution differed by tissue origin. Bladder samples were classified as two Ba/Sq tumours (n=2, 66.7%) and one LumP tumour (n=1, 33.3%), whilst upper tract samples were exclusively classified as LumP (n=4, 100%).

4.4 Discussion

4.4.1 Technical limitations in tumour sequencing

A number of caveats need to be noted as the fundamental limitations of this dataset to contextualise the findings presented in this chapter. The low RIN values observed in tumour samples indicate RNA degradation, a factor known to affect the availability of full length transcripts and the accuracy of transcript level quantification (Gallego Romero *et al.*, 2014). Furthermore, the short mean read length of 312.7 bp negates the primary advantage of long read technology for full length transcript capture, given that the median mRNA is approximately 2000 bp (Piovesan *et al.*, 2019). Consequently, this dataset suffers from considerable limitations as the combination of degraded RNA and resulting read lengths means it does not fully exploit long read sequencing capabilities for robust isoform characterisation. The mean processed read depth of 169,643 reads per sample with four samples containing fewer than 68,000 reads is very low for robust differential expression detection (Dong *et al.*, 2021). To contextualise this constraint the mean raw sequencing depth for these cancer samples (1.12 million reads, excluding outlier 1525T) was strikingly lower than the 24.3 million reads achieved for the normal tissue cohort. Due to this significantly lower depth, it forced a more conservative transcriptomic divergence in the long read analysis identifying only 1.43% of tested genes as significantly differentially expressed

compared to 13.1% in the higher depth short read dataset. These constraints can be attributed to the clinical challenges of obtaining high quality RNA from tumour specimens.

High attrition across the preprocessing pipeline can be attributed to the previously mentioned low quality input RNA and short read lengths. However, when accounting for individual sample outliers the similarity in full length read recovery rates between the bladder and upper tract cohorts suggests that anatomical tissue type itself does not substantially influence full length transcript detection. The low supplementary alignment frequency of 0.2% is an encouraging metric indicating mostly unambiguous mapping across the dataset, though the reduced mapping success that was observed in samples 2179T, 1829T, and 989T was a direct consequence of their exceptionally low sequencing depth (fewer than 56,000 mapped reads). Despite the quality limitations observed in sample 1525T, it was decided that this sample be retained in order to maintain adequate statistical power given the limited sample size. This decision was directly supported by the principal component analysis where sample 1525T did not separate from the rest of the cohort, demonstrating that its higher sequencing depth did not disproportionately skew its transcriptomic similarity to the other samples. Consequently, 5,196 genes representing 91.4% of the total were excluded following filtering. As mentioned in the previous chapter, a ~2% GC content difference persists between bladder and upper tract samples, confirming the association between tissue type and GC content even in malignancy.

The mean number of transcripts per gene of approximately 1.03 across all samples, reaching only 1.09 even in the highest depth sample, indicates severely limited detection of alternative isoforms throughout the dataset. This inability to robustly capture alternative splicing events is a direct consequence of the technical constraints identified during preprocessing. Because of this, the statistical power for detecting differential expression is substantially reduced. These data must therefore be interpreted with caution as the study is limited by the lack of information on moderately or lowly expressed genes meaning that these results represent only the most prominent transcriptomic differences between bladder and upper tract tumours.

4.4.2 Suppression of tissue specific programmes

The most obvious finding to emerge from the cancer dataset is that malignant transformation substantially suppresses the tissue specific molecular programmes that distinguish normal bladder from normal ureter urothelium. This pattern is consistent across pathway analyses, gene set enrichment and individual marker assessments. Together these represent a coherent biological signal, more than the limited output of just 7 significantly differentially expressed genes might initially suggest. Although this exceptionally low yield is in part a consequence of the technical limitations and low sequencing depth described above, it is also consistent with previous reports indicating that upper tract and bladder urothelial carcinomas are transcriptionally similar. Recent single-cell sequencing has demonstrated that despite remaining separable by enrichment analysis their tumour cells subpopulations correlate strongly (Zhang *et al.*, 2024). These findings suggest that the paucity of differentially expressed genes identified in this study likely reflects a genuine biological convergence during malignant transformation rather than being solely an artefact of technical constraints.

Overall marker activity in tumour samples represents a substantial reduction from normal tissue levels suggesting a global loss of differentiation in malignancy. Furthermore, divergent expression patterns and extremes in luminal and basal phenotypes indicate considerable inter-tumour heterogeneity. This generalised loss of differentiation is reflected across multiple specific pathway analyses. Both tumour types showed reduced SHH pathway activity compared to normal bladder. As observed in the normal data, the consistent absence of *GLI1* indicates that canonical pathway signalling remains quiescent in malignancy whilst the selective upregulation of *TLE1* and *ADGRG1* may indicate non-canonical pathway activation or compensatory mechanisms. The reduction in SHH signalling in bladder tumours and minimal pathway activity in upper tract carcinomas suggests that the developmental programme maintaining urothelial stratification and barrier function is disrupted during malignant transformation and this loss may contribute to the aberrant differentiation patterns observed in urothelial carcinomas (Amin, 2009; Shin *et al.*, 2014).

Similarly, both tumour types showed reduced pathway activity compared to normal bladder and ureter, indicating the suppression of Wnt signalling in malignancy. The reduction in overall Wnt pathway activity, particularly the loss of elevated signalling observed in normal bladder tissue suggests a disruption of the developmental programmes that maintain epithelial homeostasis during malignant transformation. Both tumour types additionally showed reduced TGF-beta pathway activity compared to normal tissues, representing a loss of the elevated bladder signalling observed in short read sequencing of normal samples. Given TGF-beta's dual role as a tumour suppressor and pro-metastatic signal, this reduction may reflect an escape from growth inhibition whilst retaining selective pathway components that promote invasion (Liu *et al.*, 2026). The expression values in the FGF pathway were similarly reduced from normal bladder and ureter levels indicating widespread pathway suppression in malignancy.

Interestingly, suppression was observed in immune and interferon signalling. Both tumour types showed reduced interferon pathway activity compared to normal bladder and ureter. The loss of interferon ligand expression, particularly the complete absence of Type II interferons that were elevated in the normal bladder, suggests the suppression of immune surveillance during malignant transformation. The retention of interferon receptors in the absence of ligands may indicate disrupted signalling or compensatory mechanisms. This pattern contrasts sharply with normal tissue where the bladder had substantially higher interferon expression and the loss of this programme in bladder tumours likely contributes to immune evasion and tumour progression. The heterogeneous immune marker expression patterns in urothelial carcinomas contrast with the elevated and relatively consistent immune infiltration observed in normal bladder tissue. The sporadic nature of marker expression, where individual samples showed exceptionally high levels whilst others remained minimal, suggests highly variable immune infiltration across individual tumours rather than consistent site specific patterns. This is further supported by GSEA where the strong inflammatory signatures that characterised normal bladder tissue were absent in cancerous bladder samples. The enrichment of hypoxia and oxidative phosphorylation pathways in bladder tumours contrasts with normal tissue where inflammatory and immune response pathways were the top enrichments in bladder urothelium. This shift from immune surveillance to metabolic alterations likely reflects the altered microenvironment characteristic of malignant tissue (Sun *et al.*, 2025). Together these results suggest that whilst some molecular differences persist between bladder and upper tract tumours, malignancy reduces the

pronounced inflammatory and immune phenotype that distinguishes normal bladder from ureter urothelium.

4.4.3 Partial retention of site specific identity

Whilst suppression and loss of tissue specific programmes is supported by the results, the data also supports the partial retention of site specific molecular identity in urothelial carcinomas reflecting their tissue of origin that persists through malignant transformation, albeit with substantially greater inter-tumour variability than observed in the normal tissues.

The patterns of HOX gene expression observed in the tumour samples mirror those observed in normal urothelium, suggesting partial retention of site specific developmental signatures. However, whilst urothelial carcinomas partially retain the HOX signatures characteristic of their tissue of origin, the patterns are more variable than in normal tissue, with high standard deviations suggesting substantial inter-tumour heterogeneity. The partial retention of luminal, basal and EMT site specific signatures occurs alongside a loss of differentiation markers overall. The extremes in luminal and basal phenotypes observed in certain samples reflect inter-tumour heterogeneity as to how much tissue identity is preserved from the normal data.

Within the FGF pathway upper tract carcinomas have significantly higher pathway activity driven by *FGFR3*, consistent with the known enrichment of *FGFR3* in upper tract carcinomas compared to bladder urothelial carcinomas (Robinson *et al.*, 2019). The reduction in overall pathway activity compared to normal tissues, coupled with the preferential retention of *FGFR3*-driven signalling in upper tract carcinomas, indicates site specific alterations in FGF pathway regulation in tumour samples. Similarly, the trend towards higher Wnt pathway activity in upper tract carcinomas, whilst modest, may reflect site specific differences in the regulatory mechanisms in Wnt signalling in urothelial malignancy.

One of the more significant findings to emerge from this study is that the most consistently retained site specific marker was *ANXA2*. This encodes a calcium dependent phospholipid-binding protein involved in apoptosis and membrane repair and yielded the most robust differential expression being significantly upregulated in bladder tissue across all three analyses, normal short read, normal long read and cancer long read (Huang *et al.*, 2025). It is interesting to note that the magnitude of *ANXA2* upregulation was greater in cancer tissue compared to normal tissue, suggesting this bladder specific marker has enhanced expression in tumour tissue and may represent a marker of both tissue identity and malignant transformation. These results are consistent with those of Guo et al. (2022), who demonstrated through TCGA cohort analysis that *ANXA2* expression is significantly elevated in bladder urothelial carcinoma compared to normal bladder tissue and that this overexpression strongly correlates with higher pathological grade and poor overall survival (Guo *et al.*, 2022).

The molecular subtype analysis further supports partial retention of site specific identity. The exclusive luminal papillary classification in upper tract samples aligns with previous research showing that upper tract urothelial carcinoma is heavily enriched for luminal papillary subtypes and a T-cell depleted microenvironment.(Robinson *et al.*, 2019; Fujii *et al.*, 2021; Yu *et al.*, 2024; Xiong *et al.*, 2025). Whilst classification confidence was generally robust for

Ba/Sq tumours, it was notably lower for LumP tumours. The exceptionally low separation score observed for bladder sample 989T indicates an uncertain subtype assignment, strongly suggesting the presence of underlying molecular heterogeneity within this tumour. The Reactome pathway analysis additionally identified nominal enrichment of cytokine signalling in cancer samples suggesting the retention of some immune related signalling despite the broader immune suppression observed in individual gene analysis. The nominal enrichment of the HIV life cycle pathway actually reflects an enrichment of ubiquitination and proteasomal degradation machinery rather than viral infection (Lata, Mishra and Banerjea, 2018).

4.4.4 Tumour specific transcriptional alterations

Beyond the patterns of suppression and partial retention of tissue specific signatures, the data identifies several transcriptional alterations that appear specific to the cancer context rather than representing simple extensions of normal tissue differences. The strong enrichment of hypoxia in bladder tumours suggests elevated hypoxic signalling, whilst the enrichment of oxidative phosphorylation indicates altered mitochondrial metabolism (Chen, Zhao and Li, 2023). These metabolic pathway enrichments represent a shift away from the immune and inflammatory enrichments that characterise normal bladder and reflect the altered tumour microenvironment rather than an amplification of normal tissue programmes.

Among the few significantly differentially expressed genes identified, metallothioneins including *MT2A* and *MT1E* showed increased expression in bladder cancer compared to upper tract cancer. These small cysteine rich proteins are involved in metal ion homeostasis and protection against oxidative stress and their differential expression may indicate altered metal metabolism or differential responses to oxidative stress between tumour types (Jimenez Jimenez *et al.*, 2025). This finding is consistent with that of Li et al. (2022), who identified *MT2A* as the most significantly upregulated metal ion binding gene implicated in urothelial cancer progression (Li *et al.*, 2022). *AQP3*, the only significantly upregulated upper tract gene encodes a water and glycerol channel protein involved in cellular permeability. It has been implicated in various cancers where its expression may influence cell proliferation and migration (Satooka and Hara-Chikuma, 2016; Marlar *et al.*, 2017). This suggests that it has a distinct regulatory or functional role in upper tract urothelial carcinoma compared to bladder cancer.

Cell cycle and proliferative gene expression was similar or slightly elevated in both tumour types relative to normal tissues, indicating retention or modest amplification of proliferative programmes in malignancy. The variance in expression across the cancer cohort, with individual samples showing much higher proliferative gene expression whilst others remain low, likely reflects underlying differences in tumour grade and molecular subtype rather than a consistent site specific pattern. The absence of clear separation between bladder and upper tract samples in PCA, with PC1 contributors spanning diverse processes including endosomal sorting, cytoskeletal regulation, mitochondrial metabolism and detoxification, and no specific biological process dominating further supports the view that transcriptomic variation in this cancer cohort is driven by multiple sources including individual tumour biology and immune microenvironment. Specifically the top PC1 contributors included *CHMP4A* (endosomal sorting), *MYL12B* (cytoskeletal regulation), *HSD17B10* (mitochondrial metabolism), and *GSTP1* (detoxification), whilst PC2 captured distinct variation in cell cycle

and vesicle trafficking through genes such as *ING3*, *SEC23A* and *ALDH2 (He et al., 2001; Tsang et al., 2006; Kondo et al., 2012; Yang et al., 2012; Amodio et al., 2013; Chen, Joshi and Mochly-Rosen, 2016; Na, Oh and Jung, 2023; Gudur et al., 2024)*. *CD14* and *CD3D* among PC2 contributors likely reflect underlying differences in immune cell infiltration between individual tumour samples.

No statistically significant isoform switches were detected in the cancer long read dataset. As was observed in the normal tissue long read results, this is likely an artefact of the limitations, specifically the low read depth and fragmented nature of the RNA. The highly fragmented state of the input material directly negated the primary advantage of long read technology for full length transcript sequencing. The isoform level regulatory differences detected in normal tissues by short read analysis, particularly bladder specific preference for isoforms with reduced protein coding capacity through ORF loss, coding to non-coding switches, intron retention and NMD sensitivity, may help to explain why bladder and upper tract urothelial carcinomas exhibit distinct clinical behaviours and molecular profiles despite their histological similarity. Whether these tissue specific isoform preferences persist or are altered in tumour samples cannot be determined from this dataset and further research using higher depth long read sequencing with higher quality tumour RNA is required.

In summary, these results show that bladder and upper tract urothelial carcinomas share broadly similar transcriptomic profiles at this level of resolution, with malignant transformation disrupting the molecular differences distinguishing normal tissues. The dramatic loss of tissue specific signatures, particularly complete interferon suppression and reduction of developmental pathway activity suggests that malignancy diminishes tissue identity programmes, whilst the partial retention of HOX signatures, *FGFR3* driven signalling in upper tract carcinomas and the bladder specific *ANXA2* signature indicates that site of origin identity is not entirely erased. However, more research is needed before definitive conclusions can be drawn.

Chapter 5: Discussion

5.1 Introduction

The overarching aim of this study was to determine whether the distinct clinical behaviours observed in bladder urothelial carcinoma (BUC) and upper tract urothelial carcinoma (UTUC) stem from fundamental molecular differences in their tissues of origin. The term urothelium is used to describe the tissue which lines both the upper and lower urinary tracts (Jafari and Rohn, 2022). Historically, the urothelium has been regarded clinically as a uniform tissue, with systemic treatment approaches from bladder cancer routinely extrapolated to UTUC despite limited evidence (Escobar *et al.*, 2025). However, recent research suggests this approach may be biologically flawed, as UTUC exhibits higher rates of stage matched invasion and poorer outcomes, with approximately 60% presenting with muscle invasion versus 25% for bladder cancer (Nally *et al.*, 2024).

This thesis challenges the assumption of urothelial uniformity through combined Illumina short-read and Oxford Nanopore long-read sequencing. The findings demonstrate that adult bladder and ureter urothelium retain distinct transcriptomic identities driven by divergent embryological origins, bladder from endodermal urogenital sinus, ureter from mesodermal ureteric bud (Rehman and Ahmed, 2024). The analysis of histologically normal tissue identified 3,698 differentially expressed genes. The malignant tissue analysis, whilst deeply impacted by sequencing data quality, highlighted differences in metabolic and oxidative stress pathways. This discussion interprets these findings in the context of urothelial development, functional adaptation and implications for site specific carcinogenesis.

5.2 Retention of Embryological Identity in Adult Tissues

The bladder and ureter urothelium retain fundamentally distinct molecular identities in adulthood. The differential expression of HOX transcription factors provides evidence for the persistence of developmental programmes. In the short-read analysis ureter samples had significantly higher HOXB cluster expression and HOXD cluster, with enrichment of anterior and central positional genes. Long-read sequencing confirmed this with elevated HOXB and HOXD in ureter. This pattern is consistent with the ureter's mesodermal origin from the ureteric bud, which expresses HOXB genes during nephric duct morphogenesis (Sanchez-Ferras *et al.*, 2021). The bladder, derived from endodermal urogenital sinus, exhibited reduced HOXB, HOXC and HOXD expression whilst maintaining similar HOXA levels. These findings accord with evidence that positional identity from embryonic HOX expression is maintained in adult tissues, particularly stem cells, providing mechanisms for cell identity and fate restriction (Steens and Klein, 2022).

HOX genes establish positional coordinates, regulatory networks governing tissue specific responses to growth factors, differentiation signals, and stress stimuli (Keleş and Gunel-Ozcan, 2025). The persistence of distinct HOX patterns suggests bladder and ureter exist in different transcriptional states under homeostatic conditions, creating distinct molecular profiles that may differentially permit or restrict oncogenic pathways.

Importantly, whilst HOX differences were most pronounced in normal tissues, they persisted in malignant samples, though with reduced magnitude. Upper tract tumours maintained elevated HOXB expression, with overall HOX activity remaining higher. This retained HOXB signature in the ureter likely reflects its posterior mesodermal origin driven by retinoic acid signaling, distinct from the anterior, endodermal patterning of the bladder which relies on SHH and Wnt pathways (Bohnenpoll, Weiss, *et al.*, 2017; Liaw *et al.*, 2018). This suggests embryological patterning is not entirely erased during malignant transformation, supporting the hypothesis that tissue of origin constrains available tumourigenic trajectories. The combination of retained HOX expression and acquired oncogenic mutations may explain why morphologically similar urothelial carcinomas at different sites exhibit different clinical behaviours.

Whilst these developmental signatures were robustly detected it is important to acknowledge that this analysis relied on a modest sample size (n=3) from different clinical contexts, cystoscopy for uterine prolapse versus transplant donors in the normal urothelial samples. Although the tissues were histologically normal, not all systemic factors can be entirely excluded. Therefore, whilst the HOX and positional identity data provide a compelling molecular baseline, expansion to large-scale profiling integrating genomic and transcriptomic analyses across cohorts of over 200 samples would be beneficial to fully validate these defining characteristics of the adult human urothelium (Mer *et al.*, 2016).

5.3 Functional Specialisation and Divergent Immune Landscapes

Beyond developmental patterning, this study reveals molecular functional specialisation, most notably within immune surveillance programmes. There was systematic upregulation of immune and inflammatory pathways in bladder tissue. TNF-alpha signalling via NF-Kappa B showed strongest enrichment, with elevated Type II interferon receptors and memory CD4+ T cell markers. Overall immune marker expression was significantly higher in bladder samples. Although absolute quantification values varied between sequencing type, likely due to differences in sequencing depth and the reduced gene detection rate in the long-read dataset, the relative expression trends demonstrated strong correlation. This consistency suggests that these are true biological signatures.

This constitutive immune activation could be interpreted as an adaptive response to chronic pathogen exposure. Unlike the ureter, where urine exposure is transient during peristalsis, the bladder stores urine for longer periods, necessitating primed immune responses against pathogens and toxins (Ortega Martell, 2020; Lescay *et al.*, 2025). This elevated inflammation may also reflect the bladder microbiome, which is less diverse in the upper tract (Ackerman and Chai, 2019; Pei *et al.*, 2025). Thus, the bladder's immune rich microenvironment represents functional adaptation to potentially toxic and microbial urine storage (Huang *et al.*, 2021).

These findings provide a biological basis for the different immune contexts in BUC and UTUC. Whilst data suggest immune depletion reflects ureter baseline state, the relationship is likely mediated by molecular subtype. The upper tract preferentially develops tumours with *FGFR3* mutations and luminal-papillary characteristics, intrinsically associated with low

immune infiltration (Robinson *et al.*, 2019). Therefore, observed immune differences are likely linked to the molecular progression favoured by the upper tract environment in muscle invasive disease.

Future studies should demonstrate whether this immune active state is an intrinsic property of the endoderm derived bladder urothelium or a response to the urinary environment. Confirmation of the spatial distribution of these cells should be done through immunohistochemistry.

The immune depleted ureter baseline presents theoretical barriers to immunotherapy, aligning with challenges facing intravesical BCG in the upper tract. However, whilst UTUC is biologically colder, checkpoint inhibitor efficacy appears heterogeneous and etiology dependent (Thouvenin *et al.*, 2021; Meeks *et al.*, 2023; Chen *et al.*, 2025). Large adjuvant trials suggest worse outcomes than bladder cancer and yet aristolochic acid exposed populations have superior responses, high tumor mutation burden may overcome baseline immune depletion (Meeks *et al.*, 2023; Chen *et al.*, 2025). Lynch syndrome associated UTUC similarly rescues immunogenicity through microsatellite instability (Hall *et al.*, 2025).

Complementing immune specialisation, the bladder samples demonstrated dramatic upregulation of barrier function genes, notably *DSG3*. This highlights mechanobiological divergence where bladder must withstand sustained distension requiring robust desmosomal adhesion, whilst ureter requires flexibility for peristalsis (Lescay *et al.*, 2025). Elevated luminal markers in bladder support the expression of genes for a highly differentiated and impermeable superficial layer during prolonged storage.

5.4 Molecular Divergences in Malignancy

Malignant tissue analysis revealed different patterns. Whilst normal tissues had 3,698 differentially expressed genes (13.1% of transcriptome), only 7 genes showed significant differential expression between tumours. Technical limitations, inadequate sequencing depth (mean 169,643 reads versus recommended >5 million) and poor RNA quality (RIN 4.4-6.8), considerably reduced the power of detection in the tumour dataset (Encode Project, 2017).

Despite these constraints, identified genes offer insights. *AQP3* downregulation in bladder tumours likely reflects stage heterogeneity rather than a difference due to anatomical site. *AQP3* is known to be expressed in non-muscle invasive bladder cancer but is lost during invasion (Rubenwolf *et al.*, 2015; Breyer *et al.*, 2017).

The coordinated metallothionein upregulation (*MT2A*, *MT1E*) in bladder versus UTUC has clinical implications as these chelators are implicated in platinum resistance. Therefore, the elevated metallothionein expression in the BUC samples suggests intrinsic baseline resistance to platinum therapies, reinforcing needs for accurate molecular stratification before applying uniform chemotherapy protocols (Borchert *et al.*, 2020; Sung *et al.*, 2022).

Most revealing was the elevated FGF pathway activity in upper tract tumours reversing what was seen in the normal tissue samples. This aligns with higher *FGFR3* mutation frequencies in UTUC. *FGFR3* mutations are associated with low grade bladder disease yet UTUC is

frequently invasive, the findings found that the UTUCs were all the luminal-papillary subtype (Foth *et al.*, 2018; Tomiyama *et al.*, 2022).

5.5 The Origin Hypothesis

This thesis proposes a tissue of origin model explaining distinct clinical behaviours. This model hypothesises that baseline molecular differences established during embryogenesis and maintained throughout adulthood create distinct cellular profiles that differentially constrain or enable specific oncogenic trajectories.

Evidence supporting this includes the retention of embryological positional identity markers persisting through transformation from normal tissue into tumours and divergent functional adaptations reflecting distinct mechanical and exposure environments.

5.6 Clinical Implications

Current risk stratification relies on clinicopathological features, with molecular subtyping focused on pan-urothelial systems (Kang, Kim and Yun, 2020). However, recent attempts applying bladder subtypes to UTUC reveal important limitations, particularly the failure of basal/non-luminal signatures to predict favorable chemotherapy responses as they do in bladder cancer (Hwang *et al.*, 2024). Site specific prognostic signatures may prove more useful than pan-urothelial classifiers. Whilst the clinical standard for UTUC has shifted to adjuvant therapy, unlike the neoadjuvant standard in bladder cancer, the underlying biological assumptions remain extrapolated (Vilhais and Fontes-Sousa, 2025). The divergent therapeutic response of basal-like tumors, chemosensitive in the bladder yet resistant in the upper tract, underscores the biological distinctiveness of UTUC and invalidates the direct extrapolation of bladder derived molecular models (Hwang *et al.*, 2024). Constitutive immune differences raise questions about differential immunotherapy efficacy. Future studies should stratify by anatomical site and consider tissue specific immune biomarkers (Catalano *et al.*, 2023; Sanguedolce *et al.*, 2025).

5.7 Conclusion

This thesis indicates that the morphologically similar, bladder and ureter urothelium maintain distinct molecular identities reflecting divergent embryological origins and functional specialisations. The identification of 3,698 differentially expressed genes in normal tissues, with coordinated HOX, immune, and barrier function differences, challenges the traditional paradigm of a molecularly uniform urothelium.

Retention of positional identity markers supports the hypothesis that the tissue of origin creates distinct molecular ecosystems persisting into adulthood, constraining and enabling specific oncogenic pathways. Whilst technical limitations prevented comprehensive tumour characterisation, available evidence suggests both persistent baseline differences and cancer specific alterations contribute to distinct clinical behaviours.

Clinically, these findings highlight the potential value of conceptualising urothelial carcinoma as anatomically distinct diseases requiring site specific biomarkers, risk stratification, and

therapeutic strategies. Continued extrapolation of bladder protocols to UTUC may represent biologically flawed approaches failing to account for fundamental molecular differences.

This combination of findings provides support for the conceptual premise of the origin hypothesis. It proposes that morphologically identical tumours behave differently not only due to acquired mutations and their functions, but because of distinct embryological contexts from which these tumours arise in their distinct locations. Further studies should focus on these tissue specific developmental signatures in larger and higher quality, multi-omic cohorts. Investigating these baseline biological differences is an important issue for future research, as these findings may help to design improved site specific therapies for patients with urothelial carcinoma.

References

Acharya, P. *et al.* (2004) 'Distribution of the tight junction proteins ZO-1, occludin, and claudin-4, -8, and -12 in bladder epithelium', *American Journal of Physiology. Renal Physiology*, 287(2), pp. F305–18.

Ackerman, A.L. and Chai, T.C. (2019) 'The bladder is not sterile: An update on the urinary microbiome', *Current bladder dysfunction reports*, 14(4), pp. 331–341.

Ahsan, M.U. *et al.* (2023) 'A survey of algorithms for the detection of genomic structural variants from long-read sequencing data', *Nature methods*, 20(8), pp. 1143–1158.

Aitken, K.J. and Bägli, D.J. (2009) 'The bladder extracellular matrix. Part I: architecture, development and disease', *Nature reviews. Urology*, 6(11), pp. 596–611.

Allory, Y. *et al.* (2014) 'Telomerase reverse transcriptase promoter mutations in bladder cancer: high frequency across stages, detection in urine, and lack of association with outcome', *European urology*, 65(2), pp. 360–366.

Ament, I.H. *et al.* (2025) 'Long-read RNA sequencing: A transformative technology for exploring transcriptome complexity in human diseases', *Molecular therapy: the journal of the American Society of Gene Therapy*, 33(3), pp. 883–894.

Amin, M.B. (2009) 'Histological variants of urothelial carcinoma: diagnostic, therapeutic and prognostic implications', *Modern pathology: an official journal of the United States and Canadian Academy of Pathology, Inc*, 22 Suppl 2(S2), pp. S96–S118.

Amodio, G. *et al.* (2013) 'Endoplasmic reticulum stress reduces COPII vesicle formation and modifies Sec23a cycling at ERESs', *FEBS letters*, 587(19), pp. 3261–3266.

Andersson, K.-E. and Arner, A. (2004) 'Urinary bladder contraction and relaxation: physiology and pathophysiology', *Physiological reviews*, 84(3), pp. 935–986.

Andrews, S. (2010) *Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data*. Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (Accessed: 17 June 2025).

Arakawa, H. *et al.* (2008) 'Scent marking behavior as an odorant communication in mice', *Neuroscience and biobehavioral reviews*, 32(7), pp. 1236–1248.

Asif, S. *et al.* (2022) 'Hmgcs2-mediated ketogenesis modulates high-fat diet-induced hepatosteatosis', *Molecular metabolism*, 61(101494), p. 101494.

Aulehla, A. and Pourquié, O. (2010) 'Signaling gradients during paraxial mesoderm development', *Cold Spring Harbor perspectives in biology*, 2(2), p. a000869.

Baard, J. *et al.* (2024) 'Unveiling the challenges of UTUC biopsies and cytology: insights from a global real-world practice study', *World Journal of Urology*, 42(1), p. 177.

Babjuk, M. *et al.* (2019) 'European Association of Urology guidelines on non-muscle-invasive Bladder Cancer (TaT1 and carcinoma in situ) - 2019 update', *European Urology*, 76(5), pp. 639–657.

Bajorin, D.F. *et al.* (2021) 'Adjuvant nivolumab versus placebo in muscle-invasive urothelial carcinoma', *The New England Journal of Medicine*, 384(22), pp. 2102–2114.

Baker, S.C., Shabir, S. and Southgate, J. (2014) 'Biomimetic urothelial tissue models for the in vitro evaluation of barrier physiology and bladder drug efficacy', *Molecular pharmaceutics*, 11(7), pp. 1964–1970.

Balin, S.J. *et al.* (2018) 'Human antimicrobial cytotoxic T lymphocytes, defined by NK receptors and antimicrobial proteins, kill intracellular bacteria', *Science immunology*, 3(26), p. eaat7668.

Barak, H. *et al.* (2012) 'FGF9 and FGF20 maintain the stemness of nephron progenitors in mice and man', *Developmental cell*, 22(6), pp. 1191–1207.

Baskin, L.S. *et al.* (1996) 'Mesenchymal-epithelial interactions in the bladder', *World journal of urology*, 14(5), pp. 301–309.

Benamran, D. *et al.* (2020) 'Risk stratification for upper tract urinary carcinoma', *Translational andrology and urology*, 9(4), pp. 1799–1808.

Benítez, R. *et al.* (2023) 'Characterization of the tumor-infiltrating immune repertoire in muscle invasive bladder cancer', *Frontiers in immunology*, 14, p. 986598.

Benkhadra, R. *et al.* (2022) 'Systematic review and meta-analysis of cisplatin based neoadjuvant chemotherapy in muscle invasive bladder cancer', *Bladder cancer (Amsterdam, Netherlands)*, 8(1), pp. 5–17.

Birzele, F., Csaba, G. and Zimmer, R. (2008) 'Alternative splicing and protein structure evolution', *Nucleic acids research*, 36(2), pp. 550–558.

Biton, A. *et al.* (2014) 'Independent component analysis uncovers the landscape of the bladder tumor transcriptome and reveals insights into luminal and basal subtypes', *Cell reports*, 9(4), pp. 1235–1245.

Blencowe, B.J. (2006) 'Alternative splicing: new insights from global analyses', *Cell*, 126(1), pp. 37–47.

Blighe, K., Rana, S. and Lewis, M. (2023) *EnhancedVolcano: publication-ready volcano plots with enhanced colouring and labeling*. Available at: https://bioconductor.statistik.tu-dortmund.de/packages/3.18/bioc/vignettes/EnhancedVolcano/inst/doc/EnhancedVolcano.html (Accessed: 15 December 2025).

Bohnenpoll, T., Wittern, A.B., *et al.* (2017) 'A SHH-FOXF1-BMP4 signaling axis regulating growth and differentiation of epithelial and mesenchymal tissues in ureter development', *PLoS genetics*, 13(8), p. e1006951.

Bohnenpoll, T., Weiss, A.-C., *et al.* (2017) 'Retinoic acid signaling maintains epithelial and mesenchymal progenitors in the developing mouse ureter', *Scientific reports*, 7(1), p. 14803.

Bohnenpoll, T. and Kispert, A. (2014) 'Ureter growth and differentiation', *Seminars in cell & developmental biology*, 36, pp. 21–30.

Bolla, S.R. *et al.* (2025) 'Histology, bladder', *StatPearls*. Treasure Island (FL): StatPearls Publishing.

Borchert, S. *et al.* (2020) 'Impact of metallothionein-knockdown on cisplatin resistance in malignant pleural mesothelioma', *Scientific reports*, 10(1), p. 18677.

Breyer, J. *et al.* (2017) 'Aquaporin 3 expression loss in urothelial carcinoma: Association with tumor invasion depth, but not with grading?', *Bladder cancer (Amsterdam, Netherlands)*,

3(1), pp. 31–34.

Brunchault, M.R. *et al.* (2025) ‘Proteomics-based characterization of ribosome heterogeneity in adult mouse organs’, *Cellular and Molecular Life Sciences*, 82(1), p. 175.

Brunskill, E.W. *et al.* (2008) ‘Atlas of gene expression in the developing kidney at microanatomic resolution’, *Developmental cell*, 15(5), pp. 781–791.

Bryan, R.T. *et al.* (2010) ‘Mechanisms of recurrence of Ta/T1 bladder cancer’, *Annals of the Royal College of Surgeons of England*, 92(6), pp. 519–524.

Cancer Research UK (2022) *Survival for bladder cancer*, *Cancer Research UK*. Available at: https://www.cancerresearchuk.org/about-cancer/bladder-cancer/survival (Accessed: 5 July 2025).

Castro, F. *et al.* (2018) ‘Interferon-gamma at the Crossroads of tumor immune surveillance or evasion’, *Frontiers in immunology*, 9, p. 847.

Catalano, M. *et al.* (2023) ‘Immunotherapy-related biomarkers: Confirmations and uncertainties’, *Critical reviews in oncology/hematology*, 192(104135), p. 104135.

Chamie, K. *et al.* (2013) ‘Recurrence of high-risk bladder cancer: a population-based analysis’, *Cancer*, 119(17), pp. 3219–3227.

Chang, S.L. *et al.* (1998) ‘Role of type III collagen in bladder filling’, *Neurourology and urodynamics*, 17(2), pp. 135–145.

Chen, C.-H. *et al.* (2012) ‘Aristolochic acid-associated urothelial cancer in Taiwan’, *Proceedings of the National Academy of Sciences of the United States of America*, 109(21), pp. 8241–8246.

Chen, C.-H., Joshi, A.U. and Mochly-Rosen, D. (2016) ‘The role of mitochondrial aldehyde dehydrogenase 2 (ALDH2) in neuropathology and neurodegeneration’, *Acta neurologica taiwanica*, 25(4)(4), pp. 111–123.

Cheng, L. *et al.* (2009) ‘Staging and reporting of urothelial carcinoma of the urinary bladder’, *Modern pathology: an official journal of the United States and Canadian Academy of Pathology, Inc*, 22 Suppl 2(S2), pp. S70–95.

Cheng, T.-L. and Qiu, Z. (2014) ‘MeCP2: multifaceted roles in gene regulation and neural development’, *Neuroscience bulletin*, 30(4), pp. 601–609.

Chen, L. *et al.* (2023) ‘Platelets in the tumor microenvironment and their biological effects on cancer hallmarks’, *Frontiers in Oncology*, 13, p. 1121401.

Chen, S. *et al.* (2018) ‘fastp: an ultra-fast all-in-one FASTQ preprocessor’, *Bioinformatics (Oxford, England)*, 34(17), pp. i884–i890.

Chen, S. (2024) *fastplong: Ultra-fast preprocessing and quality control for long-read sequencing data*. Github. Available at: https://github.com/OpenGene/fastplong (Accessed: 1 July 2025).

Chen, W., Zhao, H. and Li, Y. (2023) ‘Mitochondrial dynamics in health and disease: mechanisms and potential targets’, *Signal transduction and targeted therapy*, 8(1), p. 333.

Chen, Y.-C. *et al.* (2025) ‘Efficacy of adjuvant nivolumab in patients with upper tract predominant urothelial carcinoma: A single-center real-world study’, *Urologic oncology*,

43(12), pp. 697.e11–697.e18.

Chen, Z. *et al.* (2020) ‘Single-cell RNA sequencing highlights the role of inflammatory cancer-associated fibroblasts in bladder urothelial carcinoma’, *Nature communications*, 11(1), p. 5077.

Choi, W. *et al.* (2014) ‘Identification of distinct basal and luminal subtypes of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy’, *Cancer cell*, 25(2), pp. 152–165.

Chua, E.L. *et al.* (2000) ‘Cloning of TC-1 (C8orf4), a novel gene found to be overexpressed in thyroid cancer’, *Genomics*, 69(3), pp. 342–347.

Chuang, T.-P. *et al.* (2023) ‘ALK fusion NSCLC oncogenes promote survival and inhibit NK cell responses via SERPINB4 expression’, *Proceedings of the National Academy of Sciences of the United States of America*, 120(8), p. e2216479120.

Clavel, J. *et al.* (1994) ‘Occupational exposure to polycyclic aromatic hydrocarbons and the risk of bladder cancer: a French case-control study’, *International journal of epidemiology*, 23(6), pp. 1145–1153.

Climente-González, H. *et al.* (2017) ‘The functional impact of alternative splicing in cancer’, *Cell reports*, 20(9), pp. 2215–2226.

Cohen, S.M. and Lawson, T.A. (1995) ‘Rodent bladder tumors do not always predict for humans’, *Cancer letters*, 93(1), pp. 9–16.

Conklin, D.J. *et al.* (2009) ‘Increased sensitivity of glutathione S-transferase P-null mice to cyclophosphamide-induced urinary bladder toxicity’, *The Journal of Pharmacology and Experimental Therapeutics*, 331(2), pp. 456–469.

Contreras-Sanz, A. *et al.* (2025) ‘Proteomic profiling identifies muscle-invasive bladder cancers with distinct biology and responses to platinum-based chemotherapy’, *Nature communications*, 16(1), p. 1240.

Cooper, E.H. (1972) ‘The biology of bladder cancer’, *Annals of the Royal College of Surgeons of England*, 51(1), pp. 1–16.

Cotillas, E.A. *et al.* (2024) ‘A versatile and upgraded version of the LundTax classification algorithm applied to independent cohorts’, *The Journal of molecular diagnostics: JMD*, 26(12), pp. 1081–1101.

Cross, W.R. *et al.* (2005) ‘A biomimetic tissue from cultured normal human urothelial cells: analysis of physiological function’, *American journal of physiology. Renal physiology*, 289(2), pp. F459–68.

Dai, P. *et al.* (2023) ‘A pancancer analysis of the oncogenic role of cyclin B1 (CCNB1) in human tumors’, *Scientific reports*, 13(1), p. 16226.

Dalghi, M.G. *et al.* (2020) ‘The urothelium: Life in a liquid environment’, *Physiological reviews*, 100(4), pp. 1621–1705.

Dalghi, M.G. *et al.* (2021) ‘Functional roles for PIEZO1 and PIEZO2 in urothelial mechanotransduction and lower urinary tract interoception’, *JCI insight*, 6(19), p. e152984.

Damaser, M.S. and Lehman, S.L. (1995) ‘The effect of urinary bladder shape on its mechanics during filling’, *Journal of biomechanics*, 28(6), pp. 725–732.

Damrauer, J.S. *et al.* (2014) 'Intrinsic subtypes of high-grade bladder cancer reflect the hallmarks of breast cancer biology', *Proceedings of the National Academy of Sciences of the United States of America*, 111(8), pp. 3110–3115.

Decarli, A. *et al.* (1985) 'Bladder cancer mortality of workers exposed to aromatic amines: analysis of models of carcinogenesis', *British journal of cancer*, 51, pp. 707–712.

Di Tommaso, P. *et al.* (2017) 'Nextflow enables reproducible computational workflows', *Nature biotechnology*, 35(4), pp. 316–319.

Dobin, A. *et al.* (2013) 'STAR: ultrafast universal RNA-seq aligner', *Bioinformatics (Oxford, England)*, 29(1), pp. 15–21.

van Doeveren, T. *et al.* (2021) 'Rising incidence rates and unaltered survival rates for primary upper urinary tract urothelial carcinoma: a Dutch population-based study from 1993 to 2017', *BJU international*, 128(3), pp. 343–351.

Dong, X. *et al.* (2021) 'The long and the short of it: unlocking nanopore long-read RNA sequencing data with short-read differential expression analysis tools', *NAR Genomics and Bioinformatics*, 3(2), p. lqab028.

Downes, M.R. (2022) 'Invasive urothelial carcinoma: Subtypes and divergent differentiation', *Urologic Cancers*. Exon Publications, pp. 1–12.

Duan, H. *et al.* (2024) 'MT-TN mutations lead to progressive mitochondrial encephalopathy and promotes mitophagy', *Biochimica et biophysica acta. Molecular basis of disease*, 1870(4), p. 167043.

Duke University School of Medicine (2015) *Urogenital Development*, *Embryology Learning Resources*. Available at: https://embryology.oit.duke.edu/urogenital/urogenital.html (Accessed: 28 August 2025).

Duplan, S.M. *et al.* (2009) 'Impact of Sim1 gene dosage on the development of the paraventricular and supraoptic nuclei of the hypothalamus: Sim1 gene dosage on development of the hypothalamus', *The European journal of neuroscience*, 30(12), pp. 2239–2249.

Dyrskjøt, L. *et al.* (2023) 'Bladder cancer', *Nature reviews. Disease primers*, 9(1), p. 58.

Earl, J. *et al.* (2015) 'The UBC-40 Urothelial Bladder Cancer cell line index: a genomic resource for functional studies', *BMC genomics*, 16(1), p. 403.

Eblin, K.E., Bredfeldt, T.G. and Gandolfi, A.J. (2008) 'Immortalized human urothelial cells as a model of arsenic-induced bladder cancer', *Toxicology*, 248(2-3), pp. 67–76.

Encode Project (2017) *Bulk RNA-seq Data Standards and Processing Pipeline*. Available at: https://www.encodeproject.org/data-standards/rna-seq/long-rnas/ (Accessed: 9 December 2025).

epi2me-labs (2024) *pychopper: cDNA read preprocessing*. Github. Available at: https://github.com/epi2me-labs/pychopper (Accessed: 1 July 2025).

Escobar, D. *et al.* (2025) 'Diagnosis and management of Upper Tract urothelial carcinoma: A review', *Cancers*, 17(15), p. 2467.

Espinosa, E. *et al.* (2024) 'Advancements in long-read genome sequencing technologies and algorithms', *Genomics*, 116(3), p. 110842.

Eymin, B. *et al.* (2001) ‘Distinct pattern of E2F1 expression in human lung tumours: E2F1 is upregulated in small cell lung carcinoma’, *Oncogene*, 20(14), pp. 1678–1687.

Fagerland, M.W., Sandvik, L. and Mowinckel, P. (2011) ‘Parametric methods outperformed non-parametric methods in comparisons of discrete numerical variables’, *BMC Medical Research Methodology*, 11(1), p. 44.

Fink, E.E. *et al.* (2022) ‘Single-cell and spatial mapping Identify cell types and signaling Networks in the human ureter’, *Developmental Cell*, 57(15), pp. 1899–1916.e6.

Fiuk, J.V. and Schwartz, B.F. (2016) ‘Upper tract urothelial carcinoma: Paradigm shift towards nephron sparing management’, *World journal of nephrology*, 5(2), pp. 158–165.

Foth, M. *et al.* (2018) ‘FGFR3 mutation increases bladder tumourigenesis by suppressing acute inflammation’, *The Journal of pathology*, 246(3), pp. 331–343.

Fujii, Y. *et al.* (2021) ‘Molecular classification and diagnostics of upper urinary tract urothelial carcinoma’, *Cancer cell*, 39(6), pp. 793–809.e8.

Gaffney, C.D. *et al.* (2023) ‘Bladder cancer carcinogens: Opportunities for risk reduction’, *European Urology Focus*, 9(4), pp. 575–578.

Gallego Romero, I. *et al.* (2014) ‘RNA-seq: impact of RNA degradation on transcript quantification’, *BMC biology*, 12(1), p. 42.

Garje, R. *et al.* (2020) ‘Fibroblast growth factor receptor (FGFR) inhibitors in urothelial cancer’, *The Oncologist*, 25(11), pp. e1711–e1719.

Genuth, N.R. and Barna, M. (2018) ‘Heterogeneity and specialized functions of translation machinery: from genes to organisms’, *Nature reviews. Genetics*, 19(7), pp. 431–452.

Georgas, K.M. *et al.* (2015) ‘An illustrated anatomical ontology of the developing mouse lower urogenital tract’, *Development (Cambridge, England)*, 142(10), pp. 1893–1908.

Gills, J. *et al.* (2018) ‘A patient-derived orthotopic xenograft model enabling human high-grade urothelial cell carcinoma of the bladder tumor implantation, growth, angiogenesis, and metastasis’, *Oncotarget*, 9(66), pp. 32718–32729.

Giudici, N. *et al.* (2021) ‘Characteristics of upper urinary tract urothelial carcinoma in the context of bladder cancer: a narrative review’, *Translational Andrology and Urology*, 10(10), pp. 4036–4050.

Glaser, A.P. *et al.* (2018) ‘APOBEC-mediated mutagenesis in urothelial carcinoma is associated with improved survival, mutations in DNA damage response genes, and immune response’, *Oncotarget*, 9(4), pp. 4537–4548.

Grabe-Heyne, K. *et al.* (2023) ‘Intermediate and high-risk non-muscle-invasive bladder cancer: an overview of epidemiology, burden, and unmet needs’, *Frontiers in oncology*, 13, p. 1170124.

Grelet, S. *et al.* (2017) ‘A regulated PNUTS mRNA to lncRNA splice switch mediates EMT and tumour progression’, *Nature cell biology*, 19(9), pp. 1105–1115.

Griffin, J. *et al.* (2024) ‘Verification of molecular subtyping of bladder cancer in the GUSTO clinical trial’, *The journal of pathology. Clinical research*, 10(2), p. e12363.

Gudur, R.A. *et al.* (2024) ‘Genetic polymorphisms in glutathione S-Transferase (GST) gene

and their correlation with toxicity of chemotherapy in breast cancer patients', *Asian Pacific journal of cancer prevention: APJCP*, 25(7), pp. 2271–2282.

Guo, C. *et al.* (2022) 'Higher expression of annexin A2 in metastatic bladder urothelial carcinoma promotes migration and invasion', *Cancers*, 14(22), p. 5664.

Guo, P. *et al.* (2018) 'TRIM31 is upregulated in hepatocellular carcinoma and promotes disease progression by inducing ubiquitination of TSC1-TSC2 complex', *Oncogene*, 37(4), pp. 478–488.

Hall, R. *et al.* (2025) 'Detection of urothelial carcinoma in Lynch syndrome using microsatellite instability analysis of urine cell-free DNA', *EBioMedicine*, 121(105969), p. 105969.

Halperin, R.F. *et al.* (2021) 'Improved methods for RNAseq-based alternative splicing analysis', *Scientific reports*, 11(1), p. 10740.

Hansen, K.D., Brenner, S.E. and Dudoit, S. (2010) 'Biases in Illumina transcriptome sequencing caused by random hexamer priming', *Nucleic acids research*, 38(12), p. e131.

Hardy, C.S.C. *et al.* (2022) 'Immunohistochemical assays for bladder cancer molecular subtyping: Optimizing parsimony and performance of Lund taxonomy classifiers', *The journal of histochemistry and cytochemistry: official journal of the Histochemistry Society*, 70(5), pp. 357–375.

Harish, B.S. and Uppuluri, K.B. (2018) 'Microbial serine protease inhibitors and their therapeutic applications', *International journal of biological macromolecules*, 107(Pt B), pp. 1373–1387.

Hayashida, T. *et al.* (2010) 'HOXB9, a gene overexpressed in breast cancer, promotes tumorigenicity and lung metastasis', *Proceedings of the National Academy of Sciences of the United States of America*, 107(3), pp. 1100–1105.

Hayes, B.W. and Abraham, S.N. (2016) 'Innate immune responses to bladder infection', *Microbiology spectrum*, 4(6). Available at: https://doi.org/10.1128/microbiolspec.UTI-0024-2016.

Healthdirect Australia (2025) *Urinary tract infection (UTI)*, *Healthdirect Australia*. Available at: https://www.healthdirect.gov.au/urinary-tract-infection-uti (Accessed: 22 July 2025).

Hedegaard, J. *et al.* (2016) 'Comprehensive transcriptional analysis of early-stage urothelial carcinoma', *Cancer cell*, 30(1), pp. 27–42.

van der Heijden, A.G. *et al.* (2025) 'European Association of Urology guidelines on muscle-invasive and metastatic bladder cancer: Summary of the 2025 guidelines', *European urology*, 87(5), pp. 582–600.

He, L. *et al.* (2022) 'Prognostic value of HPV E6 and APOBEC3B in upper urinary tract urothelial carcinoma', *Disease markers*, 2022, p. 2147494.

Hemelt, M. *et al.* (2009) 'The effect of smoking on the male excess of bladder cancer: a meta-analysis and geographical analyses', *International journal of cancer. Journal international du cancer*, 124(2), pp. 412–419.

He, R.-Q. *et al.* (2018) 'Prognostic signature of alternative splicing events in bladder urothelial carcinoma based on SpliceSeq data from 317 cases', *Cellular physiology and biochemistry: international journal of experimental cellular physiology, biochemistry, and*

pharmacology, 48(3), pp. 1355–1368.

He, X.Y. *et al.* (2001) 'Characterization and localization of human type10 17beta-hydroxysteroid dehydrogenase', *European journal of biochemistry*, 268(18), pp. 4899–4907.

Ho, K.-Y. *et al.* (2012) 'Cholesteatoma growth and proliferation: relevance with serpin B3', *The Laryngoscope*, 122(12), pp. 2818–2823.

Honda, Y. *et al.* (2019) 'Clinical staging of upper urinary tract urothelial carcinoma for T staging: Review and pictorial essay', *International journal of urology: official journal of the Japanese Urological Association*, 26(11), pp. 1024–1032.

Hou, C. *et al.* (2023) 'Application of microfluidic chips in the simulation of the urinary system microenvironment', *Materials Today. Bio*, 19(100553), p. 100553.

Huang, L. *et al.* (2025) 'ANXA2 in cancer: aberrant regulation of tumour cell apoptosis and its immune interactions', *Cell death discovery*, 11(1), p. 174.

Huang, X. *et al.* (2021) 'The inflammatory microenvironment and the urinary microbiome in the initiation and progression of bladder cancer', *Genes & diseases*, 8(6), pp. 781–797.

Hu, P. *et al.* (2000) 'Ablation of uroplakin III gene results in small urothelial plaques, urothelial leakage, and vesicoureteral reflux', *The journal of cell biology*, 151(5), pp. 961–972.

Hu, R. *et al.* (2024) 'Short-read RNA-Seq', *Methods in molecular biology (Clifton, N.J.)*, 2822, pp. 245–262.

Hustler, A. *et al.* (2018) 'Differential transcription factor expression by human epithelial cells of buccal and urothelial derivation', *Experimental cell research*, 369(2), pp. 284–294.

Hu, X.-H. *et al.* (2022) 'The predictors and surgical outcomes of different distant metastases patterns in upper tract urothelial carcinoma: A SEER-based study', *Frontiers in surgery*, 9, p. 1045831.

Huynh, T.N.A. *et al.* (2025) 'Accuracy and limitations of ureteroscopic biopsy in the staging and grading of upper tract urothelial carcinoma: A retrospective analysis at a large tertiary center', *Bladder*, 12(3), p. e21200048.

Hwang, M.W. *et al.* (2024) 'Upper tract urothelial cancer (UTUC) genomic profiling and correlation regarding benefit of platinum-based chemotherapy', *Exploration of targeted anti-tumor therapy*, 5(6), pp. 1261–1270.

Inamura, K. *et al.* (2007) 'HOXB2 as a novel prognostic indicator for stage I lung adenocarcinomas', *Journal of thoracic oncology: official publication of the International Association for the Study of Lung Cancer*, 2(9), pp. 802–807.

Ingersoll, M.A. and Albert, M.L. (2013) 'From infection to immunotherapy: host immune responses to bacteria at the bladder mucosa', *Mucosal immunology*, 6(6), pp. 1041–1053.

Ishioka, J. *et al.* (2015) 'Risk stratification for bladder recurrence of upper urinary tract urothelial carcinoma after radical nephroureterectomy', *BJU International*, 115(5), pp. 705–712.

Ivashkiv, L.B. (2018) 'IFNγ: signalling, epigenetics and roles in immunity, metabolism, disease and cancer immunotherapy', *Nature reviews. Immunology*, 18(9), pp. 545–558.

Jafari, N.V. and Rohn, J.L. (2022) ‘The urothelium: a multi-faceted barrier against a harsh environment’, *Mucosal immunology*, 15(6), pp. 1127–1142.

Jaimes-Parra, B.D. *et al.* (2016) ‘Ex vivo construction of a novel model of bioengineered bladder mucosa: A preliminary study’, *International journal of urology: official journal of the Japanese Urological Association*, 23(1), pp. 85–92.

Jeong, S.-H. *et al.* (2022) ‘Clinical determinants of recurrence in pTa bladder cancer following transurethral resection of bladder tumor’, *BMC cancer*, 22(1), p. 631.

Jimenez Jimenez, A.M. *et al.* (2025) ‘New insights into the role of metallothioneins in obesity and diabetes’, *International journal of obesity (2005)* , 49(10), pp. 1958–1972.

Jobczyk, M. *et al.* (2020) ‘Validation of EORTC, CUETO, and EAU risk stratification in prediction of recurrence, progression, and death of patients with initially non-muscle-invasive bladder cancer (NMIBC): A cohort analysis’, *Cancer medicine*, 9(11), pp. 4014–4025.

John, B.A. and Said, N. (2017) ‘Insights from animal models of bladder cancer: recent advances, challenges, and opportunities’, *Oncotarget*, 8(34), pp. 57766–57781.

Jubber, I. *et al.* (2023) ‘Epidemiology of bladder cancer in 2023: A systematic review of risk factors’, *European Urology*, 84(2), pp. 176–190.

Kamoun, A. *et al.* (2020) ‘A consensus molecular classification of muscle-invasive bladder cancer’, *European urology*, 77(4), pp. 420–433.

Kang, H.W., Kim, W.-J. and Yun, S.J. (2020) ‘The therapeutic and prognostic implications of molecular biomarkers in urothelial carcinoma’, *Translational cancer research*, 9(10), pp. 6609–6623.

Kara, N. *et al.* (2015) ‘Orc1 binding to mitotic chromosomes precedes spatial patterning during G1 phase and assembly of the origin recognition complex in human cells’, *The journal of biological chemistry*, 290(19), pp. 12355–12369.

Keleş, M. and Gunel-Ozcan, A. (2025) ‘HOX and MEINOX in cellular plasticity, fibrosis, and cancer’, *World journal of stem cells*, 17(9), p. 109102.

Kim, J. *et al.* (2009) ‘TC1(C8orf4) is a novel endothelial inflammatory regulator enhancing NF-kappaB activity’, *The journal of immunology*, 183(6), pp. 3996–4002.

Kim, S. *et al.* (2019) ‘Epigenetic regulation of mammalian Hedgehog signaling to the stroma determines the molecular subtype of bladder cancer’, *eLife*, 8, p. e43024.

Kolawa, A., D’Souza, A. and Tulpule, V. (2023) ‘Overview, diagnosis, and perioperative systemic therapy of Upper Tract urothelial carcinoma’, *Cancers*, 15(19), p. 4813.

Koll, F.J. *et al.* (2023) ‘Optimizing identification of consensus molecular subtypes in muscle-invasive bladder cancer: a comparison of two sequencing methods and gene sets using FFPE specimens’, *BMC cancer*, 23(1), p. 504.

Koll, F.J. *et al.* (2024) ‘Impact of consensus molecular subtypes on survival with and without adjuvant chemotherapy in muscle-invasive urothelial bladder cancer’, *Journal of Clinical Pathology*, 78(1), pp. 19–27.

Komura, K. *et al.* (2023) ‘The impact of FGFR3 alterations on the tumor microenvironment and the efficacy of immune checkpoint inhibitors in bladder cancer’, *Molecular cancer*, 22(1), p. 185.

Kondo, T. *et al.* (2012) ‘Diphosphorylated but not monophosphorylated myosin II regulatory light chain localizes to the midzone without its heavy chain during cytokinesis’, *Biochemical and biophysical research communications*, 417(2), pp. 686–691.

Kullmann, F.A. *et al.* (2017) ‘Urothelial proliferation and regeneration after spinal cord injury’, *American journal of physiology. Renal physiology*, 313(1), pp. F85–F102.

Kumar, V. *et al.* (2021) ‘The role of Notch, Hedgehog, and Wnt signaling pathways in the resistance of tumors to anticancer therapies’, *Frontiers in cell and developmental biology*, 9, p. 650772.

Kurniawan, F. *et al.* (2024) ‘Phosphorylation of Orc6 during mitosis regulates DNA replication and ribosome biogenesis’, *Molecular and cellular biology*, 44(7), pp. 289–301.

LaRosa, D.F., Rahman, A.H. and Turka, L.A. (2007) ‘The innate immune system in allograft rejection and tolerance’, *The journal of immunology*, 178(12), pp. 7503–7509.

Lata, S., Mishra, R. and Banerjea, A.C. (2018) ‘Proteasomal degradation machinery: Favorite target of HIV-1 proteins’, *Frontiers in microbiology*, 9, p. 2738.

Lee, J. *et al.* (2024) ‘Long-read RNA sequencing of archival tissues reveals novel genes and transcripts associated with clear cell renal cell carcinoma recurrence and immune evasion’, *Genome research*, 34(11), pp. 1849–1864.

Lefort, F. *et al.* (2023) ‘Clinical and biological differences between Upper Tract carcinoma and bladder urothelial cancer, including implications for clinical practice’, *Cancers*, 15(23), p. 5558.

Lescay, H.A. *et al.* (2025) ‘Anatomy, abdomen and pelvis ureter’, *StatPearls*. Treasure Island (FL): StatPearls Publishing.

Leslie, S.W., Soon-Sutton, T.L. and Aeddula, N.R. (2025) ‘Bladder cancer’, *StatPearls*. Treasure Island (FL): StatPearls Publishing.

Liang, F.-X. *et al.* (2005) ‘Cellular basis of urothelial squamous metaplasia: roles of lineage heterogeneity and cell replacement’, *The Journal of Cell Biology*, 171(5), pp. 835–844.

Liao, Y., Smyth, G.K. and Shi, W. (2014) ‘featureCounts: an efficient general purpose program for assigning sequence reads to genomic features’, *Bioinformatics (Oxford, England)*, 30(7), pp. 923–930.

Liao, Y., Smyth, G.K. and Shi, W. (2019) ‘The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads’, *Nucleic acids research*, 47(8), p. e47.

Liaw, A. *et al.* (2018) ‘Development of the human bladder and ureterovesical junction’, *Differentiation; research in biological diversity*, 103, pp. 66–73.

Libretti, S. and Aeddula, N.R. (2025) ‘Embryology, genitourinary’, *StatPearls*. Treasure Island (FL): StatPearls Publishing.

Li, H. (2018) ‘Minimap2: pairwise alignment for nucleotide sequences’, *Bioinformatics (Oxford, England)*, 34(18), pp. 3094–3100.

Li, K. *et al.* (2024) ‘Transcriptomic insights into UTUC: role of inflammatory fibrosis and potential for personalized treatment’, *Journal of translational medicine*, 22(1), p. 24.

Lin, D. *et al.* (2023) ‘Therapeutic strategies for asymptomatic upper urinary tract urothelial carcinoma’, *Wideochirurgia i inne techniki malo inwazyjne [Videosurgery and other miniinvasive techniques]*, 18(2), pp. 343–350.

Lindskrog, S.V. *et al.* (2021) ‘An integrated multi-omics analysis identifies prognostic molecular subtypes of non-muscle-invasive bladder cancer’, *Nature communications*, 12(1), p. 2301.

Li, R. *et al.* (2020) ‘Macroscopic somatic clonal expansion in morphologically normal human urothelium’, *Science (New York, N.Y.)*, 370(6512), pp. 82–89.

Liu, G. *et al.* (2019) ‘Genetic variant rs17185536 regulates SIM1 gene expression in human brain hypothalamus’, *Proceedings of the National Academy of Sciences of the United States of America*, 116(9), pp. 3347–3348.

Liu, J. *et al.* (2024) ‘Activation of Piezo1 or TRPV2 channels inhibits human ureteral contractions via NO release from the mucosa’, *Frontiers in pharmacology*, 15, p. 1410565.

Liu, J. *et al.* (2026) ‘TGF-β in tumor development and progression: mechanisms and therapeutics’, *Molecular Biomedicine*, 7(1), p. 9.

Liu, M. *et al.* (2018) ‘HOTAIR, a long noncoding RNA, is a marker of abnormal cell cycle regulation in lung cancer’, *Cancer science*, 109(9), pp. 2717–2733.

Liu, R.-H. *et al.* (2012) ‘Depletion of OLFM4 gene inhibits cell growth and increases sensitization to hydrogen peroxide and tumor necrosis factor-alpha induced-apoptosis in gastric cancer cells’, *Journal of biomedical science*, 19(1), p. 38.

Li, W.-M. *et al.* (2022) ‘High MT2A expression predicts worse prognosis in patients with urothelial carcinoma’, *Oncology*, 100(9), pp. 485–497.

Li, X., Liu, J., *et al.* (2023) ‘Analysis of the relationship between bladder cancer gene mutation and clinical prognosis by high-throughput sequencing’, *Laboratory medicine*, 54(2), pp. 142–152.

Li, X., Hu, J., *et al.* (2023) ‘Mechanotransduction in the urothelium: ATP signalling and mechanoreceptors’, *Heliyon*, 9(9), p. e19427.

Lopez-Castejon, G. and Brough, D. (2011) ‘Understanding the mechanism of IL-1β secretion’, *Cytokine & growth factor reviews*, 22(4), pp. 189–195.

Love, M.I., Huber, W. and Anders, S. (2014) ‘Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2’, *Genome biology*, 15(12), p. 550.

Luo, J. *et al.* (2017) ‘Matrilin-2 regulates proliferation, apoptosis and cell cycle during radiation-induced injury in HPAEpiC cell’, *Biochemical and biophysical research communications*, 485(3), pp. 577–583.

Luo, Q. *et al.* (2023) ‘Peroxisomal trans-2-enoyl-CoA inhibits proliferation, migration and invasion of hepatocellular carcinoma cells’, *Acta histochemica*, 125(2), p. 152002.

Mackenzie, F. and Ruhrberg, C. (2012) ‘Diverse roles for VEGF-A in the nervous system’, *Development (Cambridge, England)*, 139(8), pp. 1371–1380.

Magyar, J.P. *et al.* (1997) ‘Myelin and lymphocyte protein (MAL/MVP17/VIP17) and plasmolipin are members of an extended gene family’, *Gene*, 189(2), pp. 269–275.

Mahadevan, V. (2016) ‘Anatomy of the lower urinary tract’, *Surgery*, 34(7), pp. 318–325.

Mahmud, S.A., Manlove, L.S. and Farrar, M.A. (2013) ‘Interleukin-2 and STAT5 in regulatory T cell development and function’, *JAK-STAT*, 2(1), p. e23154.

Ma, J. *et al.* (2024) ‘Long-term recurrence rates of low-risk non-muscle-invasive bladder cancer-how long is cystoscopic surveillance necessary?’, *European Urology Focus*, 10(1), pp. 189–196.

Malone, E.R. *et al.* (2020) ‘Molecular profiling for precision cancer therapies’, *Genome medicine*, 12(1), p. 8.

Marasco, L.E. and Kornblihtt, A.R. (2023) ‘The physiology of alternative splicing’, *Nature reviews. Molecular cell biology*, 24(4), pp. 242–254.

Mar, J.C. (2019) ‘The rise of the distributions: why non-normality is important for understanding the transcriptome and beyond’, *Biophysical Reviews*, 11(1), pp. 89–94.

Marlar, S. *et al.* (2017) ‘Aquaporin-3 in cancer’, *International journal of molecular sciences*, 18(10), p. 2106.

Maroof, H., Paramore, L. and Ali, A. (2024) ‘Theories behind Bacillus Calmette-Guérin failure in high-risk non-muscle-invasive bladder cancer and update on current management’, *Cancer pathogenesis and therapy*, 2(2), pp. 74–80.

Marzouka, N.-A.-D. *et al.* (2018) ‘A validation and extended description of the Lund taxonomy for urothelial carcinoma using the TCGA cohort’, *Scientific reports*, 8(1), p. 3737.

Masson-Lecomte, A. *et al.* (2025) ‘European Association of Urology guidelines on upper urinary tract urothelial carcinoma: Summary of the 2025 update’, *European urology*, 87(6), pp. 697–716.

Meeks, J.J. *et al.* (2021) ‘Tumor subtyping: Making sense of heterogeneity with a goal toward treatment’, *Bladder cancer (Amsterdam, Netherlands)*, 7(1), pp. 1–11.

Meeks, J.J. *et al.* (2023) ‘Checkpoint inhibitors in urothelial carcinoma-future directions and biomarker selection’, *European urology*, 84(5), pp. 473–483.

Meirelles, P.M. *et al.* (2024) ‘Optimizing next-generation sequencing efficiency in clinical settings: analysis of read length impact on cost and performance’, *BMC genomics*, 25(1), p. 856.

Mer, A.S. *et al.* (2016) ‘Study design requirements for RNA sequencing-based breast cancer diagnostics’, *Scientific reports*, 6(1), p. 20200.

Meuser, M. *et al.* (2022) ‘FGFR2 signaling enhances the SHH-BMP4 signaling axis in early ureter development’, *Development (Cambridge, England)*, 149(1), p. dev200021.

Moinuddin, Z. and Dhanda, R. (2015) ‘Anatomy of the kidney and ureter’, *Anaesthesia & intensive care medicine*, 16(6), pp. 247–252.

Mo, Q. *et al.* (2018) ‘Prognostic power of a tumor differentiation gene signature for bladder urothelial carcinomas’, *Journal of the National Cancer Institute*, 110(5), pp. 448–459.

Morera, D.S. *et al.* (2020) ‘Clinical parameters outperform molecular subtypes for predicting outcome in bladder cancer: Results from multiple cohorts, including TCGA’, *The journal of urology*, 203(1), pp. 62–72.

Mori, K. *et al.* (2022) 'Discordance between clinical and pathological staging and grading in Upper Tract urothelial carcinoma', *Clinical Genitourinary Cancer*, 20(1), pp. 95.e1–95.e6.

Mortazavi, A. *et al.* (2008) 'Mapping and quantifying mammalian transcriptomes by RNA-Seq', *Nature methods*, 5(7), pp. 621–628.

Mullenders, J. *et al.* (2019) 'Mouse and human urothelial cancer organoids: A tool for bladder cancer research', *Proceedings of the National Academy of Sciences of the United States of America*, 116(10), pp. 4567–4574.

Nagalakshmi, U., Waern, K. and Snyder, M. (2010) 'RNA-Seq: a method for comprehensive transcriptome analysis', *Current Protocols in Molecular Biology*, Chapter 4(1), p. Unit 4.11.1–13.

Nagalakshmi, V.K. and Yu, J. (2015) 'The ureteric bud epithelium: morphogenesis and roles in metanephric kidney patterning', *Molecular Reproduction and Development*, 82(3), pp. 151–166.

Na, K., Oh, B.-C. and Jung, Y. (2023) 'Multifaceted role of CD14 in innate immunity and tissue homeostasis', *Cytokine & growth factor reviews*, 74, pp. 100–107.

Nally, E. *et al.* (2024) 'Upper Tract urothelial carcinoma (UTUC): Prevalence, impact and management challenge', *Cancer management and research*, 16, pp. 467–475.

Nassour, A.-J. *et al.* (2023) 'Relative risk of bladder and kidney cancer in Lynch syndrome: Systematic review and meta-analysis', *Cancers*, 15(2), p. 506.

National Cancer Institute (2023a) *Bladder Cancer Prognosis and Survival Rates*, *National Cancer Institute*. Available at: https://www.cancer.gov/types/bladder/survival (Accessed: 5 July 2025).

National Cancer Institute (2023b) *Cancer of the Urinary Bladder - Cancer Stat Facts*, *SEER*. Available at: https://seer.cancer.gov/statfacts/html/urinb.html (Accessed: 30 June 2025).

National Disease Registration Service (2024) *Cancer data interactive dashboards*, *NDRS*. Available at: https://digital.nhs.uk/ndrs/data/data-outputs/cancer-data-hub (Accessed: 17 June 2025).

Necchi, A. *et al.* (2021) 'Comprehensive genomic profiling of Upper-tract and bladder urothelial carcinoma', *European urology focus*, 7(6), pp. 1339–1346.

Nguyen, H.T. *et al.* (2000) 'Cyclic stretch activates p38 SAPK2-, ErbB2-, and AT1-dependent signaling in bladder smooth muscle cells', *American journal of physiology. Cell physiology*, 279(4), pp. C1155–67.

Nicolai, L., Pekayvaz, K. and Massberg, S. (2024) 'Platelets: Orchestrators of immunity in host defense and beyond', *Immunity*, 57(5), pp. 957–972.

Nogueira, A. *et al.* (2010) 'Influence of DNA repair RAD51 gene variants in overall survival of non-small cell lung cancer patients treated with first line chemotherapy', *Cancer chemotherapy and pharmacology*, 66(3), pp. 501–506.

Ogobuiro, I. and Tuma, F. (2025) 'Physiology, renal', *StatPearls*. Treasure Island (FL): StatPearls Publishing.

O'Meara, S. *et al.* (2024) 'Mechanical characteristics of the ureter and clinical implications', *Nature reviews. Urology*, 21(4), pp. 197–213.

Ortega Martell, J.A. (2020) ‘Immunology of urinary tract infections’, *GMS infectious diseases*, 8, p. Doc21.

Oyasu, R. (1995) ‘Epithelial tumours of the lower urinary tract in humans and rodents’, *Food and chemical toxicology: an international journal published for the British Industrial Biological Research Association*, 33(9), pp. 747–755.

Paduthol, G. *et al.* (2026) ‘A microphysiological human mini-bladder reveals urine-urothelium interplay in tissue resilience and UPEC recurrence in urinary tract infections’, *Nature Communications*, 17(1), p. 2322.

Paner, G.P. *et al.* (2018) ‘Updates in the eighth edition of the Tumor-Node-Metastasis staging classification for urologic cancers’, *European urology*, 73(4), pp. 560–569.

Pan, Q. *et al.* (2008) ‘Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing’, *Nature genetics*, 40(12), pp. 1413–1415.

Pan, S. *et al.* (2025) ‘Carcinogen metabolism and bladder cancer: role of gut microbiota in disease and prevention’, *Frontiers in Cellular and Infection Microbiology*, 15(1727550), p. 1727550.

Paraskevopoulou, V., Papafotiou, G. and Klinakis, A. (2016) ‘KRT14 marks bladder progenitors’, *Cell cycle (Georgetown, Tex.)*, 15(23), pp. 3161–3162.

Park, J.-S., Valerius, M.T. and McMahon, A.P. (2007) ‘Wnt/beta-catenin signaling regulates nephron induction during mouse kidney development’, *Development (Cambridge, England)*, 134(13), pp. 2533–2539.

Paysan-Lafosse, T. *et al.* (2025) ‘The Pfam protein families database: embracing AI/ML’, *Nucleic acids research*, 53(D1), pp. D523–D534.

Pei, X. *et al.* (2025) ‘What is the relationship between microorganisms in the human body and upper tract urothelial carcinoma?’, *Frontiers in immunology*, 16(1636782), p. 1636782.

Pellarin, I. *et al.* (2025) ‘Cyclin-dependent protein kinases and cell cycle regulation in biology and disease’, *Signal transduction and targeted therapy*, 10(1), p. 11.

Pertea, M. *et al.* (2015) ‘StringTie enables improved reconstruction of a transcriptome from RNA-seq reads’, *Nature biotechnology*, 33(3), pp. 290–295.

Piovesan, A. *et al.* (2019) ‘Human protein-coding genes and gene feature statistics in 2019’, *BMC research notes*, 12(1), p. 315.

Porras-Tobias, A.L., Caldera, A. and Castro-Piedras, I. (2025) ‘Intron Retention: A reemerging paradigm in RNA biology and post-transcriptional gene regulation’, *Genes*, 16(8), p. 986.

Powles, T. *et al.* (2020) ‘Avelumab maintenance therapy for advanced or metastatic urothelial carcinoma’, *The New England Journal of Medicine*, 383(13), pp. 1218–1230.

Powles, T. *et al.* (2024) ‘Enfortumab vedotin and pembrolizumab in untreated advanced urothelial cancer’, *The New England Journal of Medicine*, 390(10), pp. 875–888.

Quarles, J. *et al.* (2021) ‘Educational case: Bladder urothelial cell carcinoma TNM stage, prognosis and management’, *Academic pathology*, 8(23742895211022256), p. 23742895211022256.

Rasouly, H.M. and Lu, W. (2013) 'Lower urinary tract development and disease: LUT development and disease', *Wiley interdisciplinary reviews. Systems biology and medicine*, 5(3), pp. 307–342.

Rebouissou, S. *et al.* (2014) 'EGFR as a potential therapeutic target for a subset of muscle-invasive bladder cancers presenting a basal-like phenotype', *Science translational medicine*, 6(244), p. 244ra91.

Rehman, S. and Ahmed, D. (2024) 'Embryology, kidney, bladder, and ureter', *StatPearls*. Treasure Island (FL): StatPearls Publishing.

van Rhijn, B.W.G. *et al.* (2020) 'FGFR3 mutation status and FGFR3 expression in a large bladder cancer cohort treated by radical cystectomy: Implications for anti-FGFR3 treatment?', *European urology*, 78(5), pp. 682–687.

Richters, A., Aben, K.K.H. and Kiemeney, L.A.L.M. (2020) 'The global burden of urinary bladder cancer: an update', *World journal of urology*, 38(8), pp. 1895–1904.

Riedel, I. *et al.* (2005) 'Urothelial umbrella cells of human ureter are heterogeneous with respect to their uroplakin composition: different degrees of urothelial maturity in ureter and bladder?', *European Journal of Cell Biology*, 84(2-3), pp. 393–405.

Robertson, A.G. *et al.* (2017) 'Comprehensive molecular characterization of muscle-invasive bladder cancer', *Cell*, 171(3), pp. 540–556.e25.

Robinson, B.D. *et al.* (2019) 'Upper tract urothelial carcinoma has a luminal-papillary T-cell depleted contexture and activated FGFR3 signaling', *Nature communications*, 10(1), p. 2977.

Rodríguez-Rojas, K. *et al.* (2024) 'A novel role of IGFBP5 in the migration, invasion and spheroids formation induced by IGF-I and insulin in MCF-7 breast cancer cells', *Breast cancer research and treatment*, 208(1), pp. 79–88.

Rossi, M.R. *et al.* (2001) 'The immortalized UROtsa cell line as a potential cell culture model of human urothelium', *Environmental health perspectives*, 109(8), pp. 801–808.

Rouprêt, M. *et al.* (2015) 'European Association of Urology guidelines on upper urinary tract urothelial cell carcinoma: 2015 update', *European Urology*, 68(5), pp. 868–879.

Rouprêt, M. *et al.* (2018) 'European Association of Urology guidelines on Upper Urinary Tract Urothelial Carcinoma: 2017 update', *European urology*, 73(1), pp. 111–122.

Rouprêt, M. *et al.* (2023) 'European Association of Urology Guidelines on Upper Urinary Tract Urothelial Carcinoma: 2023 Update', *European Urology*, 84(1), pp. 49–64.

Rubenwolf, P. *et al.* (2015) 'Loss of AQP3 protein expression is associated with worse progression-free and cancer-specific survival in patients with muscle-invasive bladder cancer', *World Journal of Urology*, 33(12), pp. 1959–1964.

Saginala, K. *et al.* (2020) 'Epidemiology of bladder cancer', *Medical sciences (Basel, Switzerland)*, 8(1), p. 15.

Saint Luke's (2025) *Bladder Cancer*, *Saint Luke's*. Available at: https://www.saintlukeskc.org/specialties-services/cancer-care/bladder-cancer (Accessed: 28 August 2025).

Salmanidis, M. *et al.* (2013) 'Hoxb8 regulates expression of microRNAs to control cell death

and differentiation', *Cell death and differentiation*, 20(10), pp. 1370–1380.

Sanchez-Ferras, O. *et al.* (2021) 'A coordinated progression of progenitor cell states initiates urinary tract development', *Nature communications*, 12(1), p. 2627.

Sanderson, K.M. *et al.* (2007) 'Upper tract urothelial recurrence following radical cystectomy for transitional cell carcinoma of the bladder: an analysis of 1,069 patients with 10-year followup', *The journal of urology*, 177(6), pp. 2088–2094.

Sanguedolce, F. *et al.* (2025) 'Diagnostic, prognostic, and predictive tissue biomarkers in urothelial carcinoma in situ: A narrative review', *Diagnostics (Basel, Switzerland)*, 15(17), p. 2163.

Sano, T. *et al.* (2004) 'Immunohistochemical expression of 14-3-3 sigma protein in various histological subtypes of uterine cervical cancers', *Pathology international*, 54(10), pp. 743–750.

Satooka, H. and Hara-Chikuma, M. (2016) 'Aquaporin-3 controls breast cancer cell migration by regulating hydrogen peroxide transport and its downstream cell signaling', *Molecular and cellular biology*, 36(7), pp. 1206–1218.

Scimeca, M. *et al.* (2024) 'Molecular profiling of a bladder cancer with very high tumour mutational burden', *Cell death discovery*, 10(1), p. 202.

Sekulovski, S. *et al.* (2021) 'Assembly defects of human tRNA splicing endonuclease contribute to impaired pre-tRNA processing in pontocerebellar hypoplasia', *Nature communications*, 12(1), p. 5610.

Sfakianos, J.P. *et al.* (2021) 'Genetic differences between bladder and upper urinary tract carcinoma: Implications for therapy', *European urology oncology*, 4(2), pp. 170–179.

Shariat, S.F. *et al.* (2011) 'Gender differences in radical nephroureterectomy for upper tract urothelial carcinoma', *World journal of urology*, 29(4), pp. 481–486.

Sharon, D. *et al.* (2013) 'A single-molecule long-read survey of the human transcriptome', *Nature biotechnology*, 31(11), pp. 1009–1014.

Shen, W., Sipos, B. and Zhao, L. (2024) 'SeqKit2: A Swiss army knife for sequence and alignment processing', *iMeta*, 3(3), p. e191.

Shin, K. *et al.* (2011) 'Hedgehog/Wnt feedback supports regenerative proliferation of epithelial stem cells in bladder', *Nature*, 472(7341), pp. 110–114.

Shin, K. *et al.* (2014) 'Hedgehog signaling restrains bladder cancer progression by eliciting stromal production of urothelial differentiation factors', *Cancer cell*, 26(4), pp. 521–533.

Shi, Z. *et al.* (2017) 'Heterogeneous ribosomes preferentially translate distinct subpools of mRNAs genome-wide', *Molecular Cell*, 67(1), pp. 71–83.e7.

Siegel, F. *et al.* (2025) 'Proliferation and regeneration of the healthy human urothelium: A multi-scale simulation approach with 16 hypotheses of cell differentiation', *PloS one*, 20(6), p. e0325132.

Sjödahl, G. *et al.* (2012) 'A molecular taxonomy for urothelial carcinoma', *Clinical cancer research: an official journal of the American Association for Cancer Research*, 18(12), pp. 3377–3386.

Skeldon, S.C. *et al.* (2013) ‘Patients with Lynch syndrome mismatch repair gene mutations are at higher risk for not only upper tract urothelial cancer but also bladder cancer’, *European urology*, 63(2), pp. 379–385.

Ślusarczyk, A. *et al.* (2023) ‘Cancer-specific survival of patients with non-muscle-invasive bladder cancer: A population-based analysis’, *Annals of surgical oncology*, 30(12), pp. 7892–7902.

Southgate, J. *et al.* (1994) ‘Normal human urothelial cells in vitro: proliferation and induction of stratification’, *Laboratory investigation; a journal of technical methods and pathology*, 71(4), pp. 583–594.

Squires, P. *et al.* (2025) ‘Recurrence and its association with overall survival (OS) in muscle-invasive bladder cancer (MIBC) patients undergoing radical cystectomy (RC): A real-world analysis’, *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, 43(5_suppl), pp. 705–705.

Staack, A. *et al.* (2005) ‘Molecular, cellular and developmental biology of urothelium as a basis of bladder regeneration’, *Differentiation; research in biological diversity*, 73(4), pp. 121–133.

Stecca, C., Abdeljalil, O. and Sridhar, S.S. (2021) ‘Metastatic Urothelial Cancer: a rapidly changing treatment landscape’, *Therapeutic advances in medical oncology*, 13, p. 17588359211047352.

Steens, J. and Klein, D. (2022) ‘HOX genes in stem cells: Maintaining cellular identity and regulation of differentiation’, *Frontiers in cell and developmental biology*, 10, p. 1002909.

Strandgaard, T. *et al.* (2024) ‘Field cancerization is associated with tumor development, T-cell exhaustion, and clinical outcomes in bladder cancer’, *European urology*, 85(1), pp. 82–92.

Streuli, C.H. *et al.* (1993) ‘Extracellular matrix regulates expression of the TGF-beta 1 gene’, *The journal of cell biology*, 120(1), pp. 253–260.

Sugawara, E., Koyama, K. and Inamura, K. (2022) ‘Molecular characterization of Upper Tract urothelial carcinoma for precision therapeutics and non-invasive diagnostics’, *Current genomics*, 23(1), pp. 2–4.

Sung, H.-C. *et al.* (2022) ‘Metallothionein 2A with antioxidant and antitumor activity is upregulated by caffeic acid phenethyl ester in human bladder carcinoma cells’, *Antioxidants (Basel, Switzerland)*, 11(8), p. 1509.

Sun, L. *et al.* (2023) ‘T cells in health and disease’, *Signal transduction and targeted therapy*, 8(1), p. 235.

Sun, Y. *et al.* (2025) ‘Dynamic evolution of the tumor immune microenvironment in malignant tumors and emerging therapeutic paradigms’, *MedComm*, 6(12), p. e70496.

Szarvas, T. *et al.* (2016) ‘Why are upper tract urothelial carcinoma two different diseases?’, *Translational andrology and urology*, 5(5), pp. 636–647.

Tacha, D., Zhou, D. and Cheng, L. (2011) ‘Expression of PAX8 in normal and neoplastic tissues: a comprehensive immunohistochemical study’, *Applied immunohistochemistry & molecular morphology*, 19(4), pp. 293–299.

Tan, G. *et al.* (2019) ‘Long fragments achieve lower base quality in Illumina paired-end

sequencing', *Scientific reports*, 9(1), p. 2856.

Teoh, J.Y.-C. *et al.* (2022) 'Recurrence mechanisms of non-muscle-invasive bladder cancer - a clinical perspective', *Nature Reviews. Urology*, 19(5), pp. 280–294.

Thouvenin, J. *et al.* (2021) 'Efficacy of immune checkpoint inhibitors in Upper Tract urothelial carcinomas: Current knowledge and future directions', *Cancers*, 13(17), p. 4341.

Tolino, M.A., Block, E.R. and Klarlund, J.K. (2011) 'Brief treatment with heparin-binding EGF-like growth factor, but not with EGF, is sufficient to accelerate epithelial wound healing', *Biochimica et biophysica acta*, 1810(9), pp. 875–878.

Tomiyama, E. *et al.* (2022) 'Comparison of molecular profiles of upper tract urothelial carcinoma vs. urinary bladder cancer in the era of targeted therapy: a narrative review', *Translational andrology and urology*, 11(12), pp. 1747–1761.

Tsang, H.T.H. *et al.* (2006) 'A systematic analysis of human CHMP protein interactions: additional MIT domain-containing proteins bind to multiple components of the human ESCRT III complex', *Genomics*, 88(3), pp. 333–346.

Tse, R.T.-H. *et al.* (2022) 'Current status of organoid culture in urological malignancy', *International Journal of Urology: Official Journal of the Japanese Urological Association*, 29(2), pp. 102–113.

Ueda, H. *et al.* (2023) 'Characterization of cytoskeletal and structural effects of INF2 variants causing glomerulopathy and neuropathy', *Scientific reports*, 13(1), p. 12003.

UK National Screening Committee (2020) *Bladder Cancer*, *Gov.uk*. Available at: https://view-health-screening-recommendations.service.gov.uk/bladder-cancer/ (Accessed: 19 May 2025).

Ullah, I. *et al.* (2019) 'Significance tests for analyzing gene expression data with small sample sizes', *Bioinformatics (Oxford, England)*, 35(20), pp. 3996–4003.

Urbanski, L.M., Leclair, N. and Anczuków, O. (2018) 'Alternative-splicing defects in cancer: Splicing regulators and their downstream targets, guiding the way to novel cancer therapeutics', *Wiley interdisciplinary reviews. RNA*, 9(4), p. e1476.

Varley, C.L. and Southgate, J. (2008) 'Effects of PPAR agonists on proliferation and differentiation in human urothelium', *Experimental and toxicologic pathology: official journal of the Gesellschaft für Toxikologische Pathologie*, 60(6), pp. 435–441.

Verwilt, J., Mestdagh, P. and Vandesompele, J. (2023) 'Artifacts and biases of the reverse transcription reaction in RNA sequencing', *RNA (New York, N.Y.)*, 29(7), pp. 889–897.

Viehweger, F. *et al.* (2022) 'Desmoglein 3 (Dsg3) expression in cancer: A tissue microarray study on 15,869 tumors', *Pathology, research and practice*, 240(154200), p. 154200.

Vilhais, G. and Fontes-Sousa, M. (2025) 'Adjuvant treatment preferences in high-risk upper tract urothelial carcinoma: the perspective of Portuguese medical oncologists', *The oncologist*, 30(11), p. oyaf365.

Vitek, R.A. *et al.* (2022) 'Fresh tissue procurement and preparation for multicompartment and multimodal analysis of the prostate tumor microenvironment', *The prostate*, 82(7), pp. 836–849.

Vitting-Seerup, K. and Sandelin, A. (2019) 'IsoformSwitchAnalyzeR: analysis of changes in

genome-wide patterns of alternative splicing and its functional consequences', *Bioinformatics (Oxford, England)*, 35(21), pp. 4469–4471.

Wang, D. *et al.* (2024) 'A real-world multi-center RNA-seq benchmarking study using the Quartet and MAQC reference materials', *Nature communications*, 15(1), p. 6167.

Wang, F. *et al.* (2022) 'Comparative analysis of differentially mutated genes in non-muscle and muscle-invasive bladder cancer in the Chinese population by whole exome sequencing', *Frontiers in genetics*, 13, p. 831146.

Wang, F., Soprano, K.J. and Soprano, D.R. (2015) 'Role of Acinus in regulating retinoic acid-responsive gene pre-mRNA splicing: Splicing of rar-regulated transcipts and acinus', *Journal of cellular physiology*, 230(4), pp. 791–801.

Wang, J. *et al.* (2018) 'Polyploid superficial cells that maintain the urothelial barrier are produced via incomplete cytokinesis and endoreplication', *Cell reports*, 25(2), pp. 464–477.e4.

Wang, L. *et al.* (2013) 'CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model', *Nucleic acids research*, 41(6), p. e74.

Wang, W., Martindale, J. and Metzger, J.M. (2009) 'Parvalbumin: Targeting calcium handling in cardiac diastolic dysfunction', *General physiology and biophysics*, 28 Spec No Focus, pp. F3–6.

Wang, X. *et al.* (2020) 'Systematic investigation of biomarker-like role of ARHGDIB in breast cancer', *Cancer biomarkers: section A of Disease markers*, 28(1), pp. 101–110.

Wang, Z., Gerstein, M. and Snyder, M. (2009) 'RNA-Seq: a revolutionary tool for transcriptomics', *Nature reviews. Genetics*, 10(1), pp. 57–63.

Wan, Q. *et al.* (2018) 'Urothelium with barrier function differentiated from human urine-derived stem cells for potential use in urinary tract reconstruction', *Stem cell research & therapy*, 9(1), p. 304.

Warrick, J.I. *et al.* (2024) 'International Society of urological pathology consensus conference on current issues in Bladder Cancer. Working group 4: Molecular subtypes of Bladder Cancer-principles of classification and emerging clinical utility: Molecular subtypes of Bladder Cancer-principles of classification and emerging clinical utility', *The American journal of surgical pathology*, 48(1), pp. e32–e42.

Weiss, R.M. *et al.* (2013) 'Brg1 determines urothelial cell fate during ureter development', *Journal of the American Society of Nephrology: JASN*, 24(4), pp. 618–626.

Wei, Y. *et al.* (2022) 'Urinary tract tumor organoids reveal eminent differences in drug sensitivities when compared to 2-dimensional culture systems', *International journal of molecular sciences*, 23(11), p. 6305.

Wezel, F., Pearson, J. and Southgate, J. (2014) 'Plasticity of in vitro-generated urothelial cells for functional tissue formation', *Tissue engineering. Part A*, 20(9-10), pp. 1358–1368.

Whyard, T. *et al.* (2020) 'Organoid model of urothelial cancer: establishment and applications for bladder cancer research', *BioTechniques*, 69(3), pp. 193–199.

Wickham, H. (2016) *Ggplot2: Elegant graphics for data analysis* [PDF]. 2nd edn. Basel, Switzerland: Springer International Publishing (Use R!). Available at: https://ggplot2.tidyverse.org (Accessed: 17 June 2025).

Wickham, H. *et al.* (2025) ‘dplyr: A Grammar of Data Manipulation’. Available at: https://dplyr.tidyverse.org.

Wickham, H., Vaughan, D. and Girlich, M. (2025) ‘tidyr: Tidy Messy Data’. Available at: https://tidyr.tidyverse.org.

Wiessner, G.B. *et al.* (2022) ‘Development, regeneration and tumorigenesis of the urothelium’, *Development (Cambridge, England)*, 149(9), p. dev198184.

Winder, M. *et al.* (2014) ‘Signalling molecules in the urothelium’, *BioMed research international*, 2014(1), p. 297295.

de Winter, J.C.F. (2013) ‘Using the Student’s t-test with extremely small sample sizes’, *Practical Assessment, Research, and Evaluation*. University of Massachusetts Amherst. Available at: https://doi.org/10.7275/E4R6-DJ05.

Witjes, J.A. *et al.* (2021) ‘European Association of Urology guidelines on muscle-invasive and Metastatic Bladder Cancer: Summary of the 2020 guidelines’, *European Urology*, 79(1), pp. 82–104.

Wu, Z. *et al.* (2024) ‘Discovery and generalization of tissue structures from spatial omics data’, *Cell reports methods*, 4(8), p. 100838.

Xie, S. *et al.* (2024) ‘Occupational exposure to organic solvents and risk of bladder cancer’, *Journal of exposure science & environmental epidemiology*, 34(3), pp. 546–553.

Xiong, S. *et al.* (2025) ‘Single-cell and spatial transcriptomic analysis reveal cellular heterogeneity and cancer cell-intrinsic major histocompatibility complex II expression in urothelial carcinoma’, *International Journal of Biological Sciences*, 21(15), pp. 6649–6673.

Yamashita, R. *et al.* (2024) ‘Cumulative incidence and risk factors for recurrence of upper tract urothelial carcinoma in patients undergoing radical cystectomy’, *BJUI compass*, 5(5), pp. 483–489.

Yang, H.-Y. *et al.* (2012) ‘Expression and prognostic value of ING3 in human primary hepatocellular carcinoma’, *Experimental biology and medicine (Maywood, N.J.)*, 237(4), pp. 352–361.

Yu, S. *et al.* (2023) ‘CSRP1 promotes colon adenocarcinoma growth and serves as an independent risk biomarker for worse prognosis’, *Genetics research*, 2023, p. 8586507.

Yu, S.H. *et al.* (2024) ‘FGFR3 mutations in urothelial carcinoma: A single-center study using next-generation sequencing’, *Journal of Clinical Medicine*, 13(5), p. 1305.

Zeber-Lubecka, N. *et al.* (2022) ‘Gene expression-based functional differences between the bladder body and trigonal urothelium in adolescent female patients with micturition dysfunction’, *Biomedicines*, 10(6), p. 1435.

Zhang, Q. *et al.* (2024) ‘Investigating cellular similarities and differences between upper tract urothelial carcinoma and bladder urothelial carcinoma using single-cell sequencing’, *Frontiers in Immunology*, 15, p. 1298087.

Zhang, Y. and Atala, A. (2013) ‘Urothelial cell culture’, *Methods in molecular biology (Clifton, N.J.)*, 1037, pp. 27–43.

Zhang, Y. and Weinberg, R.A. (2018) ‘Epithelial-to-mesenchymal transition in cancer: complexity and opportunities’, *Frontiers of medicine*, 12(4), pp. 361–373.

Zhang, Z.-T. *et al.* (1999) ‘Urothelium-specific expression of an oncogene in transgenic mice induced the formation of carcinoma in situ and invasive transitional cell carcinoma’, *Cancer research*, 59(14), pp. 3512–3517.

Zhao, X., Wang, Y. and Liang, C. (2022) ‘Cigarette smoking and risk of bladder cancer: a dose-response meta-analysis’, *International urology and nephrology*, 54(6), pp. 1169–1185.

Zhao, Y. *et al.* (2025) ‘PTPRK promotes resiniferatoxin-induced postherpetic neuralgia via activating DUSP1/p38 MAPK signaling pathway in dorsal root ganglia’, *Scientific reports*, 15(1), p. 40630.

Zhou, X. *et al.* (2017) ‘PROTOCADHERIN 7 acts through SET and PP2A to potentiate MAPK signaling by EGFR and KRAS during lung tumorigenesis’, *Cancer research*, 77(1), pp. 187–197.

Zhu, L. *et al.* (2025) ‘Quantitative biomechanical analysis of ureteral obstruction and peristalsis’, *Li xue xue bao [Acta mechanica Sinica]*, 41(4), pp. 1–11.

Zlotta, A.R., Fleshner, N.E. and Jewett, M.A. (2009) ‘The management of BCG failure in non-muscle-invasive bladder cancer: an update’, *Journal de l’Association des urologues du Canada [Canadian Urological Association journal]*, 3(6 Suppl 4), pp. S199–205.

Zupančič, D. *et al.* (2020) ‘Vitamin A rich diet diminishes early urothelial carcinogenesis by altering retinoic acid signaling’, *Cancers*, 12(7), p. 1712.