The role of Tropomyosins and their Isoforms in Urothelial Carcinoma of Bladder ('TropICana-B')

By

Ms. NADA HUMAYUN-ZAKARIA

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

Supervisors:
Prof Richard T Bryan
Dr Roland Arnold
Dr Douglas G Ward

Bladder Cancer Research Centre
Institute of Cancer and Genomic Sciences
College of Medical and Dental Sciences
University of Birmingham 2024

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

ABSTRACT

High throughput sequencing has informed molecular pathology for classification of urothelial bladder cancer (BC). Noting the prominence of 'biomarkers', they may be a key tool to avoid impaired quality of life (QoL) for patients undergoing investigation and surveillance for BC. This is predicated on the practicality of implementing such tests as a point-of-care (POC) urine-based test. It can potentially aid in early detection and diagnosis and facilitate screening pathways of patients prior to the development of T1 disease. Altered transcriptional events generate tropomyosin (*TPM*) isoforms that appear to have selective upregulation in the presence of BC. They are stable coiled-coil dimers, resistant to degradation and can be found in urine. We created a multistep verification model to assess the biomarker potential of *TPM* isoforms for detection of BC, incorporating sequencing techniques and wet lab experiments.

IsoTPM, a bespoke bioinformatics pipeline was created to identify upregulated *TPM* isoforms that contributed to overall gene expression by exploring short-read RNA sequences from 6 datasets: Cancer Cell Line Encyclopaedia (CCLE n=26, BC cell lines), West Midlands Bladder Cancer Prognosis Program (BCPP n=85, non-muscle-invasive), The Cancer Genome Atlas (TCGA n=414, muscle-invasive), a paracancerous dataset (GEO Accession GSE133624 n=29), University of York (n=8, bladder stroma and urothelium) and Genotype Tissue Expression (GTEx n=7, benign). Coupled with targeted long-read sequencing, we evaluated the expression of these target *TPM* isoforms in BCPP archived tissues and BC cell lines (n=96). Experimental validation of in silico analyses was performed on the top candidates (Tpm1.4, Tpm1.7, Tpm4.2 and Tpm 4.1). Isoform specific primer-pairs with probe-based multiplex qPCR assay was created to measure 'relative' isoform expression. Preliminary assay testing

was performed on urine supernatants from cancer patients and those with benign urological conditions to define its potential as a urine-based test.

We found 2 pairs of isoforms with strong disease diagnostic value (p-value: 0.031, 9.18e-06). Furthermore, our qPCR assay showed a 1.41-fold increase in relative isoform expression in cancer vs benign conditions (p-value: 0.00312, n=28). Our assay can effectively predict the presence of BC when compared to other benign bladder conditions (AUC= 0.82).

We have pioneered the use of relative isoform usage as a diagnostic urine-based-assay for detection of BC. Currently available detection kits utilize mutation/ methylation-based techniques incorporating molecular adaptations at the gene-level but cannot evaluate isoform expression which constitutes an additional layer of infor mation indicative of gene expression to understand the mechanisms of carcinogenesis.

Translational Relevance:

With the use of isoform evaluation and differences found in expression levels, a urine-based-diagnostic test can be utilized for creation of a point-of-care test to triage patients for further investigations, if required, via referral to the haematuria pathway, decreasing the unnecessary need for undergoing flexible cystoscopy as a first-line investigation for identifying bladder cancer.

DEDICATION

For my parents,

Who have always smiled in the face of adversity with unrelenting resilience and have taught me life lessons innumerable.

For Meesha and Yousuf,

Who have been my rock and support; I am endlessly grateful for your patience and understanding throughout this journey.

ACKNOWLEDGEMENT

I acknowledge Queen Elizabeth Hospital Birmingham charities for their financial support in funding consumables for the course of this PhD project, without whom this research would not have been made possible. I would also like to express my sincere gratitude to University of York, Wrexham Academic Unit and Human Biomaterials Resource Centre for providing me with patient samples and data. I am deeply indebted to the participants of the Bladder Cancer Prognosis Program for their willingness to share their personal health information that has enabled this research to be conducted.

I also acknowledge the invaluable assistance and contribution by the Bladder Cancer Research Centre members: Naheema Gordon, Anshita Goel, and students: Minghao Yu, and Faisal bin Humaid for library construction, data analysis and support.

I acknowledge the mentorship and support provided by DMA Wallace who provided me with motivation to pursue excellence in my endeavours.

I would also like to appreciate my family and friends for their unrelentless support, patience, and encouragement throughout my PhD journey. Their constant love and support allowed me to persevere during the challenges encountered through the research process.

Lastly, I extend my heartfelt gratitude to Professor Rik Bryan, Dr Douglas Ward and Dr Roland Arnold, my research supervisors, for their unwavering support, invaluable guidance, and insightful feedback every stage of the project. Their expertise and advice have been instrumental in shaping the research and has facilitated it to completion.

LIST OF CONTENTS

I Abstract

II Dedication

III Acknowledgement

IV List of Contents

V Abbreviations

VI List of Tables

VII List of Figures

1. Introduction

1.1 Synopsis 1

1.2 Bladder Cancer 2

1.2.1 Epidemiology 2

1.2.2 Aetiology and risk factors 3

1.2.2.1 Urothelial carcinogens 3

1.2.2.2 Smoking 6

1.2.2.3 Occupational factors 8

1.2.2.4 Genetic predisposition 9

1.2.2.5 *APOBEC*-mediated mutagenesis 10

1.2.3 Classification 10

1.2.4 Clinical presentation 15

1.2.4.1 Imaging 17

1.2.4.2 Flexible cystoscopy 19

1.2.4.3 Urine cytology 22

1.2.4.4 Urinary biomarkers 22

1.2.4.5 Other non-urine based biomarkers 31

1.2.5 Management 32

1.2.6 Risk stratification, recurrence, and progression 36

1.3 Molecular Landscape of Bladder Cancer 40

1.3.1 Muscle-invasive bladder cancer 41

1.3.2 Non-muscle invasive bladder cancer 46

1.4 Transcription, the splicing process, & the mRNA 49

1.4.1 Alternative and aberrant splicing 50

1.5 Tropomyosins 53

1.5.1 Background and diversity 53

1.5.2 Structure and functional impact on other actin-binding proteins 54

1.5.3 Splice-form variants and Isoforms 59

1.6 Tropomyosin profiling in bladder cancer 64

1.6.1 Tropomyosin expression and signalling pathways 64

1.6.2 'Hallmarks of cancer' applied to urothelial carcinogenesis 65

1.6.3 Isoform switching due to alternative promoter usage 69

1.7 Isoform and gene quantification 70

1.7.1 Next generation sequencing 70

1.7.2 Gene vs Isoform level quantification of RNA expression 70

1.8 Hypothesis 75

1.9 Aims 75

2. Materials And Methods

2.1 Bioinformatics 76

2.1.1 Acquisition of RNA-seq datasets 76

2.1.1.1 BCPP 76

2.1.1.2 CCLE 76

2.1.1.3 TCGA 77

2.1.1.4 Non-cancer GEO GSE133624 77

2.1.1.5 Non-cancer from University of York 78

2.1.1.6 Non-cancer from GTEx 78

2.1.2 Computational software 78

2.1.2.1 BlueBEAR 78

2.1.2.2 Python, Biopython 79

2.1.2.3 R, Rstudio 79

2.1.2.4 Salmon 80

2.1.3 Annotation 80

2.2 Lab materials 82

2.2.1 Bladder cancer cell lines 82

2.2.2 Bladder cancer tissues 83

2.2.2.1 Archived BCPP samples 83

2.2.2.2 Fresh samples from Wrexham 84

2.2.3 Archived HBRC urine samples 84

2.2.4 Antibodies 85

2.2.4.1 Primary antibodies 85

2.2.4.2 Secondary antibodies 86

2.2.4.3 Recombinant protein 86

2.3 Use of custom python scripts to identify unique primers and peptide sequences for experimental validation 87

2.4 RNA extraction and sequencing 88
2.4.1 RNA extraction and cDNA synthesis from cell lines and tissues 88
2.4.2 PCR and agarose gel electrophoresis 88
2.4.3 Purification of DNA from PCR products and Sanger sequencing 89
2.4.4 Quantitative Polymerase Chain Reaction and Taqman probe assay 90
2.4.5 RNA extraction and quantification from urine 91
2.4.6 Bioanalyzer 91
2.4.7 Long-read RNA sequencing library preparation: PCR 91
2.5 Protein expression quantification protocols 95
2.5.1 Mass spectrometry 95
2.5.2 Western blotting 95

3. Characterization Of Tropomyosin Isoform Expression *In Silico*

3.1 Background 97
3.2 Aims and objectives 99
3.3 Development of IsoTPM 101
3.3.1 Analysis of datasets for shortlisting of isoforms 104
3.4 Isoform curation from short-read RNA seq 105
3.4.1 Tropomyosin isoform expression in bladder cancer cell lines 105
3.4.2 Tropomyosin isoform expression in urothelial carcinoma of bladder tissues 117
3.4.2.1 BCPP 117
3.4.2.2 TCGA 131
3.4.2.3 Other datasets 136

3.5 Summary 146

4. Biomarker potential of Tropomyosin isoforms

4.1 Aims and objectives 148

4.2 Relative isoform expression levels as an indicator of disease 149

4.3 Correlations with molecular subtypes 154

4.3.1 TPM gene expression in UROMOL 2021 NMIBC subtypes 154

4.4 Isoform inclusion levels (PSI) and relative isoform usage are not prognostic for disease progression 159

4.5 Immune infiltration and TPM isoform expression 166

4.6 Summary 169

5. Characterization of Tropomyosin isoform expression using Long-Read RNA Sequencing

5.1 Background 171

5.2 Aims and objectives 173

5.3 Long-read sequencing sample selection 174

5.3.1 PacBio SMRTBell sequencing and the Iso-Seq pipeline 174

5.3.2 SMRTBell library preparation 175

5.4 Long-read RNA-seq data analysis 182

5.5 Summary 196

6. High throughput assay development for biomarker use

6.1 Aims and objectives 198

6.2 PCR 199

6.3 Optimization of qPCR assay 204

6.4 qPCR assay testing in BCPP tissue samples 208

6.5 qPCR testing in urine supernatants 214

6.6 Summary 216

7. Tropomyosin protein isoforms as biomarkers for bladder cancer

7.1 Background 218

7.2 Aims and objectives 218

7.3 Mass spectrometry 220

7.4 Western blotting 225

7.5 Summary 227

8. Discussion

8.1 Bioinformatics pipeline development & comparison with others 229

8.2 Tropomyosin isoform expression in urothelial carcinoma of bladder 231

8.2.1 Overview 231

8.2.2 Co-expression of *TPM4* isoforms in MIBC 232

8.2.3 Immune correlation with *TPM* isoform expression 233

8.3 Clinical significance and relevance 235

8.3.1 qPCR assay as biomarker tool for detection of urothelial carcinoma of bladder 235

8.3.2 Detection in urine 236

8.4 Our findings and the *IsoTPM* pipeline are of relevance to other cancers 237

8.4.1 Prostate cancer 238

8.4.2 Malignant Breast Epithelial cells 238

8.4.3 Renal cell cancer 239

8.4.4 Oesophageal cancer 239

8.5 Limitations 240

8.6 Further experiments 242

8.6.1 Antibody development 242

8.6.2 Validation studies for qPCR biomarker assay 243

8.6.3 Whole gene transcriptomic profiling and co-expression analyses with *TPM* 243

9. Conclusion

10. References

11. Appendix

- Annotation database with TSL tags
- Bioinformatics coding scripts
- qPCR data
- MS data

ABBREVIATIONS

Ab	Antibody
ABP	Actin-binding protein
ADF	Actin depolymerization factor
AS	Alternative splicing
BC	Bladder Cancer
BCG	Bacillus Calmette-Guérin
BCPP	Bladder Cancer Prognosis Program
CCLE	Cancer Cell Line Encyclopaedia
CIS	*Carcinoma in situ*
CT	Computed tomography
EAU	European Association of Urology
EORTC	European Organization for Research and Treatment of Cancer
HG	High grade
HMW	High molecular weight
Indel	Insertion deletion
LG	Low grade
LMW	Low molecular weight
LR-seq	Long read RNA sequencing
MDT	Most dominant transcript
MIBC	Muscle-invasive bladder cancer
mRNA	messenger Ribonucleic acid
MS	Mass spectrometry
NBI	Narrow band imaging

NGS	Next generation sequencing
NICE	National Institute of Health and Care Excellence
NMIBC	Non-muscle-invasive bladder cancer
NPA	Naphthylamine
NVH	Non-visible haematuria
PAH	Polyaromatic hydrocarbons
PDD	Photodynamic diagnosis
QC	Quality control
RNA	Ribonucleic acid
RNA seq	RNA sequence
SD	Standard deviation
SR-seq	Short read RNA sequencing
TCGA	The Cancer Genome Atlas
TNM	Tumour Nodes Metastases
qTPM	quantified Transcript Per Million
TPM or Tpm	Tropomyosin
TURBT	Transurethral resection of bladder tumour
US FDA	United States Food and Drug Administration
VH	Visible haematuria
WHO	World Health Organization
WLC	White-light cystoscopy

LIST OF TABLES

1.1 Pooled performance indicators for 4 diagnostic (DNA- & RNA- based) biomarker assays used to assess recurrence of NMIBC during follow up.

1.2 Summary of active clinical trials with the biomarker assay being tested.

1.3 EAU NMIBC Prognostic factor risk groups

1.4 *TPM* genes and their isoforms with corresponding Ensembl and RefSeq IDs.

1.5 Differences between two principal options of read-alignment techniques.

2.1 Bladder cancer cell lines (included in CCLE database) and their characteristics.

2.2 List of BCPP tissue samples used in the experiments performed in this project.

2.3 List of tropomyosin targeted antibodies

2.4 *TPM* Isoforms and Primer sequences used for experimental analysis (RT-PCR).

2.3 *TPM* isoform Primer and Probe sequences for Taqman qPCR.

2.6 Primer names and their sequences (5' to 3') used for PacBio sequencing.

2.7 Thermocycler program settings for preparing SMRTBell library.

3.1 Comparison of quantification methods available for RNA-seq quantification.

3.2 List of isoforms with high-expression in SR-seq RNA of BC cell lines from CCLE.

3.3 Comparison of mean isoform expression values for *TPM1* in the BCPP cohort.

3.4 Comparison of mean isoform expression values for *TPM2* in the BCPP cohort.

3.5 Transcript isoforms of *TPM3* with mean expression levels in the BCPP cohort.

3.6 Transcript isoforms of *TPM4* with mean expression levels in the BCPP cohort.

4.1 *TPM* whole gene expression in individual Uromol 2021 NMIBC classes.

4.2 Isoform with highest expression in the Uromol class listed, with associated p-value.

5.1 Comparison of Third generation sequencing platforms.

6.1 RT-PCR products that were sequenced were ‘BLAST’ with NCBI BLAST-n.

7.1 List of peptide sequences from bladder cancer cell lines.

7.2 List of peptide sequences from BCPP tissues.

LIST OF FIGURES

1.1 Metabolism of Aromatic amines and polycyclic hydrocarbons.

1.2 Smoking status significantly affects the probability of NMIBC recurrence.

1.3 TNM Staging of Bladder Cancer.

1.4 WHO 1973 2004/2016/2022 Classification of histopathological grading of urothelial neoplasm.

1.5 Management pathway of patients presenting with haematuria in the UK.

1.6 Schematic representation of VIRADS scoring model.

1.7 Comparison of WLC (a) with PDD (b), digital WLC (c) and NBI (d) to visualize urothelial tumours.

1.8 Evolution of urinary biomarker tests including US FDA approved, with sensitivity and specificity and assay targets.

1.9 Schema depicting Alternative splicing events.

1.10 Consensus classification of MIBC molecular subtypes.

1.11 UROMOL 2021 NMIBC molecular classification.

1.12 Tropomyosin-actin composition in non-muscle cells.

1.13 Tropomyosin-Actin regulation and dynamics with other ABPs.

1.14 Schematic representation of Human *TPM* genes with their isoforms and exons.

1.15 Hallmarks of cancer in context of *TPM*s and other genes.

3.1 Schema- Automated segment of *IsoTPM* pipeline.

3.2 (a-d) Differential *TPM1-4* isoform expression in BC cell lines from the CCLE dataset.

3.3 (a-d) Abundance levels of *TPM1-4* isoforms in bladder cancer cell lines.

3.4a, 3.5a, 3.6a & 3.7a Differential expression of isoforms of *TPM1-4* in BCPP cohort.

3.4b, 3.5b, 3.6b & 3.7b Isoform expression levels for *TPM1-4* in BCPP tissue samples.

3.8 (a-d) TCGA data with differential isoform expression of *TPM1-4*.

3.9 (a-d) (i) *TPM1-4* differential expression of isoforms
(ii) Pairwise comparisons of highly expressed isoforms using Wilcoxon rank sum test with continuity correction (p-adjusted Bonferroni).

4.1 Relative isoform expression (ratio) of Tpm1.7 to Tpm1.4 in BCPP(a) and TCGA(b).

4.2 Relative isoform expression (ratio) of Tpm4.2 to Tpm4.1 in BCPP(a) and TCGA(b).

4.3 (a)*TPM* whole gene expression in BCPP tissues according to UROMOL subtypes
(b) Isoform expression observed in UROMOL subtypes.

4.4a Kaplan-Meier curves illustrating recurrence-free survival (RFS), progression-free survival (PFS) and overall survival (OS) in Tpm1.7 (i), Tpm1.4 (ii) and ratio of isoform expression of *TPM1* (iii)

4.4b Kaplan-Meier curves illustrating recurrence-free survival (RFS), progression-free survival (RFS) and overall survival (OS) in Tpm4.2 (i), Tpm4.1 (ii) and ratio of isoform expression of *TPM4* (iii)

4.5 (a) Association of *TPM* gene expression with immune composition of BCPP tissues
(b) *TPM* gene expression levels in relation to specific immune cells.

5.1 Agarose GE of PCR1 products: RT-PCR with target specific primers.

5.2 QC check- Agarose GE of PCR1 products.

5.3 QC check- Agarose GE with PCR2 products.

5.4 QC check- Agarose GE of pooled PCR2 products.

5.5 Electropherogram traces from the Bioanalyzer of the 3 pooled samples and the final pool (Mix).

5.6 (i) *TPM* isoform distribution in samples from LR-seq: *TPM4*(a) *TPM1*(b) *TPM3*(c) *TPM2*(d)
(ii) Correlation between SR-seq and LR-seq of isoforms of *TPM4*(a) *TPM1*(b) *TPM3*(c) *TPM2*(d).

6.1 Agarose gel electrophoresis for validation of *TPM* expression using primer pairs targeted to exons 1 and 9 from *TPM1*.

6.2 Agarose gel electrophoresis for validation of isoform-specific primers of *TPM1-4.*

6.3 dC_T values of *TPM* isoform expression relative to GAPDH.

6.4 Dilution charts for qPCR Taqman assay.

6.5 Isoform expression in BCPP tissues and cell lines indicated by C_T values.

6.6 Fold-change expression of *TPM1*(a) and *TPM4*(b) isoforms.

6.7 qPCR fold-change expression of *TPM* isoforms according to Uromol subtypes

6.8 qPCR assay testing in urine samples

6.9 ROC curve- qPCR assay analysis using *TPM4* isoform expression tested in urine samples.

7.1 MS data for Tpm1.7

7.2 MS data for Tpm4.2

7.3 Antibody detection of Tpm protein isoforms.

Some of this work has been previously published (1, 2). For this work I was responsible for the writing of the text and the preparation of any figure(s) reused in this thesis. I declare that, apart from the assistance acknowledged and sources quoted, all the work presented in this thesis is my own work and has not been submitted in candidature for any other degree.

Unless stated otherwise, all the figures in the thesis are created using Biorender.com.

Chapter 1

INTRODUCTION

1.1 Synopsis

An overview of the relevant basic knowledge of the clinical aspects of bladder carcinoma is provided including epidemiology, aetiology, and classification. Clinical features of the disease, diagnostic modalities with subsequent management, prognosis and risk stratification measures are also discussed.

The concept of alternative splicing is explained that leads to generation of multiple isoforms from the same gene. A detailed discussion on tropomyosin isoforms and their complex gene family is provided including structure, function, exon usage, nomenclature and functional impact on cellular morphogenesis.

The molecular pathology that underpins the development of bladder cancer subtypes is highlighted with focus on the use of tropomyosin isoforms in the tumorigenesis of bladder cancer.

Lastly, sequencing techniques have been discussed along with available bioinformatics tools to quantify isoforms for evaluation of gene expression.

1.2 Bladder cancer

1.2.1 Epidemiology

Every year c. 20,500 new cases of bladder cancer (BC) are diagnosed in the UK, which accounts for 4.5% of all new cancer diagnoses (3, 4). It is the 7th most frequent cause of cancer deaths in the UK with nearly 5500 deaths per year (4). In 2020, worldwide statistics estimated 212,536 cancer deaths resulting from BC, making it the 10th most common cause of cancer deaths globally (5). Geographical distribution indicates an increased incidence in North America, Europe and Western Asia with 90-95% of cases presenting as urothelial (transitional cell) carcinoma (6). Men are affected three to four times more often than women; however, women often have more advanced disease at presentation with less favourable outcomes (7, 8). The disparity in gender incidence is complex; however, it is ascribed to a multitude of factors. Environmental factors like exposure to carcinogens (occupational/ smoking, more common in men) and delay in clinical management have been associated with the gender differences in incidence and survival. Other factors include differences in metabolism of sex steroid hormones with direct effects on the bladder and local immune milieu, in addition to tumour microenvironment, and response to intravesical/ chemotherapeutic drug treatments (9-12). BC is a disease affecting older persons with a median age incidence of 70 years, although it can also occur in younger individuals (as young as 45 years) (4, 13).

The second most common histological type of bladder cancer is squamous cell carcinoma (5%). Squamous cell carcinoma (SCC) is associated with parasitic infection of the bladder caused by *Schistosoma haematobium* (bilharzial), or chronic inflammation secondary to bladder stones or long-term catheterization (non-bilharzial)

(14). It is endemic in the Middle East, Sub-Saharan-Africa, South-East Asia and South America (15). In regions where there is increased incidence of bilharzial-SCC, an Egyptian study found that it contributed to at least 59% of associated SCC-cases (15).

Other rarer types of bladder cancer include adenocarcinoma (1-2%), sarcoma (<1%), melanoma (<1%), lymphoma (<1%) and small cell-neuroendocrine carcinoma (<1%) (13, 16).

1.2.2 Aetiology and Risk factors

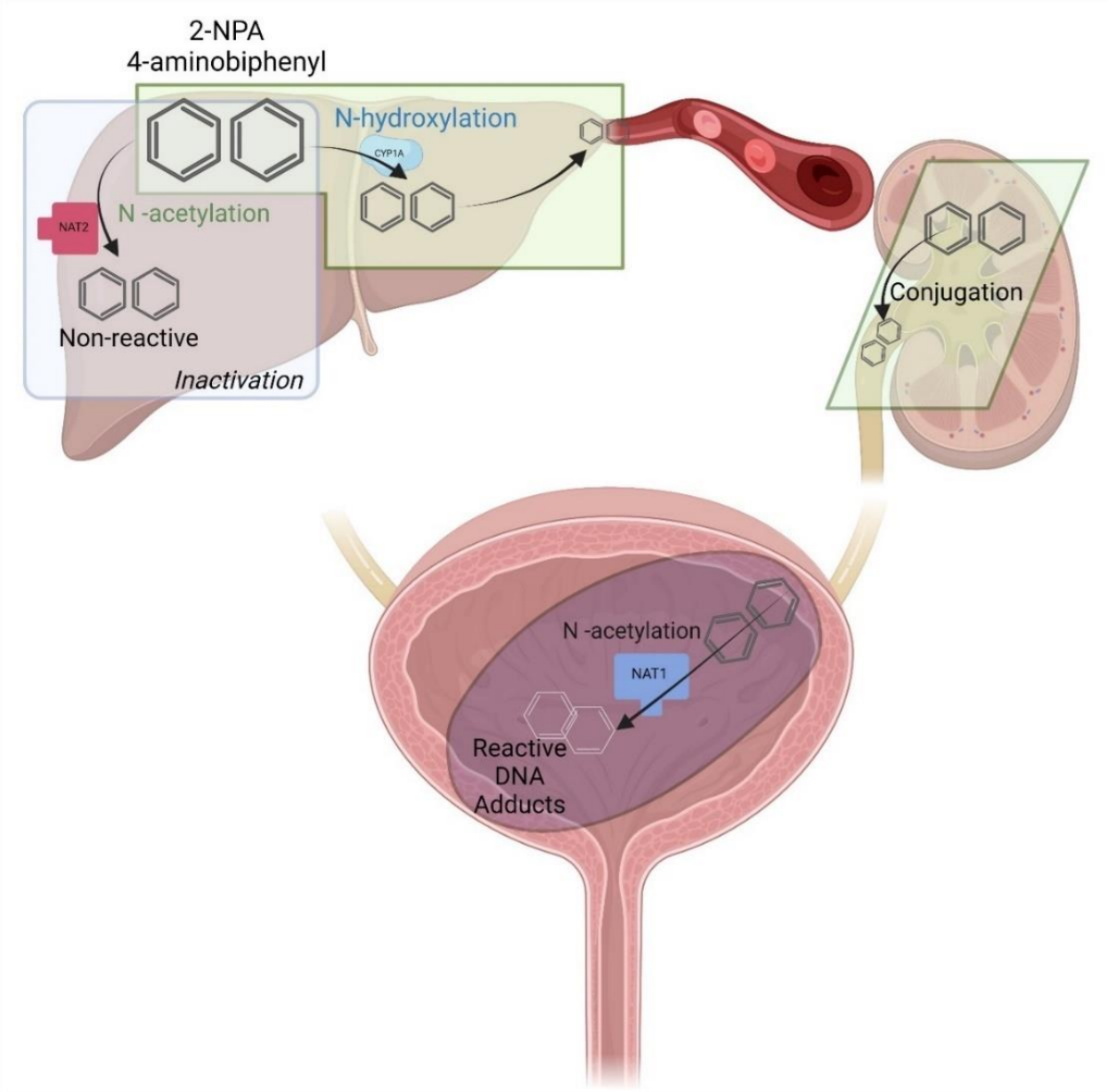
Risk factors for BC can be differentiated into modifiable risk factors due to lifestyle habits or external exposures, or unmodifiable risk factors such as gender and certain genetic (inherited) predisposition. Understanding the causative agent(s), the urothelial carcinogen(s), is vital in the primary prevention of BC and reducing disease-specific mortality. Nonetheless, all risk factors pose individual risks to the development and pathogenesis of BC. The *Attributable Risk* of BC is calculated as a fraction or percentage of the risk factor exposure responsible for the development of disease, of which cigarette smoking is the primary causative factor, followed by occupational exposure to chemicals and dyes.

1.2.2.1 Urothelial carcinogens

Continued exposure to urothelial carcinogens triggers the process of carcinogenesis by inducing genotoxicity directly or causing downstream disruptions in the cellular metabolic pathways. A vast number of carcinogens have been discovered in factories,

chemicals and dyes, but those posing greatest risk to BC development are polycyclic aromatic hydrocarbons (PAH) and aromatic amines (17, 18). The International Agency for Research on Cancer has categorized carcinogens based on toxicity with Class 1 as *definitely carcinogenic*, Class 2A *probably carcinogenic*, Class 2B *possibly carcinogenic* and Class 3 *non-carcinogenic in humans* (19). Urothelial carcinogens do not exist in their activated forms but are 'activated' by a series of steps that includes N-hydroxylation and N-acetylation assisted by the liver enzymes cytochrome P450 1A and N- acetyltransferase 2 (Figure 1.1).

***Figure 1.1 Metabolism of Aromatic amines and polycyclic hydrocarbons**. Urothelial carcinogens like 2-NPA or 4-aminobiphenyl are present in their inactive form. Within the liver, these compounds are first hydroxylated by CYP1A and then transported to the kidney in their conjugated form before being excreted in the urine. On interaction with the urothelial lining, they are metabolised by NAT1 to highly reactive carcinogens/ DNA adducts. Conversely, the liver inactivates the carcinogens by acetylation using NAT2. 2-NPA = 2-naphthylamine, CYP1A = Cytochrome P450 1A, NAT1 = N-acetyltransferase-1, NAT2 = N-acetyltransferase-2.*



1.2.2.2 Smoking

Consumption of tobacco/ cigarette smoking is the single predominant modifiable risk factor for the development of BC accounting for up to 50% of the presenting cases. Incidence rates of BC are higher in geographical regions where smoking is prevalent, and studies have reported a positive correlation of tobacco use with BC in men and women ($r = 0.20$, $r = 0.67$, respectively) (20). The risk is significantly greater for heavy smokers who have had higher cumulative smoking exposure (>30-pack year), with worse prognosis (21).

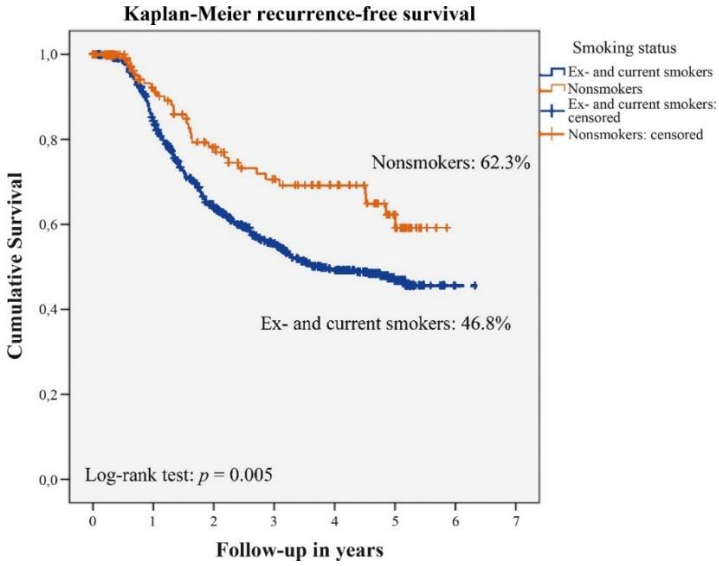
Tobacco smoke constituents are aromatic amines, particularly 2-naphthylamine (NPA) and PAH such as 4-aminobiphenyl (IARC Class 1). The composition of NPA and PAH varies in the type of tobacco smoked; blond tobacco appears to have a lesser content of NPA and PAH compared to black tobacco, which directly impacts the risk of BC (22, 23).

Use of electronic cigarettes (e-cig) has been marketed as a safer alternative to tobacco smoking and has gained much popularity, especially amongst younger individuals. It uses a heating element to vaporize a liquid solution containing flavoured nicotine dissolved in propylene glycol or vegetable glycerine solvents, which is inhaled by the user in aerosol form. Potentially it avoids the toxic substances that are generated through conventional tobacco smoking, albeit e-cigs may also harbour other (unknown) carcinogenic compounds released via heating the solvents. Preliminary studies performed on mice models (and humans) to investigate association of e-cigs with bladder cancer have shown an increase in bladder hyperplasia (and adenocarcinoma of lung), caused by DNA damages induced through release of mutagenic substances

within urine (24, 25). These results are alarming and further prospective studies are mandated to validate the safety profile of e-cigs.

The incidence of BC has slowly declined over the last decade with greater awareness of the cause-effect relationship between smoking and cancer. For smoking, there is a c. 30-year lead time between first exposure and development of BC. In estimates from early 2000s the European population attributable risk (PAR) of BC had notably declined to 43% (from 50%) in men and 23% (from 35%) in women, respectively for smokers (26). Conversely, with rapid industrialization taking place in developing countries (and less stringent regulation of known BC carcinogens), an increase in the number of new cases is observed (r=0.48 in men, r=0.44 in women) (20). Smoking cessation appears to have better disease-free and recurrence-free survival compared to smokers indicating smoking status to be an independent prognostic risk factor for BC (27, 28) (Figure 1.2).

Figure 1.2 Smoking status significantly affects the probability of NMIBC recurrence (log-rank test p= 0.005). *Figure reproduced from (27), license number: 5839320629249.*



Whilst smoking remains the number one risk factor for development of BC (PAR = 36.8%), it is likely there are a great number of (other) drivers for the disease (29, 30). Furthermore, the current understanding of the classification of bladder cancer subtyping indicates there may be differing mutational pathways (Section 1.3).

1.2.2.3 Occupational factors

Dye manufacturing industries became infamous for harbouring urothelial carcinogens present in chemical aniline dyes as discovered in the first clinical observational study by Rehn in 1895 (31). To date, occupational exposure is the second most common modifiable risk factor for BC (PAR = 5.7%) (30). Exposure to these compounds have been found in industries handling paint, metal, dye, and petroleum products. In recent years, increasing public awareness of these hazardous substances have prompted implementation of workplace safety regulations. Aromatic amines like 4-aminobiphenyl and 2-NPA have been found in rubber, plastic, and painting materials, these fall into the category of IARC Class 1. Workers (including firefighters) in close proximity to smoke produced from incomplete combustion of fuels have an increased risk of developing BC due to inhalation of carcinogenic PAHs like benzopyrene (summary incidence risk estimate (SIR) = 1.12; 95% CI: 1.04 to 1.21) (32). A Nordic Occupational Cancer study found that occupational exposure to the aforementioned carcinogens significantly increases the risk of development of BC, irrespective of smoking status, as was noted in chimney sweeps (SIR = 1.25; 95% CI: 1.05 to 1.56), waiters (SIR = 1.22; 95% CI: 1.07 to 1.38), hairdressers (SIR = 1.14; 95% CI: 1.02 to 1.26), cooks and stewards (SIR = 1.12; 95% CI: 1.01 to 1.25), printers (SIR = 1.11; 95% CI: 1.04 to 1.18) and seamen (SIR = 1.09; 95% CI: 1.03 to 1.14) (33).

1.2.2.4 Genetic predisposition

Though there is no evidence to suggest that BC has a familial inheritance pattern, genome-wide association studies have identified genomic variants that may increase susceptibility of BC development by conferring additional risk on exposure to carcinogenic environments. The balance between DNA damage by carcinogens and detoxification by a host of enzymes is delicately maintained. Phase I and phase II metabolic enzymes are mostly polymorphic, and two genes with associated single nucleotide polymorphisms (SNPs) have been implicated in increased bladder cancer risk: *NAT2* and *GSTM1* (34, 35).

N-acetyl transferase 2 (*NAT2*) governs the acetylation and bioconversion of aromatic amines to non-reactive compounds. Polymorphic nucleotide substitution produces four variant alleles, the presence of at least two gives rise to the 'slow' variant with decreased or absent catalytic activity. Smokers with slow acetylator status/ variant have prolonged exposure to carcinogens in comparison to carriers of wild-type rapid variant and thus have a higher attributable risk to developing BC (36, 37).

GSTM1 encodes for Glutathione S-transferase M1, which is responsible for the detoxification of PAH found in tobacco smoke via conjugation with glutathione. Homozygous deletion polymorphism (null alleles) results in complete loss of enzyme activity. This gene defect has been found to have a strong association with an invasive BC phenotype (38).

Lynch syndrome is an autosomal dominant disorder whereby pathogenic variants in DNA mismatch repair genes (*MSH2*, *MSH6*, *PMS2* and *MLH1*) promote development of colorectal and endometrial carcinoma. Extracolonic cancers are uncommon; however, urothelial carcinoma of the upper urinary tract in 9% (and bladder- 1%) has

a strong association with mutations in the *MSH2* gene with an increased female preponderance at an earlier age of onset (39-41).

1.2.2.5. APOBEC-mediated mutagenesis

Although smoking, occupational factors and the aforementioned genetic variants provide correlational evidence as established risk factors for BC, there remains a disconnect between smoking signature and BC per se, and the smoke carcinogens do not appear to have a direct effect as noted in the case of lung or oral cancer. Differing molecular pathways with gene mutations drive the carcinogenesis of urothelial bladder carcinoma (discussed in more detail in section 1.3), and the mutations can occur as a result of ageing or *APOBEC*-mediated (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like) mutagenesis (42). The *APOBEC* family of enzymes assist in the innate immune response particularly providing protection against retroviruses and infection. However, they can also cause C→T mutations within the genome by deamination of cytosine residues (43). The most frequent *FGFR3* mutation found in BC has a strong *APOBEC* mutational signature, suggesting it to be the driving force for BC development, in comparison to other tumours without *APOBEC* motifs (44).

1.2.3 Classification

The urinary bladder is a hollow muscular organ located in the pelvic region that stores urine before it is expelled from the body through micturition. The epithelial lining of the urinary tract, including the bladder, is comprised of transitional cells (urothelium) derived from the embryonic endoderm. This specialized stratified cell stratum is

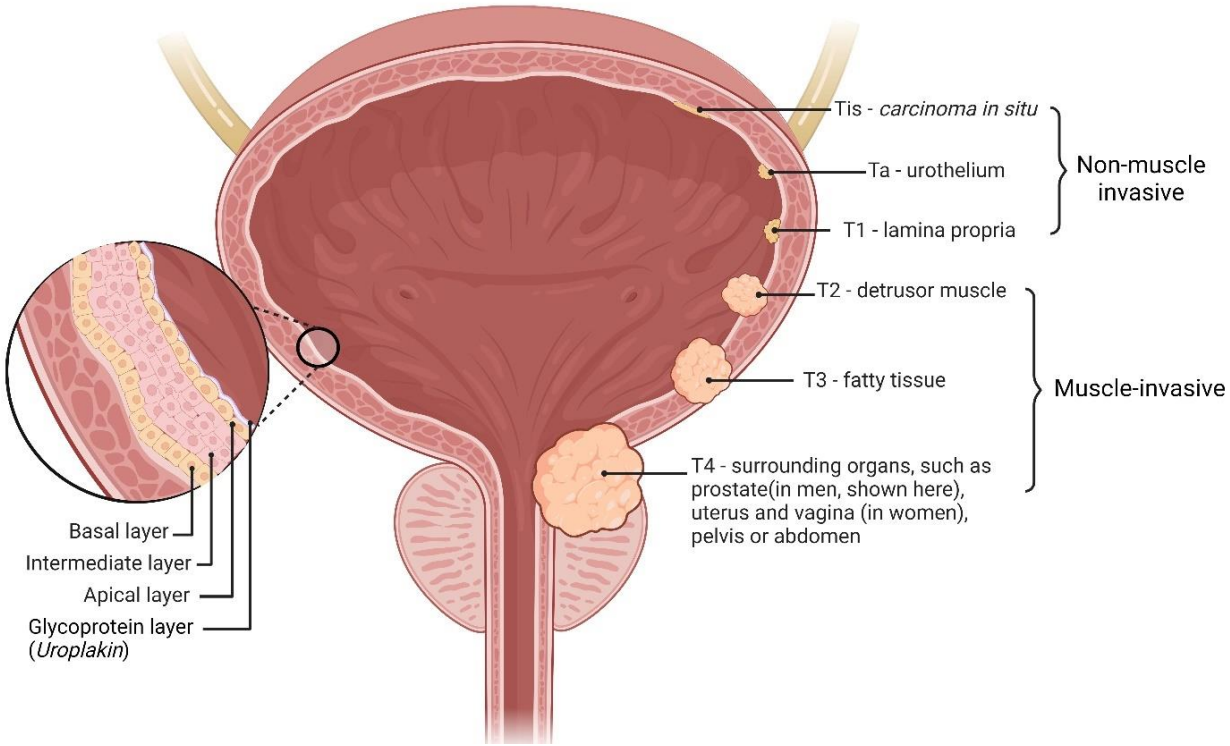
composed of three distinct layers. The innermost apical layer is formed by dome-shaped umbrella cells with impermeable tight junctions, that become flat on bladder distention. A protective plaque of glycoproteins (uroplakins) overhangs the apical layer and separates the urine from the urothelium. The intermediate (multi-) layer is made of polygonal cells and a single cell (basal) layer of mono-nucleated cuboidal cells keeps the layers anchored to the basement membrane. These basal cells are progenitor urothelial stem cells and serve as markers for their identification (45). The transitional urothelium gains its name from the ability of the bladder to stretch under tension (to accommodate urine) whilst maintaining a strong osmotic barrier which remains urine-proof (46). The urothelial cells remain in a quiescent state and have a very slow turnover- it takes ~200 days for a urothelial stem cell to mature and eventually be shed. However, any breach in the epithelium (caused by inflammation, injury or cancer) instigates the process of rapid rejuvenation to stabilize and completely restore the urothelial lining (47, 48).

Underlying the surface epithelia is a layer of fibrous connective tissue called the lamina propria, which separates the urothelium from the detrusor muscle. An outermost layer of fatty connective tissue surrounds the detrusor muscle.

The Union for International Cancer Control Tumour, Node, Metastasis (UICC TNM, 8th edition) Staging classification is used to assess the extent of invasion of cancerous cells within the bladder wall (and/ or beyond) (Figure 1.3). Tumours arising from the innermost urothelial lining can occur as flat or papillary lesions. If they are confined to the urothelium and/ or lamina propria, they are categorized as non-muscle-invasive bladder cancer (Stages Ta/T1). Tumours that traverse into the muscle layer or into surrounding fatty tissue or nearby organs are categorized as muscle-invasive bladder

cancer (Stages T2/T3/T4). Highly malignant urothelial cells that appear as flat lesions confined to the urothelium are termed *carcinoma in situ* (*CIS*; Stage Tis). These lesions are usually invisible, rarely appear as solitary lesions (1-2%) and are usually found to occur concomitantly with other visible tumours. Nodal status is determined by the presence or absence of metastases within the regional lymph nodes: single lymph node with size <2cm (N1), single lymph node with size >2cm but <5cm or multiple lymph nodes <5cm in size (N2), or single/multiple lymph nodes >5cm in greatest length (N3).

Figure 1.3 TNM Staging of Bladder Cancer. *Stage Ta: papillary tumours are confined to the inner most lining of the bladder wall called the urothelium, comprised of specialized transitional epithelial cells (enlarged schema shows the epithelial lining of the bladder with three distinct layers: apical, intermediate, and basal, with a glycoprotein layer of uroplakins that separates the urothelium from the urine). Stage T1: tumour cells breach the basement cell membrane and invade underlying lamina propria. Stage T2: bladder cancer cells invade into muscular layer. T3: malignant tumour cells traverse through the bladder wall and invade the adipose (fatty) layer of connective tissue. T4: Extra-vesical growth of BC cells into the surrounding organs, pelvic cavity and/ or abdomen.*



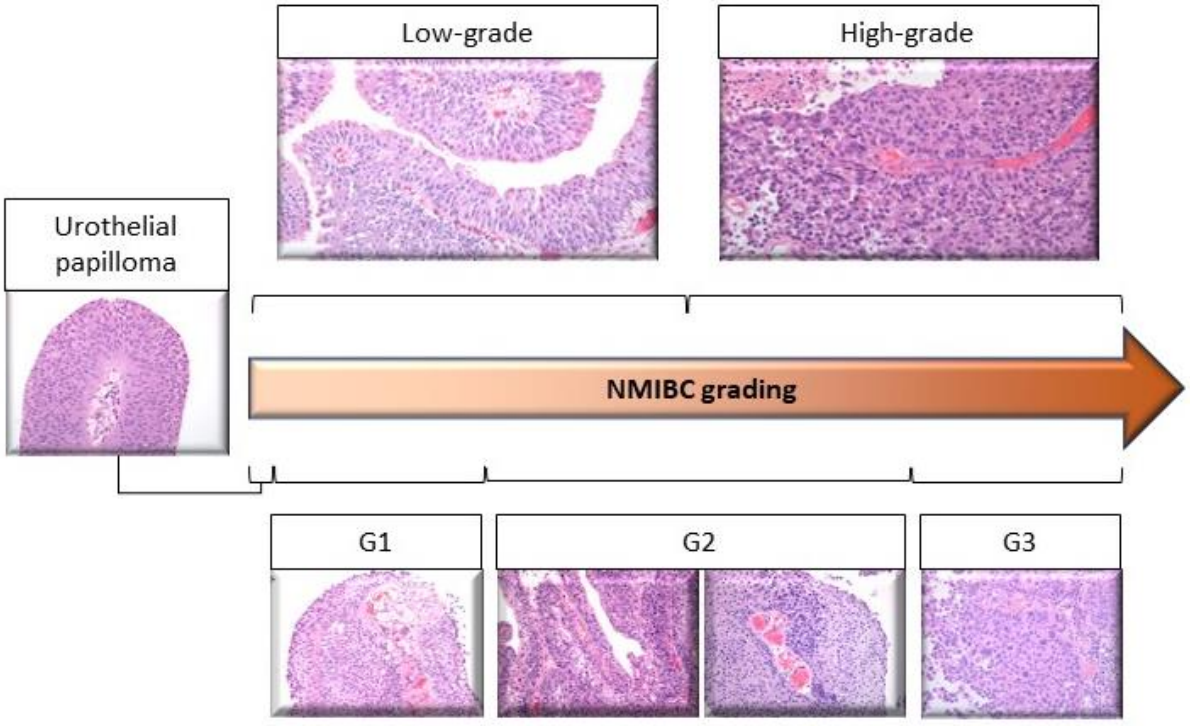
Microscopically, assessment of the cellular morphology of the cancerous tissue determines the grade of the disease. Two systems of histological grading system are used to evaluate the prognostic index of NMIBC tumours. The 1973 WHO Classification uses a numerical system to differentiate between Grade 1 well-differentiated (G1), Grade 2 moderately-differentiated (G2) and Grade 3 poorly-

differentiated cells (G3) (49). In 2004, a revised version of this WHO classification introduced the use of high-grade (HG) and low-grade (LG) tumours based on nuclear pleomorphism and cellular atypia. A further category of papillary urothelial neoplasm of low malignant potential (PUNLMP) was added. Accordingly, in the revised 2016 edition, G1 represents PUNLMP and low-grade tumours, G2 represents low-grade and some high-grade tumours and G3 represents high-grade tumours only. In the current edition of the WHO 1973 2004/2016/2022 Classification (Figure 1.4), the categories are divided into 4-tier combinations: Low-grade/ G1, Low-grade/ G2, High-grade/ G2 and High-grade/ G3 (50). This covers all G2 tumours by subdivision of G2 class into LG and HG based on disease prognosis. *CIS* lesions are categorized as High-grade/G3. In a recent study by Sylvester et al, it was observed that lesions categorized as PUNLMP and those patients presenting with LG-Ta disease had similar prognoses, therefore the use of urothelial hyperplasia or dysplasia is not considered as a unique term but its clinical and biological significance as early presentation of LG-Ta is recognized (51). In clinical practice, histopathologists use the four-tiered classification to define the tumour characteristics to aptly stratify tumours according to risk for further management (discussed later) (52).

Figure 1.4 WHO 1973 2004/2016/2022 Classification of histopathological grading of urothelial neoplasm*. Urothelial papilloma (hyperplasia/ dysplasia) is not included in the 4-tier grading system. Urothelial papilloma: Hyperplasia of urothelial cells is observed, with no nuclear atypia but streaming due to vertical polarization. G1-Low-grade: Variable thickness of urothelium with slightly enlarged nuclei but limited variation in nuclear size. G2-Low-grade: Variation in nuclear size with hyperchromasia and crowding is noted. G2-High-grade: Mix of disordered cellular differentiation and G1 features, with increase in cellular atypia and hyperchromatic nuclei. G3-High-grade: Severe cellular atypia with pleomorphism, hyperchromasia and variability in nuclear size. CIS lesions are included in*

the High-grade category (not shown). Low-grade: Papillary formation of varying size urothelial cells with enlarged nuclei surrounding central fibrovascular core. Variations in thickness of urothelial lining is observed but the apical layer of umbrella cells remain intact. High-grade: Disordered cellular architecture with variation in nuclear size, increased nuclear: cytoplasmic ratio with atypia, and lack of differentiation of apical cell layer. Histology slides for the figure are adapted from van der Kwast T. Grading-bladder. PathologyOutlines.com website.

https://www.pathologyoutlines.com/topic/bladdergrading.html.



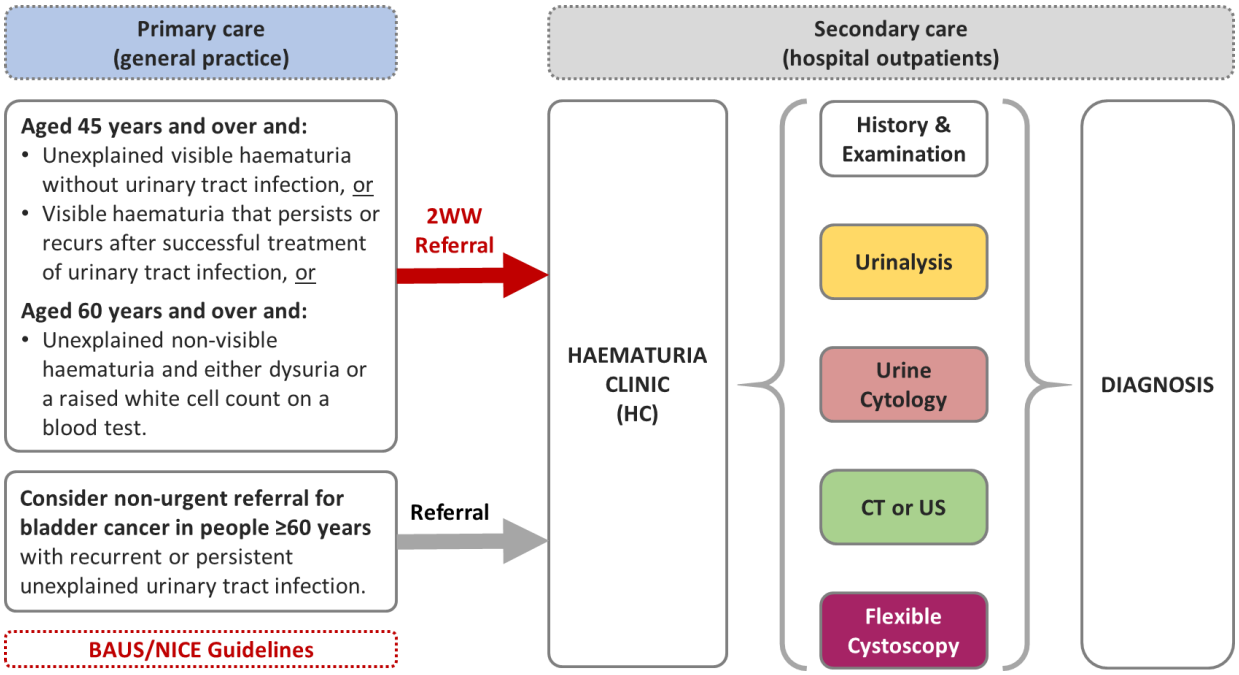
1.2.4 Clinical presentation

In ~90% of cases, the most common symptom of bladder cancer is blood in the urine, or *haematuria*. It can appear grossly (visible haematuria, VH), or be detected on a urine dipstick test (nonvisible haematuria, NVH) performed by a healthcare professional. By urine microscopy, the presence of at least 2-3 red blood cells per high powered field is equivalent to a positive dipstick test (53). NVH has a low positive predictive value (1-

2%) such that at most 1 in 20 individuals with NVH will be diagnosed with BC. Other causes of NVH may include urinary tract calculi (stones), urinary tract infections or, in some cases, have no identifiable cause. However, VH confers a PPV of 2.8% in patients who are diagnosed with BC, nearly double the risk of cancer when compared to NVH (54). A single-centre prospective study analysed the rationale for investigating haematuria with a protocol-driven assessment and found that, from a cohort of 4020 patients, 18.9% with VH and 4.8% with NVH had BC (55).

If found to have haematuria in the primary care setting, a prompt (*2-week-wait*) referral is made for further assessment to be carried out at a specialist review clinic according to national guidelines (Figure 1.5). Practise varies across the UK (and internationally), but most tertiary care hospitals have a one-stop haematuria clinic for evaluation. Diagnosis is made following a complete but focused urological history along with a repertoire of investigations to assess the urinary tract to exclude evidence of malignant disease (of the bladder and upper tract) that may present as haematuria. Imaging studies include either an ultrasound scan of the kidneys, ureter and bladder (US KUB), or computed tomography (CT) (Section 1.2.4.1). CT utilizes serial images in multiple phases, pre- and post- injection of a contrast agent, to highlight any structural deformities within the urinary tract. In addition, the lower urinary tract including the bladder, prostate and urethra is directly visualized by flexible cystoscopy (section 1.2.4.2).

***Figure 1.5 Management pathway of patients presenting with haematuria in the UK**. As depicted in the illustration, patients are seen within a primary care setting initially, with subsequent referral to secondary care for investigation and diagnosis of bladder cancer. Current BAUS/ NICE guidelines suggest that if patients have a high suspicion of bladder cancer based on the criteria listed, they fall within the 2-week-wait category and are referred on an urgent basis to the haematuria clinic. At this clinic, a urology specialist (clinician or nurse) conducts the investigation with the aid of available diagnostic tools including a comprehensive urological history and abdominal examination. A preliminary diagnosis is made awaiting further management. 2WW= 2 week wait, BAUS= British Association of Urological Surgeons, CT= Computed Tomography, NICE= National Institute of Health and Care Excellence, US= Ultrasound scan. Figure kindly donated by Professor Richard T Bryan.*



1.2.4.1 Imaging

Computed tomography (CT) urography is utilized to view contrast enhanced images of the urinary tract to identify any 'filling defects' that would indicate intraluminal papillary tumours of the bladder and/or ureter. Additionally, CT imaging can identify other

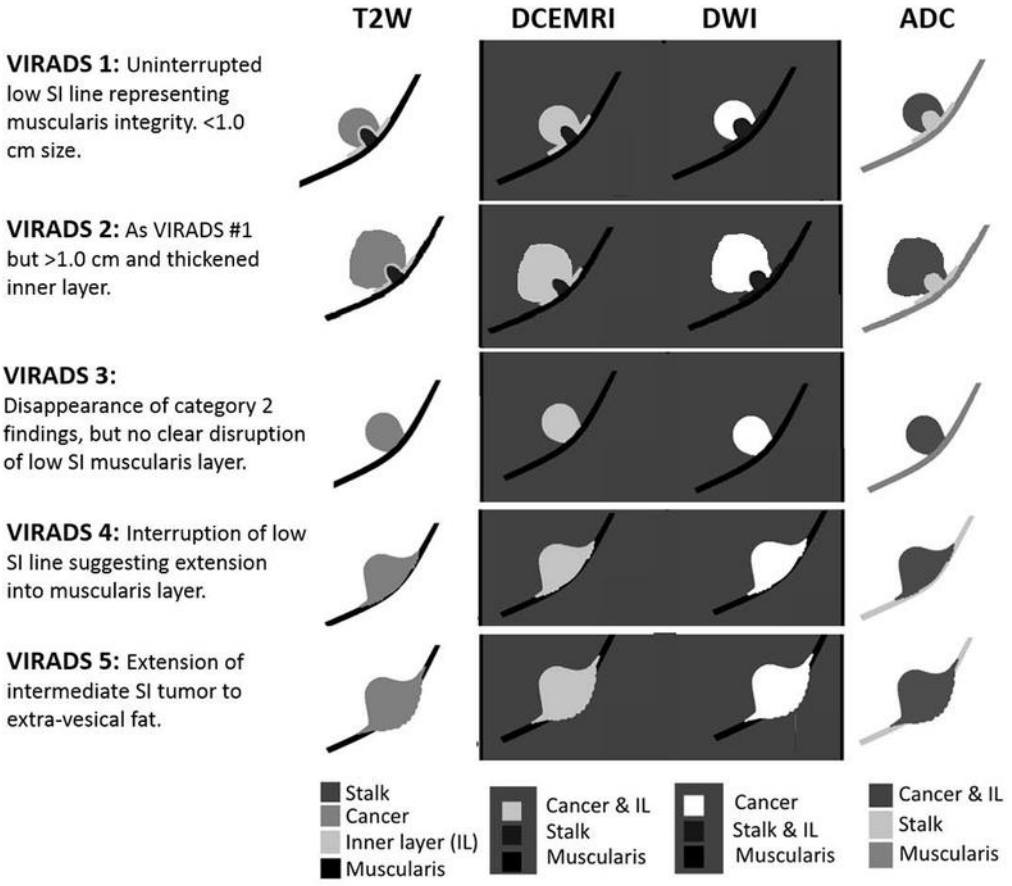
lesions present within the upper urinary tract (renal parenchyma, intrarenal pelvicalyceal system and upper and/ or mid-ureters) and presence of any nodal disease involvement or distant spread to other intra-abdominal organs (56). The presence of a concomitant upper tract urothelial tumour increases by 7.5% if a bladder tumour is found at the trigonal area, therefore mandating further imaging with CT (57).

Alternatively, US KUB can be performed to detect abnormalities present within the kidneys such as calculi or hydronephrosis (which can result due to obstruction distal to the kidneys), to characterise any lesions, and/ or presence of any masses within the bladder. US is highly operator-dependent and cannot reliably distinguish ureteric tumours as effectively as CT (58).

The role of multiparametric magnetic resonance imaging (mpMRI) is currently being explored as an enhanced imaging modality that provides better resolution of the anatomy of the tissue planes in comparison to traditional CT. It has the added advantage of null radiation exposure and can be used to assess disease extent pre- and post- treatment. A standardised MRI reporting technique- Vesicle Imaging Reporting and Data System (VI-RADS) has been developed based on a 5-point score calculated from T2-weighted imaging, diffusion weighted imaging (DWI) and dynamic contrast enhancement (DCE), which is able to appreciate the ‘invasiveness’ and extension of BC (59, 60) (Figure 1.6).

It must be noted that no imaging modality can detect *CIS* which, at present, can only be diagnosed by cystoscopy and biopsy.

Figure 1.6 Schematic representation of VIRADS scoring model. *The multiparametric MRI utilizes T2-weighted imaging (T2W), dynamic contrast enhancement (DCE), diffusion weighted imaging (DWI) and apparent diffusion coefficient (ADC) to interpret a VIRADS score between 1-5. IL= inner layer, SI= signal intensity. Figure reproduced from Panebianco et al. Multiparametric Magnetic Resonance Imaging for Bladder Cancer: Development of VI-RADS (Vesical Imaging-Reporting And Data System). Eur Urol. 2018 Sep;74(3):294-306. doi: 10.1016/j.eururo.2018.04.029. Epub 2018 May 10 (59), license number: 5839321110800*



1.2.4.2 Flexible cystoscopy

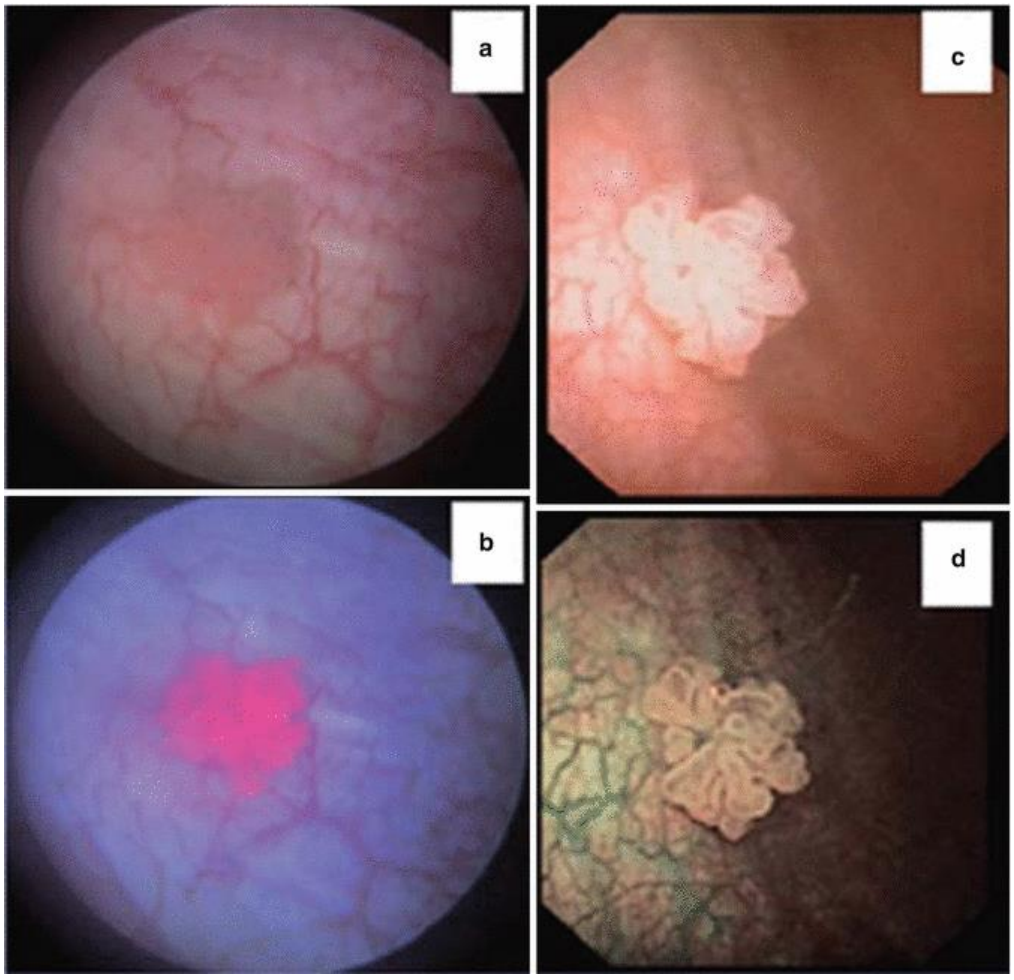
Flexible cystoscopy is the principal diagnostic technique used for inspection of the urinary bladder for suspected malignancy. It utilizes a flexible fibre-optic tubing with a camera tip connected to a monitor to produce real-time imaging, or an eyepiece to

allow visualization of bladder mucosa directly. Conventional white-light cystoscopy (WLC) has a sensitivity of 85% and specificity of 87%, although this is highly operator-dependent (61). Specific shortcomings include misdiagnosing ‘red patches’ as tumours and overlooking flat *CIS* lesions. To overcome the limitations of conventional WLC, adaptations have been made to improve and enhance cystoscopy imaging with the use of technological adjuncts viz. Narrow Band Imaging (NBI) and blue light cystoscopy (photodynamic diagnosis, PDD).

NBI technology is based on splitting white light into two colour bandwidths: blue (415nm) and green (540nm) wavelengths. Malignant bladder tissue is surrounded by dense capillary networks. In the NBI mode, these areas appear to have enhanced colour contrast as brown or green structures against a lighter background when compared to WLC (61-64). In comparison, blue-light fluorescence cystoscopy (PDD) is carried out following the instillation of a prodrug (Hexa-aminolevulinic acid, HexVix) and utilizes selective accumulation of photoactive protoporphyrin IX within BC cells compared to the adjacent normal urothelium. Tumour cells appear bright pink when illuminated by blue light (65, 66).

The performance of NBI and PDD both appear to be superior to WLC in their diagnostic proficiency, with sensitivities of 96% and 93%, respectively (67) (Figure 1.7). However, clinical uptake of these optical enhancement technologies has not been universal, with WLC remaining the mainstay of BC diagnosis. Furthermore, the practicalities (and expense) of instilling HexVix one hour prior to cystoscopy means that PDD can be applied in specialized centres only and not be part of a regular outpatient diagnostic setting.

Figure 1.7 Comparison of WLC (a) with PDD (b), digital WLC (c) and NBI (d) to visualize urothelial tumours. *Using these optical enhancements greatly improves accuracy of diagnosis particularly of CIS lesions, which may otherwise go overlooked. WLC = White-light cystoscopy, PDD = Photodynamic diagnosis, NBI = Narrow-band imaging. Figure is reproduced from Abdominal Key, Enhanced Imaging: NBI, PDD, SPIES* https://abdominalkey.com/enhanced-imaging-nbi-pdd-spies/#Fig5



In addition to the costs, patients experience significant anxiety associated with having a 'procedure' being carried out in the early diagnostic stage to determine their disease status. Cystoscopy is invasive, uncomfortable and carries a risk (albeit small) of morbidity. Self-limiting after-effects (with almost all patients) include slight burning/ dysuria, increased frequency of urination and/ or bleeding. Nearly 5-6% may also

develop urinary tract infection requiring treatment with antibiotics (68-70). Rarer complications such as urinary retention (inability to pass urine) and insertion of a temporary urethral catheter can occur in 1 in 250 patients, particularly those that have an enlarged prostate.

Thus, there is a pressing need for better diagnostic procedures, e.g., urinary biomarkers (section 1.2.4.4).

1.2.4.3 Urine cytology

Voided urine cytopathology is considered an important non-invasive adjunct to cystoscopy and is useful for diagnosing high-grade disease (sensitivity 84%) but lacks accuracy for low-grade tumours with sensitivity ranging between 16-48% (71). Cells from urine are observed under the microscope directly to identify any loss of cellular differentiation. Although it remains a candidate on the BC evaluation panel due to very high specificity even for low-grade tumours (~98%), there is significant interobserver variability (72, 73). Additionally, with the updated EAU guidelines suggesting use of urinary biomarkers as adjuncts in follow-up of BC patients (discussed in the next section), many units are excluding the use of urine cytology from their diagnostics panel.

1.2.4.4 Urinary biomarkers

Current guidelines continue to advocate the use of ‘gold standard’ cystoscopy as first-line diagnostics, for its sensitivity (87-100%) and specificity (64-100%), despite the

associated economic implications and burden to healthcare services and patients alike (74-76). There is an urgent unmet need for alternative non-invasive approaches, e.g. urinary biomarkers, that can be incorporated within the diagnostic and surveillance pathways for BC. Recent technological advancements in high-throughput DNA and RNA sequencing have facilitated understanding of the molecular biology and pathology of BC, with mutations, gene expression profiling, and transcriptomic analyses paving the way for *'precision medicine'*. This has allowed for investigation of potential biomarkers that could substantially improve the diagnostic and screening modalities for BC by reducing the dependence on cystoscopy, and the utilization of non-invasive techniques. Commercially available biomarkers have been proposed as an adjunct to cystoscopy for the follow-up of patients but there is apparent hesitance in implementation within clinical practice due to the lack of appropriate clinical validation and perceived inability to match the accuracy of cystoscopy. As a means of *liquid biopsy*, urine is a logical choice: it is in immediate contact with BC tumour tissue and it serves as an excellent reservoir of DNA, RNA and/or protein that can be sourced from exfoliated tumour cells or found as cell-free constituents. Sample collection is inexpensive, reproducible, and non-invasive. In fact, the latest iteration of the EAU guidelines for NMIBC mentions use of promising novel urinary biomarkers that may be used in the follow-up/ surveillance of BC patients that have low/ intermediate-grade disease that are averse to cystoscopy due to patient preference or their performance status (77).

The clinical application of a urinary biomarker tool can be classed under the following categories (albeit they are not mutually exclusive):

- one that can be provided *en masse* for the large-scale screening of selected populations (*screening*),
- or that which has a technically proficient but sophisticated system available for multiple uses, such as triaging haematuria patients for cystoscopy (*diagnostic*),
- providing an adjunct to cystoscopy for NMIBC surveillance (replacing some cystoscopies for low-risk/ intermediate-risk NMIBC, or the intervals in between check cystoscopies for high-risk NMIBC) (*monitoring*).
- They may also provide molecular-genomic tumour characteristics to guide the urgency of investigation or indicate optimal treatment modalities (*prognostic, predictive*),
- or to assess treatment response (*efficacy*).

When investigating a potential biomarker, multiple factors influence its performance such as the characteristics of the biomarker itself, i.e. whether it is elevated or reduced in cancer (affected by the protein concentration or gene expression) or whether it is present only in cancer (due to DNA mutation). Additional factors that may influence data analysis include sample collection and processing via analytical methods (e.g. preservatives or immediate processing), the study design, sample size, patient characteristics (demographics, age, sex), patient selection and control subjects. The clinical utility of a biomarker can be measured via its ability to correctly identify a high proportion of true positives with the disease (*sensitivity*), a high proportion of true negatives without the disease (*specificity*) and its predictive value. An ideal biomarker is difficult to achieve (due to the confounding factors as mentioned); it may possess high sensitivity at the expense of low to moderate specificity and vice versa. Additionally, benign (inflammatory) conditions of the bladder may also influence the

results yielding false positives (*Type I error*), or LG BC tumours may not meet the threshold for quantifiable results to generate false negatives (*Type II error*).

From a practical standpoint, clinicians can estimate the probability of BC diagnosis based on a multitude of factors which includes patient characteristics, history, risk factors and assisted imaging (Figure 1.5). National centres have also published their *'risk of BC'* algorithms to assist in the management of these patients (78). The IDENTIFY study, a research initiative by British Urology Researchers in Surgical Training (BURST), was conducted as a multi-centre prospective international observational study to evaluate diagnostic practise for suspected urinary tract cancer with the aim to develop a standardised risk-based protocol for assessment of haematuria to predict likelihood of malignancy (79-81). Additionally, biomarkers may be used in the surveillance setting to monitor patients for disease recurrence or progression. However, it is unlikely that HG BC or high-risk patients will be on surveillance solely dependent on biomarker assessments. The use of biomarkers in this instance would provide useful genomic information and assist in treatment planning or escalation.

A comprehensive literature search on biomarker discovery revealed how technological advancements have revolutionized the search for the optimal substrate from single protein immunoassays to whole-genome sequencing (2). Studies in the late 1990s to 2000s investigated urinary proteins as biomarkers, leading to the development of *Urothelial Bladder Carcinoma* – antigen assay (UBC), and other protein-based bladder cancer detection kits such as *NMP22* and *BTA* (82-84). The latter two tests can be measured quantitatively using Enzyme linked immunosorbent assay (ELISAs) or as qualitative point-of-care (POC) tests. At least 6 urine-based tests have been approved

by the US FDA that may be used in conjunction with cystoscopy: *NMP22*, *NMP22 BladderChek Test* and *UroVysion*® for diagnostic and surveillance use, whereas *BTA TRAK, BTA Stat,* and *ImmunoCyt* for surveillance only.

Nuclear matrix proteins (NMP) responsible for DNA replication and gene regulation within the nucleus are shed into urine upon apoptosis of cells and reflect increased mitotic activity. *NMP-22 ELISA* (quantitative) and *NMP-22 BladderChek* (qualitative POC) have been proposed as diagnostic and surveillance tests for BC (85) (Figure 1.8).

Bladder tumour antigen (BTA) *Stat* and *TRAK* kits are assays that detect presence of human complement factor H – related protein in urine. They are available as qualitative POC and quantitative tests respectively, to be used in a surveillance setting in addition to cystoscopy (86).

Validation studies have not proven satisfactory sensitivity and specificity for these protein-based assays (*NMP22* and *BTA*) to be used as stand-alone tests for primary diagnosis (or surveillance) of BC, as they may also be found elevated in benign bladder conditions and cannot distinguish bladder cancer from other diseases (82, 87-95) (see Figure 1.8 for sensitivity and specificity for each test).

Combining protein biomarkers into multi-analyte tests has largely failed to increase sensitivity and specificity although it remains possible that tumour specific variants of proteins (e.g., novel splice variants or glycoforms) could offer improved performance.

UroVysion utilizes fluorescent in situ hybridization (FISH) to identify aneuploidy of chromosomes 3, 7, 17 and/ or loss of heterozygosity (LOH) 9p21. This test offers high

sensitivity for HG disease ranging between 83-100% and can be used for surveillance of patients to check treatment response to immunotherapy (96, 97).

Immunocyt (also referred to as uCyt+) targets the glycoform carcinoembryonic antigen and sulphated mucins found on exfoliated bladder cancer cells detected by three fluorescently labelled monoclonal antibodies. However, like other protein-based tests, false positives are obtained with Immunocyt in benign urological conditions including urinary tract infections, urinary tract stones and/ or benign enlargement of the prostate (98).

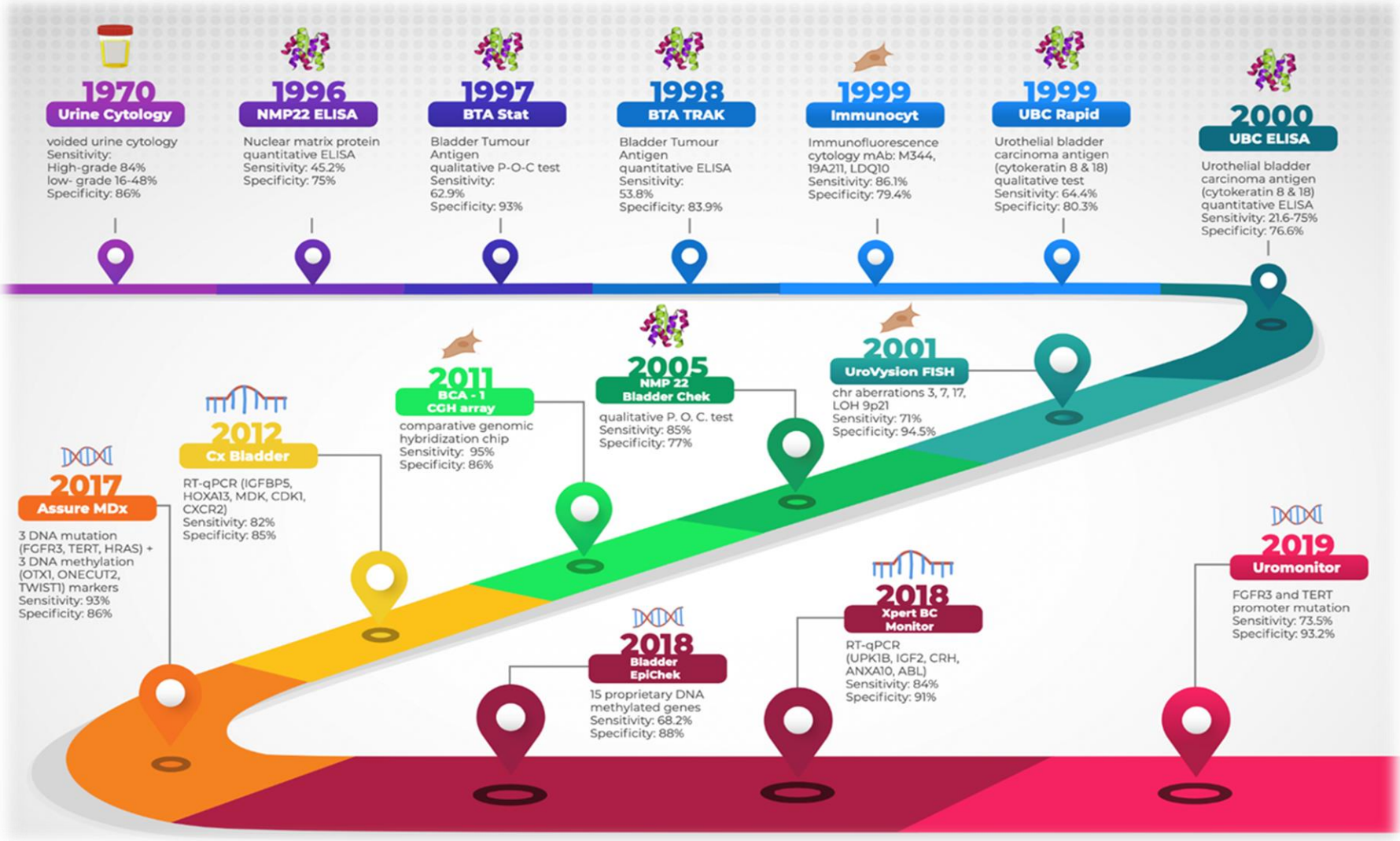
Other protein assays have also come forth that utilize detection of cytoskeletal proteins i.e. cytokeratins which are released (cytokines) on cellular breakdown of BC epithelial cells. These include cytokeratins 8, 18, 19 and 20; the **UBC** assay is based on detection of fragments of cytokeratin 8 and 18 in urine and is available as quantitative ELISA and qualitative POC tests (87, 99, 100). More recently, cytokeratin 17 has been implicated in detection of bladder cancer with the UroX biomarker study underway for validation and implementation in clinical practise (Table 1.2) (101).

Biomarkers that are informative as diagnostic indicators of cancer-disease-burden include genomic data (DNA from tumour cells) and/ or transcriptomic information/ gene expression data (shed/ cell-free RNA). This has been tested utilizing polymerase chain reaction (PCR), quantitative PCR (qPCR), and other highly sensitive analytical techniques such as digital droplet PCR (ddPCR) or high throughput sequencing. DNA copy number, mutation and methylation specific changes have been observed in HG BC (and LG BC) and these phenomena have been utilized to create commercially available BC detection assays. ‘Next-generation’ tests that have gained approval from

the US FDA have targeted DNA (such as Assure MDx and Bladder Epichek) or RNA (CxBladder and Xpert BC Monitor) extracted from urine (Figure 1.8).

Urine-tests based on detecting tumour DNA can be utilized to detect BC (and genomic characterization). ***Assure MDx*** is a (6 target) multipaneled DNA mutation-methylation based assay that analyses a combination of mutation status of *FGFR3*, *TERT* and *HRAS*, and epigenetic profiling of 3 loci (102). Methylation status change is observed in BC and can be used as a differentiating factor between normal urothelium and carcinomatous change (2, 103-105). An example is ***Bladder Epicheck*** - a urinary biomarker based on 15 DNA methylation markers, which can be used to monitor patients for recurrence and progression. ***Uromonitor*** is a DNA-based assay that detects hotspot mutations in 3 genes: *FGFR3*, *TERT* and *KRAS*, which is useful during surveillance checks to indicate possibility of recurrences (106, 107).

Figure 1.8 Evolution of urinary biomarker tests including US FDA approved, with sensitivity and specificity and assay targets. *Figure from (2)*



RNA-sequencing (RNA-seq) provides a *(dynamic)* representation of the expressed elements as genes and their isoforms via measurement of the transcriptome (108-111). Even though most RNA may be degraded by RNAses, appreciable amounts can still be detected in urine (112, 113).

Given the published data thus far, and the results of meta-analysis studies, these biomarkers appear to have variable sensitivities and specificities (114) (Table 1.1). The analyses are further dependent on the population characteristics of the study i.e. those with HG BC may be more sensitive compared to LG BC.

Table 1.1 | Pooled performance indicators for 4 diagnostic (DNA- & RNA- based) biomarker assays to assess recurrence of NMIBC during follow up. The number of studies for meta-analysis of each biomarker is listed. PPV= positive predictive value, NPV= negative predictive value. (114)

Biomarker test	Target genes	Studies	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Xpert BC	5x RNA: *ABL1, ANXA10, CRH, IGF2, UPK1B*	10	72	76	43	92
CxBladder monitor	5x RNA: *CXCR2, CDC2, HOXA13, MDK, IGFBP5*	2	91	61	16	98
Uromonitor	2x DNA: *FGFR3, TERT*	2	93	79	67	96
Bladder Epichek	15 DNA methylation loci	5	74	84	48	94

Preliminary results have prompted large-scale prospective clinical trials to validate the use of these biomarkers in clinical practise such as the ‘STRATA Study’ for Cxbladder,

and ‘ReplaceCysto’ for comparing Bladder Epichek and Xpert BC (see Table 1.2 for additional studies underway).

Table 1.2 | Summary of active clinical trials with the biomarker assay being tested. NCT= national clinical trial

Study Name	Biomarker	Location	Clinical Trial no.
STRATA	CxBladder	USA	NCT03988309
ReplaceCysto	Bladder Epichek, Xpert BC	USA	NCT05796375
URISAFE	Urine DNA methylation analysis	China	NCT05337189
SeARCH	DNA-based multitarget	Netherlands	NCT06026189
PENK	EarlyTect® DNA Methylation	South Korea	NCT05220189
UroX biomarker bladder cancer	UroX™ immunobiomarker, binds to oncoprotein Keratin17	UK	NCT03973307

1.2.4.5 Other non-urine-based biomarkers

Most urine-based biomarker tests discussed in the previous section are based on detection of DNA (methylation, mutation) extracted from either pelleted cells from urine (cpDNA) or from the urine supernatant as cell-free DNA (cfDNA). Similarly, a high turnover of cells can be detected as an abundance of plasma-borne tumour DNA known as circulating tumour DNA (ctDNA), which is directly associated with prognosis and overall survival (115-117). Sampling and measurement of mutation markers in ctDNA would aid in predicting recurrence, identifying metastatic disease or guiding

treatment strategies using targeted therapy regimens for supposed clinical trials (118). However, the somatic mutations that may be found in ctDNA (plasma) or cfDNA (urine) can come from shed cells that may appear (phenotypically) non-cancerous histologically generating false positive results due to the '*field effect*' of carcinomatous changes within the urothelium.

1.2.5 Management

NMIBCs are managed by an initial transurethral resection of bladder tumour (TURBT) under general or regional anaesthetic followed by a single shot instillation (SSI) of intravesical chemotherapy where indicated (as per EAU guidelines) (77). The rationale for use of SSI is to eradicate any residual 'floating' tumour cells after performing TURBT, that may otherwise implant into the urothelium to trigger new growth, and to have an ablative effect on any other lesions that have microscopic disease (119). Furthermore, the greatest risk of recurrence of BC is in the initial 3-months following TURBT, and SSI has shown to reduce the recurrence rate (by 39%) from preventing any reseeding of tumour cells post initial-TURBT (120). Following TURBT, EAU/ EORTC risk scores (Table 1.3) guide further management, adjuvant therapies and the schedule of cystoscopic surveillance. Adjuvant long-term intravesical instillation of therapeutic agents decreases the risk of recurrence (at 3 years) by 86.1% compared to short-term intravesical therapy particularly for intermediate and high-risk NMIBC (121). Mitomycin C (MMC, chemotherapy) and Bacillus Calmette-Guerin (BCG-immunotherapy) are the two main therapeutic agents used for NMIBC intravesical therapy. Alternative antineoplastic chemotherapeutic agents include Epirubicin,

Pararubicin and Gemcitabine. EAU/ EORTC risk scores for NMIBC are based upon clinico-pathological staging with TNM, tumour grade and anatomy to calculate risk score for administering definitive therapy (section 1.2.6).

Adjuvant intravesical BCG-immunotherapy is more effective in reducing the rate of recurrence of BC (in high-risk BC) than TURBT alone or even TURBT with SSI of intravesical chemotherapy with MMC (122-124). Standard optimal BCG instillation regimen includes empirical six induction doses given once-weekly followed by full-dose or reduced dose maintenance therapy given three-weekly spread out over a period of three years. The maintenance dosage schedule is variable, but it is reported that full-dose BCG therapy including maintenance at three-weekly instillations up to three years (at scheduled intervals) is far superior in preventing BC recurrence (and progression to MIBC) compared to reduced-dose regimen in high-risk patients (125) *(see Table 1.3 for progression/ recurrence rates at 1-, 5- and 10- year intervals).* Notwithstanding, <5% of patients may develop serious side-effects from BCG therapy (126-128). The side-effects profile includes systemic symptoms with low severity e.g. general malaise, fever, arthralgia to more adverse effects e.g. BCG sepsis, and local effects to the bladder (cystitis, haematuria, symptomatic granulomatous prostatitis). If complications develop, careful manipulation of delivery of reduced dose-BCG is required to increase better tolerance, reduce toxicity, and optimise efficiency.

In some cases, tumours may be unresponsive to BCG or have a BCG-relapsing course with recurrence of HG tumours. Therefore, it is not unreasonable to propose early radical cystectomy (surgical removal of bladder with surrounding lymph nodes, and prostate in men, or cervix and uterus in women, RC), to limit disease spread and progression. However, in instances where patients are not suitable candidates for RC

(due to comorbidities, frailty, social factors, or patient's wishes!), bladder preservation approaches such as intravesical and/ or systemic chemotherapy, chemoradiotherapy or use of trimodality treatments are offered. Intravesical treatment using combination chemotherapy with Gemcitabine-Docetaxel (Gem-doce) is an alternative to BCG-unresponsive NMIBC and has shown to mitigate resistance and enhance therapeutic efficacy with better recurrence-free survival than BCG (HR-0.56; 95%CI: 0.34-0.92) (129, 130). Other treatment options include neoadjuvant chemotherapy, chemoradiotherapy, or a combination of both, to contain the disease and manage symptoms.

Management guidelines for MIBC includes RC as definitive therapy. For those presenting with organ-confined disease, aggressive treatment with RC can provide 10-year recurrence-free survival of 78-86% for pTa-T1, and 76-85% for pT2-T3a disease (131). However, locally-advanced BC (pT3b-T4) treated with RC alone provides a 10-year recurrence-free survival of 45-61% (131). Therefore, it is recommended that neoadjuvant chemotherapy (NAC) in the form of cisplatin-based chemotherapy is given to improve OS outcome. It is reported that use of trimodality therapy (maximum efficacious TURBT, followed by external beam radiotherapy in combination with chemotherapy) can effectively improve quality of life for these patients with good overall-survival (OS) outcomes comparable to that of MIBC disease undergoing cystectomy alone (60, 132, 133). For those patients that cannot undergo RC as a radical treatment, the BC2001 RCT was the largest study to evaluate bladder-sparing treatment options. It revealed that adding chemotherapy (Fluorouracil + Mitomycin-C) to targeted radiotherapy significantly improves locoregional control of disease compared to radiotherapy alone, with an increase in 5-year OS rates of 48% and 35%,

respectively (HR= 0.82; 95% CI, 0.63 to 1.09; P=0.16) (134, 135). Of recent, propensity matched analyses comparing RC with trimodality therapy have shown no difference in oncological outcomes of patients with MIBC. Instead, the 2024 updated EAU guidelines (and NICE) recommend that all patients should be offered both (radical) options, in the setting of a multidisciplinary approach to make an informed decision, and that trimodality therapy should not be offered to patients only if they are unsuitable for RC (136-138).

In a meta-analysis study, published data were analysed to report an absolute improvement of 9% in disease-free survival at 5 years in favour of NAC with platinum-based combination chemotherapy (139). These results were further consolidated by a large international phase III RCT which showed improved 10-year survival to 36% with platinum-based combination NAC (140).

Patients that are ineligible for platinum-based chemotherapy (due to poor renal function, comorbidities, age, performance status- New York Heart Association class III heart failure, or patient's wishes) can be offered neoadjuvant immunotherapy as an alternative. Atezolizumab (which binds to PD-L1 ligand, immune checkpoint inhibitor) was evaluated in cisplatin-ineligible MIBC patients in a multicentric, single-arm, phase II trial (ABACUS) and appeared to have a 2-year disease-free survival of 85% in patients with a complete pathological response (31%) (141, 142). Whereas the results of the phase II, single-arm trial (PURE-01) using Pembrolizumab (which binds to PD-1 on T-cells) reported 54% complete pathological response and 42% with pathological remission (143). Tissue sampling from initial TURBT can be used for PD-1/ PD-L1 biomarker analysis to identify responding tumours.

Following RC, or initial platinum-based chemotherapy/ RT for patients unfit for RC (or with locally-advanced/ metastatic disease), switch maintenance therapy with avelumab is recommended, after patients have achieved stable disease status. In the JAVELIN Bladder 100 study, the efficacy of Avelumab (following combination platinum-chemotherapy) was recognised as the OS rates improved from 14.3 months (best supportive care only) to 21.4 months (HR- 0.69; 95%CI: 0.56-0.86) (144, 145).

Disease spread beyond the extravesical margins incorporating nodal involvement or distant metastases is more challenging and requires a multidisciplinary approach with a team of oncologists and palliative care. However, immune checkpoint inhibitors (mentioned before) may be changing this paradigm with treatment options to restrict disease and prevent further progression. The use of platinum-based combination chemotherapy is the recommended first-line treatment, with immunotherapy as second-line for platinum-fit patients, and avelumab as maintenance. For platinum-unfit patients, the outcome is generally poor with best supportive care as the available options. Some trials have reported that use of immunotherapy (Enfortumab vedotin and/ or Pembrolizumab) as first-line may be feasible for this patient cohort, but further studies are mandated for conclusive outcomes (146, 147).

1.2.6 Risk stratification, recurrence and progression

Given the complex characteristics of BC, there exists significant overlap and heterogeneity within the broadly divided classes of NMIBC and MIBC with distinct prognoses within the molecular subtypes (148, 149) (Section 1.3). Nearly 35-50% of patients diagnosed with NMIBC will have recurrence of disease and up to 15% can

progress to MIBC (118). Patients need to be stratified categorically considering all clinicopathological parameters, in addition to risk of recurrence and/ or progression, to be able to provide an appropriate management strategy. Multiple prediction tools have been made available to facilitate the decision-making process and tailor management accordingly. The European Organization for Research and Treatment of Cancer (EORTC) devised and developed a prognostic factor risk scoring model (Table 1.3) for prediction of short- and long-term disease recurrence and progression. Patients are scored according to the clinical risk factors (age, gender, prior recurrence), anatomy of tumours (number, size >3cm, concurrent *CIS*) and histopathological grading system adapted from WHO 1973/2021 classification (51, 150, 151). Based on the cumulative score, they are stratified in either of four groups: *Low, Intermediate, High,* or *Very High*, with recommendations for subsequent perioperative treatment following initial TURBT and SSI intravesical chemotherapy.

Table 1.3 | EAU NMIBC Prognostic factor risk groups, with probability of progression based on **WHO2004/2016** and **WHO 1973** classification.

Risk group	Scoring model	Probability of progression & confidence interval (CI)		
		1 year	5 years	10 years
Low	Primary, single, (LG) G1Ta/T1 < 3cm diameter without CIS + age < 70 years	0.06% (0.01%–0.43%)	0.93% (0.49%–1.7%)	3.7% (2.3%–5.9%)
	Primary (LG) G1Ta tumour without CIS + ONE additional clinical risk factors[1]	0.12% (0.02%–0.82%)	0.57% (0.21%–1.5%)	3.0% (1.5%–6.3%)
Intermediate	Patients without CIS not included in either low, high, or very high-risk groups	1.0% (0.50%–2.0%)	4.9% (3.4%–7.0%)	8.5% (5.6%–13%)
		0.65% (0.36%–1.2%)	3.6% (2.7%–4.9%)	7.4% (5.5%–10%)
High	All (HG) G3/T1 without CIS, EXCEPT those in very high-risk group All CIS patients, EXCEPT those in very high-risk group	3.5% (2.4%–5.2%)	9.6% (7.4%–12%)	14% (11%–18%)
	(LG) G2Ta or G1T1, no CIS + all 3 risk factors (HG) G3Ta or (LG) G1/G2 T1, no CIS + ≤2 risk factors G2T1, no CIS + ≤1 risk factor	3.8% (2.6%–5.7%)	11% (8.1%–14%)	14% (10%–19%)
Very High	G2T1 and CIS + ≤2 risk factors (HG) G3Ta and CIS + all 3 risk factors	16% (10%–26%)	40% (29%–54%)	53% (36%–73%)
	(HG) G3T1 and CIS + ≤1 risk factor (HG) G3T1, no CIS + all 3 risk factors	20% (12%–32%)	44% (30%–61%)	59% (39%–79%)

[1] Additional clinical risk factors include age >70 years; multiple papillary tumours; tumour diameter >3cm.

The risk group table is a useful scoring model to predict disease recurrence and/ or progression of BC patients but is not infallible. Molecular markers may be more informative particularly in cases of high-grade tumours/ more aggressive disease and could thus improve accuracy of the currently available predictive risk stratification tools. The challenge persists in accurately identifying HG disease confined to the urothelium (Stage *Ta),* but which has a higher risk of progression/ recurrence, and to provide optimal treatment either in the form of RC or intravesical BCG. Patients harbouring HG-Ta disease have a greater risk of progression (20-25%) to invasive disease than LG NMIBC and therefore should be treated with caution. Similarly, *CIS* may be considered as a precursor to the invasive subtype of BC because of its genotypic parallels with MIBC. Therefore, all patients with *CIS* are classed under High-risk and are suitable candidates for early RC. In some instances, early RC has been suggested as an over-treatment as patients may respond effectively to BCG or other intravesical chemotherapeutic agents.

Genomics-based molecular subtyping may represent a pathway to better risk stratification and more appropriate treatment strategies (see next section).

1.3 Molecular landscape of bladder cancer

The histopathological findings of the bladder cancer cell morphology indicate a temporal relationship with hyperplasia, dysplasia and *CIS*. This gave rise to the dichotomous divergent pathways of bladder tumorigenesis of papillary low-grade and non-papillary high-grade BC (152, 153). Papillary (LG) NMIBC tumours were proposed to arise from epithelial hyperplasia, whereas *CIS* was due to dysplastic cell proliferation, as a precursor to infringement and invasion of lamina propria into muscle (MIBC) to possible dissemination beyond the bladder wall. It was postulated that HG NMIBC tumours may be a consequence of mixed urothelial atypia, dysplasia and hyperplasia giving rise to a papillary design of urothelial cell carcinoma (153).

Over the last few decades research into bladder cancer genetics, epigenetic modifications and gene expression has informed and expanded on the simplistic approaches to classify BC. Following the 2-pathway theory, the earliest available cytogenetics studies established an association with allelic loss of chromosome 9 as a primary event to the development of hyperplasia leading to papillary NMIBC lesions (154, 155). Loss of heterozygosity (LOH) studies and comparative genomic hybridization (CGH) array analyses on chromosome 9 have further revealed multiple loci present on 9p (*TSC1*) and/ or 9q (*CDKN2A*) that can inherently affect the (negative) regulatory mechanisms in the p53/ Rb cell signalling pathways (leading to dysregulation of cell cycle control). *CDKN2A* deletions cause loss of/ insufficient expression of $p16^{INK4a}$ and $p14^{ARF}$ inhibiting cell cycle arrest and promoting cellular proliferation (156). These copy number alterations can provide prognostic information on stage, grade, and outcome via classification into molecular subtypes (157).

As a consequence of activating mutations within the *FGFR3* gene, *Ta* (LG) tumours are characterized as indolent and genetically stable that may recur but are slow-growing with a very low risk of progression to muscle invasion (158, 159). *FGFR3* regulates the *RTK/RAS/RAF/MAPK* and *PI3K/Akt/mTOR* cellular pathways that mediate mutually exclusive events, whereas *PI3K* and *FGFR3* are found to exist concomitantly in at least 25% of NMIBC tumours (160, 161). The more advanced stage (*CIS*/ HG) aggressive cancers are found to have mutations resulting in loss of action of the *TP53* (tumour suppressor) gene (162-164). Concomitant loss of *CDKN2A* and *RB1* with upstream overexpression of cyclin D1 (*CCND1*) results in dysregulation of the *RB1* pathway, which is associated with higher tumour grade, metastasis and poor prognosis (165).

1.3.1 Muscle-invasive bladder cancer

Muscle-invasive bladder cancer is characterized by complex genetic alterations responsible for disease and progression with extensive heterogeneity. Gene expression profiling studies have aided in establishment of molecular subtyping of BC with the earliest work performed by the group at the Lund University (166). Based on whole gene clustering and mutation analyses of 144 BC samples (of all grades and stages) they identified two subtypes: MS1 and MS2, with distinct gene mutations (*FGFR3, PIK3CA, RAS, CDKN2A, TSC1, RB1, E2F3, TP53, MDM2*), transcriptional activity and genomic instability (166). Damrauer and colleagues proposed a similar classification of basal and luminal subtypes of HG MIBC based on cellular differentiation, drawing parallels from epithelial breast carcinoma (167). Basal tumours

were found to have increased expression of *EGFR*, *CCND1*, *E2F3*, *CCNE1* and alterations of the *RB1* pathway, whereas luminal subtypes were typical of *FGFR3* and *TSC1* alterations. *TP53* mutations appeared similar in both subtypes. Additionally, increased high molecular weight keratins and *CD44* were found in basal like tumours (basal urothelial cells) whereas uroplakins and low molecular weight keratins were a property of the umbrella (luminal) cells of the urothelium (167).

Further work by Sjodahl et al from the Lund group on global transcriptomic profiling of 308 BC samples expanded on the two classes (MS1 and MS2) to develop a new classification with 5 subtypes: urobasal A (uroA), genomically unstable (GU), urobasal B (uroB), squamous cell carcinoma like (SCC-like), and infiltrated (which appeared to be a heterogenous mix of other subtypes that was infiltrated by non-tumour cells) (168). GU, SCC-like and infiltrated subtypes make up the MIBC samples, whereas uroA and uroB had a mixed subtype of NMIBC and MIBC samples (169). The proposed classification was aimed to align gene expression levels with immunohistochemistry (based on tumour biopsy features) independent of pathological staging. Subtype groups were identified through clustering according to gene signatures for immune (infiltration) and/ or cell cycle and cell adhesion (proliferation). UroA subtype expressed early cell cycle signature for *KRT13*, *15* and *17*, along with cell adhesion proteins and housed activating mutations for *FGFR3*. In the GU cluster, late cell cycle gene signature was more pronounced than *FGFR3*, with increased *ERBB2* expression and *TP53* mutations. Increased keratin gene expression with squamous differentiation were features of SCC-like subtype. The uroB subtype was considered a variant of uroA as it had similar *FGFR3* mutation-related gene signature along with features of SCC-like subtype. Survival analyses based on this classification showed that uroA was

associated with good prognosis, GU and infiltrated had intermediate, and uroB and SCC-like had poor prognosis (168). Subsequent efforts to identify minimal signatures associated with subtypes prompted exploration for transcription factor binding sites at the promoter regions of the genes involved to reduce the clusters to three distinct groups: urothelial-like (uroA and uroB), genomically unstable and basal/ squamous-like (Ba/Sq) (170). Urothelial-like subclass of tumours expressed high levels of *RXRA*, *PPARG*, *FOXA1*, *GATA3*, and anterior *HOXA* and *HOXB*, whereas GU tumours were enriched by the DREAM complex transcription factors which includes *FOXM1* (activated by *PLK1*), *E2F4* and *MYBL2* that promote proliferation. Besides the DREAM complex, keratinization signature for *STAT3* binding was upregulated for Ba/Sq subtype along with downregulation of the *PPARG/ RXRA* and *FOXA1/ GATA3* axes (170).

As part of The Cancer Genome Atlas (TCGA) project, urothelial tumours were grouped into 5 categories based on gene clustering of RNA expression: luminal-papillary, luminal-infiltrated, luminal, basal/ squamous and neuronal (156). Luminal subgroups had increased expression of uroplakins and urothelial differentiation markers (characteristic of normal umbrella cells), whereas basal/ squamous had high expression of squamous differentiation of basal and stem-like markers. The luminal-papillary subtype was characterized by *FGFR3* mutations, activation, and chimeric fusions with *TACC3*. These papillary structured tumours had low *CIS* content with active sonic-hedgehog signalling. They were found to have slow progression with decreased response rate to NAC (171). The luminal-infiltrated subtype had a low cellular purity with high expression markers for EMT, myofibroblasts and *miR200*. Additionally, this subtype was found to be resistant to cisplatin-based therapy, instead

a better response was noted with atezolizumab (172). Basal/squamous differentiation tumours had higher expression for *PDL1* and *CTLA4*, and basal keratin markers with immune infiltration. This subtype would respond to either cisplatin-based NAC and/ or immune checkpoint immunotherapy (173). The neuronal subtype was found in a low proportion of samples (5%) with neuronal and neuroendocrine expression (*SOX2*, *E2F3*), atypical of the morphological features of neuroendocrine tumours. High expression of uroplakins (*UPK1A*, *UPK2*) and luminal markers were characteristic of the luminal subtype, which accounted for only 6% of the tumours.

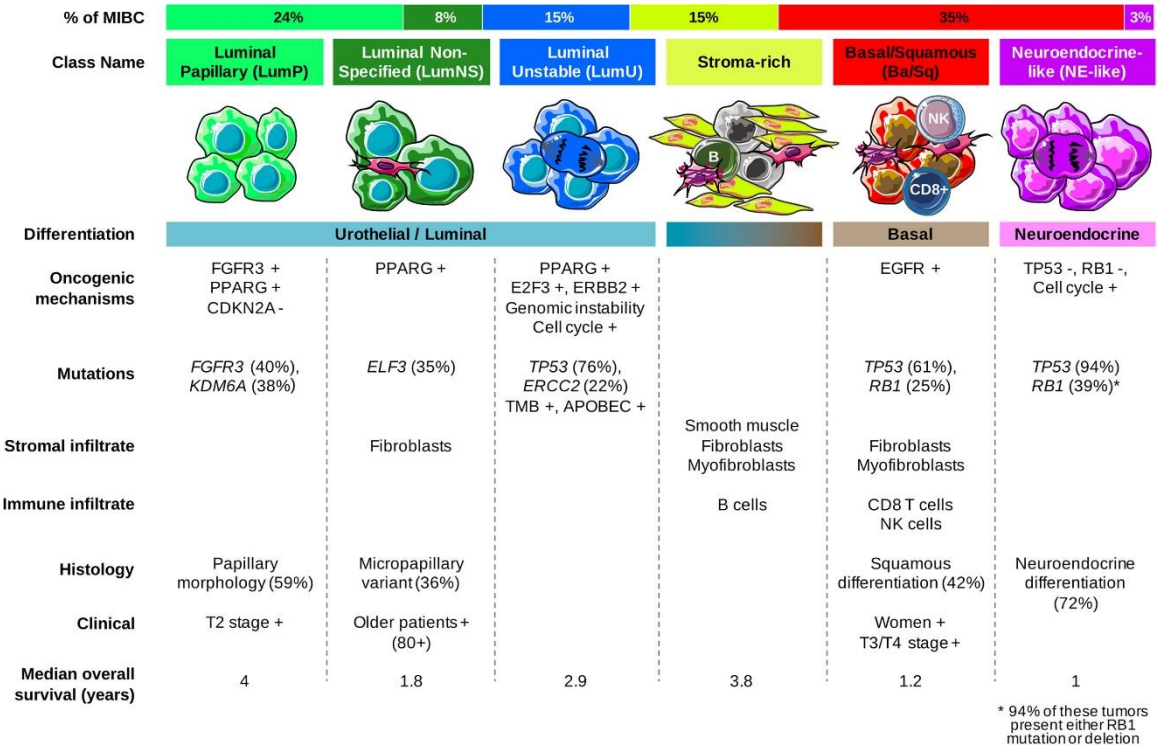
As multiple studies revealed comparable classifications for subtyping BC, Kamoun and colleagues proposed an international consensus classification based on combined analysis of 1750 transcriptomic profiles of MIBC that were assessed by 6 independent models (148) (Figure 1.9). The resulting 6-molecular subtype classification differed in their underlying molecular pathognomonic, histological and clinical features, immune association, and outcomes: luminal papillary (LumP), luminal non-specified (LumNS), luminal unstable (LumU), stroma-rich, basal/squamous (Ba/Sq) and neuroendocrine-like (NE-like). Luminal-classified tumours (LumP, LumNS and LumU) were derived from luminal papillary urothelium with increased expression of uroplakins, keratins and urothelial differentiation markers. These tumours differed in their subsets according to their gene expression signatures, *FGFR3* activity, prognosis, and overall survival. Whilst LumP tumours were found to have more frequent deletions of *CDKN2A* (12%) and were enriched for mutations of *FGFR3* (33%), *KDM6A* (38%) and *STAG2* (23%) (compared to other groups), with good prognosis and overall survival, LumU appeared genetically unstable with high tumour mutational burden, *HER2* amplification and *APOBEC* signature, *PPARG* and *ERCC2* (22%) mutations, increased aggressiveness,

and poor prognosis. These 2 subsets appear to have originated from the papillary and non-papillary pathways and have distinct genomic properties. LumNS exhibited a mix of LumP and LumU characteristics with worst overall prognosis (of the luminal group), having micropapillary variant histological features with genomic alterations of regulatory factors (*ELF3* and *PPARG*) and immune infiltration by fibroblasts. Frequently altered genes in the Ba/Sq subtype included *TP53* (60.8%) and *Rb1*, with associated downregulation of the urothelial differentiation markers (*GATA3*, *FOXA1* and *PPARG*) and upregulation of *KRT5/6*, *14* and *CD44* to exhibit a basal/ squamous (stem cell) differentiation. These tumours also had an increased immune-stromal infiltrate by myofibroblasts, fibroblasts, CD^{8+} T cells and NK cells, with overexpression of epithelial-mesenchymal transition (EMT) factors increasing potential for metastatic dissemination. Increased stromal and immune infiltration typified by smooth muscle cells, fibroblasts, myofibroblasts and B-cells were characteristic of the stroma-rich subtype. NE-like tumours displayed high cell cycle activity with co-existing mutations of *TP53* and *Rb1*, with morphological features of urothelial differentiation exhibiting atypical neuroendocrine markers suggestive of small cell carcinoma (of bladder).

Concerning invasion and metastases, EMT plays an important role in the transformation of malignant BC cells to breach the basement membrane and promote tumour cell invasion and spread. It has been identified in previous studies that downregulation of the cell adhesion molecule (E-cadherin, which maintains normal urothelial cell-cell adhesion) with alternative upregulation of the N- and P-cadherin molecules encourages a more invasive phenotype. Due to the expression of P-cadherin within the basal layer, Ba/Sq tumours usually present with advanced stage

disease, having accommodated the EMT state with activation of *STAT3* (and overexpression of *TWIST*) (174-176).

Figure 1.9 Consensus classification of MIBC molecular subtypes. *The main characteristics of each subtype is detailed in the table below the schematic representation of the tumour cells. Figure reproduced from Kamoun et al under the license CC BY-NC-ND 4.0 (148).*



% of MIBC	24%	8%	15%	15%	35%	3%
Class Name	Luminal Papillary (LumP)	Luminal Non-Specified (LumNS)	Luminal Unstable (LumU)	Stroma-rich	Basal/Squamous (Ba/Sq)	Neuroendocrine-like (NE-like)
Differentiation	Urothelial / Luminal				Basal	Neuroendocrine
Oncogenic mechanisms	FGFR3 + PPARG + CDKN2A -	PPARG +	PPARG + E2F3 +, ERBB2 + Genomic instability Cell cycle +		EGFR +	TP53 -, RB1 -, Cell cycle +
Mutations	*FGFR3* (40%), *KDM6A* (38%)	*ELF3* (35%)	*TP53* (76%), *ERCC2* (22%) TMB +, APOBEC +		*TP53* (61%), *RB1* (25%)	*TP53* (94%) *RB1* (39%)*
Stromal infiltrate		Fibroblasts		Smooth muscle Fibroblasts Myofibroblasts	Fibroblasts Myofibroblasts	
Immune infiltrate				B cells	CD8 T cells NK cells	
Histology	Papillary morphology (59%)	Micropapillary variant (36%)			Squamous differentiation (42%)	Neuroendocrine differentiation (72%)
Clinical	T2 stage +	Older patients + (80+)			Women + T3/T4 stage +	
Median overall survival (years)	4	1.8	2.9	3.8	1.2	1

* 94% of these tumors present either RB1 mutation or deletion

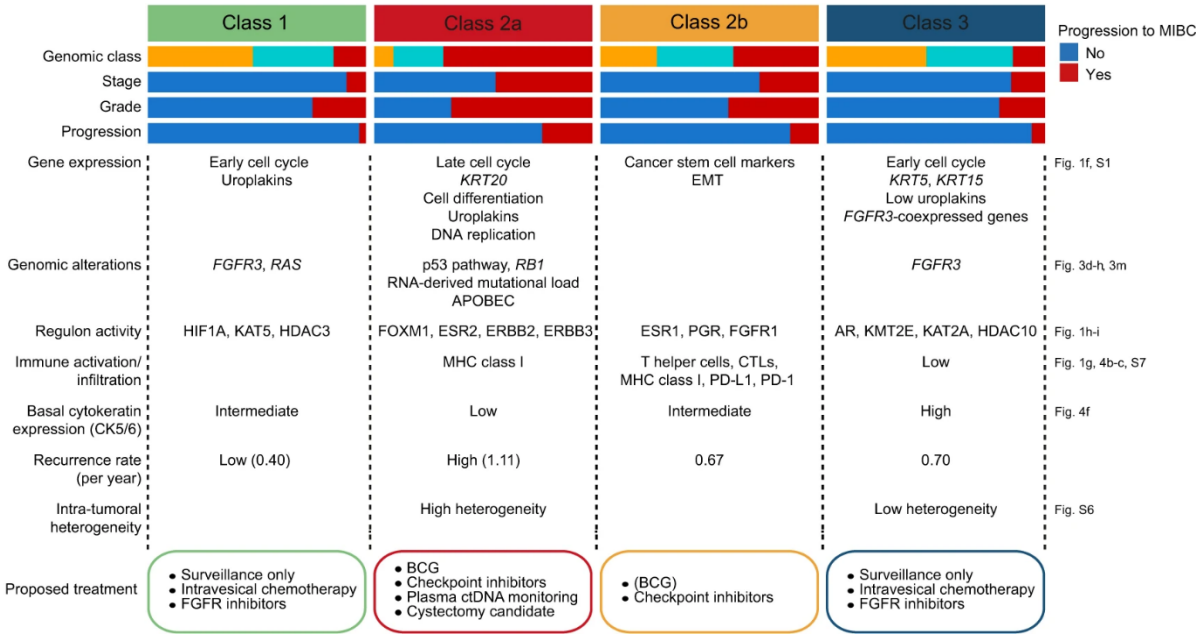
1.3.2 Non-muscle invasive bladder cancer

Like MIBC, NMIBC exhibits intra-tumoral heterogeneity which cannot be explored within the confines of histopathological classification alone. Much of the investigative work has been performed on MIBC or a combination of MIBC and NMIBC tumours (156, 168, 170). More recently, Hedegaard et al. explored 460 early-stage bladder tumours (Ta/T1) to propose a 3-class NMIBC schema based on luminal- and basal-

subtyping with associated outcomes (149). Class I and Class II had luminal-like differentiation but with significant variability within the two classes to create two distinct groups. Associated with good prognosis, low-grade Ta tumours with *FGFR3* mutations and high expression of early cell cycle genes (*CCND1*) and uroplakins were grouped under Class I. Class II was typified by high-risk Ta/ T1 tumours with elevated expression of late cell cycle regulators (*ERBB2*, *ERBB3*, *FOXM1*, *ESR2*), *APOBEC* mutation signature, EMT transition with increased risk of progression to MIBC and evidently had a poor prognosis. Basal-like features (high expression of *KRT5*, *14*, *CD44*) were found in Class III but differentiated them from MIBC tumours due to increased *GATA3* activity (feature of NMIBC tumours only) with co-expression of *FGFR3*. Although these tumours appeared indolent due to decreased cell cycle activity, if progressed they switched to Class II with *CIS* differentiation.

Expanding on the 3-class model, the UROMOL NMIBC classification found two subgroups of Class II viz. 2a and 2b (177) (Figure 1.10). Class IIa was considered high risk due to high tumour mutational burden, increased *APOBEC* gene signature, enhanced DNA replication activity (*p53/ Rb1* pathway) leading to multiple progression events. In comparison, Class IIb displayed elevated cancer stem cell and EMT markers with increased immune infiltration and was associated with a low risk of progression. Further analyses revealed Class I had more methylated gene promoter regions than Class III. Genomic alterations consistent with Class I affected *RAS/MAPK/ERK* pathways and *FGFR3*, making it suitable for therapy with *FGFR* inhibitors.

Figure 1.10 UROMOL 2021 NMIBC molecular classification. *Summarized characteristics of the 4 classes are detailed with associated molecular features and proposed therapeutic options. Figure reproduced from Lindskrog et al under the license CC BY-NC-ND 4.0 (177).*



The available classification systems for MIBC and NMIBC provide a roadmap for molecular subtyping, gene signatures, histopathological subdivisions and risk stratification models to guide clinical management and aid in further research in bladder cancer. However, it must be noted that these classes are based on gene-level activity and any important changes in isoform/ transcript usage may not be evaluated and the true reflection of isoform activity could be missed.

1.4 Transcription, the splicing process, & the mRNA

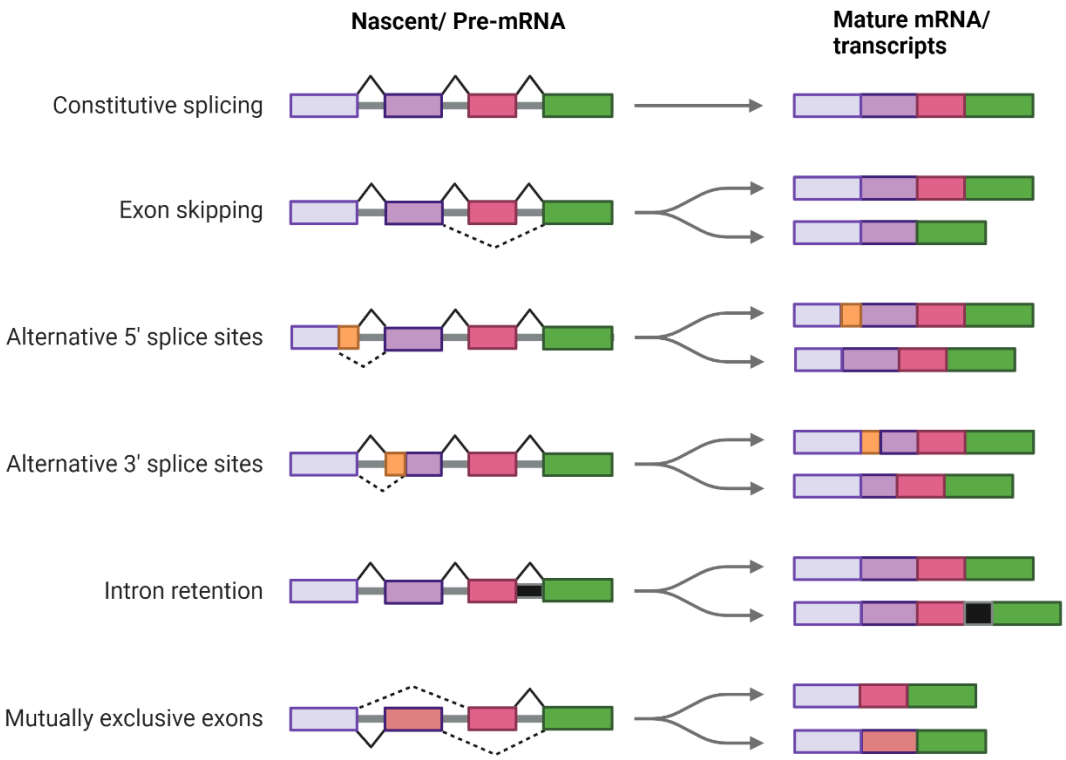
Eukaryotic genes typically contain multiple exons (protein-coding regions) interspersed with introns (non-coding regions excised from mRNA) that are transcribed into pre-mRNA which is processed into messenger RNA (mRNA) sequences (called isoforms) that can be translated into functional proteins. However, not all exons have protein-coding sequences but may have 5' or 3' untranslated regions (UTRs). Transcription occurs within the nucleus where nascent transcripts undergo capping with 7-methylguanosine at the 5' end and polyadenylation at the 3' end to create a polyA (adenosine) tail, to enhance stability of the mRNA and promote export of the mature mRNA from the nucleus to the cytoplasm (178). Constitutive splicing is a tightly regulated (post- or co-) transcriptional event that involves sequential selection and ligation of exons, with removal of introns from pre-mRNA to generate the mature mRNA (179, 180). The part of the DNA segment where transcription is initiated is called the transcription start site (TSS) and is regulated by interaction between *cis*-elements upstream and *trans*-acting factors that promote or repress the process of exon ligation. *Cis*-acting elements (transcription factor binding sites, TFBS) bound to positive *trans*-acting factors (serine/ arginine- rich nuclear phosphoproteins, SR proteins) work as intronic and/ or exonic splicing enhancers (ISE, ESE respectively), whereas negative *trans*-acting factors (such as heterogenous nuclear ribonucleoproteins, hnRNP) include intronic and/ or exonic splicing silencers (ISS, ESS respectively). The binding of transcription factors to specific DNA sequences of TFBS are the primary drivers that regulate gene expression in a tissue-specific manner and determine whether the gene is switched 'on' or 'off'. The spliceosomal machinery, which is a large ribonucleoprotein complex (comprised of 5 small ribonucleoproteins: U1, U2, U4, U5 and U6, and >100

associated proteins), mediates smooth manipulation of the transcriptional process to remove introns from the nuclear RNA (181). The splice donor site which is almost invariably a ‘GU’ sequence at the 5’ end of an intron and the splice acceptor site (‘AG’ sequence) at the 3’ end of an intron are recognized by the spliceosome to undergo splicing from the pre-mRNA to create the mature mRNA. Mutations within the splice donor or acceptor sites can alter the splicing process with generation of aberrant transcripts leading to a diseased state. Variations to the splicing process also exist whereby two exons can be ligated that have originated from two different pre-mRNAs that can result in fusion transcripts. An example is seen with *FGFR3* and *TACC3* in the TCGA cohort which causes constitutive activation of the oncogenic signalling pathway (161).

1.4.1 Alternative and Aberrant Splicing

Aberrant (alternative) splicing (AS) differs from the conventional splicing process whereby there may be exon skipping, use of alternative TSS at 5’ and/ or 3’ sites, mutually exclusive exons, or intron retention, that form the sequence of transcripts (182) (Figure 1.11). These events can modulate the cells to perform multiple operations within tissues thereby causing a diseased or malignant state for the cells. Gene expression can be represented by a combination of transcripts that moderate protein abundance, but for most genes, there is one most dominant transcript (MDT, also termed the *canonical isoform*) that governs whole gene expression. Transcripts with retained introns lead to decreased protein abundance since they are likely to undergo non-sense mediated decay (180).

Figure 1.11 Schema depicting Alternative splicing events. *Figure created using Biorender.com.*



Dysregulation of AS events can produce multiple isoforms that can be responsible in mediating the pathway for carcinogenesis, including BC (183-185). This can occur by way of mutations within the genome itself at the donor or acceptor sites (via substitution, insertion/ deletions- indel, or duplication), or within the spliceosome, TSS, or stop codon. For example, in the instance of colorectal carcinoma, the isoform expression of $VEGF_{165}$ resulting from AS affects angiogenesis promoting growth of tumour whilst its variant isoform $VEGF_{165}b$ is anti-angiogenic (186). Similarly, Pawlak et al found increased expression of low molecular weight isoforms of tropomyosin in BC cell lines, thought to be contributing to increased cellular motility compared to the high molecular weight isoforms that maintain adherence and attachment to the basement membrane (187) (discussed later, Section 1.5.2).

Aberrant splicing may give rise to novel isoforms that are significant in cancer-mediated cell-signalling pathways or disrupt normal protein-protein interactions within cells (184). These isoforms may themselves have oncogenic properties, thereby providing a stimulus for continued growth and progression of the tumour. Use of alternative promoters within an upstream sequence can generate different versions of transcripts with uncontrolled replication in an oncogenic environment and switching between isoforms may be required to maintain survival of the cancer cell. Therefore, it is crucial to accurately quantify isoform expression to determine their role in tumour development.

1.5 Tropomyosins [(1)]

1.5.1 Background and diversity

The cytoskeleton is made up of three types of fibres: large microtubules (tubulins), intermediate filaments (keratin), and (actin) microfilaments. Actin exists as G- actin monomers that bind together to form polymeric F- actin (filamentous). Tropomyosins (Tpm) are actin-binding proteins (ABP) that are encoded by a multigene family (*TPM1-4*). Muscle Tpms form the myofibrillar structural component of tropomyosin-actin microfilaments that recruit myosin to facilitate synchronous movements of the contraction-coupling mechanism (188). Originally thought to exist only in muscle cells (cardiac, skeletal and/ or smooth), Lazarides found that Tpms also play a pivotal role in the cytoskeletal composition of all cells (189). In non-muscle cells, they bind to the actin microfilaments to support the cellular architecture to maintain viability and spatial profiling. Tpms also regulate the interaction of F-actin filaments with other binding proteins to influence their assembly kinetics and moderate key intracellular functions such as vesicle transportation, cellular motility, cytokinesis (190). The Tpm- actin composition involved in these processes have variable organization, cytosol locations and Tpm-ABP interactions.

Diversity in tissue-specific functions and alterations in expression by Tpm isoform selection influence cellular kinetics and differentiation by regulating actin (190-192). Thus, in a tumour environment, *TPM* gene expression can affect cellular polarity, adhesion to the basement membrane and extra-cellular interactions determining the fate of the cancer cell - survival, proliferation, migration and/or metastases.

1.5.2 Structure and functional impact on other actin-binding proteins

The Tpm protein is arranged as a coiled-coil dimer of 2 alpha-helical chains forming a continuous head-to-tail polymer that binds to the grooves of the actin filament through ionic interactions (193). These stable proteins are classed under two categories according to molecular weight found on sodium dodecyl sulfate–polyacrylamide gel electrophoresis SDS-PAGE gels: high-molecular-weight (HMW) weighing 34-40kDa and low-molecular-weight (LMW) with a weight of 28-33kDa (194). HMW Tpms are ~284 amino acids in length that span 7 actin monomers and LMW are ~248 amino acids in length spanning 6 actin monomers (191, 193). Except for the *TPM2* gene (that translates for HMW isoforms only), all other *TPM* genes encode for both LMW and HMW isoforms.

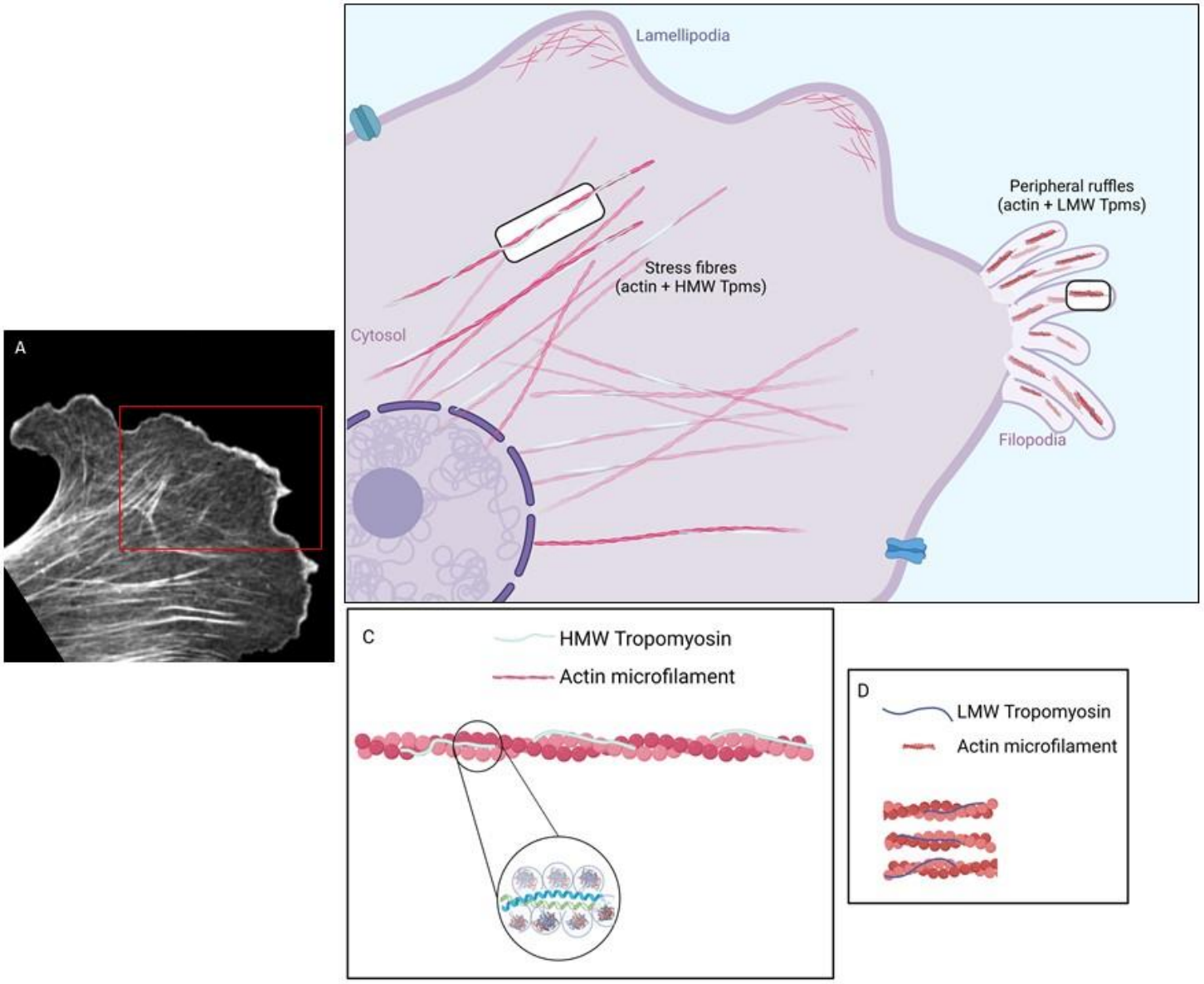
HMW Tpms from *TPM1* and *TPM2* are found in abundance in all cells. HMW isoforms are more stable and have a higher affinity for binding to actin compared to LMW isoforms. This characteristic stabilizes actin at the pointed end of the filament (which is otherwise prone to depolymerization) aiding in the formation of *stress fibres* (195) (Figure 1.12). Stress fibres are cross-linked actin filaments that are higher order contractile cytoskeletal structures. They are found abundantly in endothelial cells, myofibroblasts and epithelial cells (196). HMW isoforms bound to actin create a state of equilibrium between the contractile force and adhesion strength of the stress fibre to maintain a constant length under tension to uphold cellular structure, adhesion and morphogenesis (197).

LMW Tpms are found within diffuse intracellular dot locations, peripheral ruffles (lamellipodia or filopodia) or bundled/ branched stress fibres (Figure 1.12). However,

when not bound to actin microfilaments, LMW Tpms have higher solubility and are less prone to degradation by proteasomes compared to HMW Tpms. It is hypothesised that, in addition to decreased expression, HMW isoforms are more prone to degradation by the proteomes, resulting in an overall decrease in HMW isoforms in transformed cells (191).

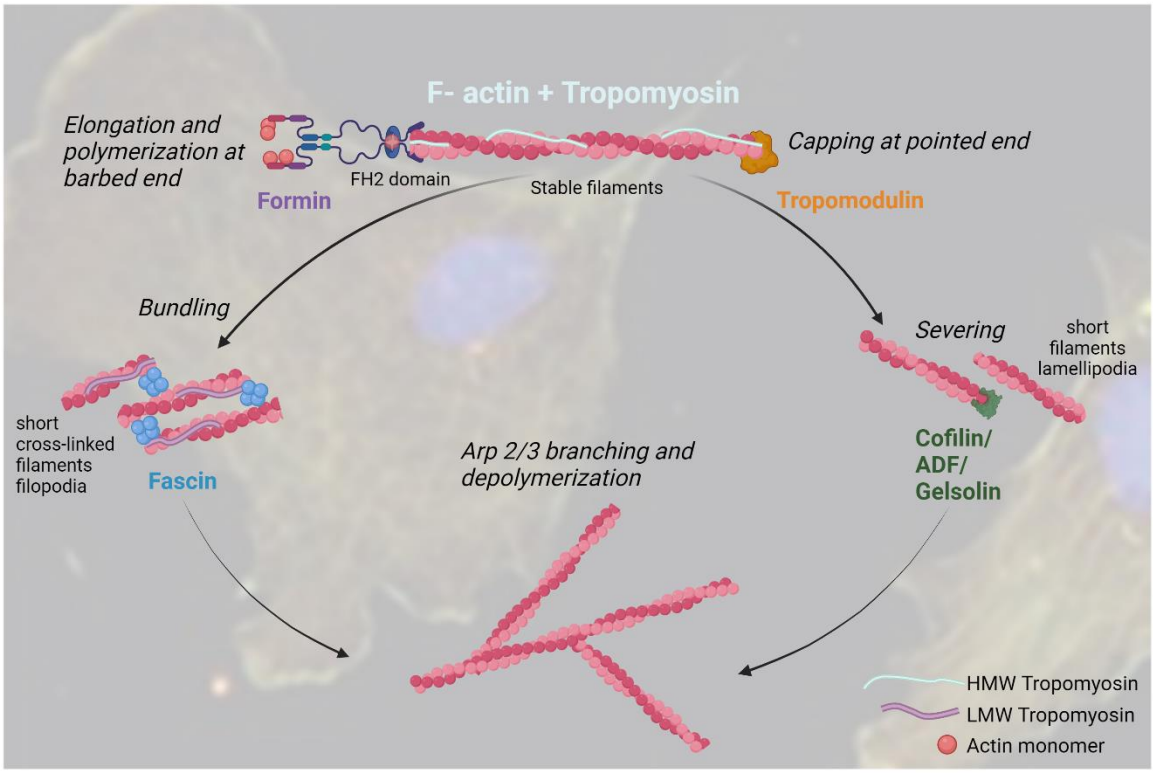
In previous studies on bladder cancer cell lines, it has been found that there is a downregulation of HMW isoforms with alternate upregulation of LMW isoforms that alters cellular configuration to promote carcinogenesis (187). Pawlak et al investigated the expression of tropomyosins in T24 bladder cancer cell line, tissue samples from bladder cancer patients and normal bladder urothelial mucosa. They found HMW isoforms Tpm2.1 and Tpm1.6 were expressed at high levels in normal urothelial cells but markedly reduced in tumour samples, with reduction levels of at least 61% and 55%, respectively. Additionally, they discovered that inactivation of the mitogen-activated protein kinase signalling pathway (discussed later) in T24 cell line restored expression of these isoforms with concomitant deployment of stress fibres. Dysregulation of Tpm isoform expression appears to be an early event in the morphogenesis of bladder cancer rather than a sequela of oncological events.

Figure 1.12 Tropomyosin-actin composition in non-muscle cells. *A) Staining of stress fibres (actin bundles) within lamellipodia of fibroblasts, image adapted from (198). B) Schematic representation of Tpm-actin placement in non-muscle cell. Stress fibres are stable HMW Tpms bound to actin, that span the length of the cell to maintain its shape and attachment to basement membrane. HMW Tpms have much slower cellular dynamics and only slight changes to structural conformity are permitted. Towards the periphery of the cell, multiple Arp-2/3 branched actin polymers are situated within membrane ruffles. Filopodia have smaller units of actin polymers bound to LMW Tpms (D), these allow for increased motility and enable cellular migration. Here it is depicting the leading edge of the cell that extends and protrudes the lamellipodium to invade into ECM. C) HMW Tropomyosin are 40 nm-long coiled coil dimers that span the length of 7 actin monomers and are bound by ionic interactions to sit within the grooves of the actin microfilament.*


A
Lamellipodia
Peripheral ruffles
(actin + LMW Tpms)
Stress fibres
(actin + HMW Tpms)
Cytosol
Filopodia
C
HMW Tropomyosin
Actin microfilament
D
LMW Tropomyosin
Actin microfilament

Tpms are the key regulatory proteins that oversee actin filament composition within cells and monitor interactions of actin to other ABPs (Figure 1.13). At the pointed end of the Tpm-bound actin filament, capping proteins such as Tropomodulin prevent further elongation and depolymerization (199). Tropomodulin has two binding sites at its N-terminus that binds to at least 2 Tpms whereas the C-terminus is bound to actin. Nucleation at the barbed end of the actin filament is facilitated by formin via their FH2 domains that assist in polymerization. Actin de-polymerising factor (ADF)/ cofilin proteins compete with HMW Tpms for binding sites on actin filaments, which destabilize the polymers and promotes Arp2/3-dependent branching (200). Activation of the Arp2/3 cascade causes disassembly of the stress fibres and promotes enucleation, shortening the actin filaments (200). Formation of lamellipodia aids in cellular motility and migration and is achieved by competitive binding of Gelsolin to actin via severing (201, 202). Conversely, Fascin is another actin-binding protein that acts synergistically with Tpms to bundle actin and induce formation of filopodia to enhance cellular motility (203, 204).

Figure 1.13 Tropomyosin-Actin regulation and dynamics with other ABPs. *Tpms bind to F- actin to stabilize these filaments. Tropomodulin prevents elongation and depolymerization by capping whereas Formin assists in addition of further ATP-actin monomers at the barbed end for de novo polymerization. Cofilin, gelsolin and/ or ADF cause severing of the actin filaments creating shorter filaments for Arp2/3 branching found in lamellipodia. Fascin acts as an actin-bundling protein to cross-link short filaments to form migratory structures at the cell periphery and contribute to migration and metastasis.*



1.5.3 Splice-form variants and Isoforms

With use of different promoter sequences through alternative splicing, the 4 *TPM* genes (viz. α-alpha, β-beta, γ-gamma, and δ-delta or *TPM1, TPM2, TPM3* and *TPM4*, respectively) produce around 100 transcripts, that encode for more than 40 (protein-coding) isoforms (Figure 1.14). The gene loci are found at 4 different autosomal sites with no interlinkage assembly with the following coordinates: *TPM1:* 15q22, *TPM2:*

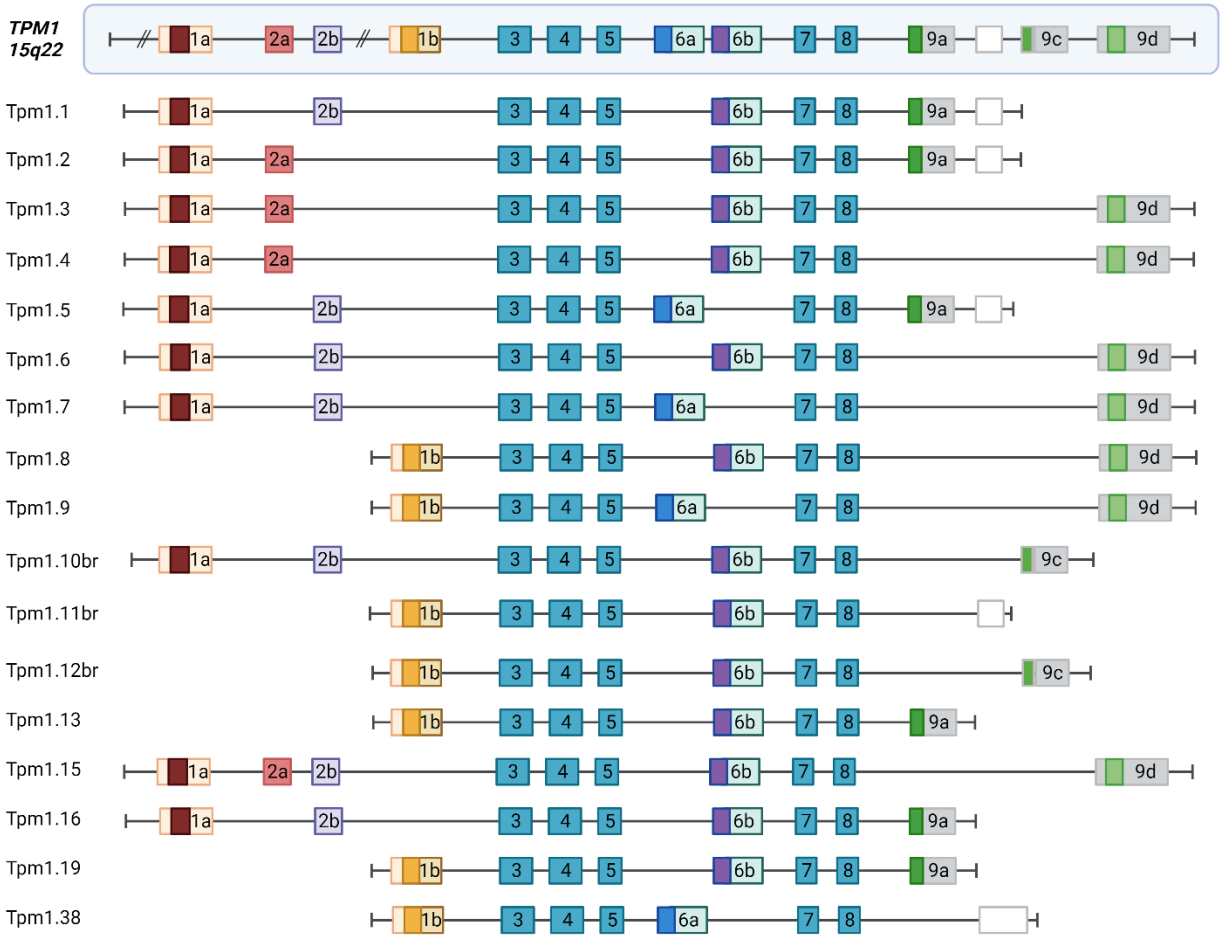
9p13, *TPM3:* 1q22 and *TPM4:* 19p13. All Tpm splice-forms have a transcription start site (TSS) placed in exon 1. HMW isoforms are generated by usage of exon 1a (5' end) and differential splicing of exon 2 (2a or 2b). LMW isoforms have an initial (alternative) promoter sequence in exon 1b. Alternative splice sites present in exon 9 at the 3' end (9a, 9b, 9c or 9d) along with end processing with poly A addition or untranslated regions (3'UTR) encode for variable C-termini. Furthermore, mutually exclusive usage of exon 6 (6a or 6b) encodes for differing peptide segments between both HMW and LMW isoforms.

The different isoforms are found in multiple intracellular locations with distinct functionality in concordance with filamentous actin. However, not all splice-forms (mRNAs) are translated into functional proteins, and a few *TPM* isoforms reported contain retained introns, and others are likely destined for non-sense mediated decay (NMD). The protein-coding (and other non-protein coding with retained introns, and NMD) isoforms have been curated through RNA-sequencing or supported by expressed sequence tags and are found to have tissue-specific functions (205).

Historically, many different names have been allocated to the isoforms investigated, thereby causing discrepancy and incoherence within studies. The 4 *TPM* genes have also been addressed as different isoforms (incorrectly) prior to development of the human genome project, thought to be arising from one gene. Thus, NCBI created a standardised dictionary and approved nomenclature for the complex isoform family of Tropomyosins (206). Table 1.4 shows the old nomenclature used for *TPM* isoforms with the updated names, corresponding ENSEMBL IDs and RefSeq IDs. The data have been collated from NCBI and Ensembl genome browser.

Figure 1.14 Schematic representation of Human TPM genes with their (protein-coding) isoforms and exons. HMW isoforms have a 5' TSS at exon 1a (red) with mutually exclusive usage of 2a or 2b. LMW isoforms have TSS at exon 1b (yellow-shorter length). Exons 3, 4, 5, 7 and 8 are the same across all the isoforms. The C-terminal is dependent on mutually exclusive usage of 9a, 9c, or 9d with a poly A addition (longer length) or UTR (white).

**br= found in neurological tissues only, // TSS-Transcription start site*



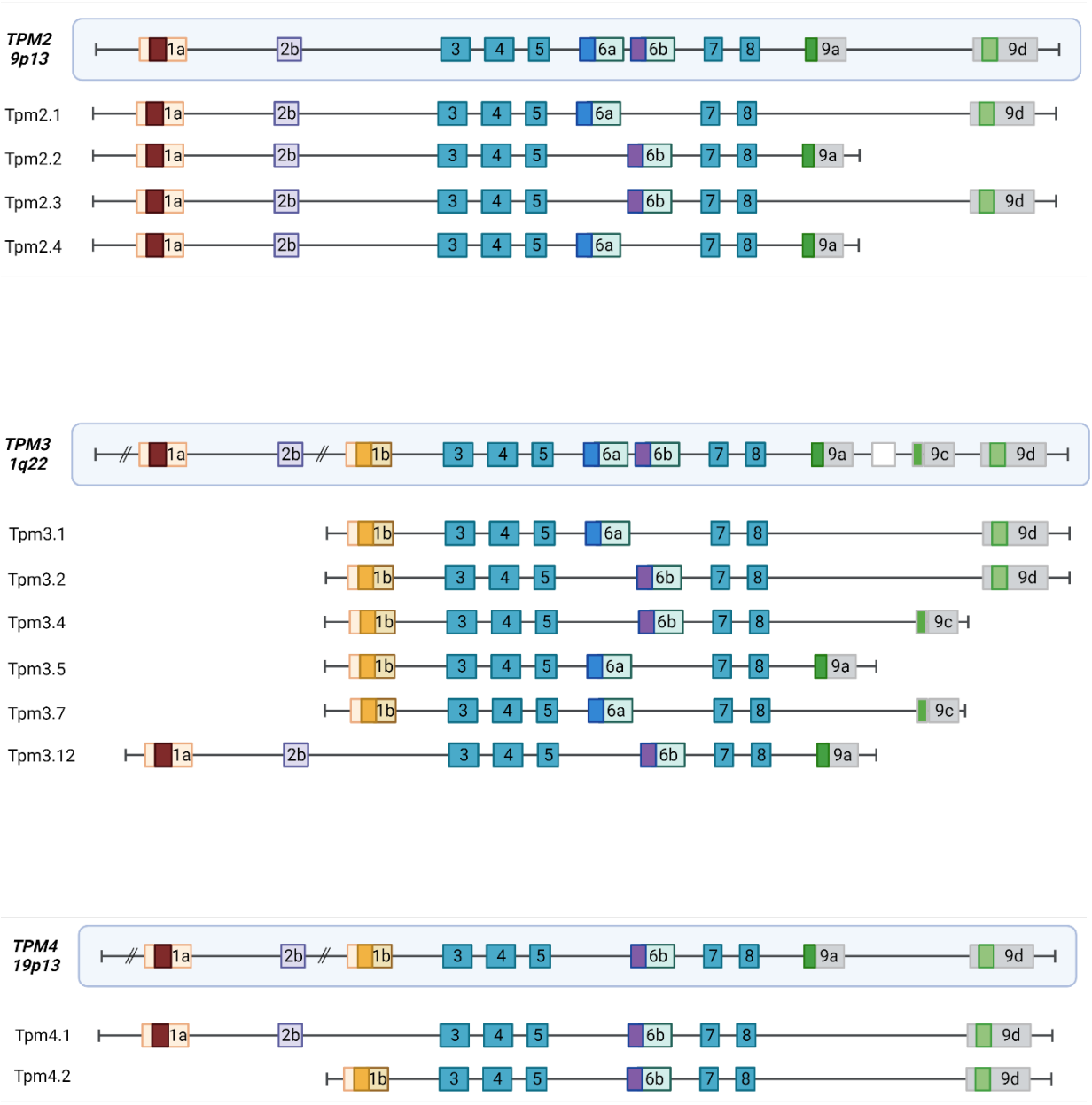

TPM2
9p13
1a
2b
3
4
5
6a
6b
7
8
9a
9d
Tpm2.1
Tpm2.2
Tpm2.3
Tpm2.4
TPM3
1q22
1b
9c
Tpm3.1
Tpm3.2
Tpm3.4
Tpm3.5
Tpm3.7
Tpm3.12
TPM4
19p13
Tpm4.1
Tpm4.2

Table 1.4 | *TPM* genes and their isoforms with corresponding Ensembl and RefSeq IDs

TPM Gene	Isoform (Old name)	Isoform name	ENSEMBL ID	RefSeq ID
1	Tmskα1	1.1	ENST00000403994	NM_001018005.2
1	Tmskα1-1	1.2	ENST00000561266	NM_001301244.2
1	Tm6(Tmsmα)	1.3	ENST00000559397	NM_001018020.2
1	Tmsmα-1	1.4	ENST00000267996	NM_001018007.2
1	Tm3-1	1.5	ENST00000288398	NM_000366.6
1	Tm2	1.6	ENST00000559556	NM_001018004.2
1	Tm3	1.7	ENST00000358278	NM_001018006.2
1	Tm5a	1.8		NM_001301289.2
1	Tm5b	1.9	ENST00000404484	NM_001330344.2
1	TmBr1	1.10br		NM_001365776.1
1	TmBr2	1.11br	ENST00000651590	NM_001365780.1
1	N/A	1.12	ENST00000317516	NM_001330351.2
1	TmBr3	1.12br	ENST00000334895	NM_001018008.2
1	N/A	1.13	ENST00000559281	NM_001330346.2
1	N/A	1.15	ENST00000357980	NM_001365778.1
1	N/A	1.16	ENST00000560970	NM_001365779.1
1	N/A	1.19	ENST00000558910	NM_001365782.1
1	N/A	1.38	ENST00000560959	NM_001407340.1
2	Tm1	2.1	ENST00000378292	NM_213674.1
2	Tm skβ	2.2	ENST00000645482	NM_003289.4
2	N/A	2.3	ENST00000329305	NM_001301226.2
2	N/A	2.4	ENST00000647435	NM_001301227.2
3	Tm5NM1	3.1	ENST00000368533	NM_153649.4
3	Tm5NM2	3.2	ENST00000330188	NM_001043351.2
3	Tm5NM4	3.4	ENST00000323144	NM_001043353.2
3	Tm5NM5	3.5	ENST00000328159	NM_001278189.2
3	Tm5NM7	3.7	ENST00000368531	NM_001043352.2
3	Tm skα-slow (αsTm)	3.12	ENST00000651641	NM_152263.4
4	Tm4HMW	4.1	ENST00000646974	NM_001145160.2
4	Tm4	4.2	ENST00000643579	NM_003290.3

1.6 Tropomyosin profiling in bladder cancer

1.6.1 Tropomyosin expression and signalling pathways

Carcinogenic transformation of cells occurs as a result of activation of proto-oncogenes (often by activating mutations) or loss of tumour suppressor genes (by non-sense mutations). In turn, activation of cancer signalling pathways can alter expression of wild-type genes. In the case of tropomyosins, this is seen as phenotypic (cytoskeletal structural) changes in transformed cells subsequent to selective *TPM* isoform expression. Tpm isoforms control the actin-based cellular mechanics by interactions with actin itself, as well as other ABPs, thereby affecting important cellular pathways for apoptosis, cytokinesis, cell motility and prevention of tumour growth *in vivo* (190, 195, 207, 208). Decreased expression of HMW Tpms destabilizes stress fibre formation and disrupts focal adhesions. These structural changes are unable to maintain cellular polarity, but instead promote anchorage-independent growth and tumorigenicity within altered cells. An upregulation/ maintained expression of LMW Tpm isoforms facilitates intracellular short transport of organelles and functional proteins for continued ‘cycling’ (replication process).

Growth receptor signalling is activated by ligand binding to tyrosine kinase- associated cell surface receptors that transmit signals to the nucleus for cell cycle progression and proliferation. Phosphorylation of the Mitogen-activated protein kinases (*MAPK*) -also known as extracellular signal-receptor kinase (*ERK*), initiates the cascade of complex enzymatic activation of (at least 3) kinases to translocate signals to the nucleus to increase transcription and expression of genes necessary for cell differentiation, proliferation and growth (209). Activation of the *MAPK* signalling pathway is known to be the primary driver of cell growth in cancer (210, 211). Overexpression of epidermal

growth factor receptor (*EGFR*) and fibroblast growth factor receptor (*FGFR*) have been extensively studied and found to be associated with the development of aggressive MIBC disease (with increased mortality) and papillary Ta tumours, respectively (158-160, 212-214). Mutations within GTPases/ Ras proteins (*HRAS*/ *KRAS*) have downstream effects on *MAPK* pathway inducing limitless replicative potential and cell survival in BC cells (159, 215). LMW Tpm isoforms assist the synthesis 'S-phase' of the (transformed) cell cycle by *ERK*, via nuclear translocation of phosphorylated-*ERK*, as observed in transgenic mouse embryo fibroblasts harbouring increased replicative capability (216). Conversely, tumour suppressor genes like the superfamily of transforming growth factor – beta (*TGF-β*) cytokines act as growth-inhibitory, pro-apoptotic proteins in normal epithelial cells (217, 218). By way of *Smad* signalling and *p38MAPK*, *TGF-*β upregulates expression of ABPs particularly of HMW Tpms (219). The oncogenic *Ras/MEK/ERK/MAPK* (*RAS/MAPK* pathway) antagonizes normal functioning of *TGF-*β. Inhibition of *TGF-β* induction of HMW Tpms (which is crucial for stress fibre formation) is significantly associated with an invasive phenotype of BC cells (219). Paradoxically, as BC cells continue to grow, the immune-suppressive properties and increased angiogenesis triggered by *TGF-β* favour continued growth and tumour expansion, thereby prompting EMT transition and changes within the tumour microenvironment (219).

1.6.2 'Hallmarks of cancer' applied to urothelial carcinogenesis

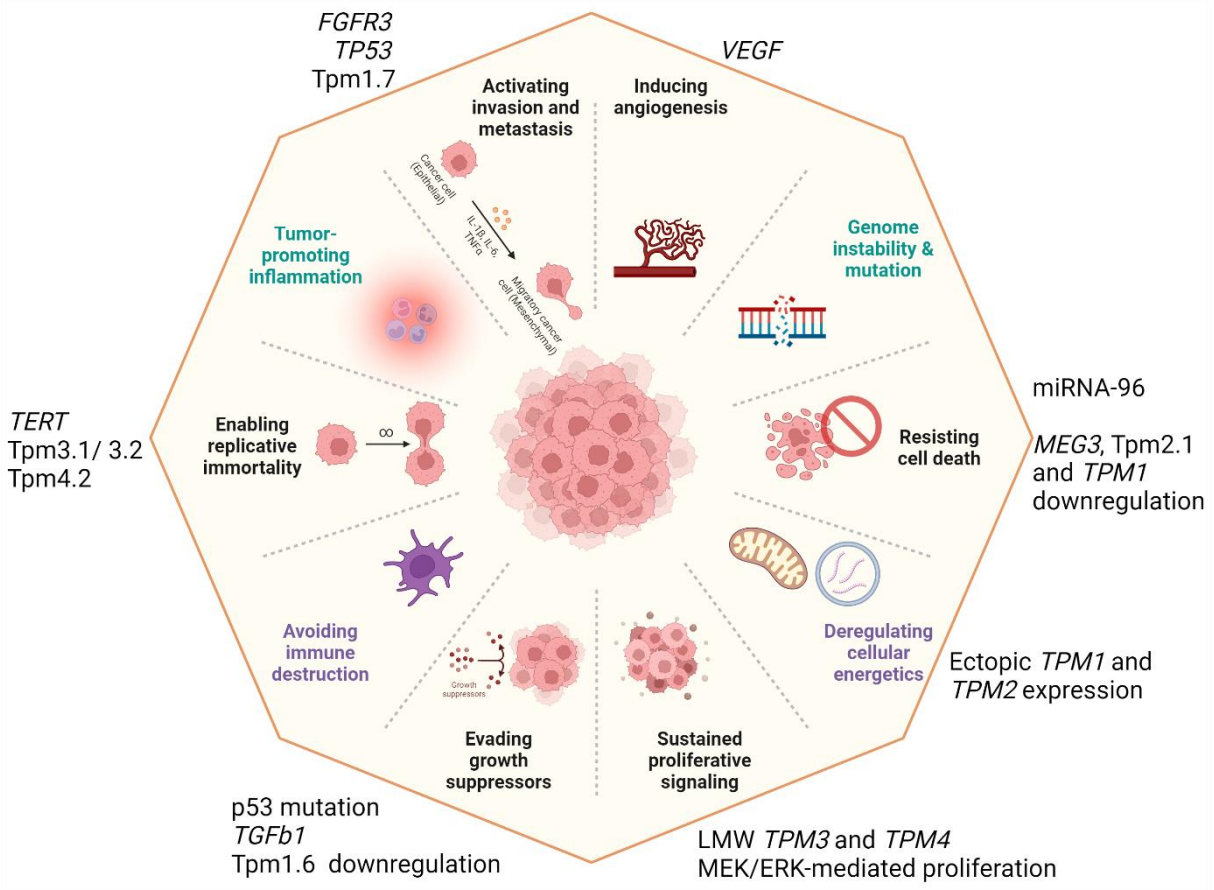
Hanahan and Weinberg first published the 'hallmarks of cancer' detailing the general underlying principles involved in the complexities of carcinogenesis with subsequent

revisions and expansion on the tumour microenvironment (220, 221). Applying the defined hallmarks in the context of BC development, the contribution, and effects of Tpm isoform usage are described (Figure 1.15).

Typically, there is an abundance of HMW isoforms of *TPM1* and *TPM2* genes found in normal urothelial cells (187, 195, 222). HMW isoforms are recruited by actin filaments to maintain the structural integrity, attachment to the basement membrane or the extracellular matrix (ECM) and conserve the normal functioning of cells. Epithelial zonula adherens is responsible for maintaining cell-cell adhesion. Additionally, cell adhesion and cytokinesis are established by Tpm2.1, which is involved in the process of anoikis (programmed cell death induced by detachment from the ECM) (223).

DNA damage can lead to aberrant splicing and dysregulation of isoform expression. *TPM1*, acting in synergy with long non-coding RNA *MEG3* (Maternally Expressed Gene 3) arrests the cells in the G0/G1 phase (with dampening of G2/M) not allowing passage through the restriction point by activation of checkpoint inhibitor. Furthermore, it activates Bcl-2 associated X (*BaX*) and *caspase-3* to promote apoptosis (with downregulation of Bcl2 and cyclin D1) in BC cells (224).

Figure 1.15 Hallmarks of cancer in context of TPMs and other genes. *Alternative splicing events can utilize different Tpm isoforms (and some other genes as shown) which can control the process involved in ongoing bladder tumorigenesis.*



It is hypothesized that the early events in the tumorigenesis of BC are triggered by oncogenic stimulus with concurrently reduced expression of HMW isoforms Tpm1.6 and Tpm2.1 (187). With ongoing mutagenic insults by either oncogenes, carcinogens, oncogenic DNA and/or RNA viruses, immunity to antigrowth signals is achieved. The transformed cells no longer require external stimuli but can propagate their own

intracellular signalling pathways. These cascades push the cell into G2/M phase with notable downregulation of *TPM1* and long non-coding RNA *MEG3* (Maternally expressed gene 3), and upregulation of *miR-96* (microRNA 96) driven by *TGF-b*. Masquerading as an antiapoptotic gene targeting *FOXQ1*, miR-96 favours BC cell survival inducing EMT and promoting the progression of disease (225). Suppression of *TPM1* and *TPM2* genes is a hallmark of EMT: loss of polarity and disassembly of cell junctions as a precursor to disease progression (224).

In BC, the *MAPK* pathway promotes limitless mitogenic capability, resulting in uncontrolled cellular proliferation, growth, and invasion (211, 226, 227). *MAPK* activation induces selective *TPM* isoform expression (through alternative promoter usage) and maintains the ability of transformed cells to express high levels of LMW Tpms (Tpm3.1 and Tpm4.2). While it is possible to ascertain the localization of these isoforms and its effects on the cell, the processes by which this is achieved is still not understood. Tpm3.1 is an important LMW isoform associated with multiple cancers (such as esophageal and ovarian cancer (228, 229)) and is involved in a multitude of cellular processes. Tpm3.1 regulates the uptake of glucose within cells via the insulin-stimulated GLUT4 transporter in the plasma membrane - glucose is needed for the energy source supply to the cancer cells, generation of bundled stress fibres, and *MEK/ERK*-mediated regulation of cell proliferation (230). Tpm3.1 also blocks the recognition of actin filaments by MyoIb (unconventional myosin Ib, a key intracellular organelle transporter), disrupting the cellular structure (231). In addition, degradation of matrix and invasion into the ECM is propagated by the formation of lamellipodia, resulting from the overexpression of LMW isoforms (223, 232). These transformed cells eventually invade and escape from the ECM to metastasize to distant sites.

Interestingly, through the ectopic expression of cell-stability isoforms Tpm1.6 and Tpm2.1 in cancerous cells, sustainability is regained in a simulated normal-like morphology, which promotes anchorage-independent growth.

Tpms in conjunction with other members of the cytoskeletal machinery, can mediate differential and alternative cellular capabilities for malignant transformation and invasion into (and escape from) the ECM. Furthermore, due to their many isoforms and tissue- and compartment-specific regulation, Tpms have the potential to support malignant transformation simply by aberrant regulation. This aberrant regulation may be mediated through alternative TFBS usage, as indicated by Savil et al using a non-cancerous cell-line (233).

1.6.3 Isoform switching due to alternative promoter usage

Isoform switching in carcinoma is observed by aberrant splicing events in transformed cells. The switch is due to dysregulation of *TPM* isoform expression with contemporaneous selection by alternative transcriptional activation. *TPM1* promoter sites indicate shared presence of TFBS with Retinoid X receptor alpha (*RxRα*), whereas *TPM3* and *TPM4* have TFBS for *E2F1* on the promoter regions of exons encoding for LMW isoforms. *RxRα* gene expression profiles were found in the TCGA cohort belonging to the squamous cell carcinoma-like (Lund classification, TCGA: neuronal) and urobasal subtypes (Lund classification, TCGA: basal-squamous), as discussed in previous sections (148, 234). *E2F1* is a transcription factor, known to perform as an oncogene as well as tumour-suppressor, playing a crucial role in cell cycle moderation (235).

1.7 Isoform and gene quantification

1.7.1 Next-generation sequencing

Massively parallel sequencing (Next generation sequencing, NGS) has been employed to investigate the genome diversity via high-throughput sequencing of nucleic acids including DNA and/ or RNA fragments. NGS is available for a diverse range of sequencing modalities, with each technique catered to answer specific research questions, a few have been outlined below.

RNA sequencing involves sequencing the entire bulk of the RNA within a sample, which may include mRNA, microRNA, and long non-coding RNA. Available sequencing technologies such as Illumina and Ion torrent have been used for sequencing 200-400 bp length of RNA, whereas Sanger and Nanopore can sequence lengths of up to 1000 and 10-100k bp, respectively. These tools have enabled quantification of isoforms (for gene expression) and highlight any notable splice events resulting in novel isoforms, which cannot be achieved with microarray.

1.7.2 Gene vs Isoform level quantification of RNA expression

RNA expression can be quantified at the gene level which refers to overall gene expression measured as cumulative abundance of transcript isoforms, or by isoform expression whereby individual transcripts are analysed to understand which isoforms are present. Gene level quantification does not aptly capture variability in transcript abundance and reflects overall activity of a gene indicated by alterations in regulation, gene expression or copy number variations. However, while often informed by known exon boundaries/splice-sites, mapping reads to a reference genome does not require

prior knowledge of TSS or splice junctions thus allowing for recognition of novel splicing events (or transcripts). Quantification of transcript isoforms can be challenging given the short lengths of the sequencing reads used to infer the whole and accurate isoform sequence, varying accuracy of read alignment, overlapping regions with increased numbers of exons, and minimal counts of reads mapped to lowly expressed isoforms of the same gene. Rapidly evolving computational tools are being crafted to overcome these difficulties and increase precision of short-read RNA-seq analysis. The available bioinformatics applications use either of the two forms of read-alignment techniques for isoform detection and quantification (see Table 1.5):

1. True/ Splice-awareness alignment where the short-read sequences are mapped directly to the reference genome considering splice/ exon-intron junctions.
2. Pseudo-alignment method which uses a reference transcriptome (instead of the genome) to map reads judged on compatibility with transcripts purely for the purpose of ‘counting’ to quantify expression.

Table 1.5 | Differences between two principal options of read-alignment techniques

True alignment	Pseudo-alignment
Mapped directly to reference genome	Mapped to reference transcriptome
Splice-aware due to full-alignment	Reads assigned to sets of transcripts based on compatibility
Requires large amount of memory	Less memory
Relatively slow and time-consuming	Increased speed

Can identify novel isoforms due to full alignment, often described as 'local events' (Exon skipping, Intron retention etc.)	Equivalent or surplus accuracy for counting number of reads to quantify expression, however, cannot identify novel isoforms

There are several splice-aware tools available to align short read RNA-seq to a reference genome, such as:

- **STAR (Spliced Transcripts Alignment to a Reference)**: The STAR Aligner efficiently aligns short read RNA seq to a reference genome via rapid retrieval of potential alignment locations. It utilizes a two-step approach of 'seed-and-stitch' where the reads are first mapped to the longest matching genome sequence (called maximal mappable prefix, MMP) with subsequent refinement of (local) alignment positions accounting for any splice junctions (based on canonical splice signals). The 'seeds' are clustered and stitched together based on a scoring achieved via best alignment for the read. Any reads that do not align to the exon-exon boundaries are identified as alternative splice sites (with novel splice junctions) (236).
- **HISAT (Hierarchical Indexing for Spliced Alignment of Transcripts):** HiSAT is another software tool that utilizes a reference genome for alignment of reads with focus on identifying splice junctions. It employs an efficient hierarchical indexing algorithm that enables handling of large RNA datasets: a whole genome index is used for initial alignment with extensions based on local indexes (237).
- **TopHat:** TopHat is a spliced-aligner tool that is capable of aligning reads to a reference genome, capable of aligning variable lengths of RNA sequences allowing for

indels and *de novo* assembly to identify novel splice sites (238). It has largely been superseded by Bowtie due to longer time requirements for processing and decreased accuracy.

Following alignment of reads using the above-mentioned aligner tools, data can be analysed for differential gene expression using tools like DESeq2, edgeR and limma (239-241).

Contrastingly, pseudo-alignment methods offer a fast alignment-free approach to estimate abundance of short-read RNA-seq reads referenced to a known transcriptome and is generally well-suited for large-scale RNA-seq datasets. Two most common software tools available are *Salmon* and *Kallisto* (discussed in Section 3.1).

In this project, the aim was for fast and reliable quantification of isoform expression of all the *known* transcript isoforms of Tropomyosins. The isoform repertoire of Tpms is well described allowing for a dictionary-based approach as by pseudoalignment to the transcript annotation. Isoform reconstruction methods (*de novo* assembly as used by StringTie) need sufficient expression and/ or deep sequencing of the transcriptome with adequate reads covering the length of the isoforms. The original StringTie publication reports around 55% sensitivity to discover isoforms in simulated data and 80% sensitivity is reached only if full-length coverage of the isoforms is achieved (242). Albeit this situation may improve with longer read lengths and deeper sequencing, lowly expressed isoforms can be missed in a *de novo* assembly approach. Splice-aware aligners utilize the genomic sequence and are not designed to quantify transcript isoforms. The resulting read counts can be tracked to individual exons, but complete isoform expression quantification requires additional mathematical steps

(provided in pseudo-alignment approaches) since exons can be shared between isoforms (or across genes of the same family). Internal simulation experiments by the Arnold lab (unpublished) using simulated data showed good performance of Salmon as a quantification method.

1.8 Hypothesis

Dysregulation of transcriptional events generate specific Tropomyosin isoforms that contribute to cancer cell survival and propagation in urothelial bladder cancer. Cancer-associated Tpm isoforms may thus represent promising diagnostic or prognostic biomarkers detectable in urine.

1.9 Aims

This project aims to study the usage of Tropomyosin isoforms and their expression in urothelial bladder cancer, and their potential as diagnostic urinary biomarkers.

Transcriptomic profiling through the process of RNA-sequencing (RNA-seq) has become the preferred method for evaluation of gene expression due to its high read accuracy and ability to assess isoform abundance as a reliable quantitative tool. This study aims to evaluate whether RNA-seq can identify the expression of the complex gene family of *TPMs* and their isoforms. Additionally, it seeks to evaluate if the same isoforms can be inferred using different RNA sequencing platforms and how they compare with experimental validation of *in silico* results.

Furthermore, the potential isoforms of interest are investigated to ascertain whether they can be utilized in the development of a biomarker assay for accurate diagnosis and risk stratification.

Chapter 2

MATERIALS AND METHODS

2.1 Bioinformatics

2.1.1 Acquisition of RNA-seq datasets

2.1.1.1 BCPP

Existing short-read RNA-seq (SR-seq) data was obtained from archived Bladder Cancer Prognosis Program (BCPP) NMIBC and MIBC tissue specimens collected from patients between 2005-2011 (Ethical approval: 06/MRE04/65) (see Section 3.2.3.1). Tissues were stored at -80°C; RNA was extracted from 25 mg frozen tissue using RNeasy kits (Qiagen, Hilden, Germany). Sequencing libraries were prepared using TruSeq® Stranded RNA LT kits (Illumina, San Diego, USA) generating ~300 base pair (bp) fragments that were sequenced using Ilumina sequencers: either HiSeq (2x100bp paired-end) or NextSeq (2x75bp paired-end). The library sizes ranged from 14.2 to 46.5 million reads, with mean and median to be 17.2 and 17.3M respectively. A total of 85 samples with paired-end reads were included (primary diagnosis of BC) that had been categorized *a priori* according to grade and stage by histopathology.

2.1.1.2 CCLE

RNA-seq data of 26 urothelial cell lines was obtained as raw reads from the Cancer Cell Line Encyclopaedia (CCLE). Data were collected from National Institute of Health National Cancer Institute Genomic Data Commons (NIH NCI GDC) legacy archive as

BAM files with paired-end alignment reads mapped to the human genome *hg38*. In CCLE, the RNA library was sequenced using Illumina HiSeq 2000 or HiSeq 2500 with sequence coverage of at least 100 million paired 101 nucleotides-long (nt) reads per sample. Pathological information indicated a set of cell lines representing 17 muscle-invasive bladder cancers (5637, 647V, BFTC905, HT1376, J82, JMSU1, KU1919, UBLC1, VMCUB1, 253J, 253J-BV, BC3C, CAL29, HT1197, TCCSuP, UMUC3), 4 non-muscle-invasive bladder cancers (SW1710, SW780, RT112, RT4), 1 squamous cell carcinoma (SCaBER), 1 metastatic bladder cancer (UMUC1), and 3 of unknown histology (639V, Hs172, KMBC2) (243).

2.1.1.3 TCGA

A total of 412 samples of MIBC and 19 adjacent normals paired-end RNA sequencing raw read data and its associated clinicopathological information were sourced from The Cancer Genome Atlas (TCGA) (https://portal.gdc.cancer.gov/), dbGaP project code: phs000178.v11.p8. Originally, RNA was extracted from high-grade muscle invasive tumours and sequenced by Illumina HiSeq resulting in paired 50bp reads (156). The reads were extracted using bioinformatics pipeline generated by Arnold Lab.

2.1.1.4 Non-cancer GEO GSE133624

Non-cancerous urothelium RNA-sequenced data was obtained from Gene Expression Omnibus Accession number: GSE133624. The sample set comprised of 29 adjacent normal surface epithelia taken from patients diagnosed with urothelial bladder cancer

who underwent radial cystectomy at Sun Yat-sen University Cancer Center (SYSUCC, Guangzhou, China). RNA-Seq libraries were constructed using KAPA mRNA Stranded mRNA-Seq Kit (KAPA) and sequenced on the Illumina HiSeq X-Ten platform at Novogene (Tianjin, CA) with paired-end 150bp reads.

2.1.1.5 Non-cancer from University of York

York Biomedical Research Institute, The University of York provided FASTQ files containing RNA-seq data from 8 samples of normal urothelium: 6 freshly isolated ureteric epithelium taken from live patient kidney transplant donors and 2 cultured bladder stromal cells taken from patients with non-cancer/ benign disease. Illumina HiSeq 2500 and Illumina HiSeq 4000 were used for RNA sequencing yielding 101bp and 75bp reads respectively. Reads were mapped to GRCh38 using HISAT2 v2.1.0.

2.1.1.6 Non-cancer from GTEx

7 samples of RNA seq data of benign bladder urothelial tissue were obtained from the Genotype-Tissue Expression online portal. Illumina TruSeq library construction protocol was followed to sequence this dataset.

2.1.2 Computational software

2.1.2.1 BlueBEAR

The data were stored and processed on a Linux high performance computing platform called BlueBEAR (Birmingham Environment for Academic Research) provided to researchers by the University of Birmingham.

2.1.2.2 Python, Biopython

Python v2.7.5 was downloaded from https://www.python.org/downloads/ and has been subsequently upgraded to v3.7.5 which was used as a programming language to script for analyses of data generated from transcriptomic profiling of RNA-seq datasets following quantification with *Salmon*. Where necessary, the add-on package tool, Biopython was used, which is specifically designed for use in computational biology. Sublime text v3.2.2. served as a text editor for encryption of Python code.

2.1.2.3 R, Rstudio

All statistical analyses and data visualization were performed on R v4.1.0, Rstudio v1.4.1106. Add on packages from R/Bioconductor included gplots, ggplot2, pheatmap, heatmap tools, Rcolorbrewer, plyr, stats, survminer, survival.

Immune cell composition was performed in house by Arnold Lab by using ConsensusTME (version 0.0.1.9). This approach utilizes a comprehensive comparison utilizing all available performance tools to enrich cells within the tumour microenvironment, outperforming all cancer-related benchmarks to provide a wide range of cell types (244). To maintain coherence within data analyses, consensusTME is set as our lab standard. Enrichment scores were calculated for 18 different immune cell types by using single-sample GSEA followed by normalization. Normalized enrichment scores were then correlated with *TPM* whole gene and isoform expression.

2.1.2.4 Salmon

Transcript quantification for all the datasets was performed using ***Salmon*** version 0.12.0 (245). It uses a lightweight ultrafast mapping technique to quantify reads, with consideration of sample-specific feature biases that include fragment GC-content and positional bias. This allows for accurate and effective quantification including correct estimates for false-discovery-rate during differential expression analyses. (Discussed more in Section 3.1)

2.1.3 Annotation

Multigene families have very similar sequences because most of the genes have arisen from duplications of the primordial gene. It is not uncommon for the gene families to have multiple exons that generate a vast array of isoforms through AS events (as discussed previously). In addition to alignment method, transcript variant abundance quantification is influenced by multiple factors including the reference model used (246, 247). Therefore, it is fundamental to categorize transcript assembly to a comprehensive and accurate annotation which is constructed using transcripts with high confidence level. Nonetheless available nucleic acid databases contain commonly expressed transcripts, and any novel splice-forms may only be found through RNA sequencing. Multiple annotation databases are available and increasing work is continuing in improving these databases using advancements in high-throughput sequencing (NGS). Nucleic acid database used for our project includes RefSeq/ GenBank- a curated collection of nucleotide sequences constructed by National Centre for Biotechnology Information (NCBI), GENCODE - a project of

ENCyclopaedia Of DNA elements (ENCODE) initially built to hold protein-coding variants only but expanded to include AS variants, non-coding and pseudogenes, and European Molecular Biology Laboratory – European Bioinformatics Institute (EMBL-EBI) platform that houses a large multi-organism genomic database collection (Ensembl) and provides other bioinformatic tools enabling data analyses such as Protein Data Bank (PDB) and Basic Local Alignment Search Tool (BLAST).

RNA-seq analysis in the initial segment of the pipeline used GENCODE v29 and reads were mapped to the human reference transcriptome (Ensembl version GRCh37). In 2020, there was an update to the Ensembl and GENCODE annotation, with new protein-coding transcripts added; some transcripts were reviewed to change biotype from protein-coding to non-sense mediated decay (NMD) or non-coding transcripts, and some transcripts were completely removed from the database. A meticulous review of the changes effecting *TPM* gene expression was carried out comparing the old GRCh37 version to the updated GRCh38 Ensembl. As the majority of splice-forms were affected by the update, it prompted a re-analysis of all the datasets using the *IsoTPM* pipeline with updated GRCh38 version. For instance, the isoform of *TPM2* gene that appeared as MDT in the first analyses, had in fact been removed in the revised version. The GENCODE v36 annotation database is comprehensive as it includes a fusion of transcripts that have been curated manually (through wet lab verification) and computational analyses/ predictions. The current 2020 version has also collaborated with RefSeq and Uniprot to reach a consensus/ convergence annotation that could be used across all platforms.

See Appendix for annotation index including TSL tags for all isoforms.

2.2 Lab materials

2.2.1 Bladder cancer cell lines

Experiments were conducted on established bladder cancer cell lines that were included in the CCLE database for corroborating isoform abundance analysis resulting from *IsoTPM* pipeline. The BC cell lines were grown and cultured using the media as described in Table 2.1.

Table 2.1 | Bladder cancer cell lines (included in CCLE database) and their characteristics, used for experimental validation of in silico analysis.

Cell Line	Grade	Stage	Gender	Growing medium	Comments	Ref.
5637	G2	MIBC	Male	90% RPMI + 10% FBS	produces growth factors: SCF, IL-1, IL-6, G-CSF, GM-CSF	(248)
RT4	G1	NMIBC	Male	90% RPMI + 10% FBS	recurrence tumour	
SW780	G1	NMIBC	Female	90% DMEM + 10% FBS		(249)
T24	G3	MIBC	Female	90% DMEM + 10% FBS		(250)
TCCSUP	G4	MIBC	Female	90% DMEM + 10% FBS		(251)
UM-UC-3	unknown	MIBC	Male	90% DMEM + 10% FBS		(252)
VM-CUB1	G3	MIBC	Male	90% DMEM + 10% FBS	p53 point mutation; cross-contaminated cell line VM-CUB3	

2.2.2 Bladder cancer tissues

2.2.2.1 Archived BCPP samples

The West Midlands Bladder Cancer Prognosis Program gained approval in 2005 and was initiated by the Cancer Research UK Bladder Cancer Group (presently, the Bladder Cancer Research Centre at the University of Birmingham) (253). Up to 1500 patients were recruited from 16 urology centres within the West Midlands region, who were newly diagnosed BC identified through suspicious cystoscopic findings at haematuria clinics, later to have confirmed disease with TURBT histopathological staging using the WHO 1973 Classification. Written informed consent was gained at the point of diagnosis prior to sample collection. Treatment-naïve biosample collection included blood, urine, paraffin-embedded tissue and fresh-frozen tumour tissue. The tumour tissues were obtained by the operating surgeon at the time of TURBT and immediately placed in cryovials and snap-frozen in liquid nitrogen. All specimens were kept in local freezers at -80 °C before being transported periodically to the BCPP storage depot following pathologist's approval. In addition to biobanking, data collected for these patients include epidemiological, clinical and follow up (5-years) data on recurrence and progression.

From the BCPP cohort, we have used 120 tissue samples for our project (Table 2.2).

Table 2.2 | List of BCPP tissue samples used in the experiments performed in this project.

Experiment	*Number of samples*
LR-seq with PacBio	96 samples of BC *(G1pTa n=9, G2pTa n=10, G3pT1 n=54, MIBC n=18, G3pTa n=5)* 16 samples of benign bladder tissue
Mass spectrometry	3 samples of NMIBC
Western blotting	4 samples of MIBC (pT2+)

2.2.2.2 Fresh samples from Wrexham

A collaborative agreement has been forged with the North Wales Clinical Research Centre, reviewed and approved by the North Wales Research Ethics Committee and by the Betsi Cadwaladr University Health Board Research and Development Committee (Ethics REC 18/WA/0046). This is a prospective study where patients have written informed consent, prior to collection of tumour tissue specimens at the point of the first TURBT. Additionally, the surgeon takes a small biopsy of adjacent healthy bladder tissue (for comparison). The samples are stored in the North Wales Clinical Research Centre labs, based in Wrexham, and will be available for up to five years until completion of the study.

cDNA from BC tumour tissues and matched paracancerous tissue biopsy specimens were provided at a concentration of 2.5ng/µl. 11 samples of BC tissue and adjacent paracancerous healthy tissue (with clinicopathological data provided) have been used in our project for LR-seq and qPCR.

2.2.3 Archived HBRC urine samples

The Human Biomaterials Resource Centre acquired ethical approval in April 2010 (North West - Haydock Research Ethics Committee; Ref 20/NW/0001) to operate under the license of the Human Tissue Authority to responsibly source biospecimens from (consented) patients to store and create a large biorepository.

For this project 32 samples of urine supernatants were used, which were collected from patients presenting to the haematuria clinic for investigation of NVH and VH. The sample set included:

Grade/Stage	*Number of samples*
G3T1	5
G3Ta	3
G2Ta	3
G1Ta	2
MIBC	3
Benign bladder condition	16

2.2.4 Antibodies

2.2.4.1 Primary antibodies

The list of anti-tropomyosin monoclonal antibodies used for Western blotting is given in Table 2.3. These antibodies were not isoform-specific but had specificity for *TPM* genes.

Table 2.3 | List of tropomyosin targeted antibodies, where they were sourced from and catalogue numbers, dilution ratios and target specificity.

Name	Dilution	Target gene	Species	Peptide sequence/ exon specificity	Source, cat no.
CG1	1:100	*TPM2*	Mouse monoclonal (IgG)	Specific to Tpm2.1	University of Iowa: AB_10573117
CGbeta6	1:100	*TPM1*	Mouse monoclonal (IgM)	Non-specific	University of Iowa: AB_10573118
CG3	1:250	*TPM3*	Mouse monoclonal (IgM)	AERLQREVEGERRAREQ Exon 1b	University of Iowa: AB_10572305

WD4/9d	1:500	*TPM4*	Rabbit polyclonal	KEENVGLHQTLDQTLNEL Exon 9d	Sigma: AB5449
Anti-*TPM4* antibody N-terminal	1:1000	*TPM4*	Mouse monoclonal (IgM)	N-terminal	ABCAM: ab230880

2.2.4.2 Secondary antibodies

Secondary antibody anti-mouse IgG was used for antibodies CG1, CGbeta6 and CG3, and secondary anti-rabbit IgG was used for *TPM4* antibodies WD4/9d and Anti-*TPM4* N-terminal antibody. Secondary antibodies were sourced from in house stock and Beggs' Lab at Institute of Cancer & Genomic Sciences, University of Birmingham.

2.2.4.3 Recombinant protein

Recombinant protein for *TPM2* and *TPM3* were kindly donated by Emeritus Professor Steve Marston, Imperial College London.

2.3 Use of custom python scripts to identify unique primers and peptide sequences for experimental validation

Online platforms available for creating primers are based on genomic data and lack specificity for sequence variation observed in transcripts. We explored the available software tools for generating specific primer sequences including Primer Quest, Primer13, Primer Blast (254, 255). However, due to the high percentage of sequence similarity across Tpm isoforms, no isoform-specific primer sequences could be created. This prompted designing a Python-based script that would search through a transcriptome index to reveal primer sequences that were either specific to-, and/ or shared with other transcript isoforms. Thus, with a combination of sequences we were able to create isoform-specific primers and probes for experimental validation of the target isoforms for qPCR and Taqman-based qPCR.

Additionally, the python script was used to interrogate data resulting from Mass spectrometry (MS) to identify any unique peptide sequences. The amino acid sequences of the peptides identified from MS were input into the script which were matched against a reference dictionary (that included all the protein sequences of Tpm isoforms). The output generated a list of isoforms that belonged to the query sequence, the exon origin and parent *TPM* gene.

2.4 RNA extraction and sequencing

2.4.1 RNA extraction and cDNA synthesis from cell lines and tissues

RNA was extracted using frozen cell pellets ($1x10^7$cells) of BC cell lines using Qiagen RNeasy® mini kit, according to manufacturer's protocol. The RNA was eluted in molecular biological water and quantified using Thermofisher Invitrogen Qubit 4 Fluorometer. The RNA high-sensitivity Qubit assay works with the fluorometer to pick up even very low counts and allows for accurate quick quantification of DNA, RNA or protein from a single sample using a minimal quantity of 1µl/ sample. Complementary DNA (cDNA) was synthesized using ProtoScript II First Strand cDNA Synthesis Kit (New England BioLabs) following the manufacturer's protocol using 500ng RNA, with a final cDNA volume of 20µl.

2-5 mm^3 tissue material from archived BCPP samples were freshly cut on ice blocks and macerated. RNA extraction was performed using Qiagen RNeasy® mini kit as per manufacturer's protocol. Eluted RNA was quantified using Thermofisher Invitrogen Qubit 4 Fluorometer. cDNA was synthesized from ~200ng RNA using Superscript VILO kit as per manufacturer's protocol, with a final volume cDNA of 20 µl.

2.4.2 PCR and agarose gel electrophoresis

The resulting cDNA was verified for presence of *TPM* isoforms using reverse transcriptase polymerase chain reaction (RT-PCR). Isoform-specific primers were designed using a combination of custom-made scripts and Primer Blast (254), sequences shown in Table 2.4. Custom scripts were able to determine if the primers were truly isoform-specific or would also amplify other isoforms that shared similar

exon sequences. PCR reaction volumes (total = 20µl) contained 5µl of cDNA, 5µl of Forward and Reverse Primer pair mix (at a final concentration of 0.5µM each) and 10µl of Phusion Flash High-Fidelity Mastermix. Using a thermocycler, PCR program was set for initial denaturation phase at 98°C for 30 seconds, followed by 35 cycles of [98°C for 15 seconds (annealing), 60°C for 30 seconds (extension), and 72°C for 30 seconds (extension)], with a final extension phase at 72°C for 5 minutes prior to infinite hold at 4°C. PCR products were analysed by electrophoresis in Standard 1.5% agarose gels, followed by visualization with UV transillumination.

Table 2.4 | *TPM* Isoforms and Primer sequences used for experimental analysis (RT-PCR).

Isoform	Forward Primer	Exon	Reverse Primer	Exon
TM1A	TGGACGCCATCAAGAAGAAGA		TGGCATGAGCCACTTTCTCT	
TM1B	TGCTGAAGCTCGACAAGGAG		TCCAGTAAAGTCTGATCCAGCA	
Tpm 1.4	AAGGAGAAGTTGCTGCGGG	2A	TCCAGTAAAGTCTGATCCAGCA	9D
Tpm 1.7	TGAGCTGGTGTCACTGCAAA	2B	TGGCATGAGCCACTTTCTCT	8/9D
Tpm 2.1	GAAGGAGGCCCAGGAGAAAC		CACAGACCTCTCGGCAAACT	
Tpm 2.3	TGATGGCCTCAGAGGAGGAG		CACAGACCTCTCGGCAAACT	
Tpm 3.2	CTGGGATCACCACCATCGAG		TCCTCTGTGCGTTCCAAGTC	
Tpm 4.2	GCGGTGAAACGCAAGATCCA	1B	ACGTTCTCTTCTTTGGCCTGGG	9D
Tpm 4.1	GGAGGCGGATAAGAAAGCCG	1A	TTTGGCCTGGGCAAGTTTCT	9D

2.4.3 Purification of DNA from PCR products and Sanger sequencing

PCR products were purified with Agencourt AMpure XP beads (bead ratio 1.8:1) prior to being sequenced to confirm identity of *TPM* spliceforms. The supernatant containing purified dsDNA was measured using Qubit DNA 4 Fluorometer. 10ng/µl of dsDNA with

0.2 µM primer concentration (Forward or Reverse) was prepared using Eurofins Genomics TubeSeq Service guidelines and sent for Sanger sequencing.

2.4.4 Quantitative PCR and Taqman probe assay development

TPM isoform-specific primer pairs were multiplexed in quantitative PCR (qPCR) assays to determine relative expression of target isoforms to each other. Taqman probes were designed, in addition to primer pairs (Table 2.5), to increase accuracy of target amplicons and to allow multiplexing of isoforms. A qPCR total reaction volume (20µl) contained 10µl Taqman mastermix, 4µl primers at 0.4µM concentration, 2µl probes at 0.25µM concentration and cDNA (20ng from tissue samples and 2ng from urine samples). The 7500 Fast qPCR thermocycler was used and the PCR reaction cycle was programmed as follows: initial holding stage (2 cycles) of 50°C for 2 mins and 95°C for 10 minutes, followed by cycling stage of 40 cycles (95°C for 15 seconds, 60°C for 60 seconds), prior to infinite hold at 4°C.

Table 2.4 | *TPM* isoform Primer and Probe sequences for Taqman qPCR

Isoform	Forward	Reverse
Tpm1.4	AAGGAGAAGTTGCTGCGGG	GTCGGAAAGGACCTTGATCTC
	Probe exon 5/6b-	AAG GCA AAT GTG CCG AGC TTG AAG
Tpm1.7	TGAGCTGGTGTCACTGCAAA	GTCGGAAAGGACCTTGATCTC
	Probe exon 6a-	AAT GGC TGC AGA GGA TAA GTA CTC GC
Tpm4.1	CGGCCACATCACCTTCAG	ACAGAGGACGAGCTGGATAA
	Probe exon 2b-	AAG CTG GAG CTC ACG GAG AAG AAG
Tpm4.2	TGGATGCGTCGGTTGAGG	GGAGGCGGTGAAACGCAAG
	Probe exon 1b/3-	AGA AAG CTG AAG GTG ATG TGG CCG

2.4.5 RNA extraction and quantification from urine

RNA was extracted from 5ml of urine supernatant using the MagMax mirVana RNA extraction kit (Applied Biosystems), as per the manufacturer's instructions. RNA was quantified using the Qubit fluorometer and the Qubit High Sensitivity RNA assay kit (Invitrogen).

2.4.6 Bioanalyzer

Following clean-up of PCR products, the Agilent High sensitivity DNA kit was used with the 2100 Bioanalyzer to assess the size of the amplicons of the PCR products using the manufacturer's protocol (256).

2.4.7 Long-read RNA sequencing library preparation: PCR

Target specific *TPM* isoform primers (Primer sequences shown in Table 2.6) were designed with 5' block and M13 tail (Ward Lab) to prevent attachment of the SMRTBell adapters to unbarcoded amplicons whilst simultaneously having the ability to be used in a multiplexed template.

Table 2.6 | Primer names and their sequences (5' to 3') used for PacBio sequencing, M13 tail and 5' block end not shown.

Primer name	Oligo sequence (5' to 3')
***TPM1-2* EX1a**	TGGACGCCATCAAGAAGAAGA
***TPM3* EX1a**	AGAAAAAGATGCAGATGCTGAAGTT
***TPM4* EX1a**	AGGCCATCAAGAAGAAAATGCAG
***TPM1* EX1b**	GAGTAGCTCGCTGGAGGC
***TPM3* EX1b**	GCTGGGATCACCACCATCGAG
***TPM4* EX1b**	CGGTGAAACGCAAGATCCAG

***TPM1* EX9a**	TATCGTTGAGAGCGTGGTCC
***TPM2* EX9a**	TCCAGTTCCTCGCTAATGGC
***TPM3* EX9a**	AGGTCATGTCATTGAGGGCG
***TPM4* EX9a**	GGTCCAGTTCCTCGCTGATAG
***TPM1* EX9c**	TAATCCTCATTCAGGGCCAGC
***TPM3* EX9c**	GCTCATTAGAAAGCAGGCGG
***TPM1* EX9d**	TCCAGTAAAGTCTGATCCAGCA
***TPM2* EX9d**	CACAGGTTGTTGAGTTCCAGC
***TPM3* EX9d**	TCATTCAGGTCAAGCAGGGTC
***TPM4* EX9d**	TCGTTTAGTGTCTGATCCAGTGT

The first PCR (PCR1) was done to generate amplicons with target-specific M13-tailed primers (Table 2.7) from 30 cycles, using a reaction mix consisting of 10µl of Phusion Flash high fidelity PCR mastermix, 5µl of primers (at a final concentration of 0.15µM), 3µl of molecular biological grade water and 2µl of cDNA.

The PCR products were reamplified in a second step PCR (PCR2) added to barcoded M13 primers to generate dual-indexed amplicons. 1µl of the PCR1 product was added to a reaction mix of 10µl of Phusion Flash high fidelity PCR mastermix, 4.5µl of molecular biological grade water, and 4.5µl of M13 barcoded primers (at a final concentration of 0.25µM) and set at 20 cycles. Generation of barcoded amplicons meant uncomplicated demultiplexing of isoform expression from each (barcoded) sample. The thermocycler protocol for PCR1 and PCR2 is shown in Table 2.7.

Table 2.7 | Thermocycler program settings for preparing SMRTBell library. The table below details the temperature settings and time duration for each phase of the PCR1 and PCR2 cycles.

PCR1- Target amplification			
		Temperature	**Time**
Primary Denaturation		95	30 sec
2 cycles	Denaturation	98	20 sec
	Annealing	60	15 sec
	Extension	72	60 sec
28 cycles	Denaturation	98	20 sec
	Annealing	65	15 sec
	Extension	72	60 sec
Final extension		72	5 min
Hold		4	∞

PCR2- Barcoded M13 primer			
		Temperature	**Time**
Primary Denaturation		95	30 sec
2 cycles	Denaturation	98	20 sec
	Annealing	60	15 sec
	Extension	72	90 sec
18 cycles	Denaturation	98	20 sec
	Annealing	65	15 sec
	Extension	72	90 sec
Final extension		72	5 min
Hold		4	∞

3 separate libraries were created of BCPP tissue samples (2µl each), HBRC urine samples (10µl each) and Wrexham tissue samples (10µl each) to have equal concentration of RNA. 50µl of each of the pooled sample was cleaned using AMpure

XP beads (Beckman Coulter) (as described in section 2.4.3), at a ratio of 1:1 beads to sample. The DNA concentrations of each of the pooled samples was measured using the Qubit fluorometer. Based on the quantification values, the volumes of the samples were adjusted, to add to the final sequencing library (combined single pooled sample), allowing for uniformity of sequencing depth across the samples. The final concentration of the sequencing library had 690 ng of DNA @13.8ng/µl.

The bioinformatics pipeline used to extract and analyse data is mentioned in detail in Chapter 5.

2.5 Protein expression quantification protocols

2.5.1 Mass spectrometry

Pelleted cells were washed with PBS, centrifuged for 5 mins at 13,000xg and resuspended in 50mM Ammonium bicarbonate. The solution was sonicated to lyse cells and heated at 95 °C for 10 minutes, followed by centrifugation for 5 mins at 13,000 xg. Acetone was added to the supernatant (SN) at 4:1 and incubated at -20 °C for one hour. The solution was centrifuged at 13,000xg for 10 mins and the SN discarded. The resulting sediment was dissolved in 50µL of diluted (1X) sample loading buffer (SLB-5X solution of 250 mM TrisHCl, 10% SDS, 30% Glycerol, 10 mM DTT, 0.05% Bromophenol Blue at pH 6.8) and run on SDS-PAGE. The Tpm isoform bands were then stained with Coomassie blue, excised, and prepped for in gel tryptic digestion. Disulfide bonds were reduced using 50mM DTT and alkylated with 100mM iodoacetamide. Gel bands were dehydrated with acetonitrile and then digested using proteomics grade trypsin ((Promega, V5111) at 1:20 dilution in 10% Acetonitrile/40mM Ammonium Bicarbonate) incubated overnight at 37 degrees Celsius. 1% Formic acid was added to tryptic peptides and analysed by LC-MS/MS using Bruker Impact Q-TOF and Orbitrap mass spectrometers. Amino acid sequences of all Tpm isoforms were curated from Ensembl and added to the existing MASCOT database to search the database for relevant isoforms.

2.5.2 Western blotting

Cells were homogenised in 1mL 50mM Tris-Urea buffer and centrifuged at 13,000 rpm x 5 mins. The protein concentrations were determined using BCA protein assay kit.

Approximately 20µg of protein from cell lysates were loaded onto SDS-PAGE. The gel was transferred onto Immunoblot membrane using BioRad Turbo Blot machine. Non-specific binding of protein was blocked with 5% semi-skimmed milk/TBST (10xTris-Buffered Saline, 0.05% Tween 20) for one hour at room temperature. The membranes were probed with primary antibodies (Section 2.2.5) and incubated overnight at 4°C, followed by secondary antibodies with an incubation period of 1-2 hours at room temperature. 4 washes x 15 mins each with TBST were carried out following each antibody incubation. Enhanced Chemiluminescence (ECL) plus western blotting reagent was applied to membranes for chemiluminescence detection and imaged using ChemiDoc CSL (Bio-Rad Labs).

Chapter 3

CHARACTERIZATION of TROPOMYOSIN ISOFORMS *IN SILICO*

3.1 Background

Quantification of isoforms using RNA-seq can be performed in either of two practises: via estimation of transcript abundance through alignment/ pseudo-alignment mapping of reads to a reference transcriptome, or detection and quantification of AS events involving *de novo* assembly. A comprehensive literature search was performed prior to finalizing which software tool would be best suited for the computational analysis (focusing on pseudo-alignment quantification with a reference transcriptome in this part of the project). Currently available computational software for RNA-seq analysis were evaluated for identifying the fastest, most effective, and accurate method to quantify transcriptomic data (Table 3.1).

Table 3.1 | Comparison of quantification methods available for RNA-seq quantification to reference transcriptome and their characteristics **(257-261)**. EM= Expectation-Maximization. Memory is graded from least required (better) = + to most required (worst) = +++.

Quantification method	Reference annotation	Concept	Alignment	Computational speed	Memory	Bias correction	Accuracy
RSEM	Transcriptome	EM algorithm	Alignment-based	Slow	++	No	High
express	Transcriptome	Online EM to reduce convergence speed	Mapping	Moderate	n/a (online)	Yes	Low with lowly expressed transcripts
TIGAR2	Transcriptome	Bayesian inference for longer reads	Alignment-based	Very slow	+++	No	Better for longer reads
Sailfish	Transcriptome	EM and k-mer statistics instead of read alignment	n/a	Fast	++	Yes	High
Salmon	Transcriptome	EM (lightweight) and quasi-mapping	Pseudo-alignment/ quasi-mapping	Ultra-fast	+	Yes	High
Kallisto	Transcriptome	De Bruijn graph	Pseudo-alignment	Fast	+	Yes	High

Salmon employs an Expectation-Maximization (EM) algorithm (increased likelihood of estimation of abundance of transcripts in presence of latent variables) to iterate data via a two-step process accounting for correction of any abundance estimate-bias including GC bias, fragment-length bias and sequence-related positional bias. Its ultrafast alignment-free method utilizes less computational time for analyses of RNA-seq data through quasi-mapping (by itself) or reading through pre-processed sequencing reads (Binary Alignment/Map, BAM files) to quantify splice-forms/ transcripts from an index transcriptome. It can perform quantification on multiple files in a single library by first indexing and then quantifying reads at random, that do not require ordering by position or targets.

Due to its many advantages (including high speed, accuracy, less memory requirement), ***Salmon*** has been used in this project as a standard method to quantify known isoforms from large datasets and forms the initial automated part of the bioinformatics pipeline.

3.2 Aims and objectives

This chapter focuses on the development of the bioinformatics pipeline created to investigate the expression of tropomyosin isoforms in bladder cancer using short-read RNA sequence (SR-seq) data. The use of bladder cancer cell lines (from the CCLE dataset) served as an experimental testbed to identify the isoforms that have potential diagnostic characteristics selective to BC. Following identification of potential isoforms from the BC cell line model, further analyses was performed in SR-seq data of patient samples (BC tissues) of NMIBC and MIBC origin.

Objectives:

- Development of a bioinformatics pipeline that enables differential isoform expression of *TPM* genes.
- Identification of potential isoforms using cell line model.
- Subsequent evaluation of disease-specific isoforms in patient samples of bladder cancer tissues (NMIBC and MIBC).
- Comparison of *TPM* isoform expression profiles of bladder cancer against a control set of normal/ benign urothelial samples.

3.3 Development of *IsoTPM*

A bioinformatics pipeline termed *IsoTPM* (see schema, Figure 3.1) was conceptualised to identify and characterize the expression of *TPM* isoforms generated through transcriptomic profiling of events resulting from AS. Pre-processing data included paired-end SR-seq RNA with high read depth (50-150bp paired-end). The GENCODE annotation database was provided to ***Salmon*** along with the reference transcriptome to align the reads to predicted isoforms. ***Salmon*** quantified the transcript variants (mRNA resulting from AS) in each sample as Transcript Per Million (herein named qTPM, not to be confused with *TPM* gene) and the output was generated in *quant.sf* files. While Salmon estimates the qTPM of isoforms/transcripts accurately even though they share a large part of their sequence, qTPM is generally calculated on the gene-level using the following formula which is given here for illustration purpose:

Step 1 – normalise for gene length by calculating reads per kilobase (RPK)

$$RPK = read\ counts\ \div length\ of\ gene\ (kilobase)$$

Step 2 – normalise for sequencing depth by calculating scaling factor

$$scaling\ factor = \sum RPK \div 1\ x\ 10^6$$

Step 3 – determine qTPM

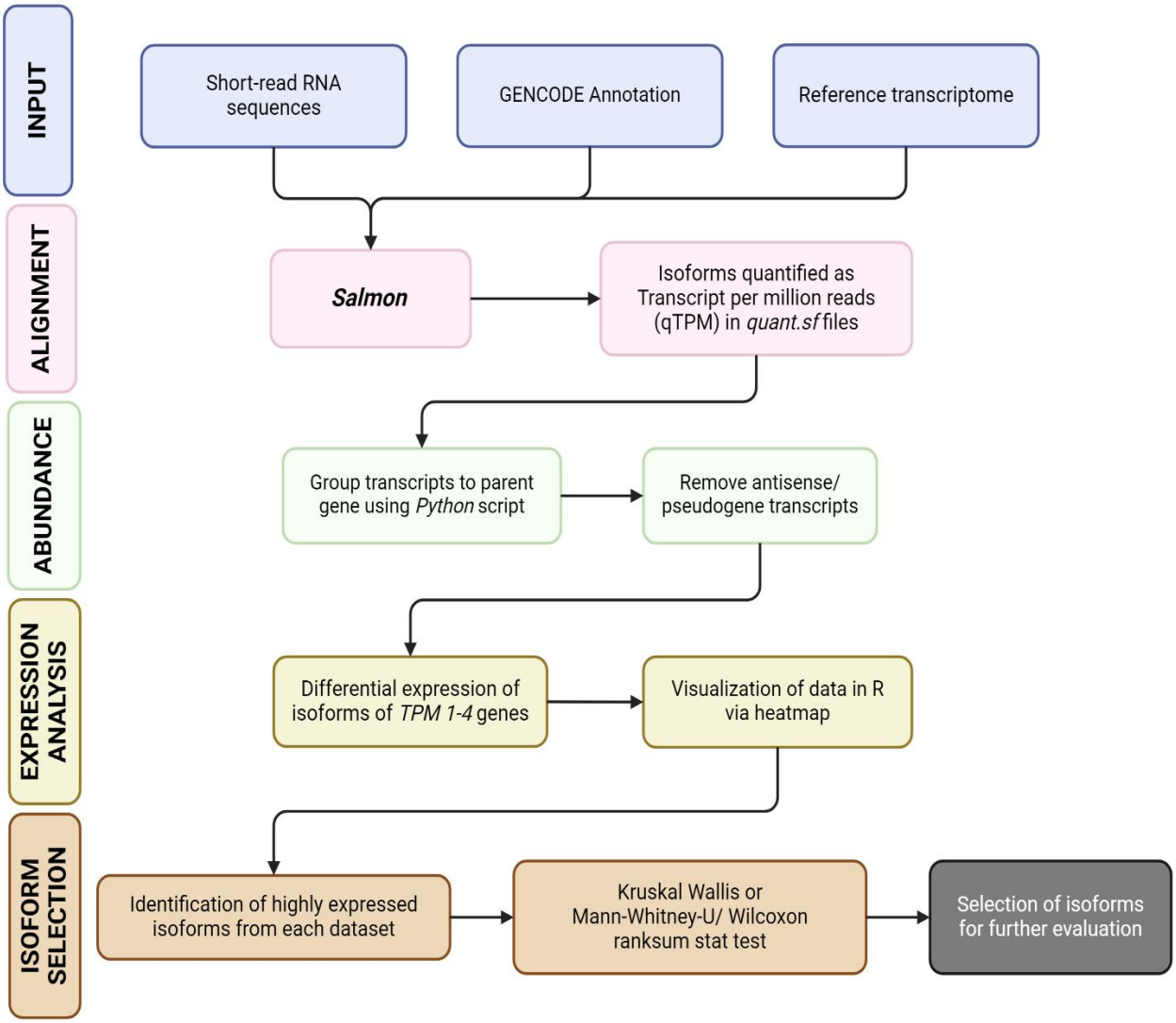
$$qTPM = RPK\ \div scaling\ factor$$

TPM isoform expression data for each Ensembl transcript were mined from RNA quantification results using custom Python scripts and annotated to their parent gene. Anti-sense or pseudogene counts were excluded and not added to qTPM of 'sense' transcripts. Statistical analyses and data visualization were performed using the

programming software- R, with heatmap.plus package (v1.3) and pheatmap package (v1.0.12). Resulting heatmaps were explored to identify the most dominantly expressed transcript (MDT) in each dataset (cancer and non-cancer- benign/ normal/ paracancerous). Relative isoform expression was calculated as a fraction of transcript abundance level over whole gene expression to give '*inclusion level*' or *'percent spliced in'* (PSI) for each transcript (0 to 1, 0= no expression, 1= gene expression wholly by isoform).

Pairwise comparison of transcript isoforms in cancer and non-cancer samples was performed to identify the potential candidates that have increased expression in BC only. Mean relative expression levels were compared across BC and non-cancer datasets using Wilcoxon rank-sum (Mann Whitney U) test. Statistically significant p-value was set at $p<0.001$.

Figure 3.1 Schema- Automated segment of IsoTPM pipeline. *Pre-processing data included paired-end short-read RNA sequences with high read depth, which was input into* ***Salmon****. This generated counts for each transcript isoform as qTPM. Differential expression for TPM isoforms were mined from RNA quantification data using custom Python scripts and analysed using (R) heatmap to identify the most dominantly expressed transcript isoform in each dataset. Following that, Kruskal-Wallis (multiple) or Mann-Whitney-U/ Wilcoxon rank sum test (pairwise) compared non-cancer vs cancer data to select transcript isoform(s) of interest.*



3.3.1 Analysis of datasets for shortlisting disease-specific isoforms

A total of 110 transcripts from all four *TPM* genes were identified and quantified by ***Salmon***. *IsoTPM* revealed variable levels of isoform abundance in all datasets. General observations indicated: For each of the *TPM* genes, apart from 2 or 3 isoforms that contributed to overall gene expression, all other transcript isoforms of the respective *TPM* genes had low to null expression (*TPM1* n= 35/38, *TPM2* n= 8/10, *TPM3* n= 22/27, *TPM4* n= 33/35). The most highly expressed isoforms were found to have a ubiquitous pattern of (high) expression identified in either of the two conditions: cancer, or non-cancer (Figure 3.2 and Table 3.2). Not all identified transcripts were protein-coding isoforms, some isoforms were non-coding with retained intronic regions and others were subjected to degradation (via nonsense mediated decay, NMD).

Expression levels of the MDTs were compared across datasets to test for significance between different sample types (NMIBC, MIBC, and benign/ normal/ paracancerous bladder tissue) using Kruskal-Wallis test for multiple testing and Mann-Whitney U/ Wilcoxon test for pairwise comparison, where applicable. Results were considered statistically significant if p-value ≤0.05*, very significant ≤0.01** and very highly significant if p-value ≤0.001***.

3.4 Isoform curation from short-read RNA seq

3.4.1 Tropomyosin isoform expression in *bladder cancer cell lines*

SR-seq RNA data of 26 bladder cancer cell lines were input into *IsoTPM* to calculate *TPM* isoform expression/ transcript abundance. Individual heatmaps for each of the four *TPM* genes were generated (Figures 3.2a-d: each *TPM* gene is discussed individually in the following pages, and Table 3.2). At least one or two transcripts from each *TPM* gene were seen to have very high expression in all the BC cell lines.

Table 3.2 | List of isoforms with high-expression in SR-seq RNA of BC cell lines from the CCLE dataset. The nomenclature of isoforms is provided alongside ENSEMBL transcripts, with mean expression level of each transcript (in order of most abundant to least abundant), relative isoform usage as a fraction of whole gene expression (0 to 1) ± standard deviation (SD) at 95% confidence intervals (CI). NMD = nonsense mediated decay

Transcript (Ensembl name)	Isoform name	Mean abundance level in qTPM	Relative isoform usage per gene (PSI) ± SD	95% CI	
				Lower	Upper
ENST00000358278	Tpm 1.7	350.7	0.575 ± 0.216	0.488	0.662
ENST00000643579	Tpm 4.2	279.7	0.716 ± 0.251	0.615	0.817
ENST00000509409	Tpm 3NMD1	234.7	0.355 ± 0.037	0.340	0.370
ENST00000330188	Tpm 3.2	172.3	0.273 ± 0.121	0.224	0.322
ENST00000378292	Tpm 2.1	119.6	0.449 ± 0.153	0.388	0.511
ENST00000509601	Tpm 3NMD2	94.7	0.126 ± 0.078	0.095	0.157
ENST00000328159	Tpm 3.5	75.3	0.119 ± 0.031	0.106	0.131
ENST00000329305	Tpm 2.3	71.5	0.349 ± 0.167	0.282	0.416
ENST00000646974	Tpm 4.1	47.9	0.167 ± 0.248	0.067	0.267
ENST00000473036	Tpm 3Intron	38.8	0.062 ± 0.049	0.042	0.081

ENST00000267996	Tpm 1.4	27.0	0.034 ± 0.033	0.020	0.047
ENST00000404484	Tpm 1.9	24.8	0.118 ± 0.136	0.063	0.173

The relative abundance of the isoforms from Table 3.2, and Figures 3.2 and 3.3 were assessed for significance levels using Kruskal-Wallis rank sum test for multiple pairwise-comparison testing across bladder cancer subtypes.

Figure 3.2 TPM gene expression in BC cell lines from the CCLE dataset. *The following heatmaps illustrate the distribution of expression of transcript variants (x-axis) of TPM1 (Fig 3.2a), TPM2 (Fig 3.2b), TPM3 (Fig 3.2c) and TPM4 (Fig 3.2d), in 26 bladder cancer cell lines (y-axis). Each row displays expression levels from the transcript variants of the parent gene in the specified cell line. Individual colour-coded blocks represent level of isoform expression using z-score normalization, transitioning from dark blue (very low abundance, below mean) to bright red (very high abundance, above mean). The cell lines are grouped according to cancer type (MIBC- muscle-invasive bladder cancer, NMIBC- non-muscle invasive bladder cancer, SCC- Squamous cell carcinoma of bladder, Met- Metastatic bladder cancer, NK- not known) and grade using the 1973 WHO Classification (G1 grade 1, G2 grade 2, G3 grade 3, G4 grade 4, NK not known). Note: Squamous cell cancer and Metastatic bladder cancer cell lines are representative of one sample each, with 4 NMIBC cell lines and 17 MIBC cell lines.*

*Figure 3.2a Isoform expression in the CCLE dataset for **TPM1** gene, transcript variants (x-axis), BC cell lines (y-axis). Met= Metastases, NK= Not known*

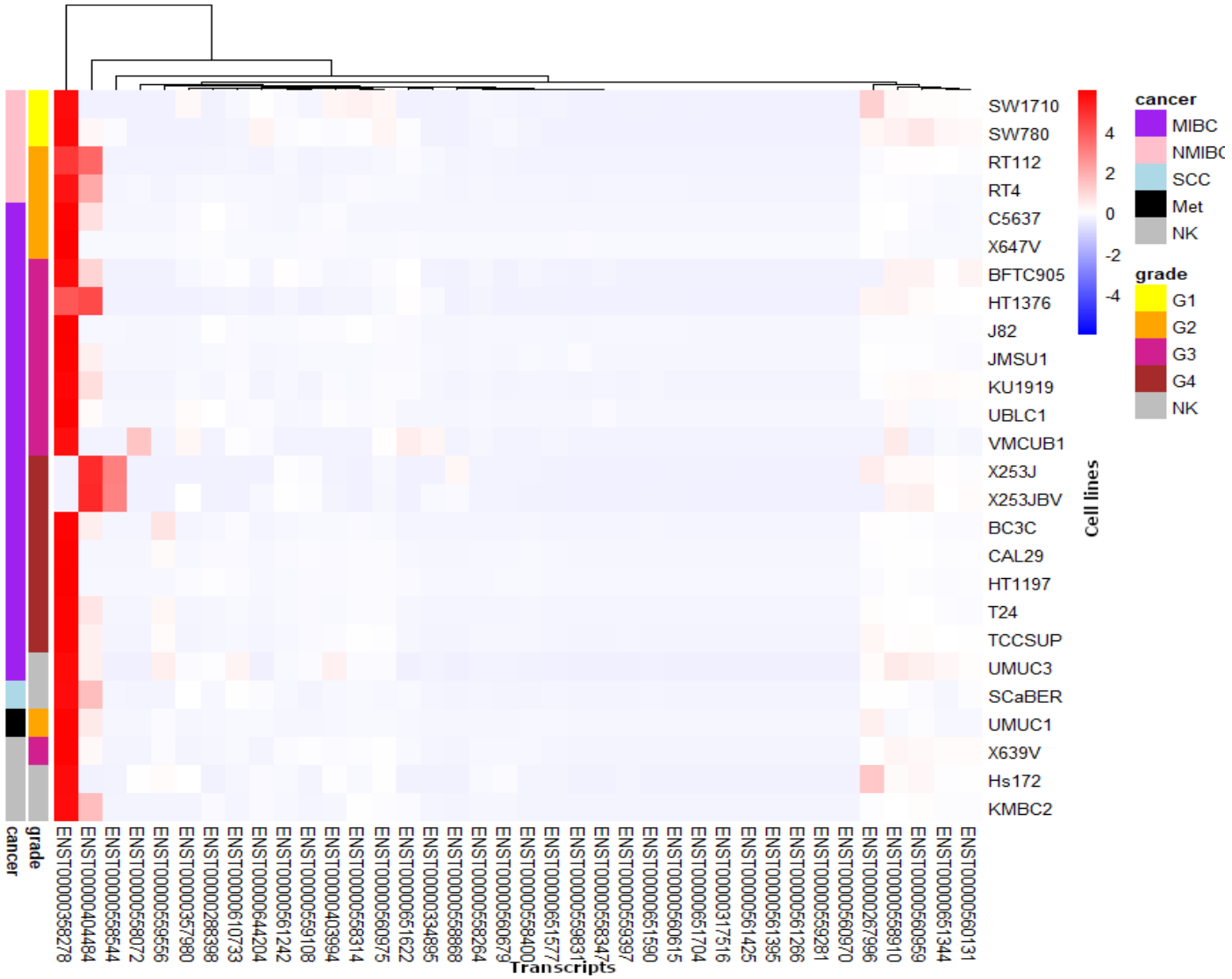


*Figure 3.2b Isoform expression in the CCLE dataset for **TPM2** gene, transcript variants (x-axis), BC cell lines (y-axis). Met= Metastases, NK= Not known*

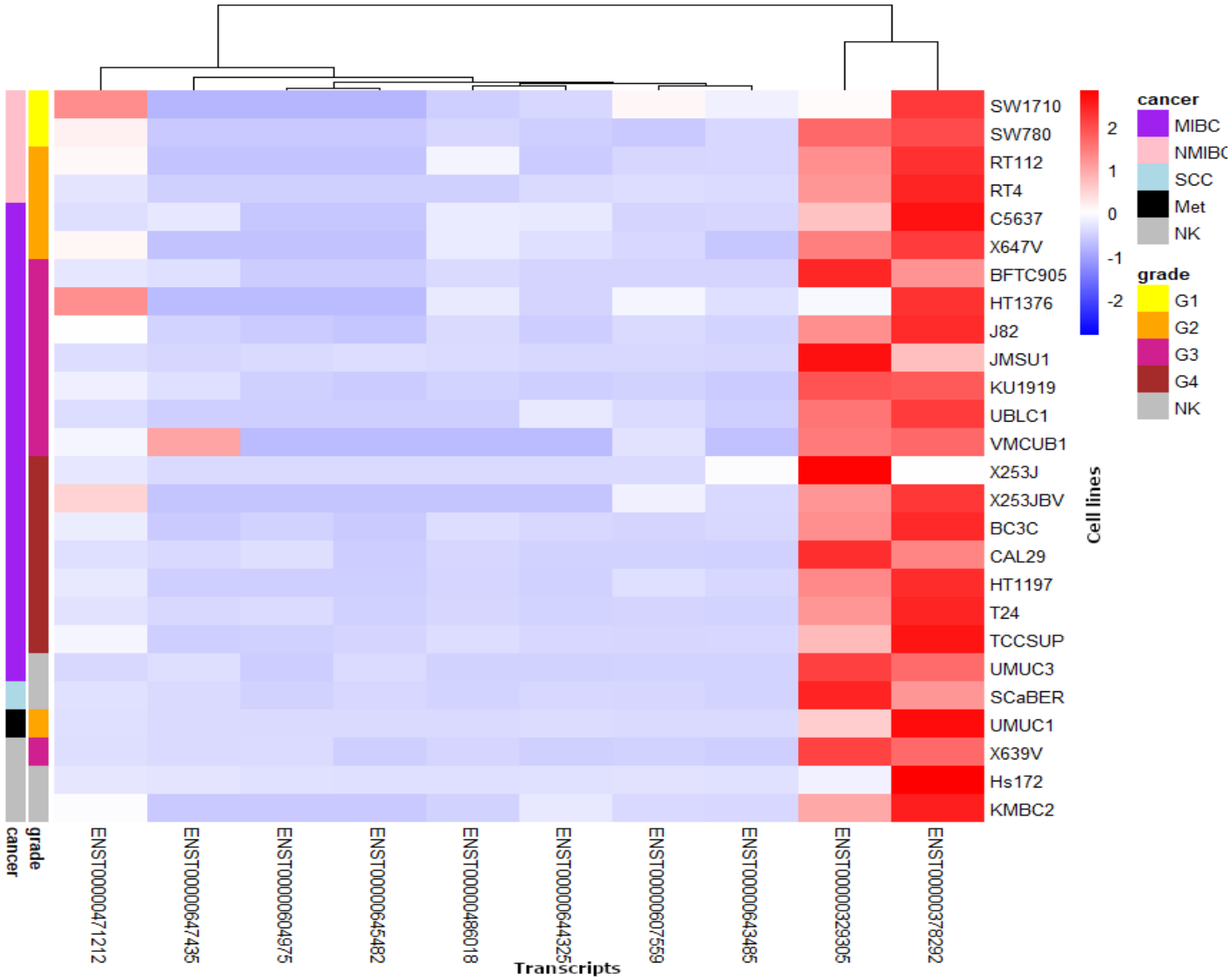


*Figure 3.2c Isoform expression in the CCLE dataset for **TPM3** gene, transcript variants (x-axis), BC cell lines (y-axis). Met= Metastases, NK= Not known*

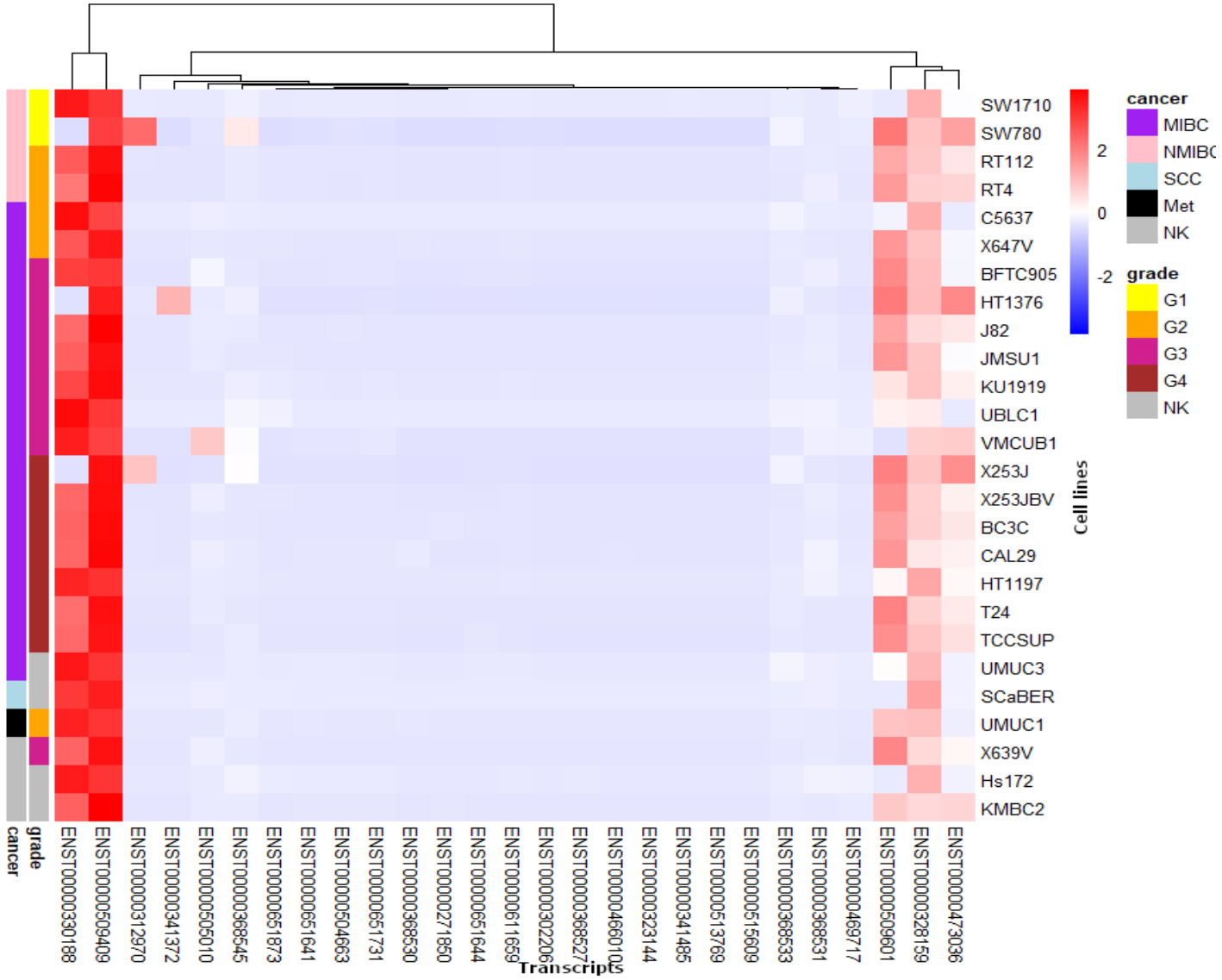
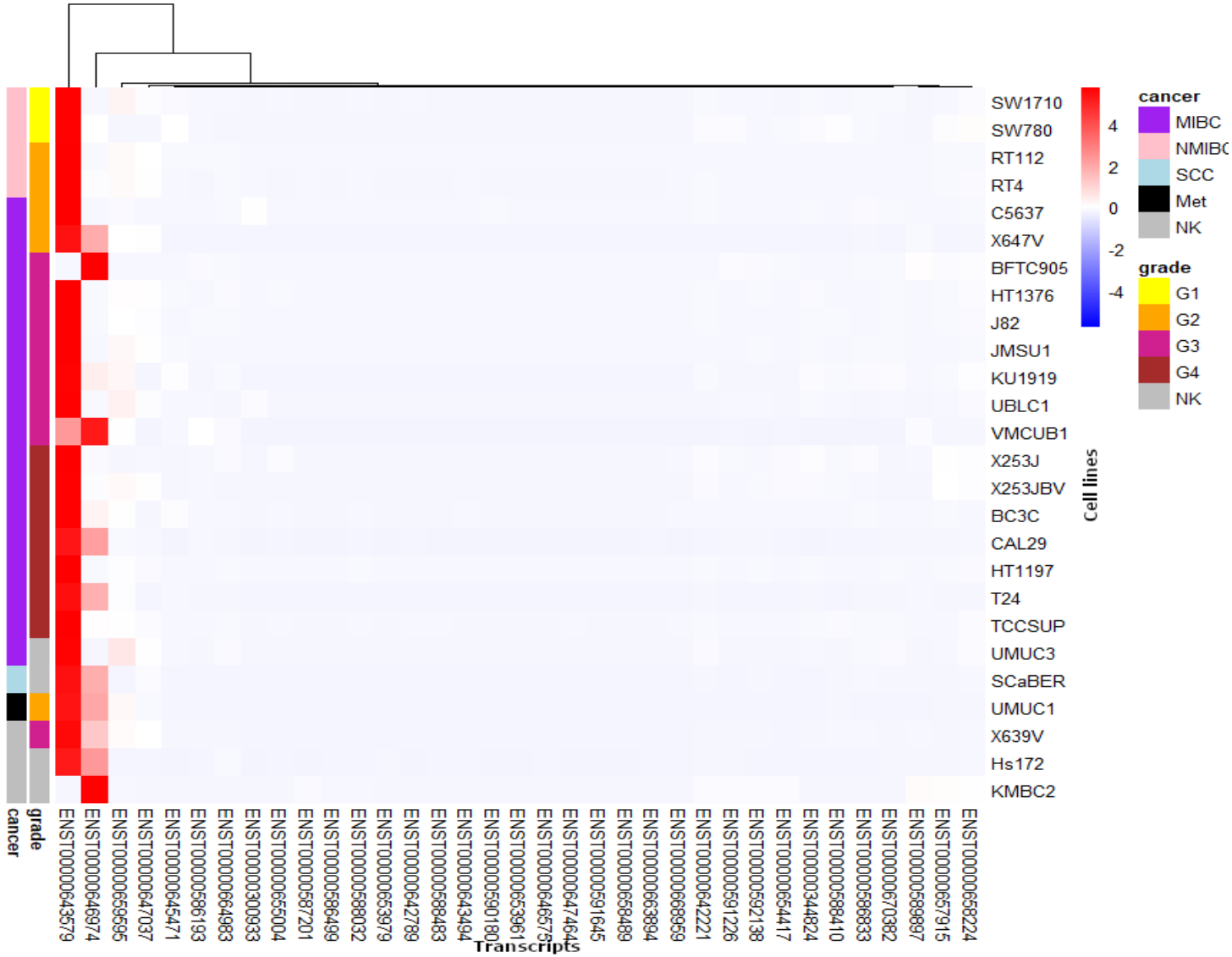


*Figure 3.2d Isoform expression in the CCLE dataset for **TPM4** gene, transcript variants (x-axis), BC cell lines (y-axis). Met= Metastases, NK= Not known*

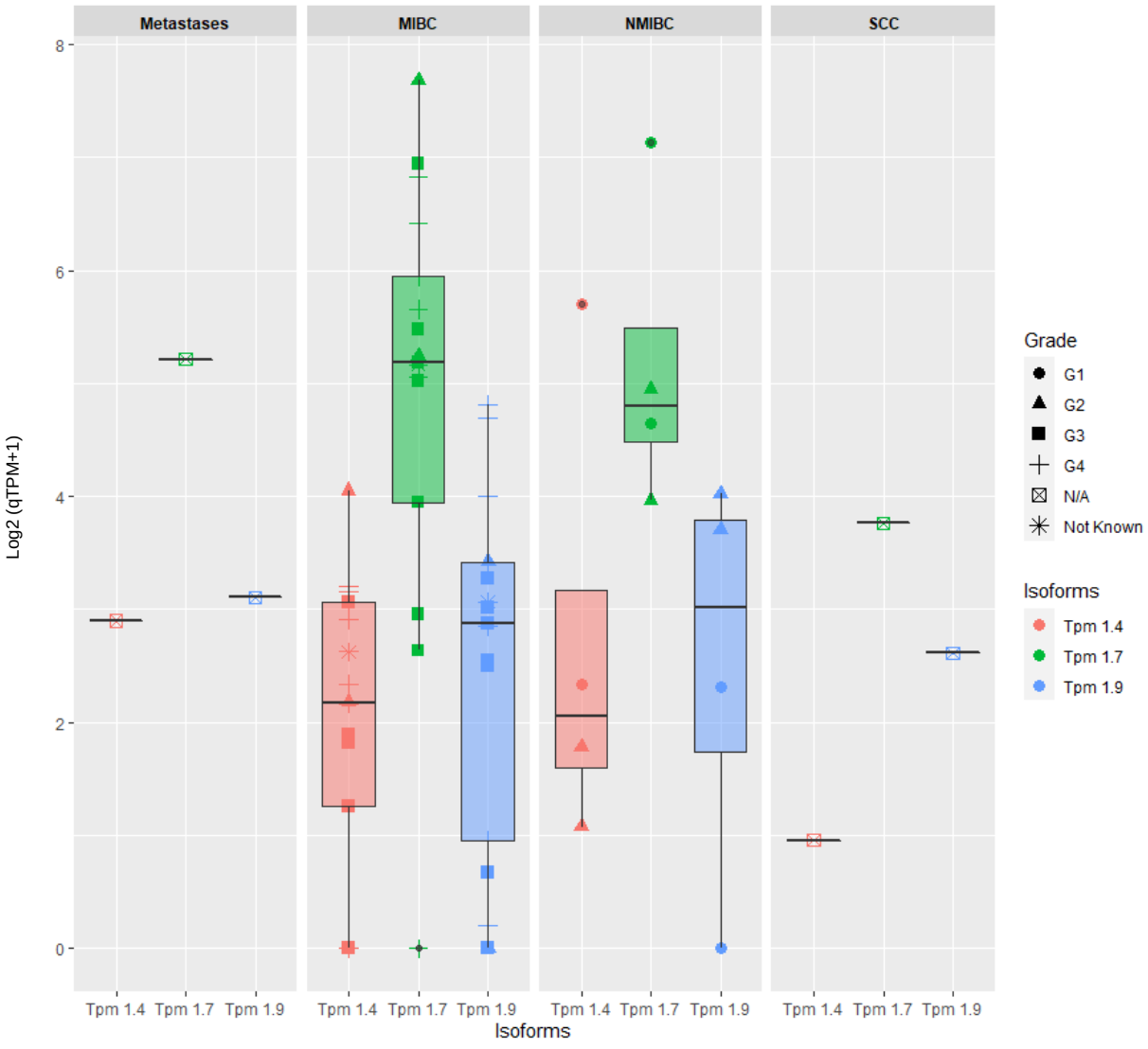


TPM1 (Figure 3.2a and Figure 3.3a)

38 transcript variants were identified for the *TPM1* gene, of which only **one** appeared to have uniform high expression in almost all BC cell lines: *ENST00000358278*, with a mean expression value of 350.7 qTPM. This transcript translates to isoform Tpm1.7. The second and third most abundant transcripts were *ENST00000267996* (Tpm1.4) and *ENST00000404484* (Tpm1.9), with mean expression levels of 27.0 and 24.8 qTPM, respectively. These three transcripts dominated gene expression for *TPM1*.

Figure 3.3a (next page) shows the expression levels of the three isoforms in the bladder cancer cell lines (grouped according to cancer subtypes). In all types of BC cell lines, HMW Tpm1.7 has the highest expression, followed by the LMW isoform Tpm1.9 and HMW isoform Tpm1.4 (Kruskal-Wallis p-value: 6.317e-06).

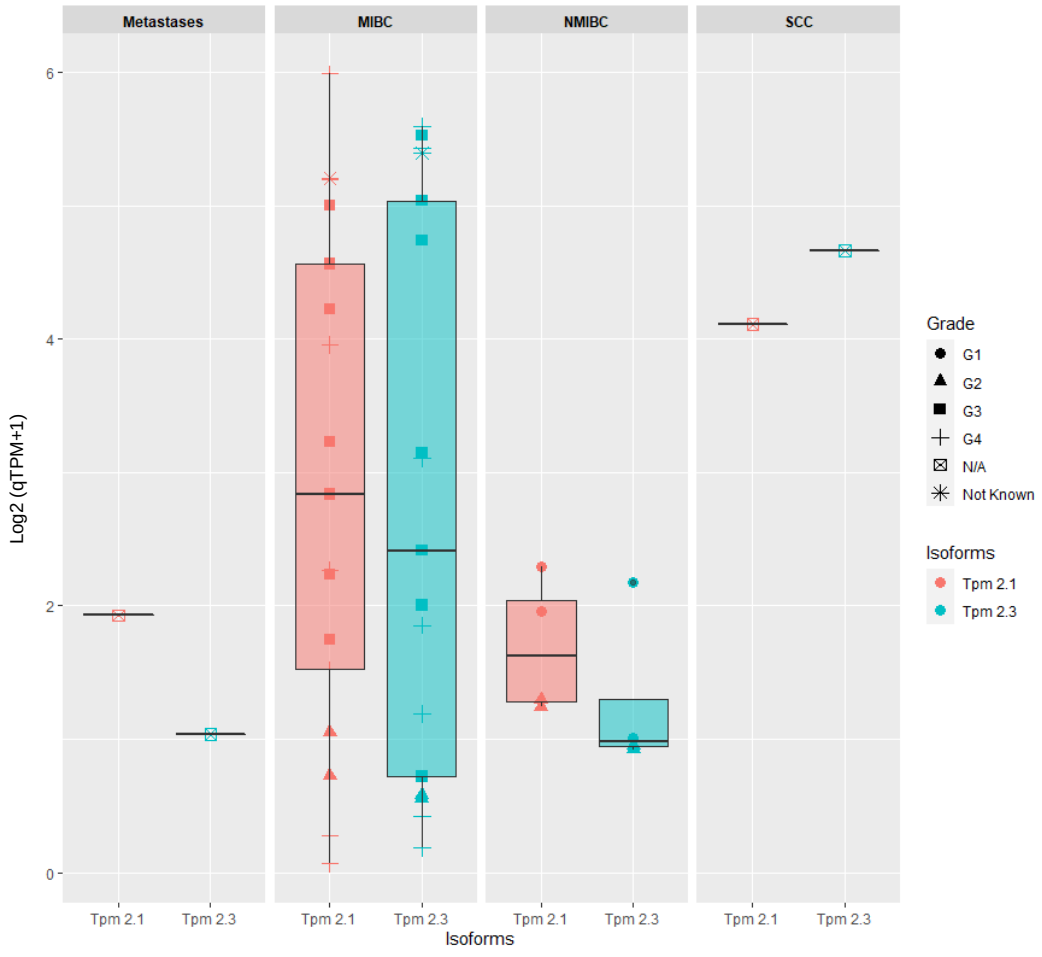
Figure 3.3a Abundance levels of TPM1 isoforms in bladder cancer cell lines. *The boxplot shows the distribution of the three isoforms that contribute to overall TPM1 gene expression (y-axis: log2(qTPM+1) values). BC cell lines are grouped according to types (Metastatic, MIBC- muscle-invasive bladder cancer, NMIBC- non-muscle invasive bladder cancer, SCC- squamous cell carcinoma). TPM1 expression appears to be increased in MIBC compared to NMIBC, of which Tpm1.7 is more dominant in high-grade disease. (Kruskal-Wallis p-value: 6.317e-06)*



TPM2 (Figure 3.2b and Figure 3.3b)

10 known isoforms were quantified for the *TPM2* gene. *ENST00000378292* (Tpm2.1) had the highest expression with a mean abundance level of 119.6 qTPM, followed by *ENST00000329305* (Tpm2.3) with a mean abundance level of 71.5 qTPM. The figure below shows the distribution of expression of the two HMW isoforms in bladder cancer cell lines. Of note, in SCC subtype, Tpm2.3 has higher expression than Tpm2.1.

Figure 3.3b Abundance levels of TPM2 isoforms in bladder cancer cell lines. *Both isoforms of TPM2 have seemingly high expression levels in SCC compared to other BC types. (Kruskal-Wallis p-value: 6.317e-06)*

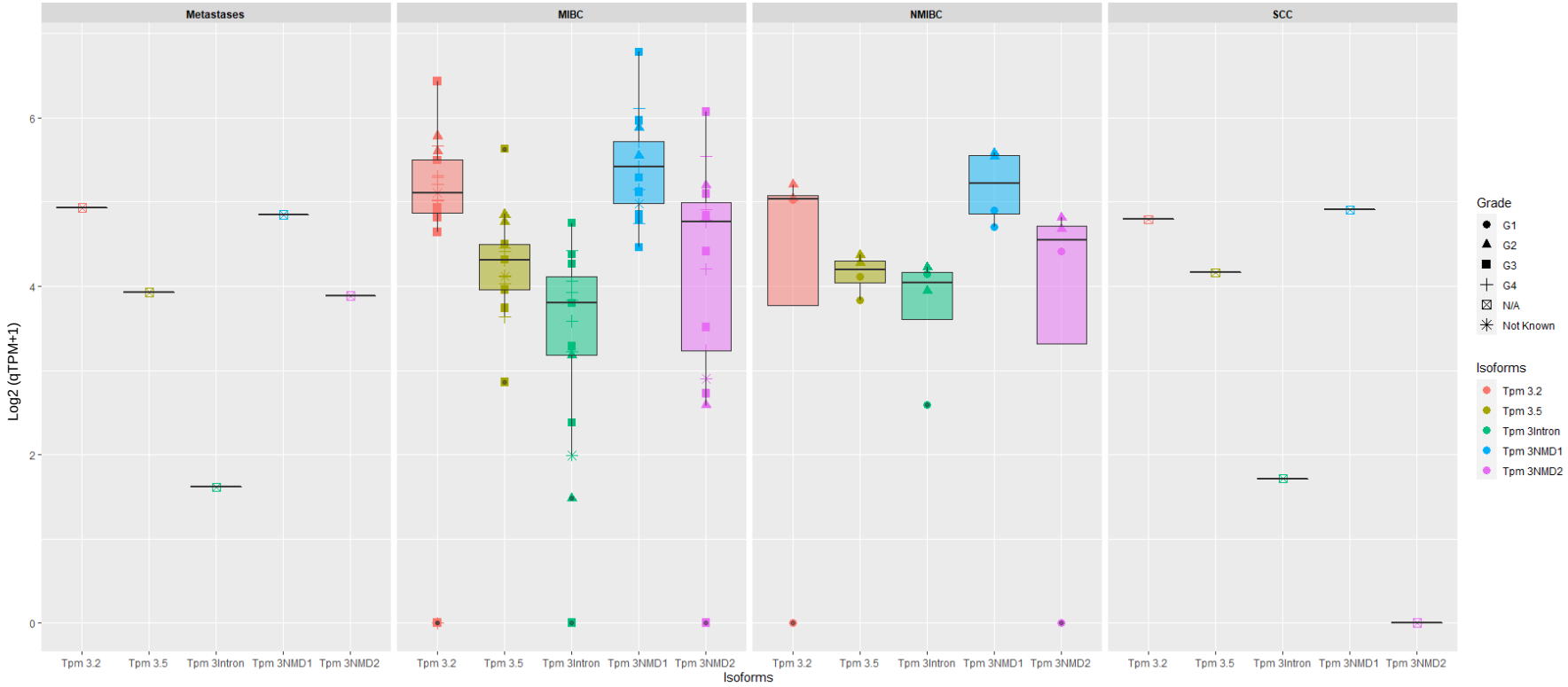


TPM3 (Figure 3.2c and Figure 3.3c)

TPM3 gene formed three clusters based on isoform expression of 27 known transcript variants. The transcripts with the highest expression that formed cluster one were *ENST00000509409* (NMD) and *ENST00000330188* (Tpm3.2) with abundance values of 234.7 and 172.3 qTPM, respectively. Cluster two with average expression levels of 94.7, 73.8 and 38.8 qTPM were contributed by *ENST00000509601* (NMD), *ENST00000328159* (Tpm3.5) and *ENST00000473036* (non-protein coding transcript which contains an intron), respectively. The remaining transcripts formed a third cluster with very low, almost null expression (≤ 1.0 qTPM).

Figure 3.3c (next page) shows the isoform expression in BC cell lines grouped according to subtypes. NMD-destined isoforms are found to have highest expression followed by the LMW isoforms Tpm3.2 and Tpm3.5 (Kruskal-Wallis p-value: 1.51e-10). No variance in pattern is noted across different cell lines, but all isoforms appear to have a similar distribution of expression.

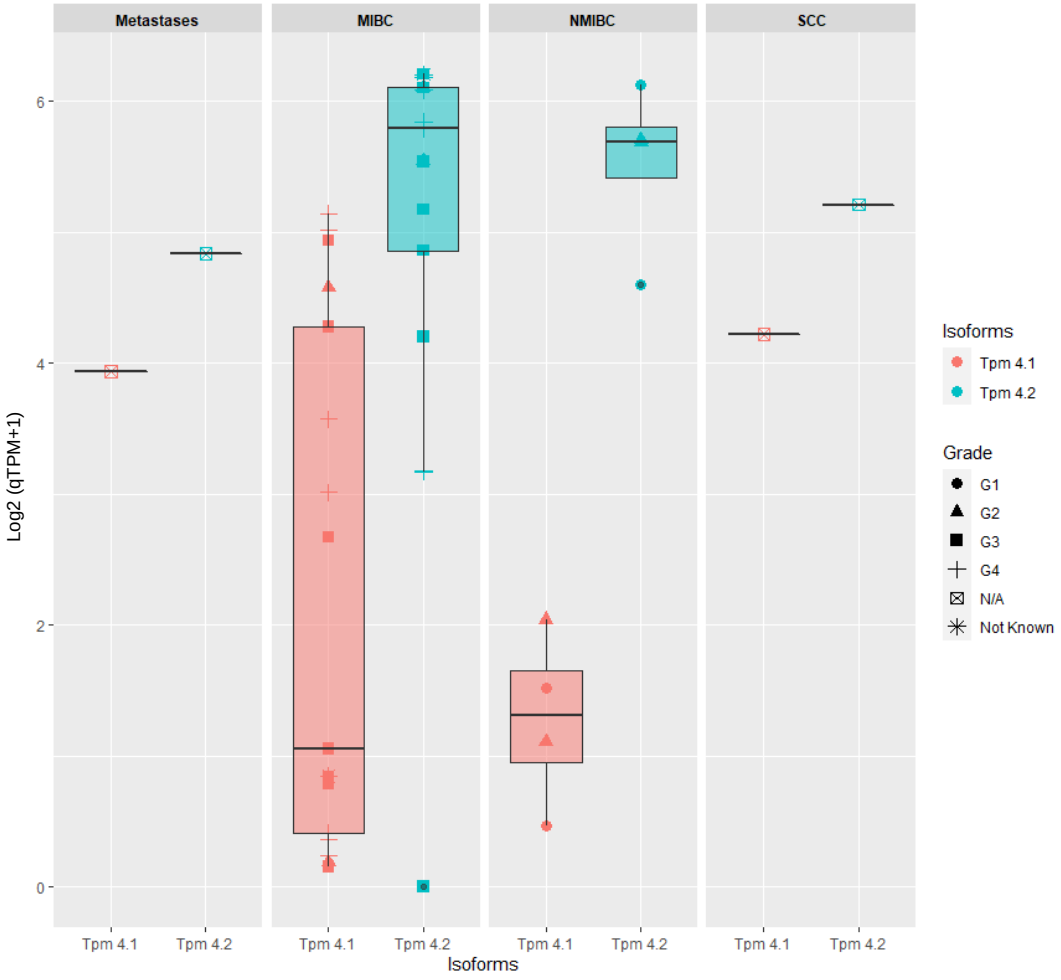
***Figure 3.3c Abundance levels of TPM3 isoforms in bladder cancer cell lines**. There are no significant differences in pattern of expression between cell lines. However, NMD isoforms appear to be upregulated in comparison to LMW isoforms Tpm3.2 and Tpm3.5 (Kruskal-Wallis p-value: 1.51e-10).*



TPM4 (Figure 3.2d and Figure 3.3d)

ENST00000643579 (LMW Tpm4.2) was found to have the highest expression for *TPM4* gene (from a total of 35 transcripts), with a mean qTPM value of 279.7. With an average qTPM count of 47.9, *ENST00000646974* encoding for HMW Tpm4.1 was the second most abundant isoform, with high read counts found in at least ten BC cell lines. Notably, these BC cell lines are of high-grade muscle-invasive origin (Figure 3.3d).

Figure 3.3d Abundance levels of TPM4 isoforms in bladder cancer cell lines*. LMW isoform Tpm4.2 (blue) has significantly higher expression levels across all BC types. Tpm4.1 has greater variability in expression levels with upregulation found in high-grade muscle-invasive cells. (Kruskal-Wallis p-value: 3.755e-06)*



Observations from the BC cell line dataset showed that the above discussed transcripts had high expression and were potential isoforms of interest expressed in BC cell lines, needing further investigation in other datasets (next section). Using *IsoTPM*, other datasets (with NMIBC and MIBC specimens) were analysed to investigate if the same pattern of *TPM* isoform expression was observed.

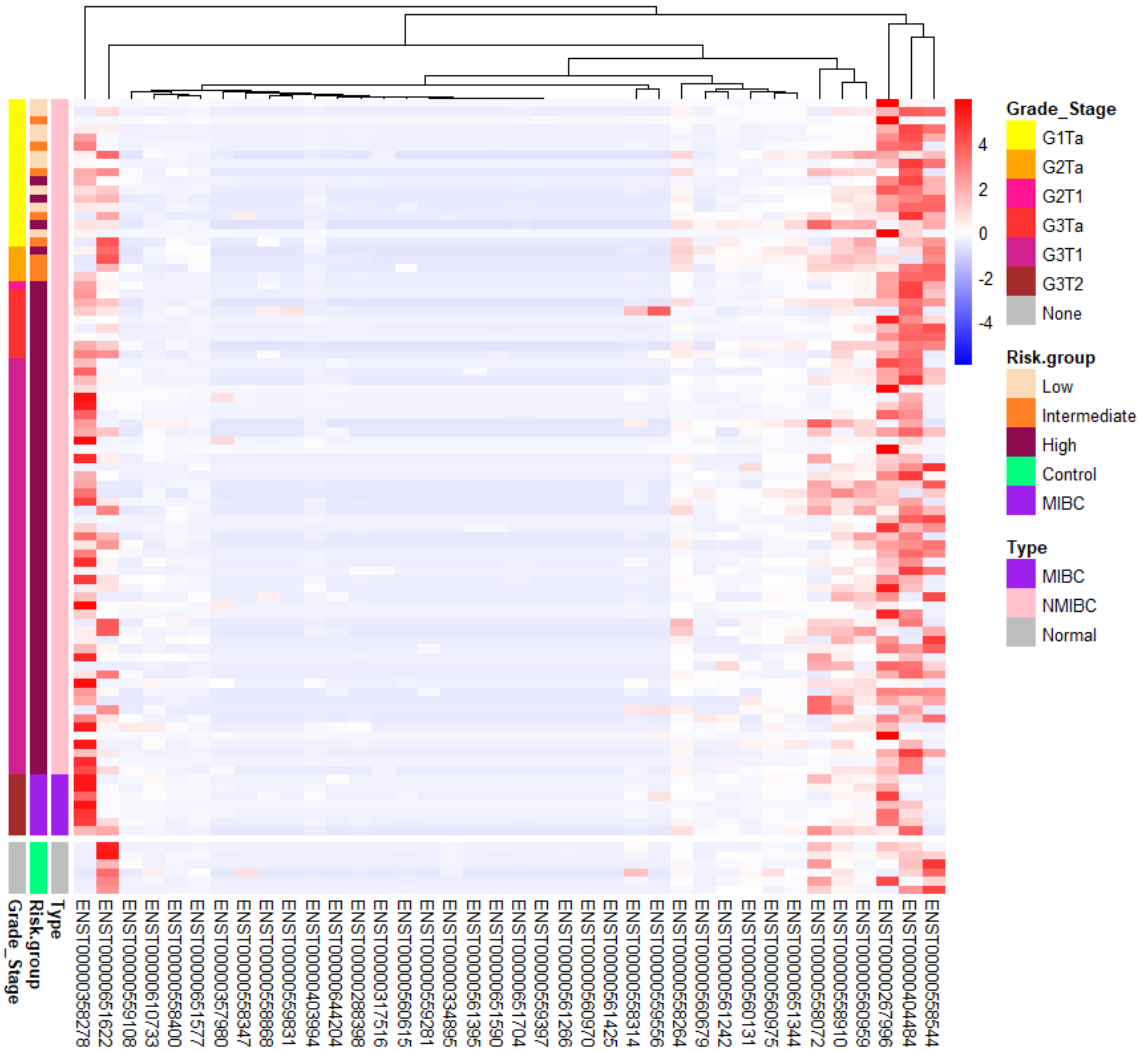
3.4.2 Tropomyosin isoform expression in *urothelial carcinoma of bladder tissues*

3.4.2.1 BCPP

TPM isoform expression was explored in SR-seq RNA datasets constructed from frozen bladder cancer tissues (TURBT specimens). The BCPP cohort provided SR-seq RNA data from 78 NMIBC and 7 MIBC tissue specimens, which included LG and HG disease (classified using the WHO 1973 grading system performed at the point of patient recruitment: G1, G2, G3). These were further identified as low-, intermediate- and/ or high-risk based on the EORTC risk scoring table (discussed earlier).
Additionally, 6 specimens of benign bladder tissue with SR-seq RNA were also used as control, for comparison of cancer vs benign conditions. *IsoTPM* was applied to quantify isoform expression for all 4 *TPM* genes.

The following heatmaps show differential isoform expression for the cancerous and benign tissue specimens of the BCPP cohort (Figures 3.4a, 3.5a, 3.6a, 3.7a) with the description for each *TPM* gene.

Figure 3.4a Differential expression of isoforms of TPM1 in BCPP cohort *including NMIBC, MIBC and benign (normal) tissue samples. The transcripts are labelled on the x-axis with the samples grouped according to Grade, Stage and Risk groups. Description on next page.*



For ***TPM1***, in addition to HMW Tpm1.7 (*ENST00000358278*), HMW Tpm1.4 (*ENST00000267966*) and LMW Tpm1.9 (*ENST00000404484*), *ENST00000558544* was found to be expressed in the tumour samples. This additional transcript is non-protein-coding and degraded by NMD (containing 7 exons and an in-frame termination codon found 50bp upstream of the final splice junction). The isoform with highest expression in the tumour samples of the BCPP dataset was LMW Tpm1.9 followed by HMW Tpm1.7 and HMW Tpm1.4 (Figure 3.4b). In contrast, the benign bladder tissue samples exhibited highest expression for another NMD-destined transcript *ENST00000651622* (Tpm1NMD) (qTPM= 50.32), with no read counts recorded for Tpm1.7 (qTPM= 0). Additionally, Tpm1.7 appeared to be greater in high-risk NMIBC (qTPM= 32.13) and MIBC (qTPM= 86.44) compared to low-risk (qTPM= 7.49) and intermediate (qTPM= 7.43) groups. Table 3.3 and Figure 3.4b show the difference in abundance levels in BC and benign bladder tissue samples for the isoforms of interest of *TPM1*.

Table 3.3 |Comparison of mean isoform expression values for *TPM1* in the BCPP cohort. The table below gives the mean expression levels in qTPM of the *TPM1* isoforms in the BCPP cohort. Mann Whitney U test was performed to test statistical significance of isoform expression within the two groups: Cancer and Benign bladder tissues.

Transcript	Isoform	Benign (n=6)	Cancer (n=85)	p-value
ENST00000267996	Tpm1.4	17.79	55.07	0.3412
ENST00000358278	Tpm1.7	0.000	31.67	0.0001***
ENST00000404484	Tpm1.9	18.30	23.47	0.5898
ENST00000651622	Tpm1NMD	50.32	6.559	<0.0001

Figure 3.4b TPM1 Isoform expression levels in BCPP *bladder cancer (blue) and benign tissue samples (salmon). Tpm1.4 and Tpm1.9 are found in cancer and benign samples, with slight upregulation in cancer comparatively. Tpm1.7 appears to be only in cancer. Tpm1NMD is found more commonly in benign tissue.*

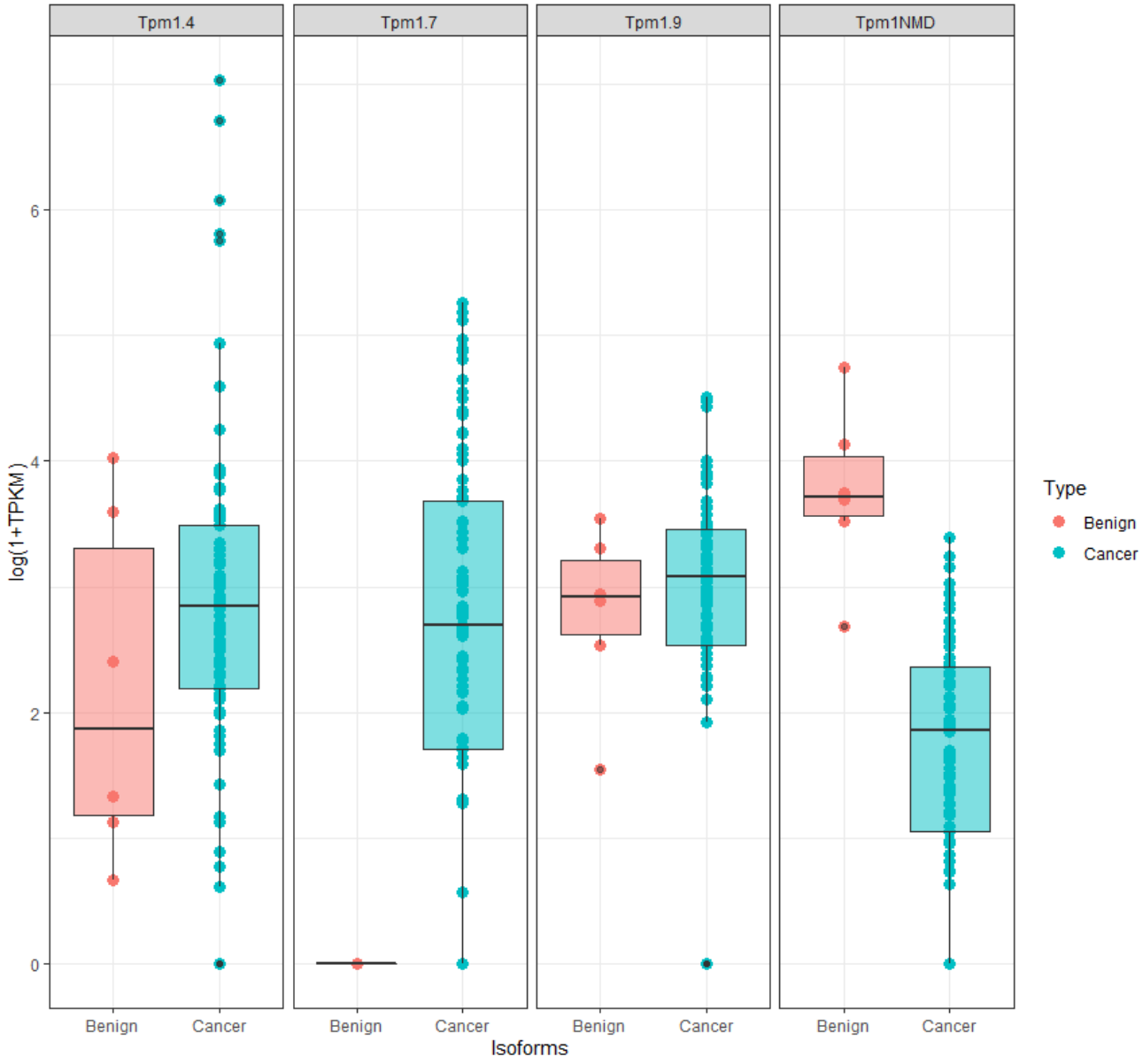


Figure 3.5a (next page) shows the heatmap for ***TPM2*** differential isoform expression, which is dominated by *ENST00000378292*, (HMW isoform Tpm2.1) in both cancer (qTPM= 84.26) and benign (qTPM= 39.53) samples. The other HMW isoform Tpm2.3 (*ENST00000329305*) also contributes to *TPM2* gene expression in at least 5 cancer samples with qTPM >20.

When comparing the mean expression levels of isoform Tpm2.1, no significant difference was noted between cancer and benign samples (p: 0.8090). Isoform expression for Tpm2.3 was observed in cancer samples, with at least one benign sample having a qTPM of 56.37 (Fig 3.5b and Table 3.4).

Figure 3.5a Differential isoform expression for TPM2 gene in BCPP cohort. *Samples include MIBC, NMIBC and benign tissue RNA seq.*

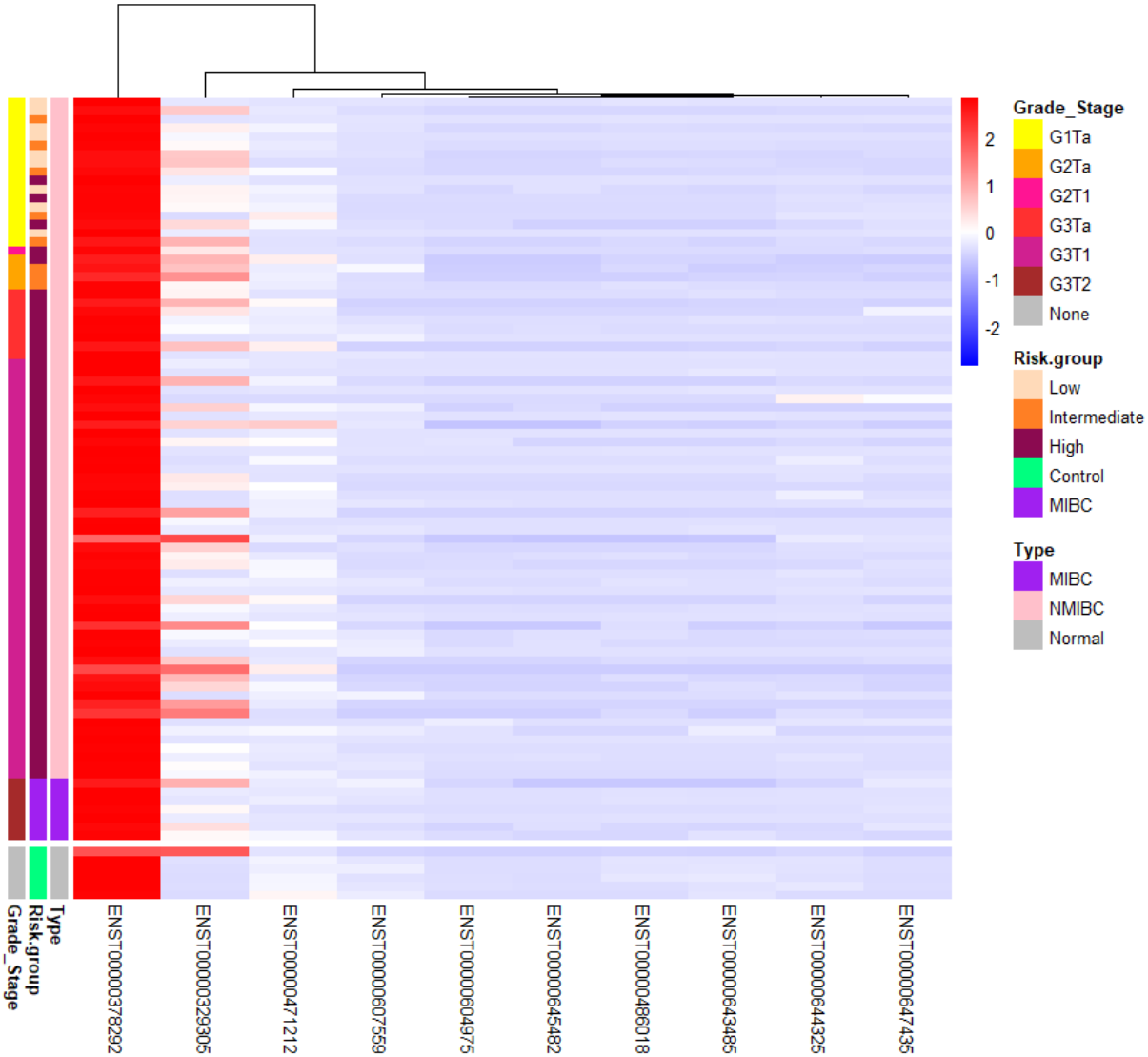


Figure 3.5b Isoform expression levels for TPM2 *in BCPP bladder cancer (blue) and benign tissue samples (salmon). No statistically significant difference is observed in abundance levels of the two isoforms between cancer and benign conditions.*

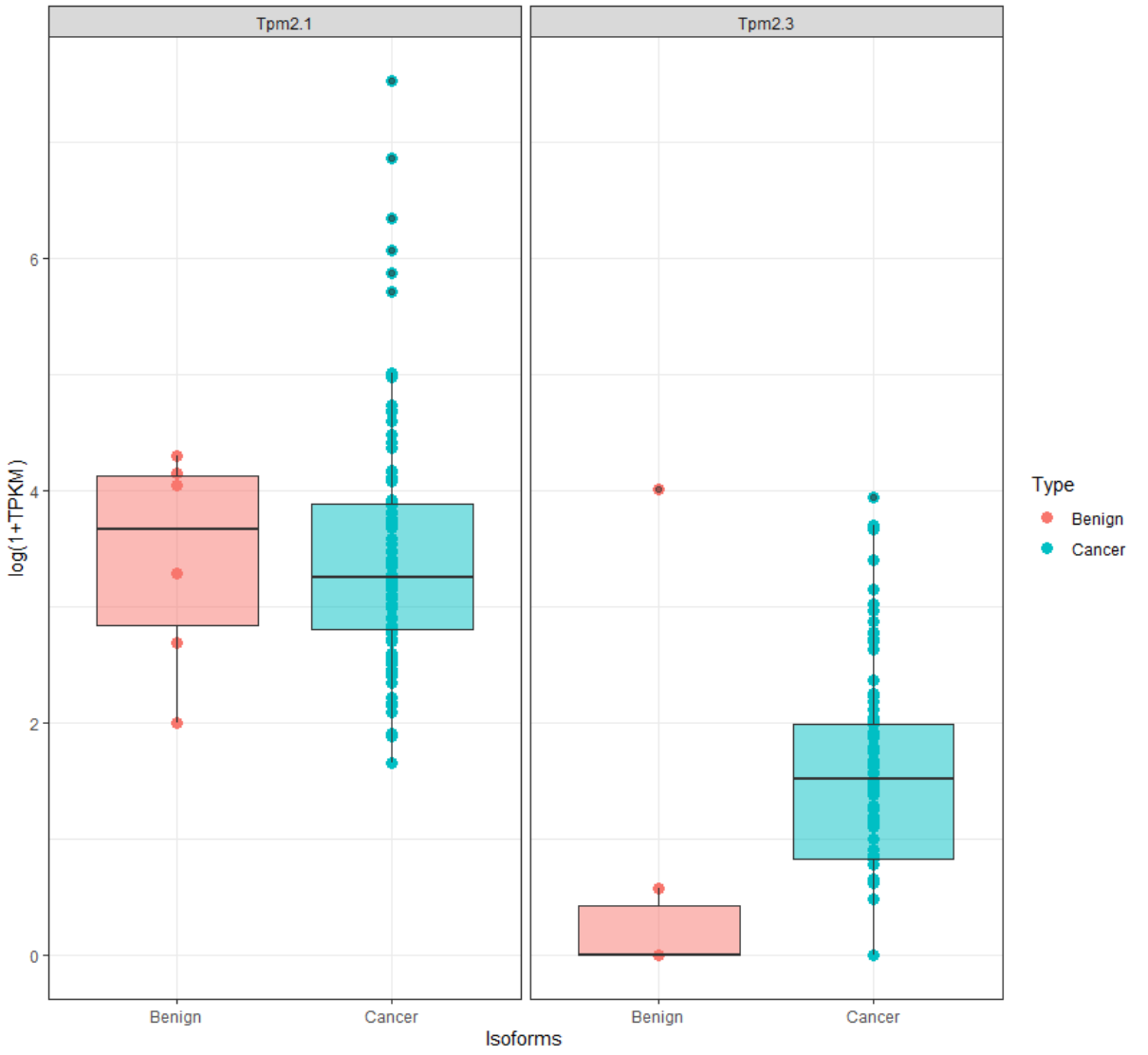


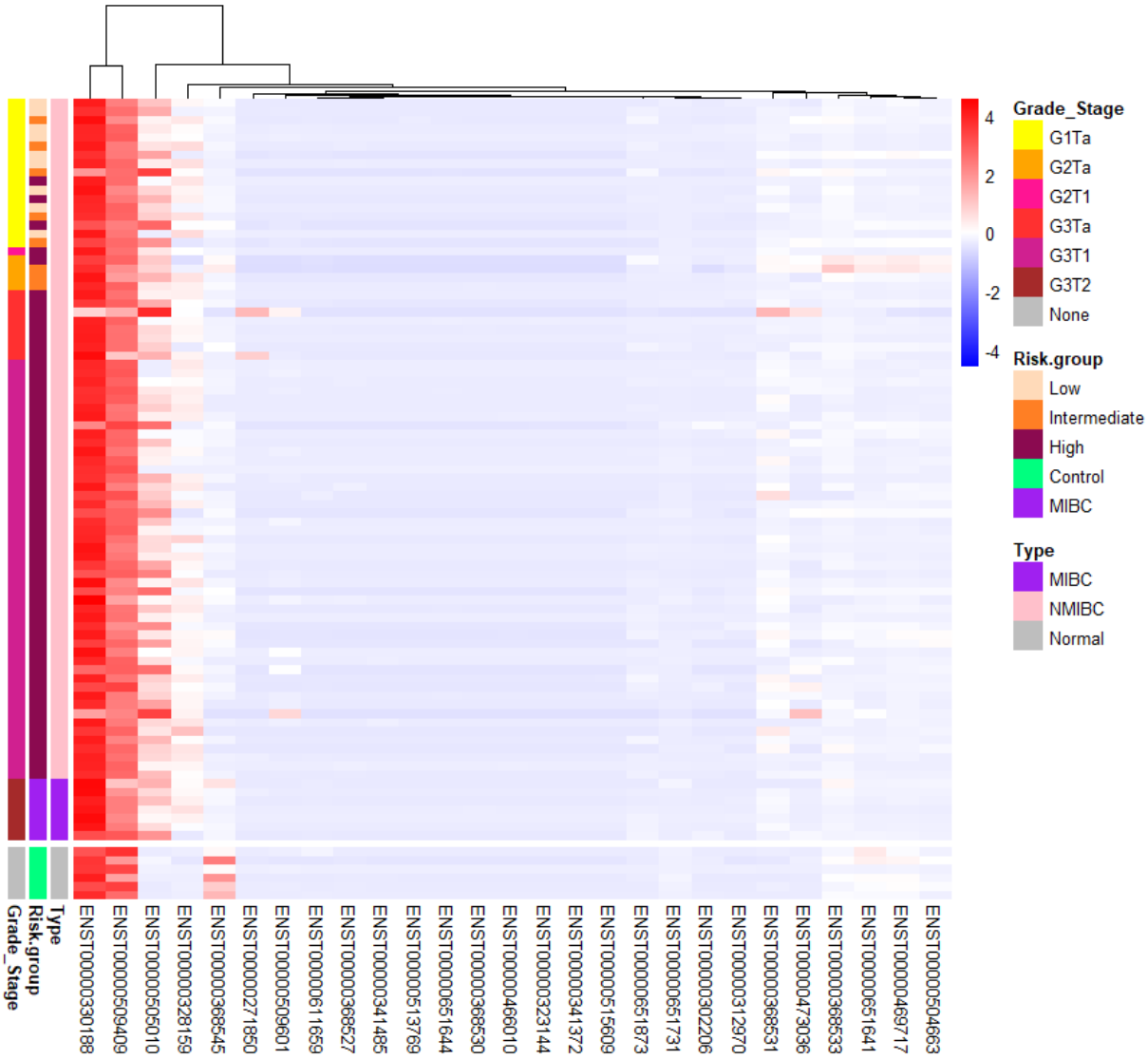
Table 3.4 | Comparison of mean isoform expression values for *TPM2* in the BCPP cohort. Mean expression values of the two conditions of the BCPP dataset are shown with associated p-value calculated using Mann Whitney U statistical test.

Transcript	Isoform	Benign (n=6)	Cancer (n=85)	p-value
ENST00000378292	Tpm2.1	39.53	84.26	0.8090
ENST00000329305	Tpm2.3	9.191	6.191	0.0573

The isoforms of ***TPM3*** did not form the same clusters as in the cell line dataset (CCLE), as shown in the heatmap (Figure 3.6a, next page). In the BCPP BC samples, cluster one formed by *ENST00000330188* (LMW Tpm3.2) and *ENST00000509409* (Tpm3NMD1) had the highest expression, with means of 133.1 and 93.2 qTPM, respectively. This cluster and pattern of expression was also observed in BC cell lines. However, additional transcripts that contributed to overall *TPM3* gene expression, in decreasing levels of abundance included *ENST00000505010*, *ENST00000328159* (Tpm3.5), *ENST00000368531* (Tpm3.7), and *ENST00000368545*. These transcript isoforms encode for LMW isoforms (Tpm3.5 and Tpm3.7) with expression levels of <10 qTPM, while others (*ENST00000505010* and *ENST00000368545*) were non-coding isoforms with retained introns.

In the benign samples, isoform distribution with the highest expression was from cluster one but included an additional intronic transcript *ENST00000368545*.

Figure 3.6a Differential isoform expression of TPM3 in BCPP cohort. *Samples include MIBC, NMIBC and benign tissues.*



Comparative analysis of the isoforms showed that LMW isoform Tpm3.2 was found in cancer and benign tissues (qTPM = 133.07 and 68.23, respectively). LMW Tpm3.5 and LMW Tpm3.7 were more prominent in cancer tissues but did not have high expression levels (qTPM= 16.88 and 8.29, respectively) (Figure 3.6b and Table 3.5).

Figure 3.6b Isoform expression levels of TPM3 in BCPP *bladder cancer (blue) and benign (salmon) tissues. Transcripts with retained introns are not included in the analysis.*

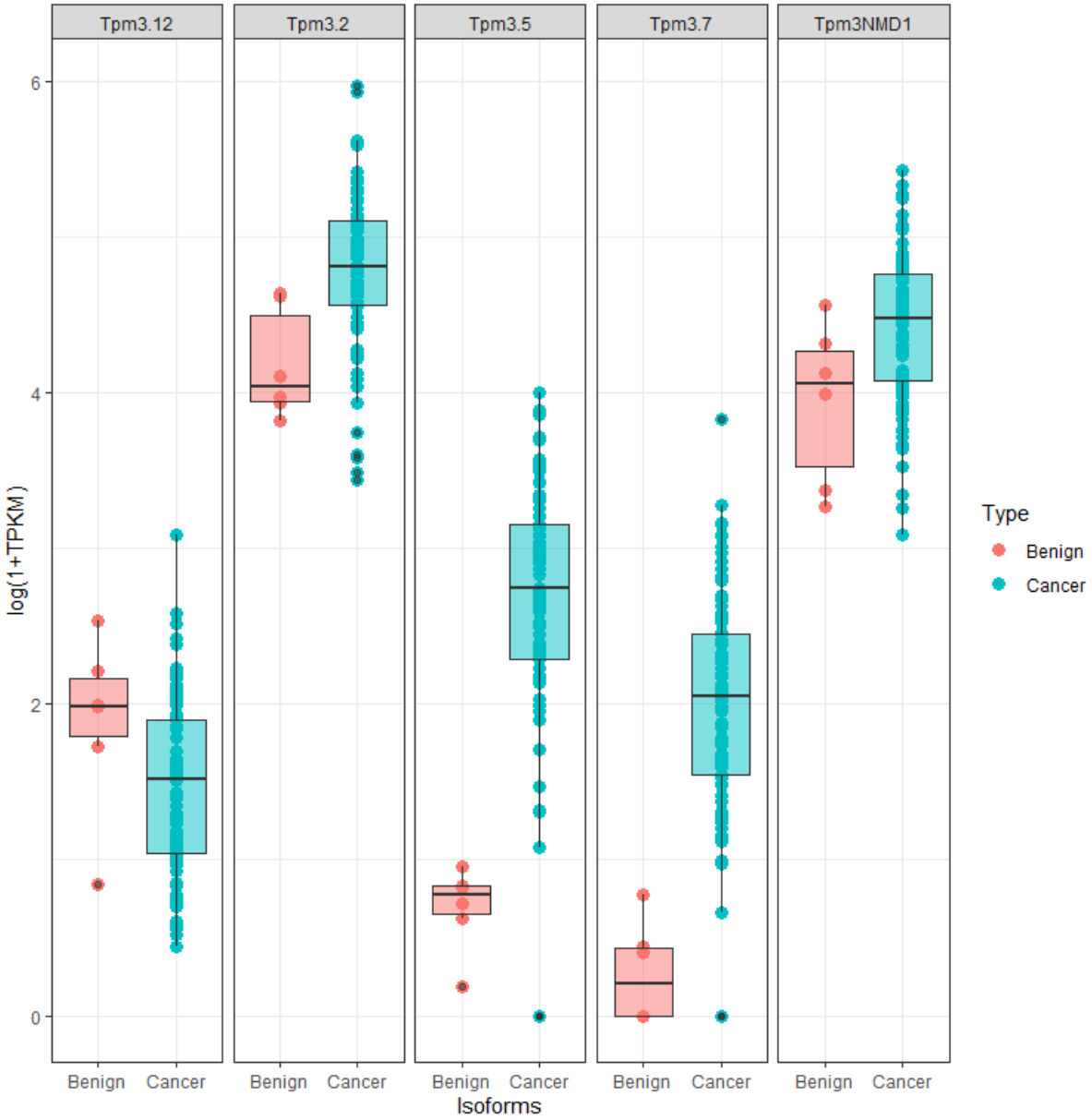


Table 3.5 | Transcript isoforms of *TPM3* with mean expression levels in the BCPP cohort. Mean expression values of the two conditions of the BCPP dataset are shown with associated p-value calculated using Mann Whitney U statistical test. (grey = not significant)

Transcript	Isoform	Benign (n=6)	Cancer (n=85)	p-value
ENST00000330188	Tpm3.2	68.23	133.1	0.0026
ENST00000509409	Tpm3NMD1	56.03	93.17	0.0345
ENST00000328159	Tpm3.5	1.054	16.88	<0.0001
ENST00000368531	Tpm3.7	0.3697	8.285	<0.0001

The expression of ***TPM4*** was dominated by *ENST00000643579* (LMW Tpm4.2) in BCPP BC specimens, followed by *ENST00000646974* (HMW Tpm4.1) (heatmap shown on next page, Figure 3.7a). The mean abundance level for Tpm4.2 was 147.02 qTPM, with at least 4 samples exhibiting values >350 qTPM, whereas Tpm4.1 mean expression value was 28.94 qTPM.

The benign tissue specimens had almost equal distribution of isoform usage shared between Tpm4.2 and Tpm4.1 with mean abundance levels of 58.01 and 50.55, respectively.

In addition, *ENST00000659595* (NMD) was also quantified in benign samples with a mean abundance level of 35.34 qTPM, with almost null expression in cancer samples (qTPM=0.02). Figure 3.7b and Table 3.6 show the distribution of isoform usage of the 3 isoforms from *TPM4* in benign and cancer specimens.

Figure 3.7a Differential isoform expression of TPM4 in BCPP cohort. *Transcript (x-axis) abundance levels are indicated as dark red cells with highest levels of expression in samples (y-axis) of muscle-invasive, non-muscle-invasive and benign type.*

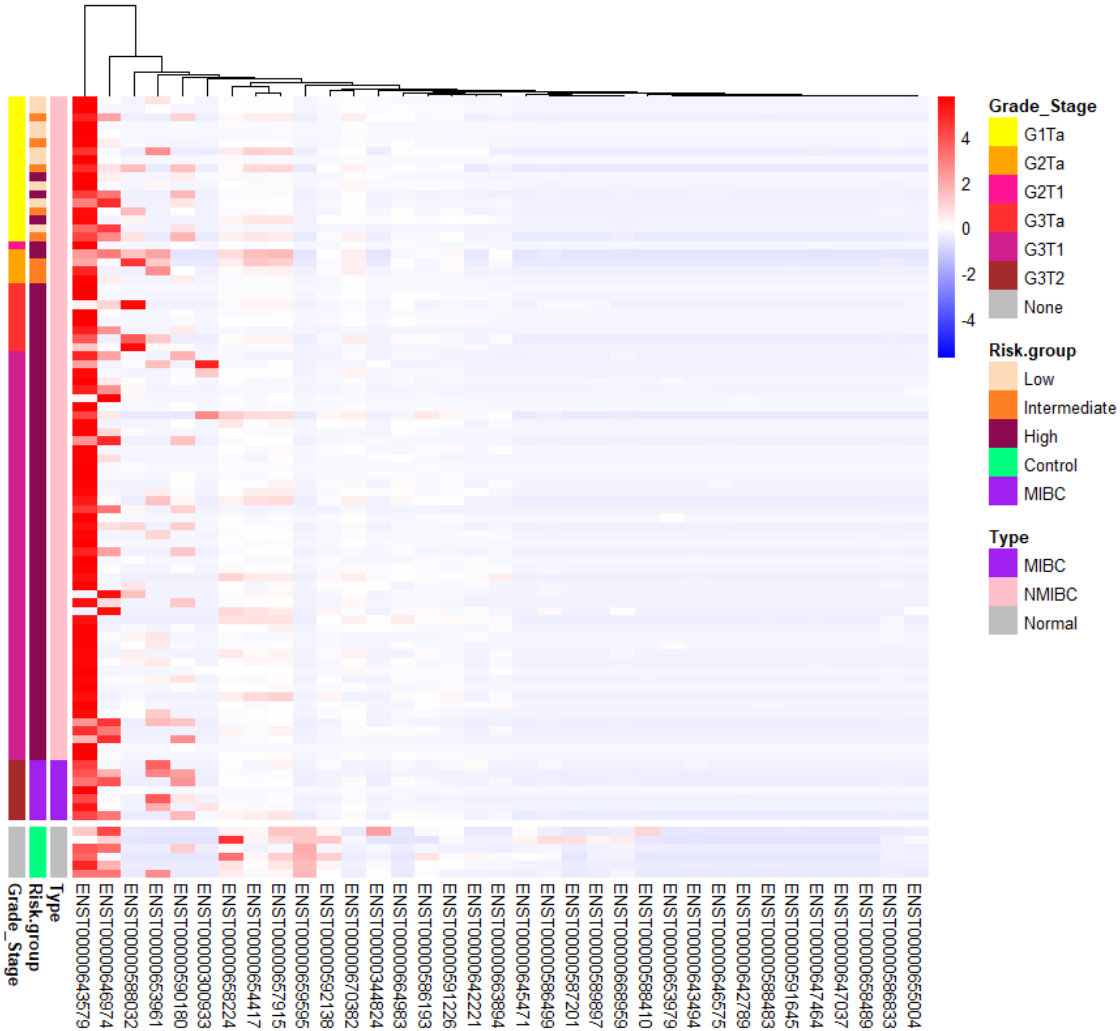


Figure 3.7b Isoform expression levels of TPM4 *in BCPP bladder cancer (blue) and benign (salmon) tissues. Tpm4.2 is predominantly expressed in cancer followed by an equal proportion of shared level expression between Tpm4.2 and Tpm4.1 in benign tissues. Tpm4NMD appears to be 'switched off' in cancer.*

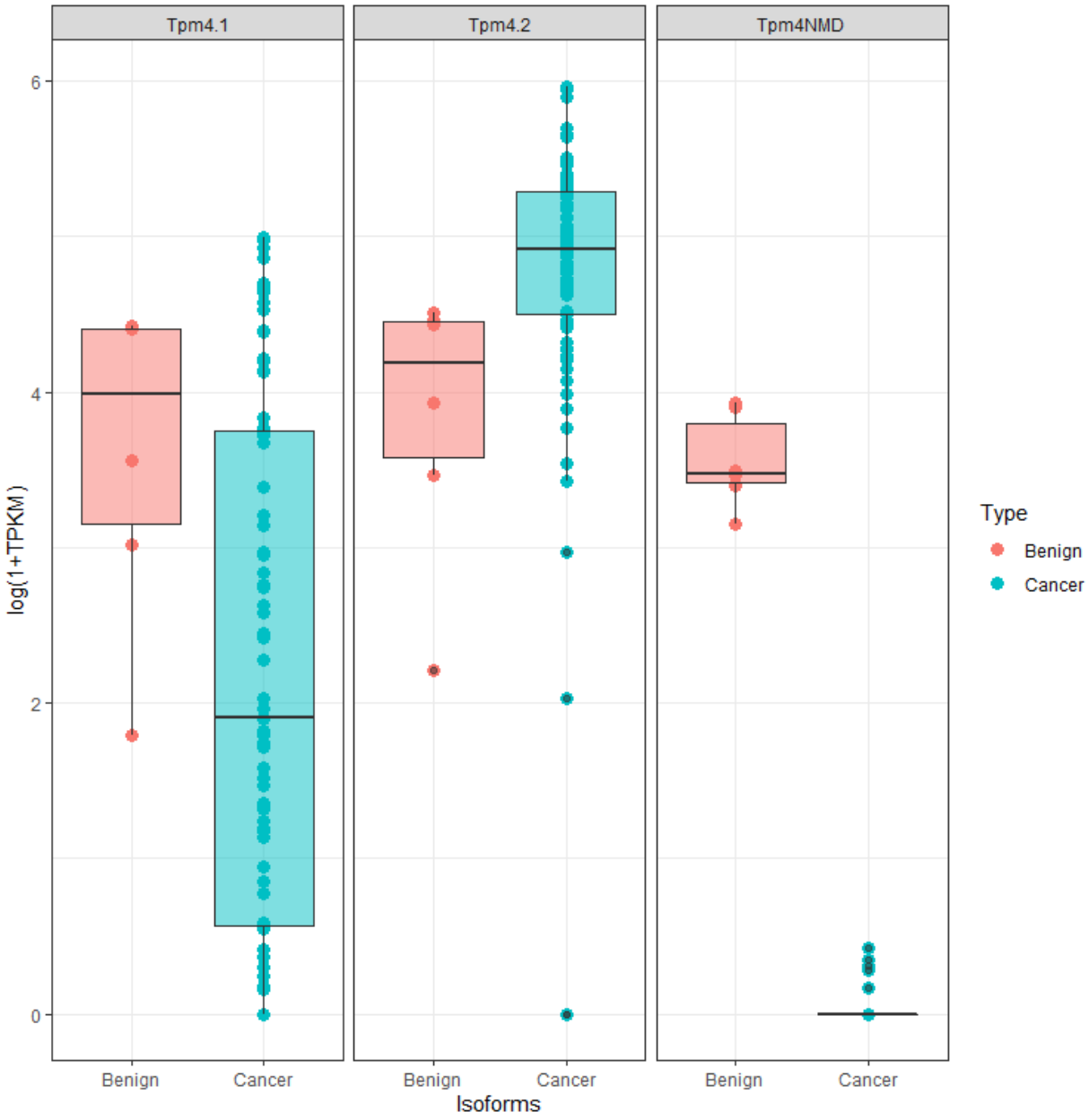


Table 3.6 | Transcript isoforms of *TPM4* with mean expression levels in the BCPP cohort. Mean expression values of the two conditions of the BCPP dataset are shown with associated p-value calculated using Mann Whitney U statistical test. (grey = not significant)

Transcript	Isoform	Benign (n=6)	Cancer (n=85)	p-value
ENST00000643579	Tpm4.2	58.01	147.02	0.0026
ENST00000646974	Tpm4.1	50.55	28.94	0.0583
ENST00000659595	Tpm4NMD	35.34	0.021	<0.0001

It is interesting to note the ratio of isoform usage of LMW Tpm4.2 to HMW Tpm4.1 is 1.1 in benign tissues, whereas in cancer there is a five-fold increase in expression of LMW Tpm4.2 to HMW Tpm4.1. However, there was no difference in expression of these two isoforms in the different risk groups: Tpm4.2 qTPM high-risk NMIBC= 151.71, Intermediate-risk NMIBC= 94.12, low-risk NMIBC= 173.42, MIBC= 132.70, and Tpm4.1 qTPM high-risk NMIBC= 28.55, Intermediate-risk NMIBC= 17.70, low-risk NMIBC= 34.00, MIBC= 38.69.

3.4.2.2 TCGA

To investigate the expression of the aforementioned isoforms in MIBC, the TCGA dataset was analysed using *IsoTPM*. The TCGA SR-seq RNA dataset comprised of 412 MIBC and 19 paracancerous (PN) tissues. The heatmaps (Figures 3.8a-d) illustrate the differential expression of isoforms for the 4 *TPM* genes in MIBC and PN specimens.

As seen in BCPP and BC cell lines, *TPM1* isoform expression (Figure 3.8a) in MIBC is contributed by the previously identified isoforms HMW Tpm1.4 (ENST00000267996), HMW Tpm1.7 (ENST00000358278), LMW Tpm1.9 (ENST00000404484), and an NMD transcript (ENST00000558544). Tpm1.4 shows the highest expression with a mean qTPM of 69.72, followed by Tpm1.7 with 50.71 qTPM and Tpm1.9 with 7.04 qTPM. The NMD transcript has a mean of 12.42 qTPM.

In the PN samples, highest expression is exhibited by Tpm1.4 with mean of 1575.25 qTPM whilst Tpm1.7 has a mean count of 22.62 qTPM.

Figure 3.8a TCGA data with differential isoform expression of TPM1. *For each sample (y-axis) the gene expression is contributed by the isoforms indicated by ENSEMBL IDs (x-axis). MIBC= purple, Paracancerous (PN)= blue*

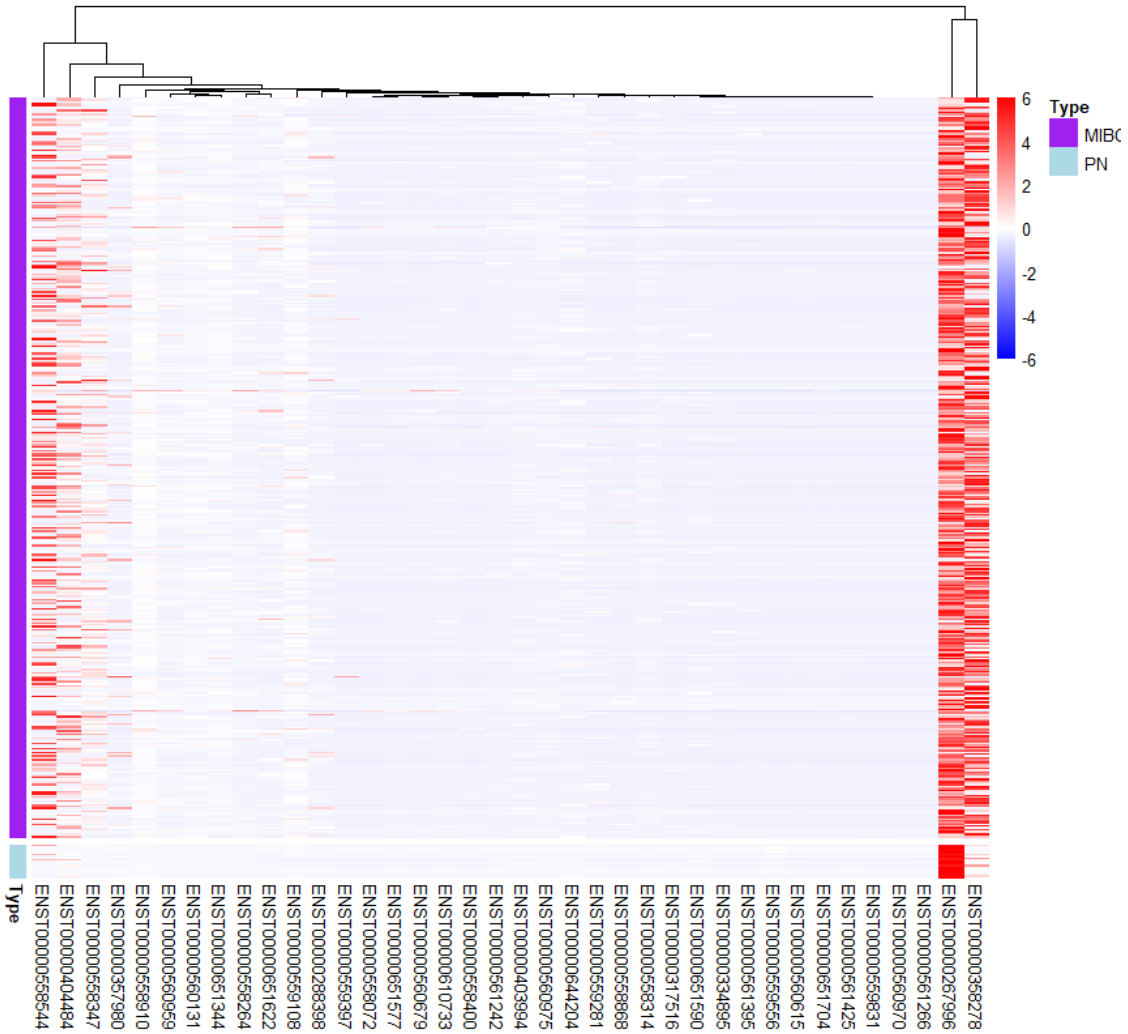
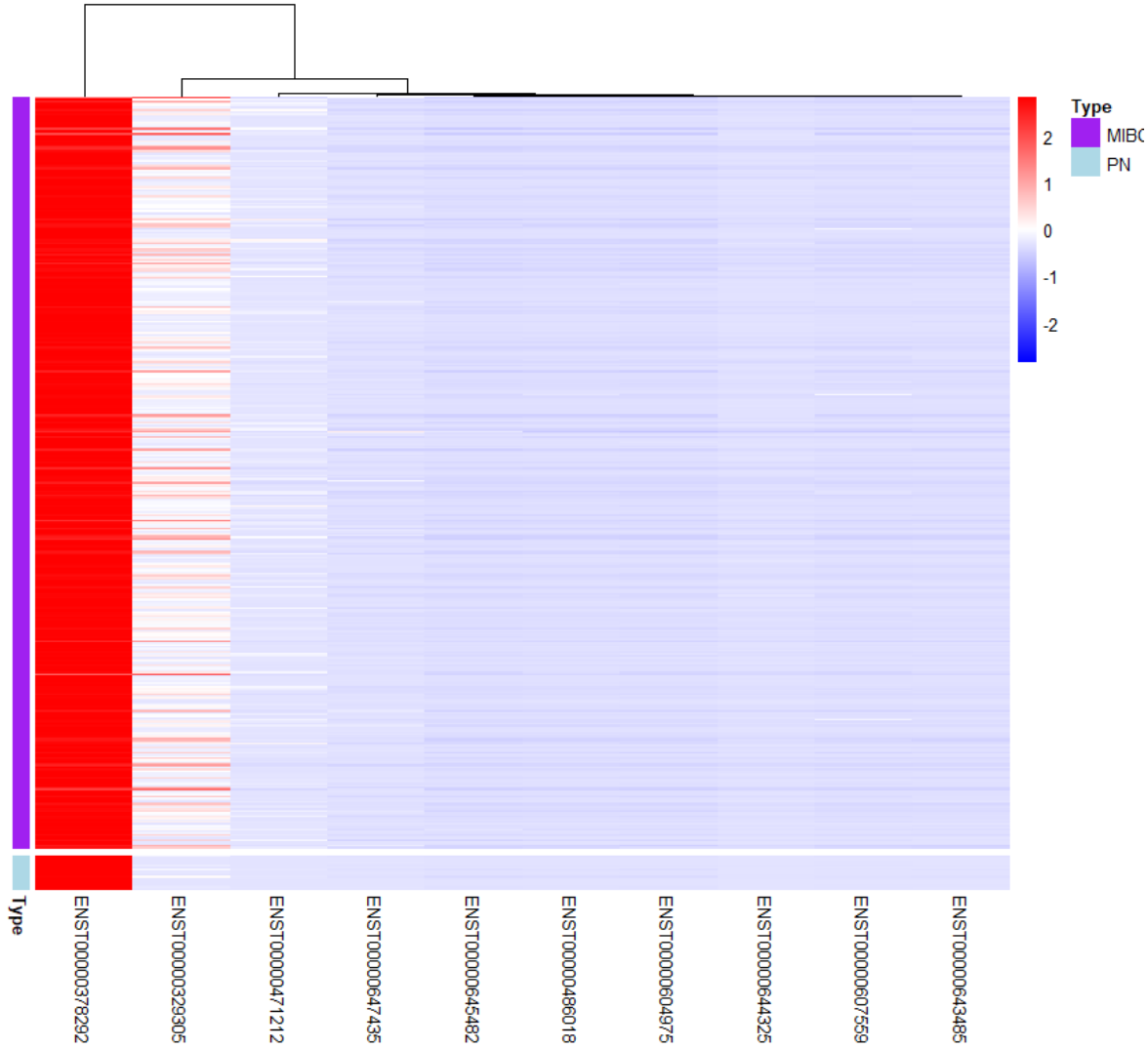
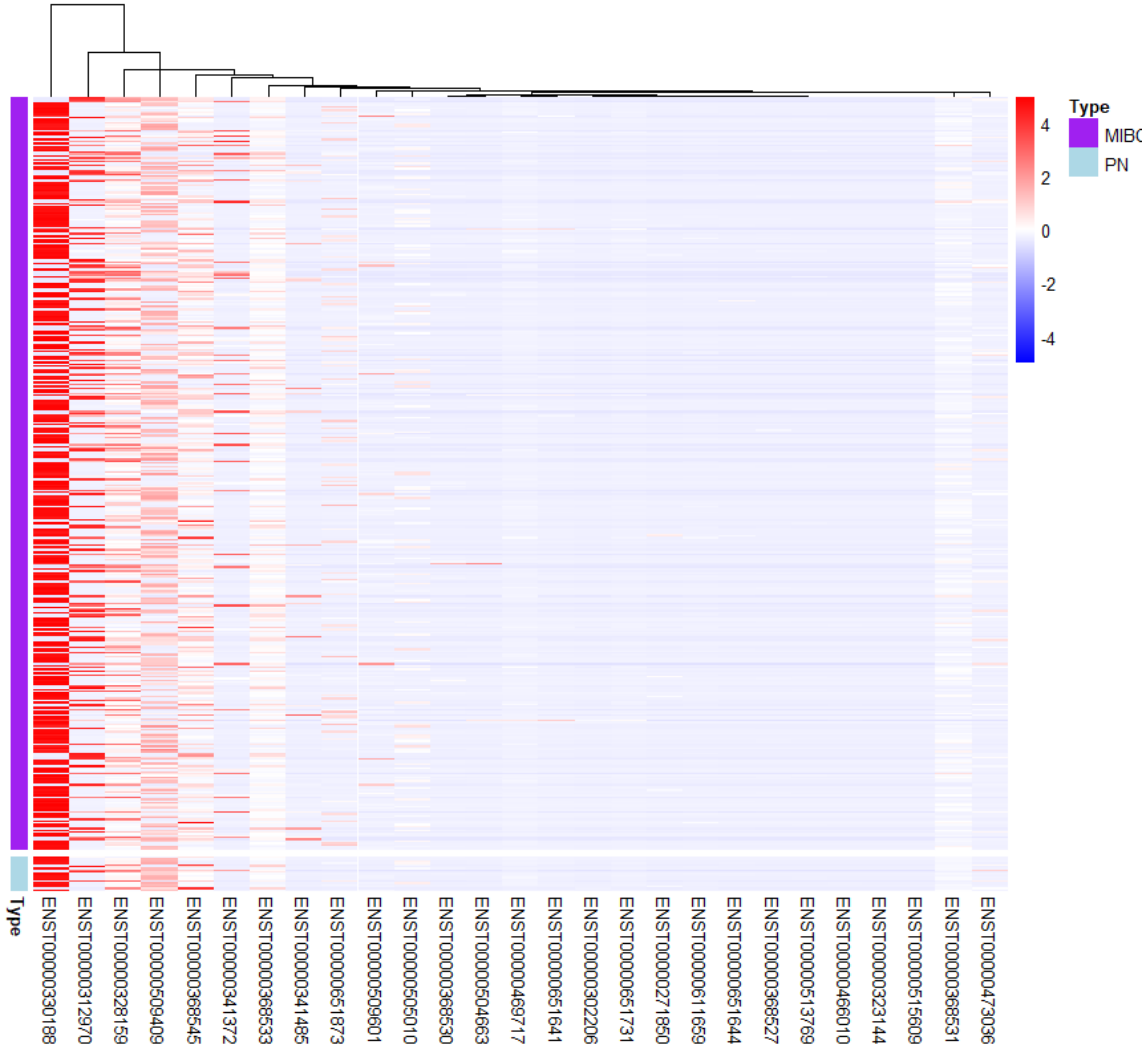


Figure 3.8b TCGA data with differential isoform expression of TPM2. *For each sample (y-axis) the gene expression is contributed by the isoforms indicated by ENSEMBL IDs (x-axis).*



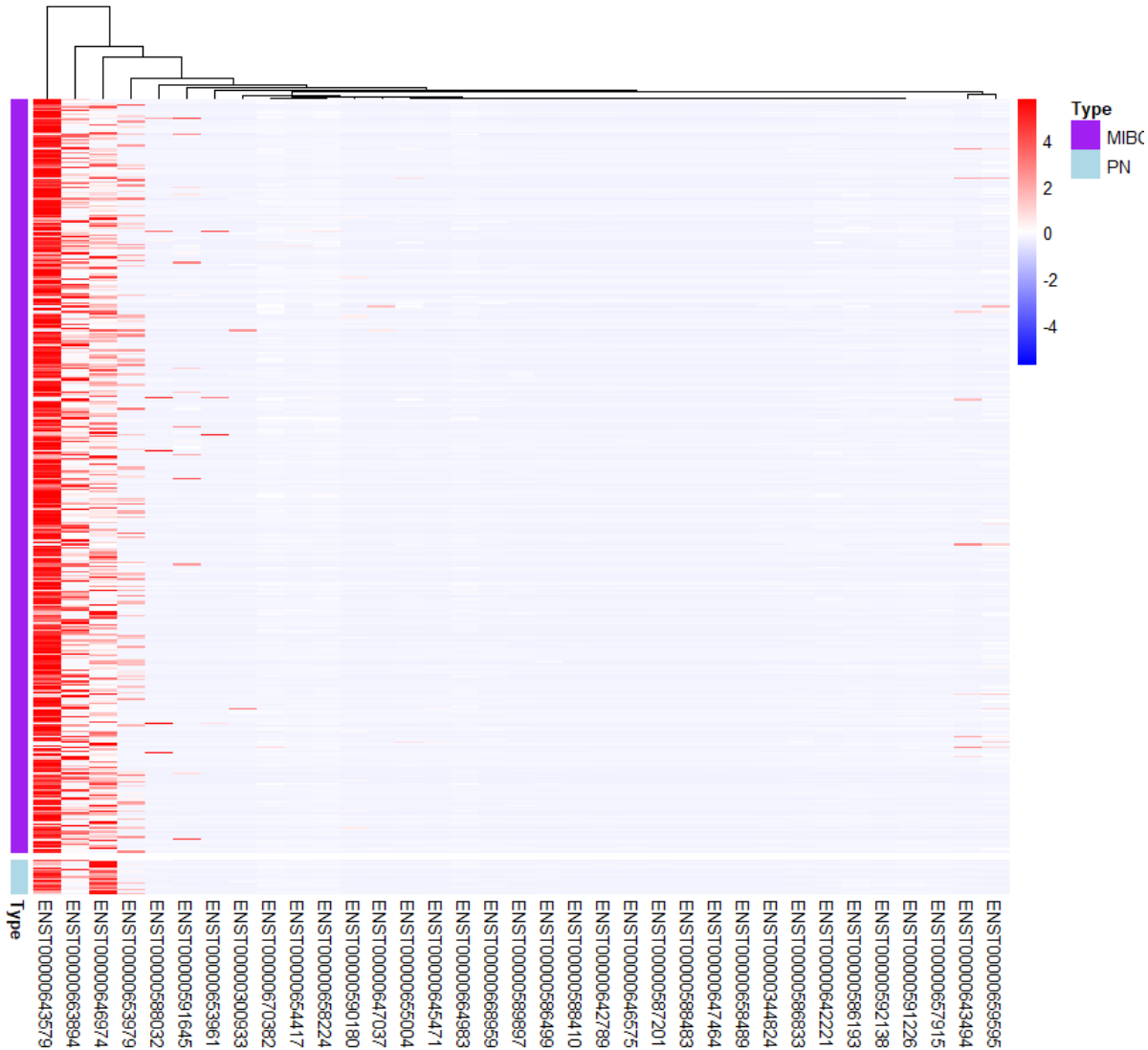
HMW Tpm2.1 (ENST00000378292) is the most dominant isoform across all samples with a mean expression value of 237.60 qTPM in MIBC and 335.60 qTPM in PN tissues. This is followed by HMW Tpm2.3 (ENST00000329305) with a mean value of 23.18 qTPM in MIBC and 22.98 qTPM in PN.

Figure 3.8c TCGA data with differential isoform expression of TPM3. *For each sample (y-axis) the gene expression is contributed by the isoforms indicated by ENSEMBL IDs (x-axis).*



TPM3 expression is represented by LMW Tpm3.2 (ENST00000330188) in both MIBC and PN, with a mean expression of 128.31 qTPM and 59.06 qTPM, respectively. The other contributory isoform is LMW Tpm3.5 (ENST00000328159) with 27.58 qTPM (MIBC) and 15.94 qTPM (PN) and two other non-protein coding transcripts (ENST00000312970 and Tpm3NMD1-ENST00000509409).

Figure 3.8d TCGA data with differential isoform expression of TPM4. *For each sample (y-axis) the gene expression is contributed by the isoforms indicated by ENSEMBL IDs (x-axis).*



LMW Tpm4.2 (ENST00000643579) appears as the most dominant isoform in MIBC tissues with a mean expression value of 171.41 qTPM, followed by HMW Tpm4.1 (ENST00000646974) with 43.38 qTPM. In addition, an intronic transcript (ENST00000663894) is also found with an expression value of 37.16 qTPM. In PN tissues, Tpm4.1 was most dominant isoform with 170.38 qTPM followed by Tpm4.2 with 155.88 qTPM.

3.4.2.3 Other datasets

To establish a control set for analysis, *IsoTPM* was applied to SR-seq RNA data from 3 other non-cancerous/ benign bladder tissue datasets: GEO GSE 133624, University of York, and GTex.

To appreciate the differences in *TPM* isoform expression in cancer and non-cancer, all the datasets were merged, normalized (as per PSI) and single heatmaps per *TPM* gene were generated. The merged dataset included 29 samples of paracancerous bladder tissue from GEO Accession number GSE133624, 7 samples of benign tissue from the GTex repository, and 8 samples of normal urothelium from York Biomedical Research Institute, in addition to 26 CCLE BC cell lines, 85 samples of BC tissue and 6 samples of benign tissue from BCPP, and 412 samples of MIBC from TCGA.

Isoform usage (as per PSI) was compared across the bladder tissue subtypes to determine which isoforms appeared dominant in BC compared to benign/ paracancerous tissues.

TPM1 revealed 4 isoforms that contributed to overall gene expression (Kruskal-Wallis p-value = 0.00124**). HMW Tpm1.4 (ENST00000267966) was found to have increased expression in benign/ normal and paracancerous samples (PSI ± SD = 0.133 ± 0.327, 0.687 ± 0.163, respectively). HMW Tpm1.7 (ENST00000358278) was expressed predominantly in BC tissues and cell lines (PSI ± SD = NMIBC: 0.114 ± 0.142, MIBC: 0.260 ± 0.214, cell line: 0.606 ± 0.216). The LMW isoform Tpm1.9 (ENST00000404484) had similar expression in all subtypes. Interestingly, Tpm1NMD (ENST00000558544) was found in NMIBC followed by benign/ normal samples (PSI ± SD = 0.101 ± 0.108, 0.0859 ± 0.146, respectively) (Figure 3.9a (i,ii)).

Figure 3.9a (i) TPM1 differential expression of isoforms (y-axis) in samples (x-axis). *The isoforms are color-coded according to MW of protein-coding isoforms (HMW and LMW), intron, NMD or canonical isoform of TPM1 gene. The samples are also grouped according to the urothelial subtype.*

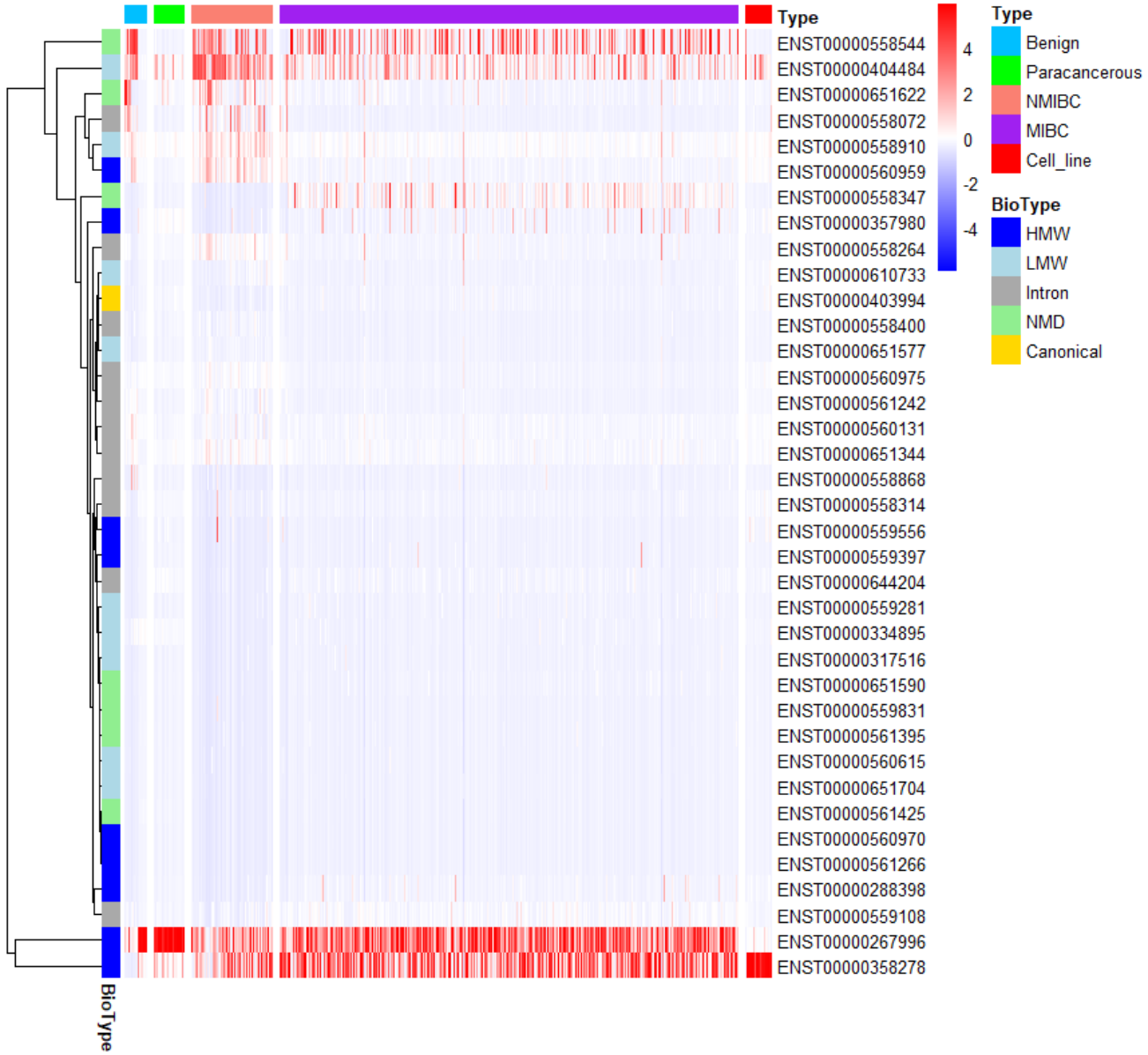
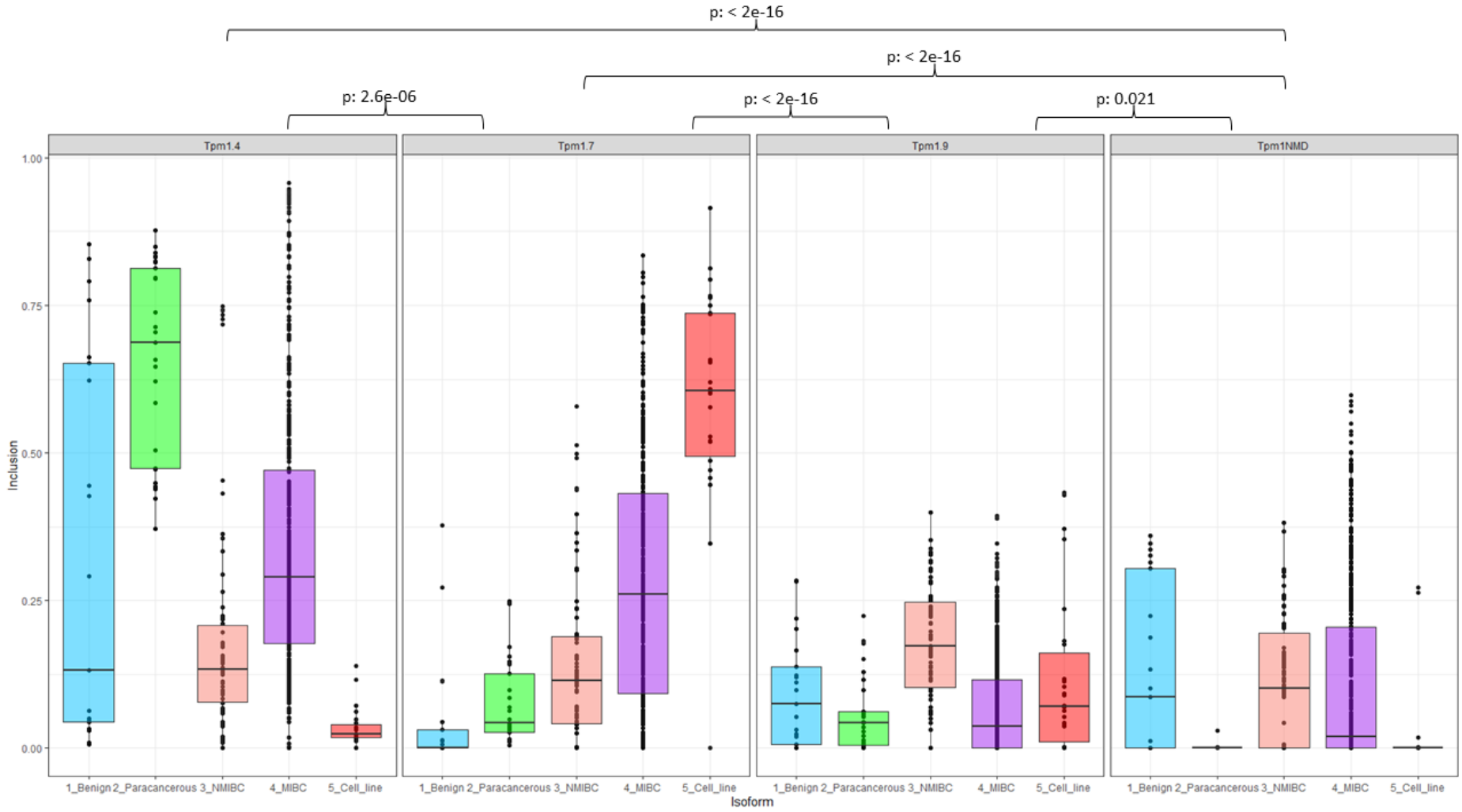


Figure 3.9a (ii) Pairwise comparisons of the top 4 highly expressed isoforms using Wilcoxon rank sum test with continuity correction (p-adjusted Bonferroni). *The y-axis represents the inclusion level (PSI) of the isoform (grey bar) in the 5 different samples (x-axis).*



No significant difference in differential isoform expression was found for *TPM2* across sample types (Kruskal-Wallis p-value= 0.7774). Isoform expression was dominated by the HMW isoform Tpm2.1 (ENST00000378292) (Figure 3.9b (i, ii)). In (7/21; GTex) 33% of benign samples, an additional isoform Tpm2.3 (ENST00000329305) also contributed to *TPM2* gene expression (PSI ± SD = 0.546 ± 0.0337). Tpm2.3 was also expressed in BC cell lines as the alternate isoform (PSI ± SD = 0.338 ± 0.173)

Figure 3.9b (i) TPM2 differential expression of isoforms *(y-axis) in the samples (x-axis) which are clustered according to urothelial subtypes. TPM2 gene encodes for HMW isoforms only, with the canonical isoform highlighted in gold. This isoform appears to dominate TPM2 gene expression.*

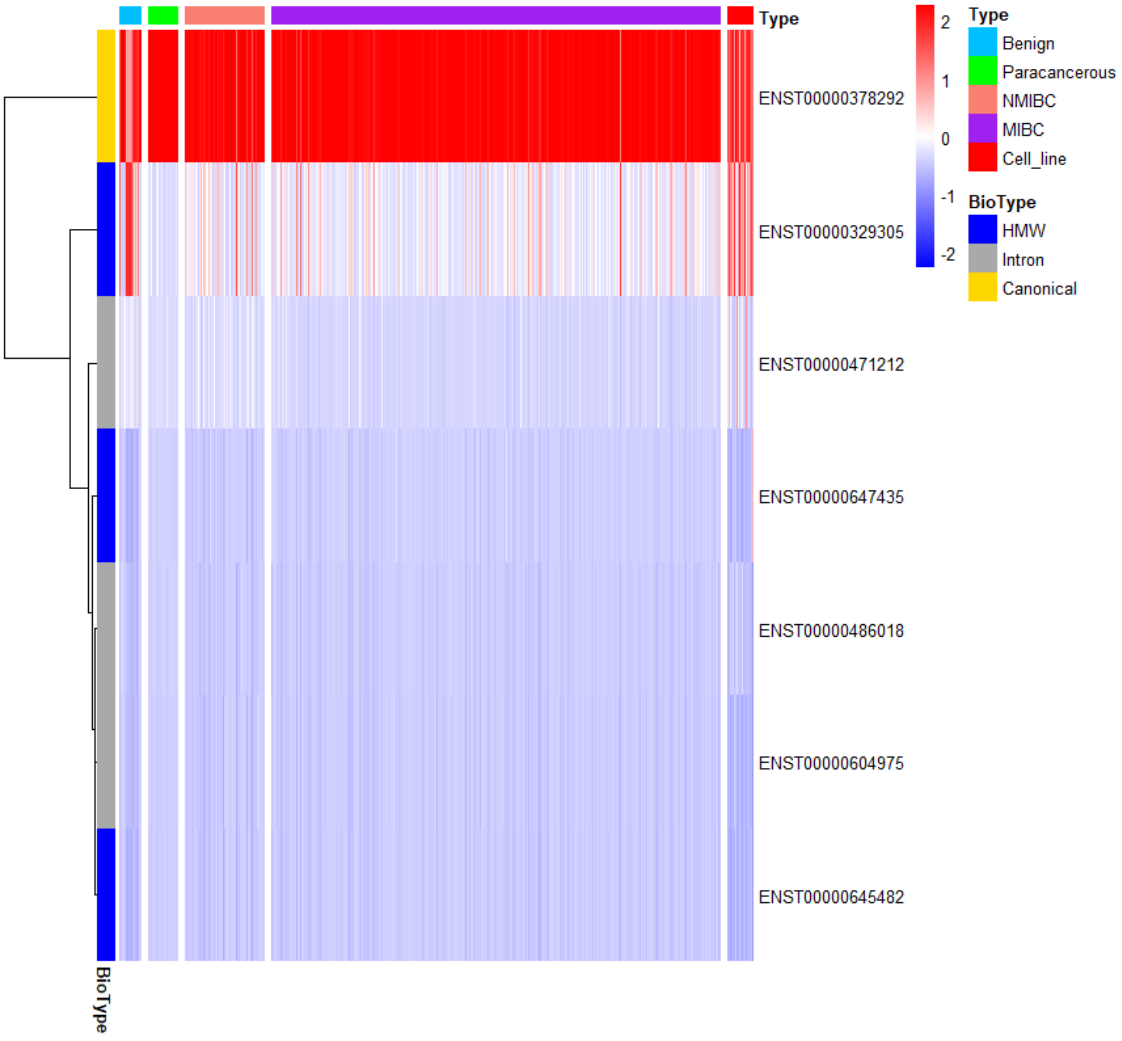
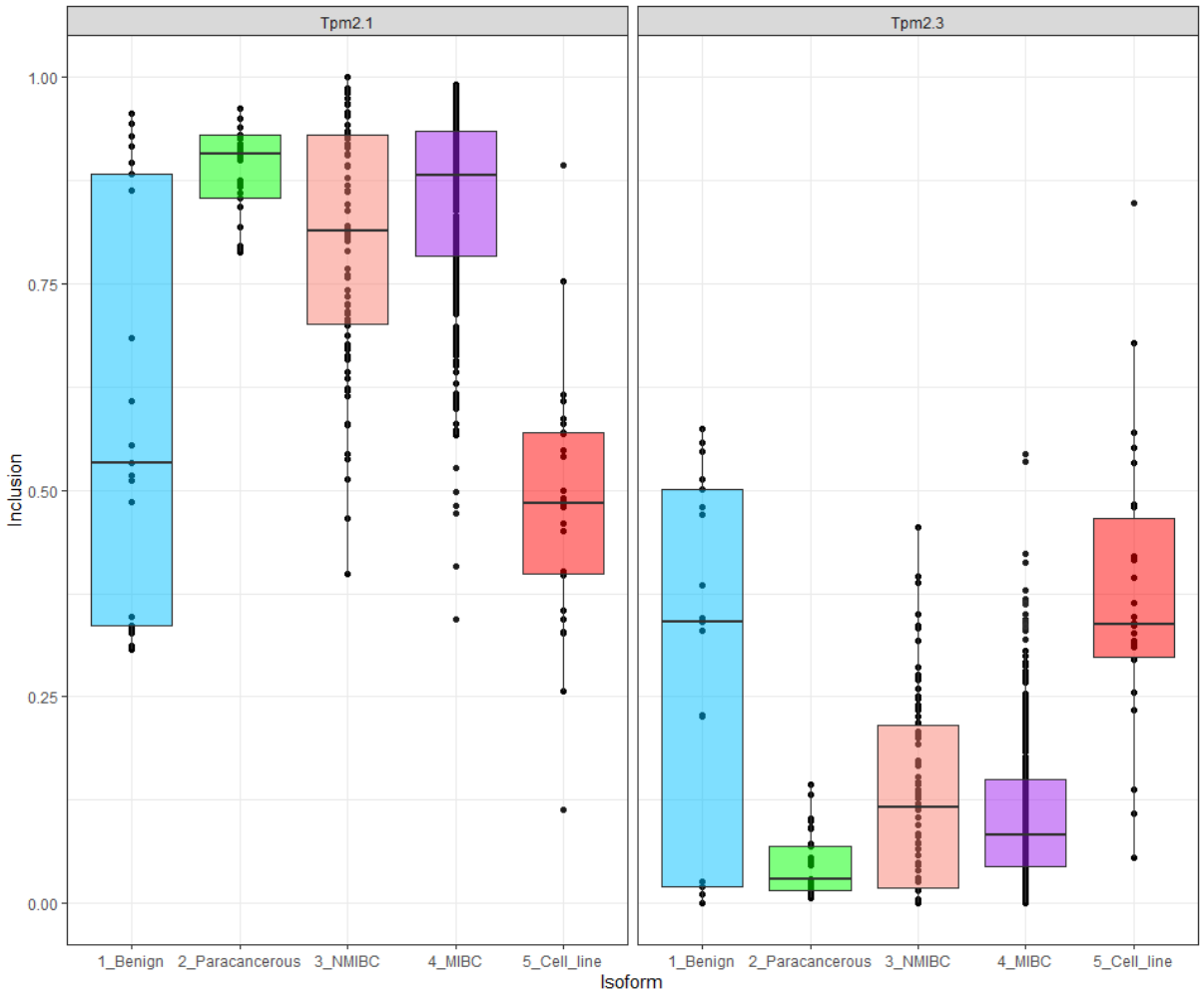


Figure 3.9b (ii) Pairwise comparisons of the highly expressed isoforms across datasets. *The y-axis represents the inclusion level (PSI) of the isoform (grey bar) in the 5 different samples (x-axis). No difference in expression of HMW Tpm2.1 was noted between benign & paracancerous samples compared to tumour samples and cell lines.*



Differential isoform usage of *TPM3* was similar across sample types. LMW isoforms Tpm3.2 (ENST00000330188) and Tpm3.5 (ENST00000328159) had seemingly higher expression in MIBC compared to other subtypes including benign samples (Kruskal-Wallis p-value 1.972e-06***) (Figure 3.9c (i,ii). Furthermore, MIBC tissues had an isoform with a retained intron (ENST00000368545) and NMIBC tissues exhibited increased activity of an NMD transcript (ENST00000509409).

Figure 3.9c (i) TPM3 gene differential expression of isoforms in urothelial sample types and BC cell lines. *Y-axis represents isoforms that have been color-coded according to protein-coding HMW, protein-coding LMW, Intron and NMD.*

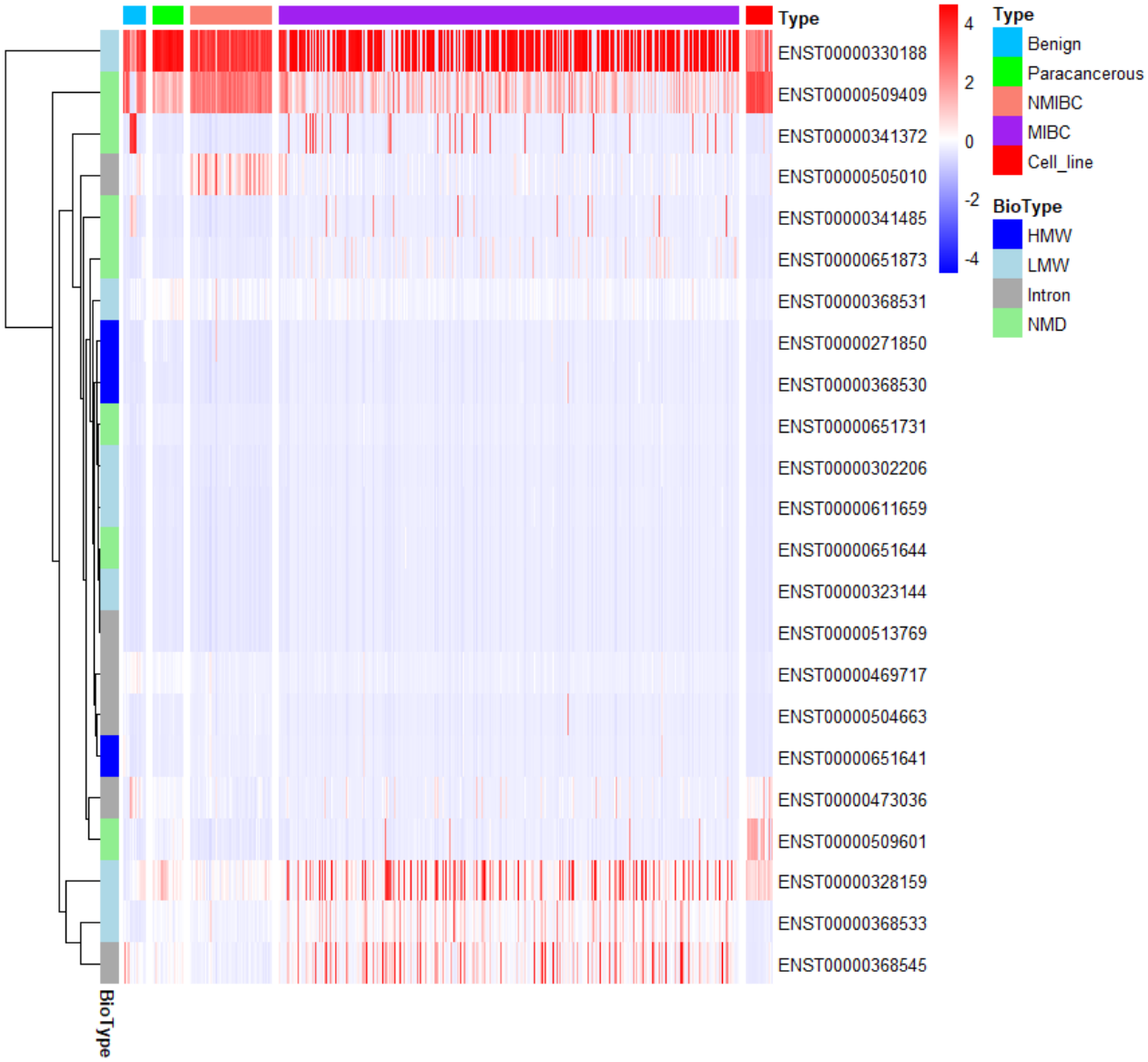
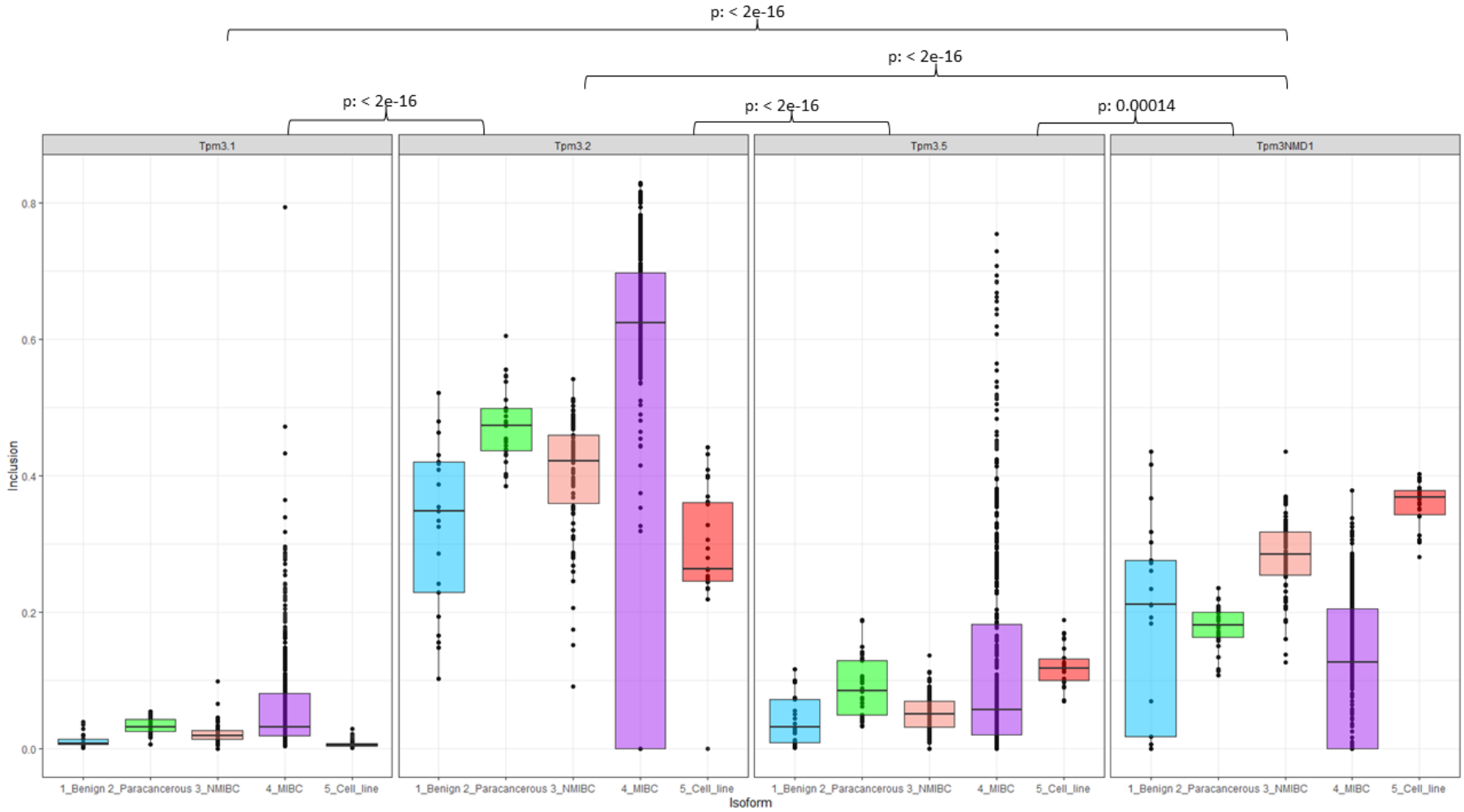


Figure 3.9c (ii) Pairwise comparisons of the top 4 highly expressed isoforms using Wilcoxon rank sum test with continuity correction (p-adjusted Bonferroni). *The inclusion levels (PSI) (y-axis) are shown for the different subtypes (x-axis) in the individual isoform (grey bar).*



TPM4 expression was dominated by LMW Tpm4.2 (ENST00000643579) in urothelial bladder cancer cell lines and tissues (Kruskal-Wallis p-value = 2.858e-05***). In benign bladder tissues and paracancerous sample set, (26/35; 74.3%) displayed increased isoform usage by HMW Tpm4.1 (ENST00000646974) comparatively (Figure 3.9d (i,ii)).

Figure 3.9d (i) TPM4 isoform distribution across the sample subtypes. *The isoforms are color-coded according to HMW and LWM protein-coding, intron and NMD.*

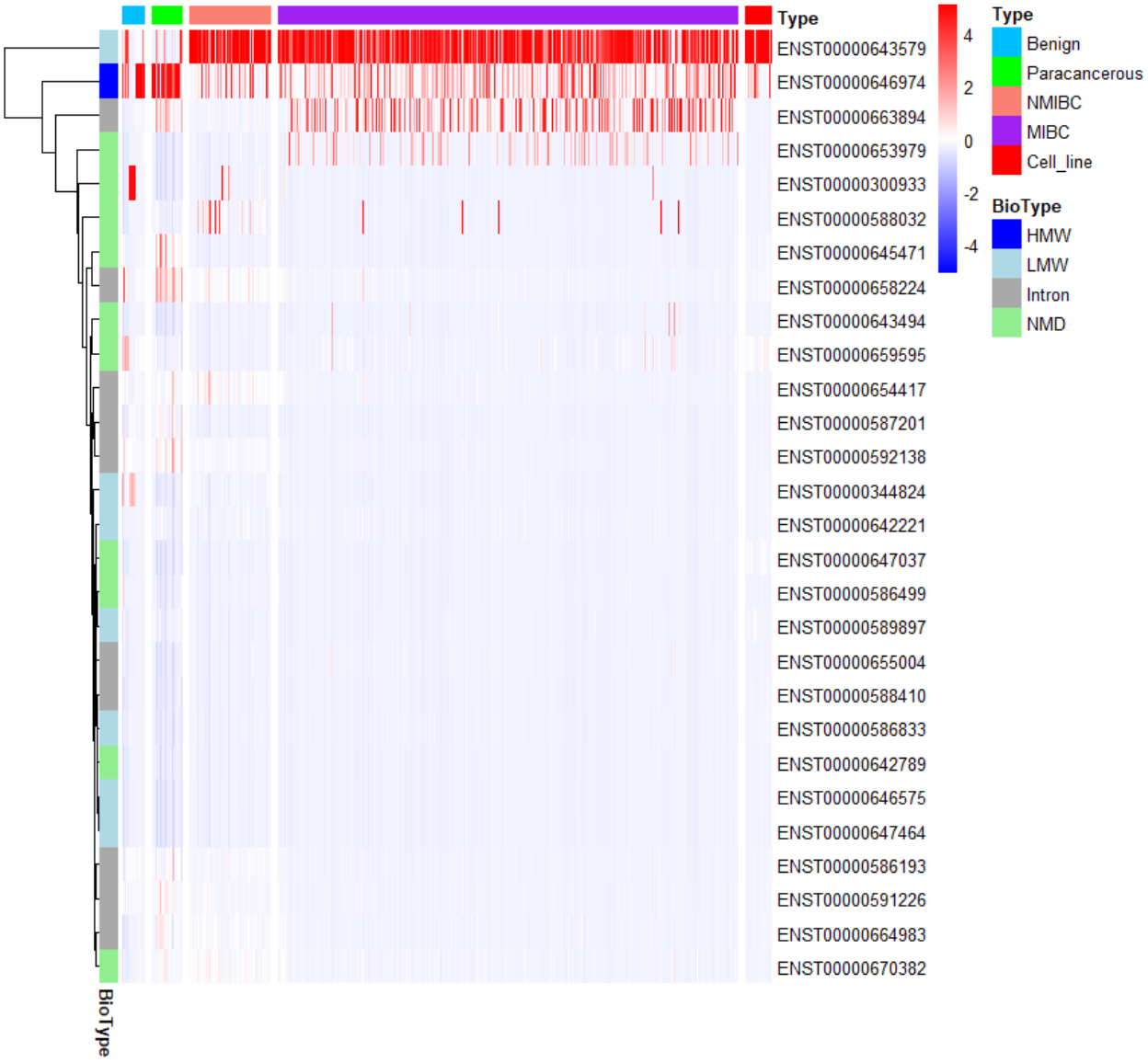
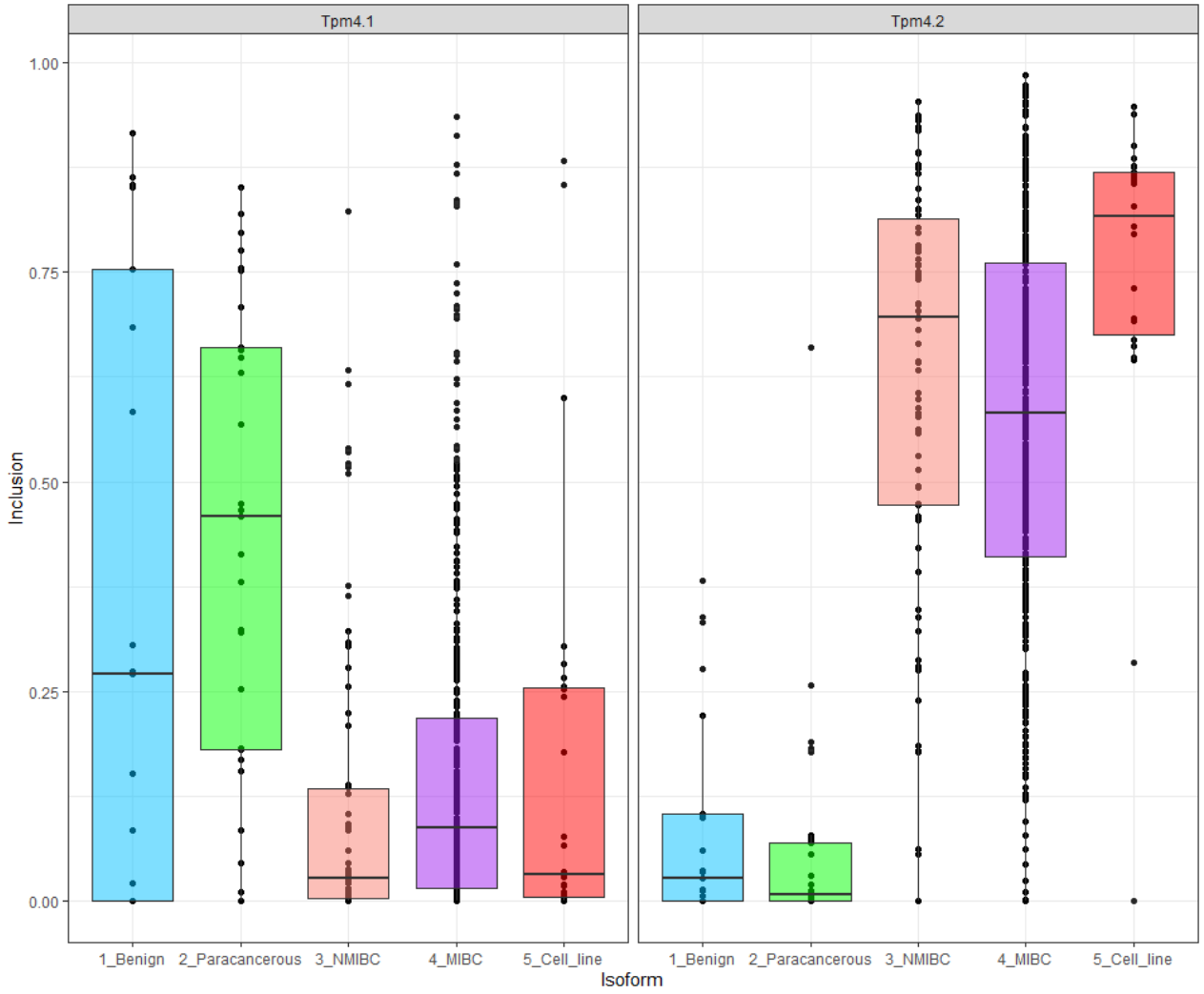


Figure 3.9d (ii) Isoform usage of TPM4 across the sample types. *Tpm4.1 has increased usage (expression) in benign and paracancerous samples, whereas in tumour samples and cells lines Tpm4.2 has greater expression, comparatively (Kruskal-Wallis p-value = 2.858e-05).*



3.5 Summary

- *Isoform quantification using SR-seq can successfully identify differential TPM isoform expression in each of the datasets analysed using IsoTPM: evaluation of TPM isoform expression and usage, interpreting qTPM values and inclusion (PSI), respectively.*
- *Whereas TPM2 and TPM3 have similar differential isoform expression in all urothelial tissue cells, TPM1 and TPM4 genes exhibit differential isoform expression patterns in urothelial bladder cancer.*
- *HMW isoform Tpm1.7 has increased expression in BC with negligible expression in non-cancerous tissues ($p < 0.001$), which otherwise show increased expression of alternative HMW isoform Tpm1.4.*
- *LMW isoform Tpm4.2 had higher expression in BC compared to HMW Tpm4.1 (which was the preferred isoform in non-cancer urothelium).*
- *HMW Tpm2.1 governs overall gene expression for TPM2, with a small contribution by HMW Tpm2.3 in non-cancerous samples and cell lines.*
- *Despite being subjected to NMD degradation, ENST00000509409 in NMIBC tissues was expressed to high levels (median, SD: 0.284 ± 0.0548) making up at least 30% of overall TPM3 gene expression. In MIBC, increased TPM3 expression is observed consequential to upregulation of LMW isoform Tpm3.2 (median, SD: 0.324 ± 0.316).*

Discussion:

AS events can lead to transcriptomic variations which reflect gene expression, and dysregulated (aberrant) splicing has been associated with multiple diseases including bladder cancer (180, 183-185, 187). Following exploration of available computational tools to quantify isoform expression from short-read RNA seq data, we constructed a bioinformatics pipeline that was best suited for quantification of large-scale RNA datasets to characterize *TPM* isoform expression. *IsoTPM* highlighted the potential isoforms in a BC cell line model: HMW Tpm1.7, LMW Tpm4.2 and LMW Tpm3.2 demonstrated high expression, which corroborated previous reports on experiments in BC cell lines (187, 262). HMW *TPM1* isoform Tpm1.4 (previously named α- *TPM*) has been associated with multiple solid tumours as a tumour suppressor (224, 263, 264), and downregulation of the isoform can prevent apoptosis of BC cells (224). Although Tpm3.1/3.2 have been found in literature to be associated with cancer (233), our findings in NMIBC and MIBC patient datasets did not show any difference in the differential usage of isoforms of *TPM3* in benign urothelial cells compared to BC. However, isoform expression for *TPM3* was upregulated in MIBC cells. Long non-coding RNA *MEG3* regulates *TPM1* expression and concomitant downregulation of both is found in BC (224); our experiments also indicated a decrease in expression of HMW isoform Tpm1.4. Furthermore, our data showed that alternative isoform usage of Tpm1.7, expressed by BC cells, was proportional to aggressiveness of disease (262). Differential expression of isoforms was similar in both cancer and non-cancerous samples with HMW isoform Tpm2.1 as the dominant isoform contributing to overall *TPM2* expression.

Chapter 4

BIOMARKER POTENTIAL of TROPOMYOSIN ISOFORMS

4.1 Aims and objectives

Discovery of disease-specific *TPM* isoforms from SR-seq RNA analysis revealed two pairs of isoforms (from *TPM1* and *TPM4*) with potential as diagnostic biomarkers for detection of urothelial bladder cancer. In this chapter, these isoforms were assessed for their utility as biomarkers and their relationship with disease progression and survival outcomes using patient data from the BCPP cohort.

Objectives

- Investigate if relative isoform expression differs in urothelial subtypes using BCPP tissue and TCGA RNA expression.
- Assess the utility of relative isoform expression as a diagnostic marker for BC.
- Assess the impact of high/ low isoform expression on survival and prognosis.
- Investigate if there is a correlation between immune infiltration and *TPM* isoform expression.

4.2 Relative isoform expression levels as an indicator of disease

We observed the pattern of differential isoform expression of *TPM1* and *TPM4* in presence of urothelial bladder cancer as follows:

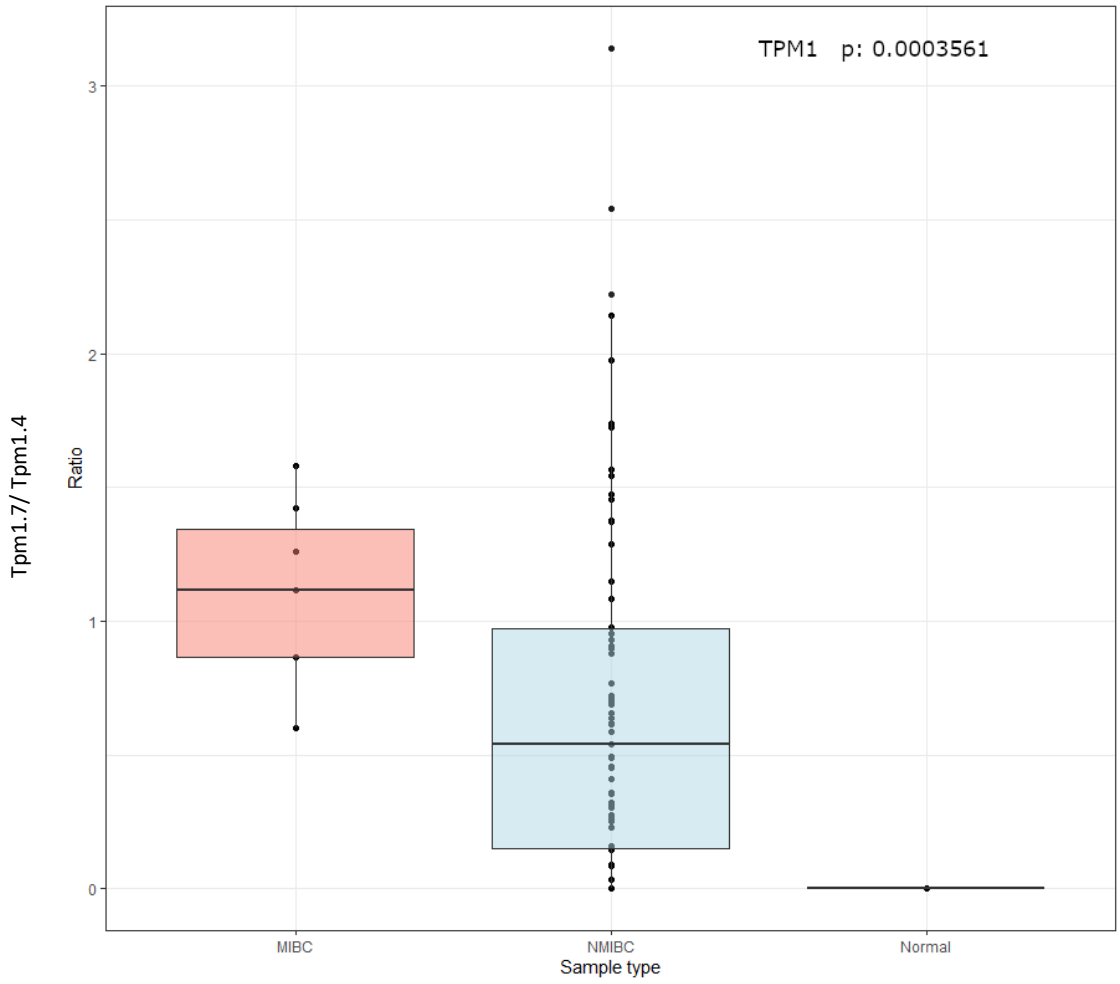
1) There was either preferential usage or upregulation of HMW isoform Tpm1.7 with concomitant decreased usage of Tpm1.4 in bladder cancer. Conversely, HMW Tpm1.4 was found to have higher expression in benign urothelium.
2) Similarly, LMW isoform Tpm4.2 had higher expression compared to HMW isoform Tpm4.1 in bladder cancer, with the reverse isoform usage noted in benign samples.

This led to the hypothesis that ratios of these isoforms might represent potential diagnostic biomarkers for urothelial bladder cancer.

To provide proof-of-concept of the above hypothesis, the relative expression (ratios) of isoform usage from *TPM1* and *TPM4* genes were assessed. We sought to identify if increased expression of Tpm1.7 and Tpm4.2 (relative to Tpm1.4 and Tpm4.1, respectively, calculated as the ratio of isoform usage) would reliably identify urothelial bladder cancer with promising sensitivity.

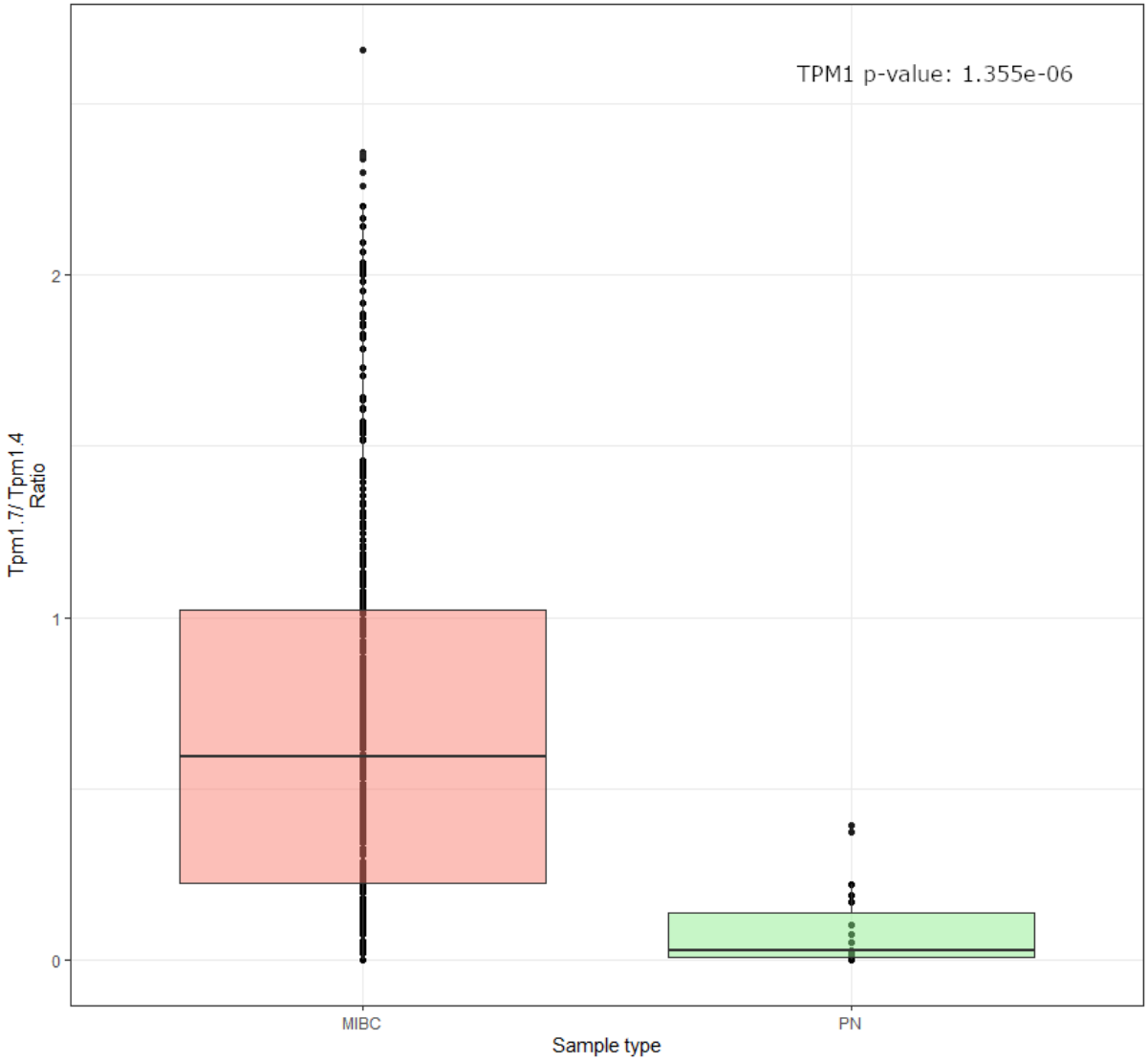
For *TPM1*, the ratio of isoform usage of Tpm1.7 to Tpm1.4 in BCPP was found to be highest in MIBC (median, SD: 1.11 ± 0.347), followed by NMIBC (median, SD: 0.54 ± 0.698) with a value of zero in normal/benign samples as no expression for Tpm1.7 was found in those samples (Kruskal wallis p-value: 0.0003561***) (Figure 4.1a). However, the separation between NMIBC and MIBC is not as clearly segregated as the difference between normal and cancerous samples.

Figure 4.1a Relative isoform expression (ratio) of Tpm1.7 to Tpm1.4 in BCPP samples. *The y-axis represents the calculated ratio of expression of Tpm1.7 relative to Tpm1.4 (qTPM) in the sample type denoted by x-axis. Whereas in normal/ benign samples the ratio is null, in MIBC and NMIBC samples it is >1 and >0.5, respectively.*



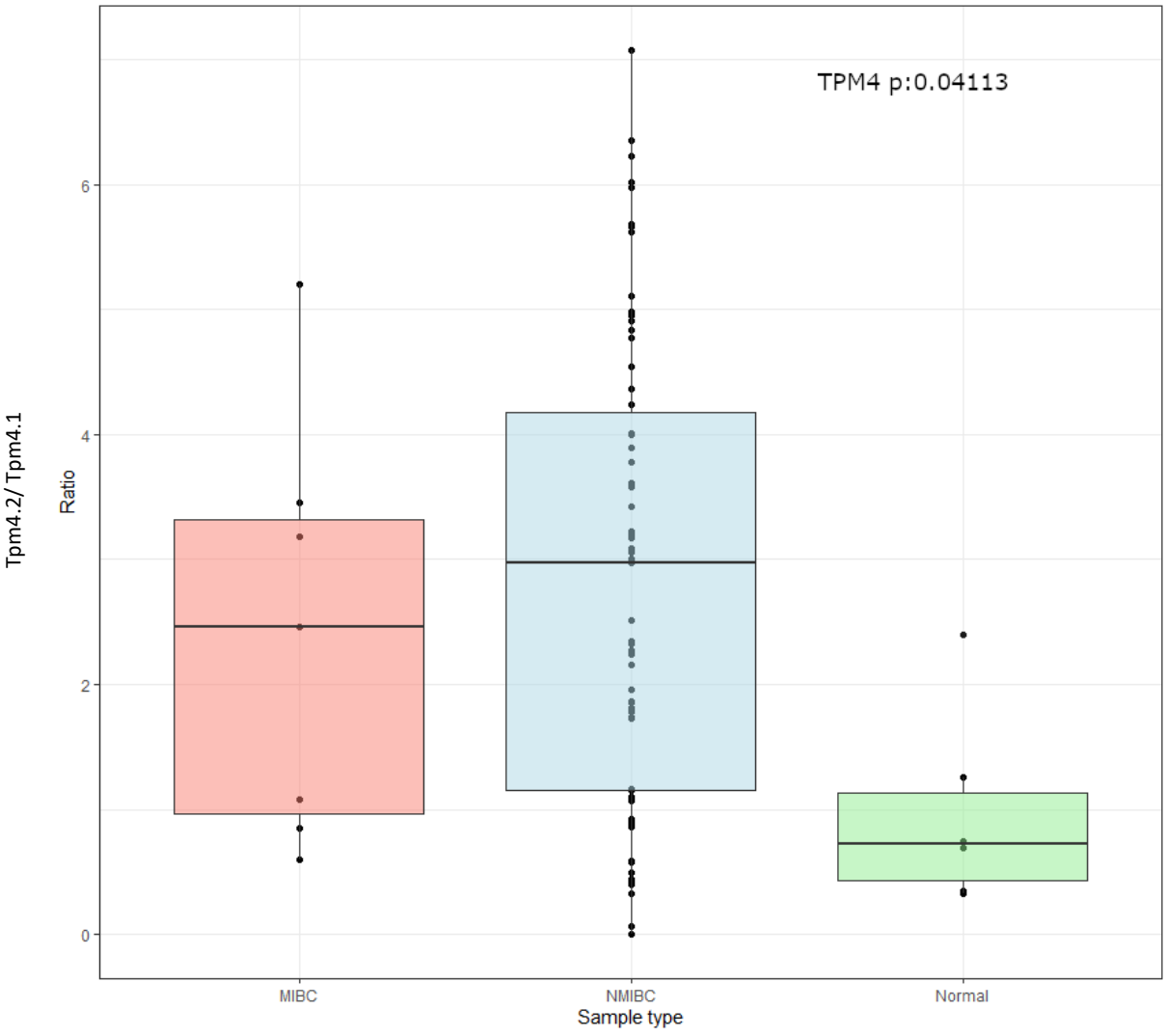
Similarly, the ratio of isoform usage of Tpm1.7 to Tpm1.4 was calculated in the TCGA cohort (Figure 4.1b), comparing MIBC samples (median, SD: 0.594 ± 0.583) with normal (PN) (median, SD: 0.0289 ± 0.122) (wilcox p-value: 1.355e-06).

Figure 4.1b Relative isoform expression (ratio) of Tpm1.7 to Tpm1.4 in TCGA samples. *The y-axis represents the calculated ratio of expression of Tpm1.7 relative to Tpm1.4 (qTPM) in the sample type denoted by x-axis.*



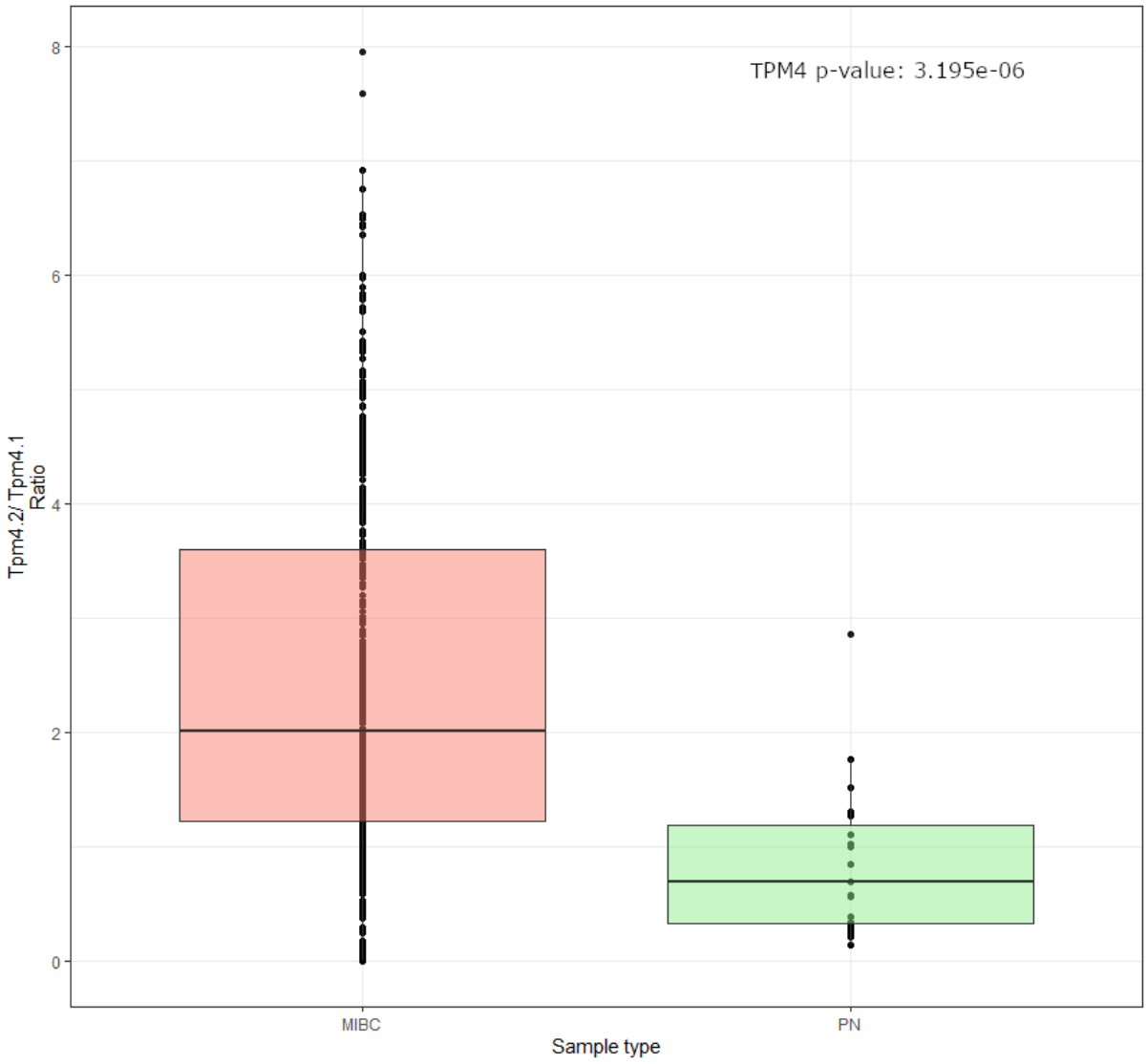
For *TPM4*, the ratio of isoform expression of LMW Tpm4.2 to HMW Tpm4.1 is highest in NMIBC (median, SD 2.97 ± 1.88), followed by MIBC (median, SD: 2.46 ± 1.68). The normal/ benign samples have a relative expression level of <1 (median, SD: 0.721 ± 0.78) (Kruskal wallis p-value:0.04113**) (Figure 4.2a).

Figure 4.2a Relative isoform expression level of Tpm4.2 to Tpm4.1 in BCPP samples. *The y-axis represents the calculated ratio of expression of Tpm4.2 relative to Tpm4.1 (qTPM) in the sample type denoted by x-axis. The levels are noted to be at least x2 greater in cancer samples compared to normal/ benign tissues.*



In the TCGA dataset, the ratio of isoform expression of Tpm4.2 to Tpm4.1 was found to be median 2.01 (SD: ± 1.65) in MIBC, whereas in PN samples it was median 0.698 (SD: ± 0.681) (wilcox p-value: 3.195e-06) (Figure 4.2b).

Figure 4.2b Relative isoform expression (ratio) of Tpm4.2 to Tpm4.1 in TCGA samples. *The y-axis represents the calculated ratio of expression of Tpm4.2 relative to Tpm4.1 (qTPM) in the sample type denoted by x-axis. TPM4 relative isoform expression ratio is higher in MIBC compared to PN samples (wilcox p-value given).*



4.3 Correlations with molecular subtypes

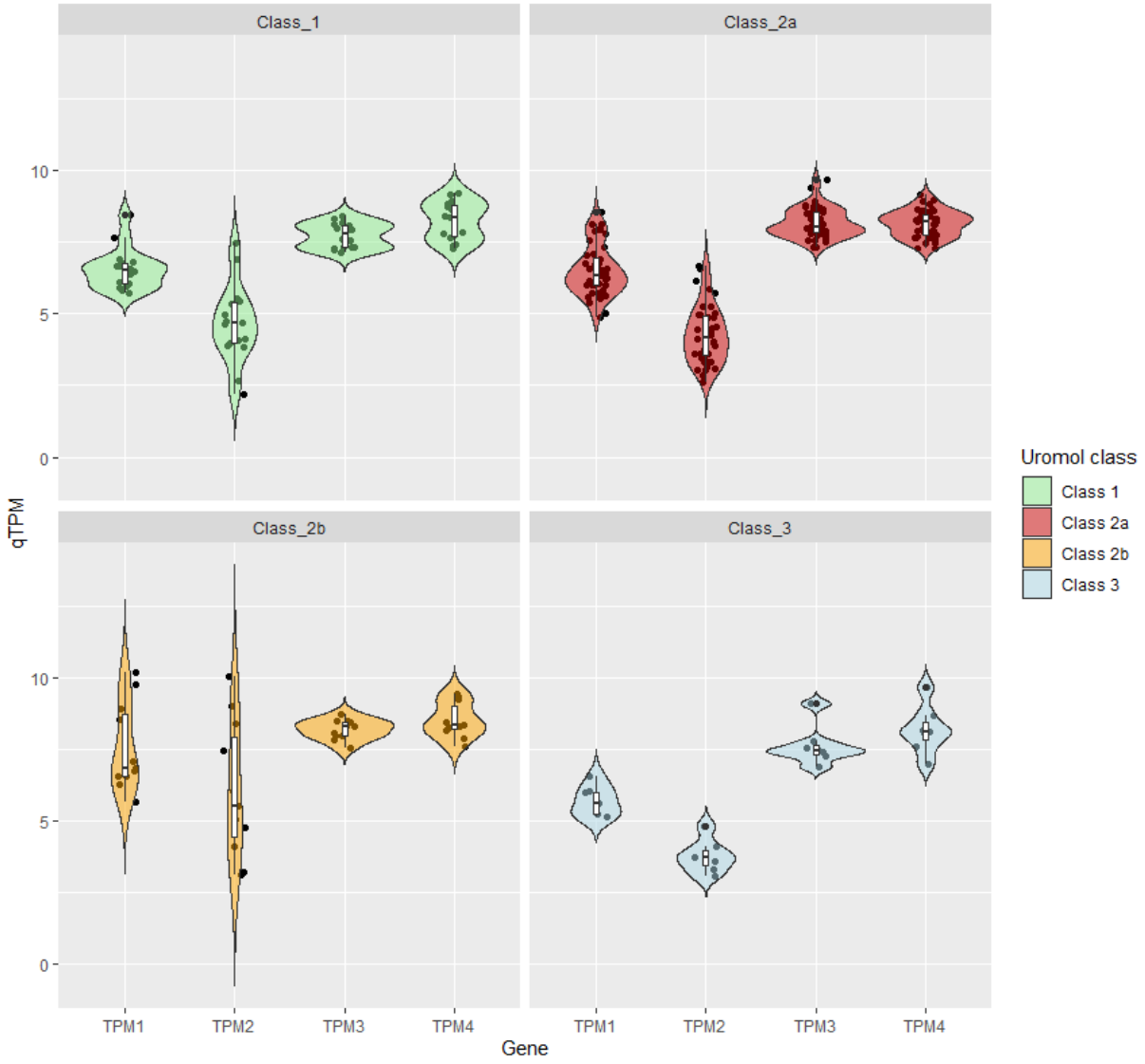
4.3.1 *TPM* gene expression in UROMOL 2021 NMIBC subtypes

The SR-seq expression of *TPM* genes was evaluated, clustering the BCPP samples according to the UROMOL 2021 NMIBC subtype classes (Class 1, 2a, 2b and 3 (177), n= 16, 44, 11 & 7 respectively) (Figure 4.3a). Class 1 included mostly pTa tumours (n= 14/16) whereas Class 2 comprised of G3pT1 tumours (2a= 38/44, 2b= 8/11), and Class 3 was characterized by low grade pTa tumours (n= 4/6). UROMOL Class 2b was found to have highest expression of *TPM1* and *TPM2* (Kruskal Wallis p: 0.003891**, 0.02562*) compared to other classes. *TPM3* gene expression was upregulated in Class 2a and 2b (Kruskal Wallis p: 0.003118**). All four classes had similar expression levels for *TPM4* (Kruskal Wallis p: 0.332) with overall increased expression compared to other *TPM* genes. Class 1 tumours were found to have upregulation of *TPM1*, *TPM3* and *TPM4*. Class 3 demonstrated highest activity by *TPM4*, followed by *TPM3*. (See Table 4.1 and Figure 4.3a)

Table 4.1 | *TPM* whole gene expression in individual UROMOL 2021 NMIBC classes. Values indicate median expression in qTPM ± standard deviation.

Gene	UROMOL Class	Median	St dev
TPM1	1	6.53	0.711
	2a	6.35	0.876
	2b	6.82	1.53
	3	5.61	0.525
TPM2	1	4.67	1.35
	2a	4.18	0.986
	2b	5.53	2.35
	3	3.73	0.577
TPM3	1	7.76	0.429
	2a	8.01	0.511
	2b	8.28	0.348
	3	7.43	0.699
TPM4	1	8.37	0.661
	2a	8.21	0.460
	2b	8.32	0.599
	3	8.09	0.840

Figure 4.3a TPM whole gene expression in BCPP tissues according to UROMOL subtypes. *TPM genes are indicated on x-axis with expression measured in qTPM (y-axis) for each of the 4 Uromol subtypes (facet grids).*



At the isoform level, Class 2b exhibited increased isoform usage of Tpm1.4 and Tpm2.1, whereas Tpm1.7, Tpm3.2 and Tpm4.2 were upregulated in Class 2a. EMT markers and immune cell infiltration were most pronounced in Class 2b. Class 1 tumours were found to have increased usage of isoforms Tpm1.4, Tpm3.2 and Tpm4.2. (See Table 4.2 and Figure 4.3b).

Table 4.2 | Isoform with highest expression in the Uromol class listed, with associated p-value given. Kruskal-Wallis chi-squared test was used to test for significance level.

Isoform	Uromol Class	Kruskal-Wallis Chi-squared test, p-value
Tpm1.4	2b	1.739e-05
Tpm1.7	2a>2b	6.308e-05
Tpm2.1	2b	0.0008
Tpm3.2	2a>2b	0.0003
Tpm3NMD1	2b=2a	0.002
Tpm4.1	2b	0.088
Tpm4.2	No difference in classes	0.514

Figure 4.3b Isoform expression observed in UROMOL subtypes. *The heatmap represents isoform expression (y-axis, normalized z-score of qTPM expression values) for BCPP NMIBC samples (x-axis) ordered according to UROMOL subtypes. Class 1 has increased Tpm4.2 expression followed by Tpm3.2 and Tpm1.4. Class 2a has upregulation of Tpm4.2, Tpm3.2 and Tpm1.7, whereas Class 2b has increased isoform usage of Tpm1.4 and Tpm2.1 in addition to Tpm4.2 and Tpm3.2. Tpm4.2, Tpm4.1 and Tpm3.2 have high expression in Class 3 compared to other isoforms. EMT markers and immune cell infiltration are most pronounced in Class 2b.*

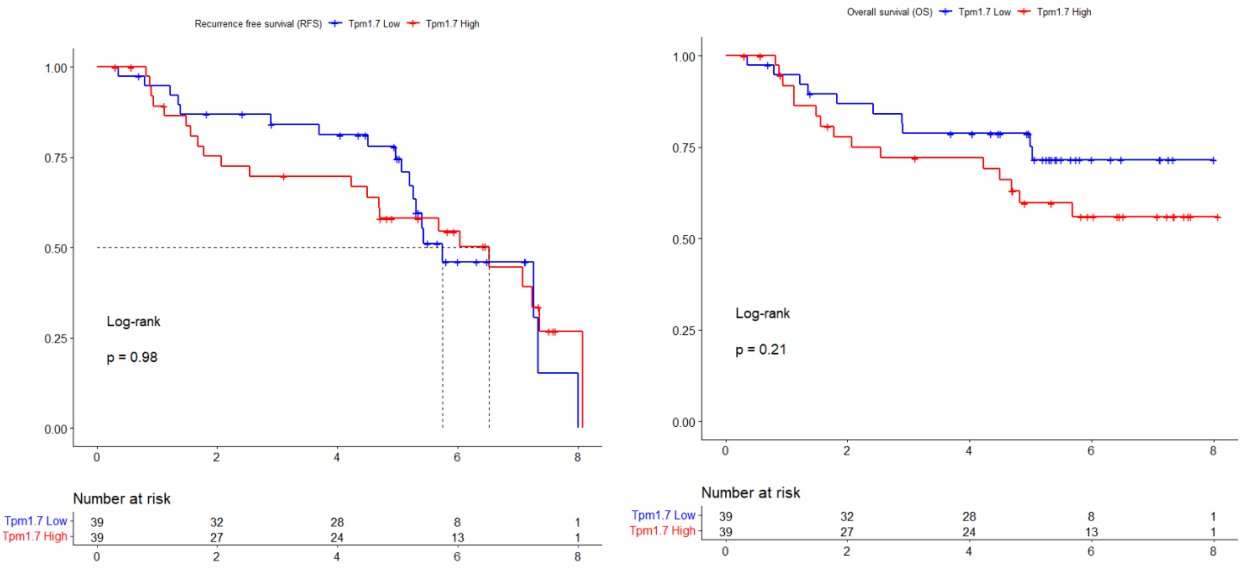


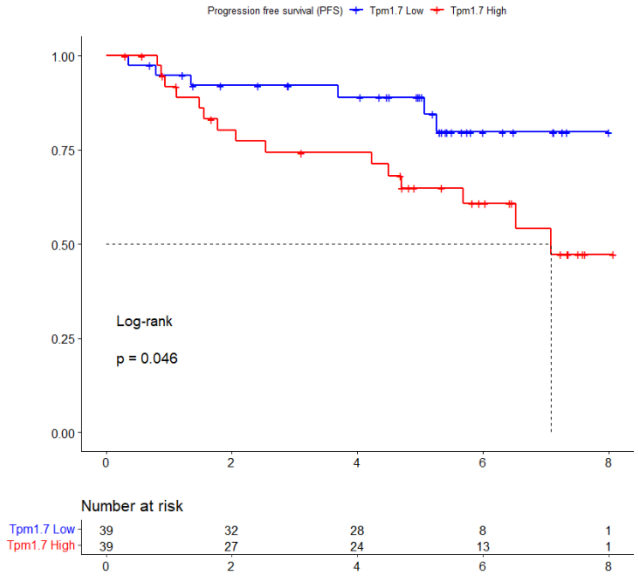
4.4 Isoform inclusion levels (PSI) and relative isoform usage are not prognostic for disease progression

The association of the target *TPM* isoforms (Tpm1.7, Tpm1.4, Tpm4.2 and Tpm4.1) with overall disease progression, recurrence and survival was assessed to investigate their use as prognostic biomarkers. Within the BCPP sample set, there were n= 41 recurrences and n= 21 had disease progression to MIBC. Univariate survival analysis was performed on NMIBC samples (n=78) to assess recurrence-free survival (RFS), progression-free survival (PFS) and overall survival (OS) using the Kaplan-Meier method and log-rank testing applied to generate p-values.

Isoform expression values of target isoforms (measured in qTPM) were split into 2 groups of ‘high’ and ‘low’ (> and < median value, respectively) (Figure 4.4a-b).

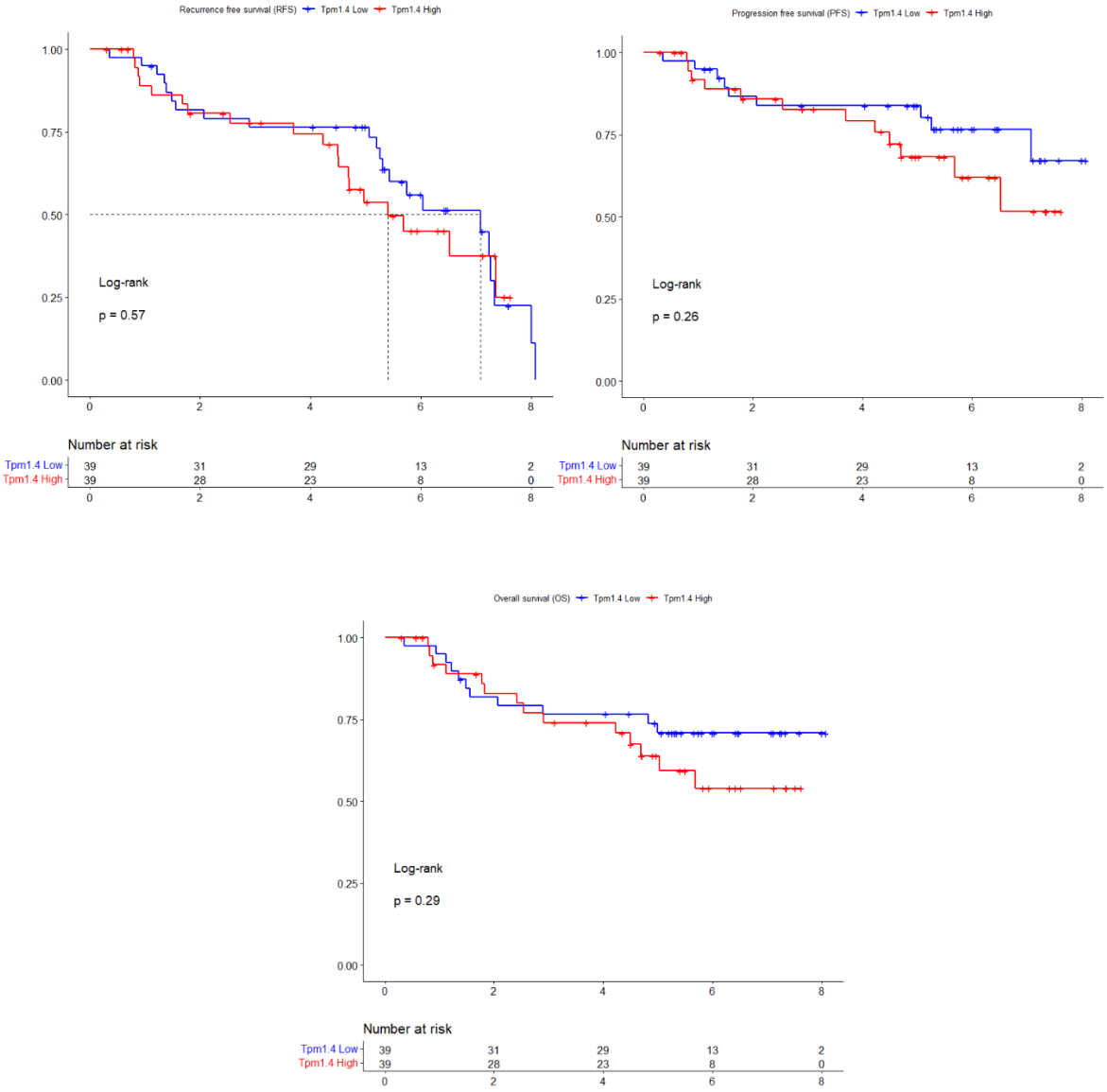
Figure 4.4a (i) Kaplan-Meier curves illustrating recurrence-free survival (RFS), progression-free survival (PFS) and overall survival (OS) for Tpm1.7. *The log-rank test p-values are given for each plot, the x-axis denotes time in years and y-axis represents survival probability.*





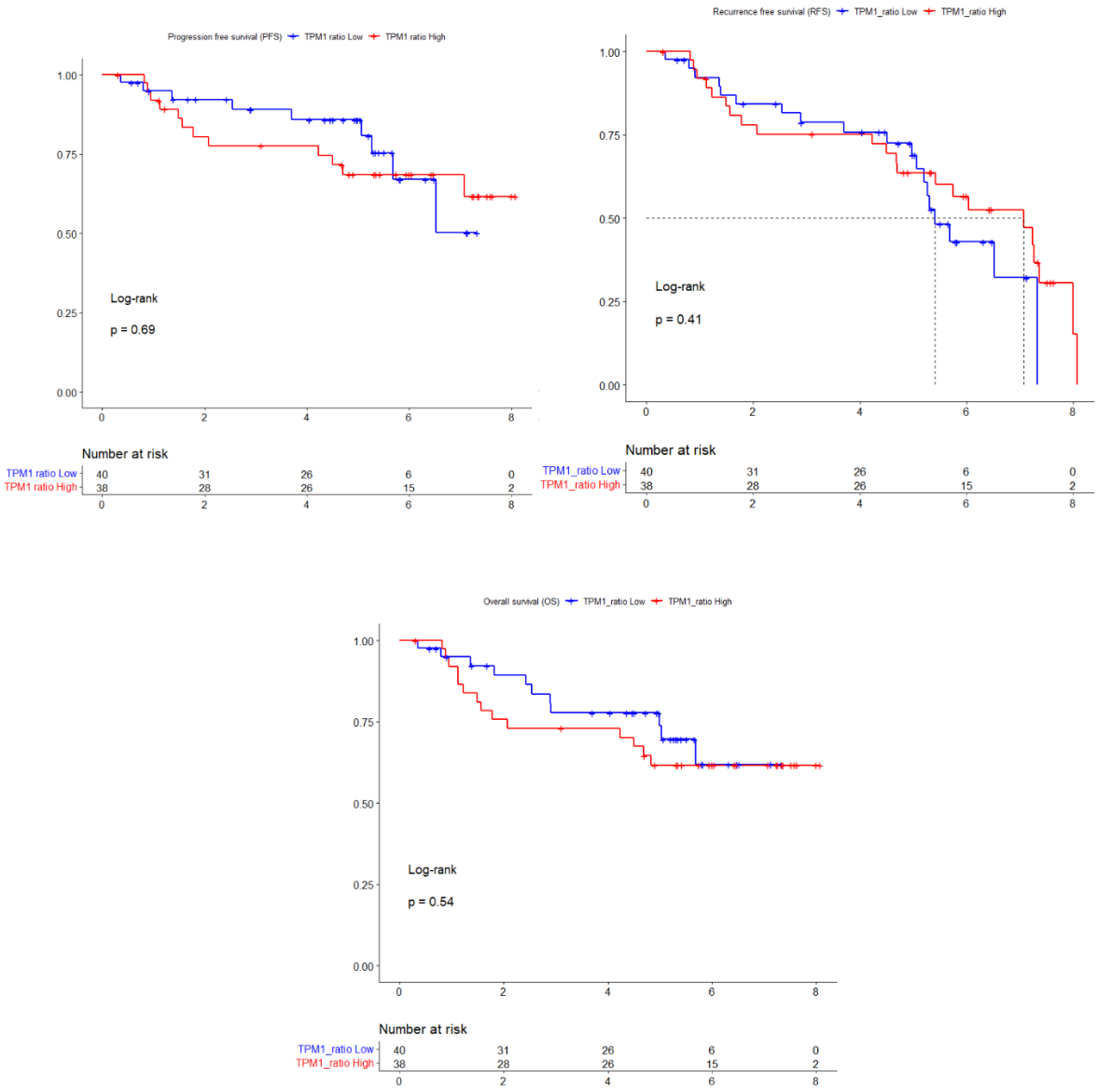
No significant differences in high or low expression of Tpm1.7 was found for NMIBC samples of the BCPP cohort in RFS (p=0.98) or OS (p=0.21). However, high expression of Tpm1.7 appears to impact progression to MIBC, PFS p= 0.046*.

Figure 4.4a (ii) Kaplan-Meier curves illustrating recurrence-free survival (RFS), progression-free survival (PFS) and overall survival (OS) for Tpm1.4. *The log-rank test p-values are given for each plot, the x-axis denotes time in years and y-axis represents survival probability.*



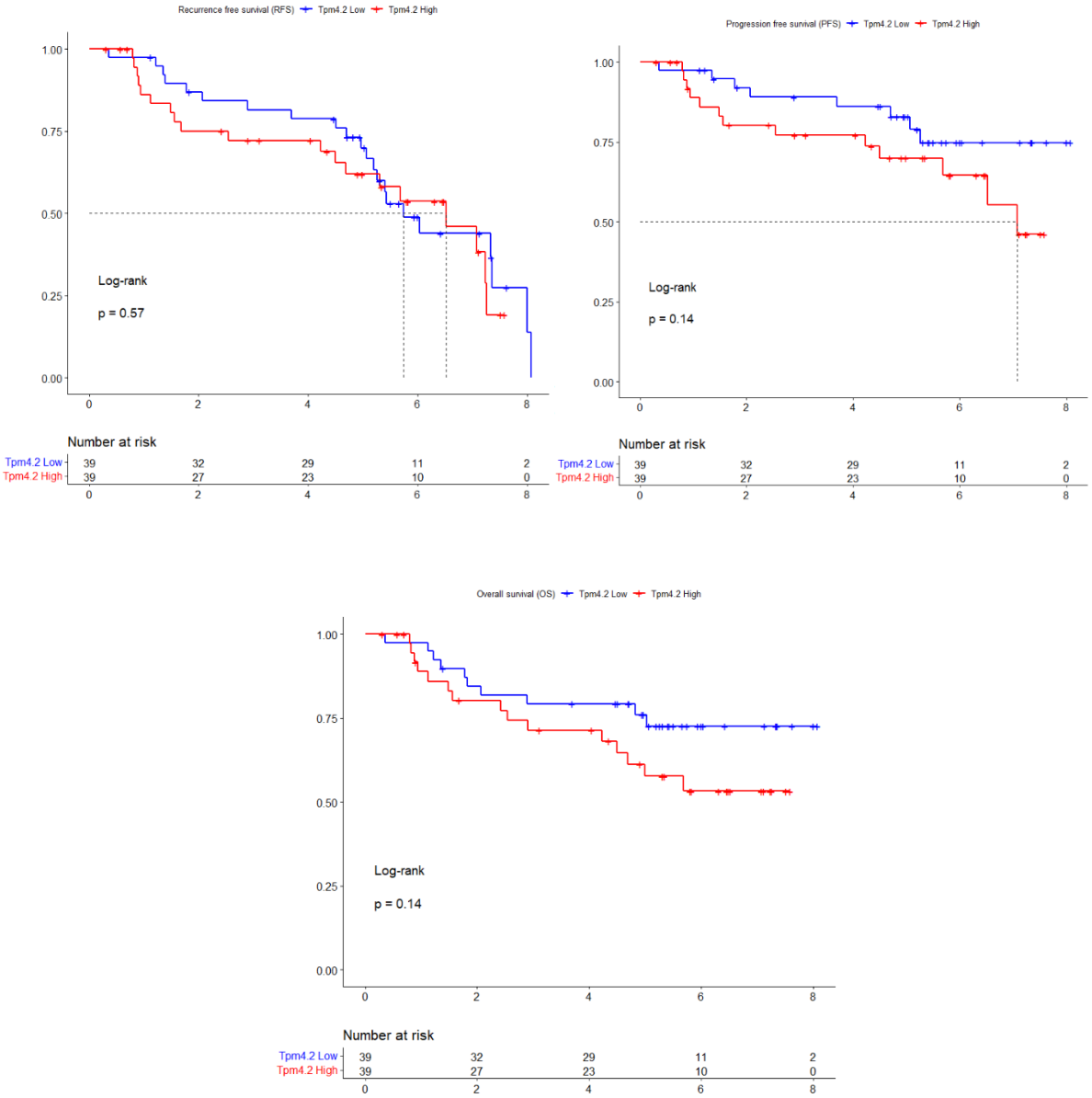
Regarding Tpm1.4, no significant differences were observed in either RFS (p=0.57), PFS (p=0.26) or OS (p=0.29).

Figure 4.4a (iii) Kaplan-Meier curves illustrating progression-free survival (PFS), recurrence-free survival (RFS) and overall survival (OS) in ratio of isoform expression of TPM1: Tpm1.7 to Tpm1.4. *The log-rank test p-values are given for each plot, the x-axis denotes time in years and y-axis represents survival probability.*



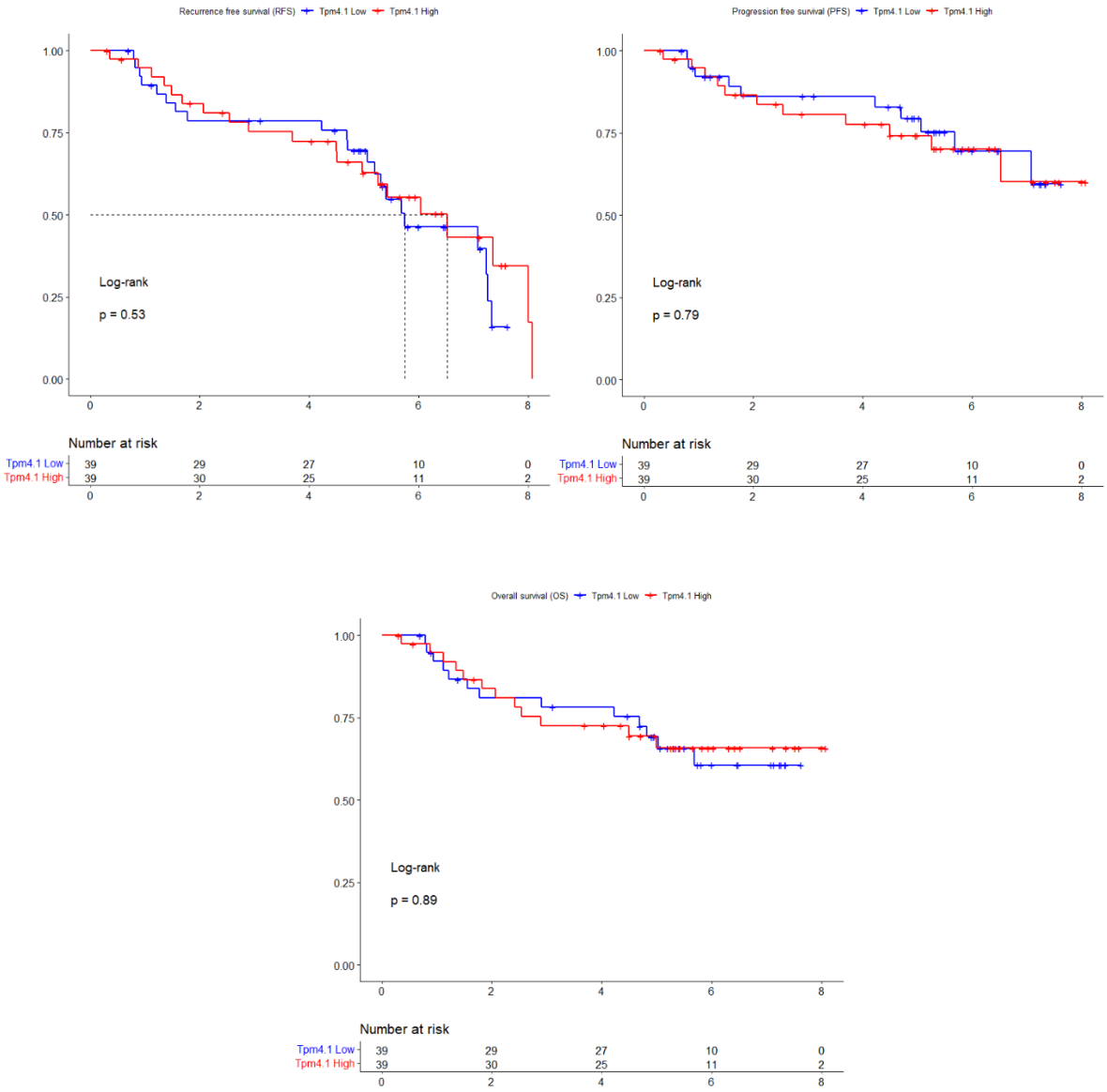
The ratio of expression of Tpm1.7 to Tpm1.4 was assessed for effects on survival probability- no differences were observed in risk of recurrence and/ or progression, indicating high *TPM1* ratio is not a prognostic factor: PFS, RFS and OS (p=0.69, 0.41 and 0.54, respectively).

Figure 4.4b (i) Kaplan-Meier curves illustrating recurrence-free survival (RFS), progression-free survival (PFS) and overall survival (OS) for Tpm4.2. *The log-rank test p-values are given for each plot, the x-axis denotes time in years and y-axis represents survival probability.*



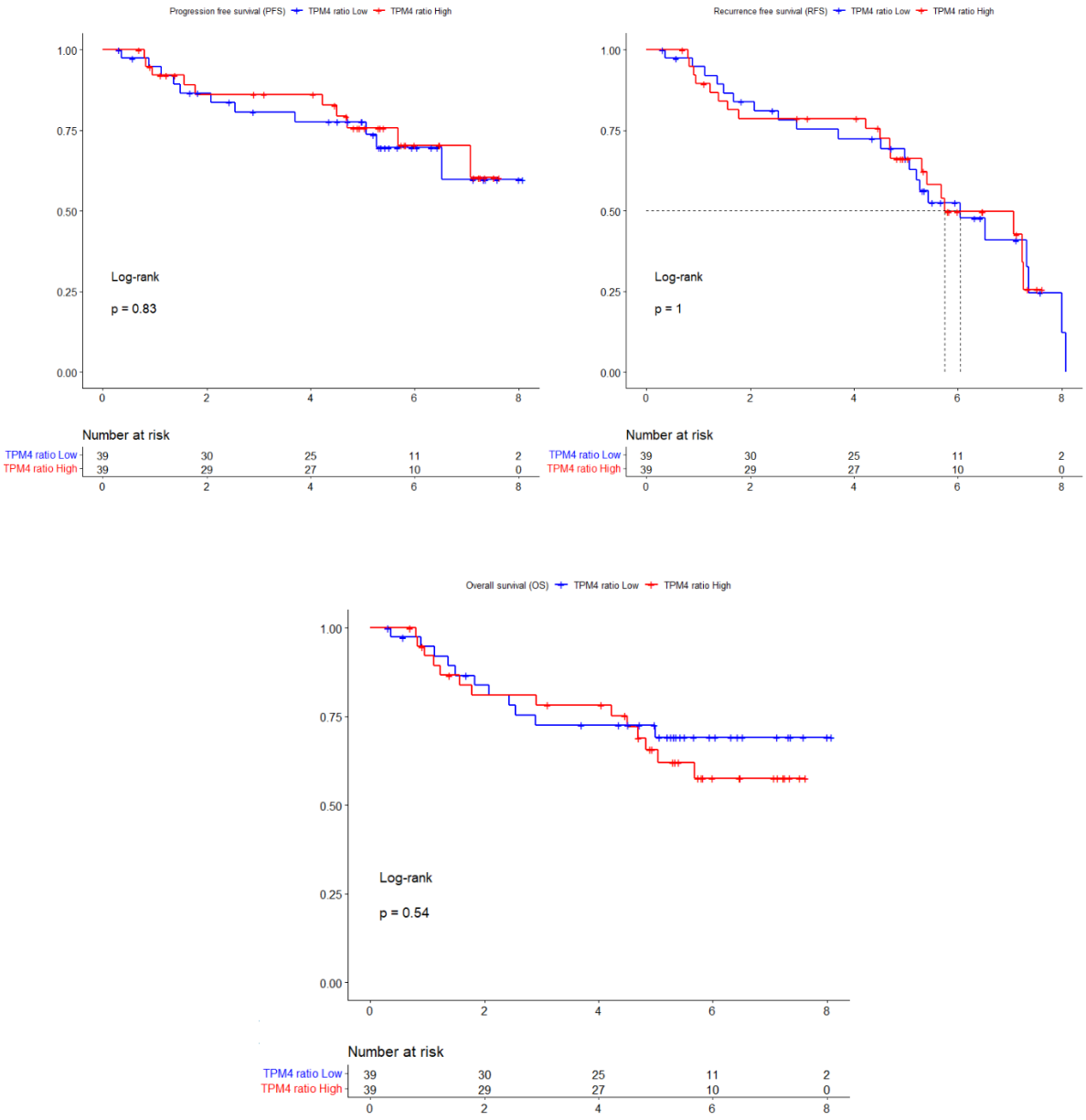
Tpm4.2 isoform expression did not illustrate any effect on survival probability as RFS, PFS and OS ratios were not significant (p= 0.57, 0.14 and 0.14, respectively).

Figure 4.4b (ii) Kaplan-Meier curves illustrating recurrence-free survival (RFS), progression-free survival (RFS) and overall survival (OS) for Tpm4.1. *The log-rank test p-values are given for each plot, the x-axis denotes time in years and y-axis represents survival probability.*



HMW Tpm4.1 isoform had similar outcomes for NMIBC patients regardless of expression, with no effects on recurrence, progression or overall survival (RFS p= 0.53, PFS p= 0.79 and OS p= 0.89).

Figure 4.4b (iii) Kaplan-Meier curves illustrating progression-free survival (PFS), recurrence-free survival (RFS) and overall survival (OS) in ratio of isoform expression of TPM4: Tpm4.2 to Tpm4.1. *The log-rank test p-values are given for each plot, the x-axis denotes time in years and y-axis represents survival probability.*



Similarly, the ratio of isoform usage of Tpm4.2 to Tpm4.1 was evaluated for prognostic utility. It appears that high *TPM4* ratio does not infer poor prognosis, PFS p= 0.83, RFS p= 1, OS p= 0.54.

4.5 Immune infiltration and TPM isoform expression[2]

The objective was to identify if the activity of certain *TPM* genes or isoform expression could be affected by, or at the least, be correlated to immune infiltration. We assessed the immune composition of the BCPP bladder cancer tissues in relation to *TPM* gene expression using the immunophenotyping values estimated from deconvolution of bulk RNA gene expression (Figure 4.5a). As noted previously, UROMOL Class 2b was found to have the highest composite immune score (overall estimate of immunophenotyping of 18 immune cells) and increased expression of Tpm1.4 and Tpm2.1. When assessing the direct correlation of immune cells to *TPM* isoform expression, we found Tpm1.4 was positively correlated with all 18 immune cell types. Similarly, Tpm2.1 was found to have a positive association with fibroblast recruitment and Tpm3.2 had a strong association with mobilization of neutrophils to tumour tissues (Figure 4.5b).

[2] These analyses were performed with assistance provided by a BCRC PhD student, Minghao Yu.

Figure 4.5a Association of TPM gene expression with immune composition of BCPP tissues. *Whereas TPM3 has strong association with B cells, Macrophages, NK cells and T cells (p<0.05), TPM1 and TPM2 have correlation with increased fibroblasts (p<0.05).*

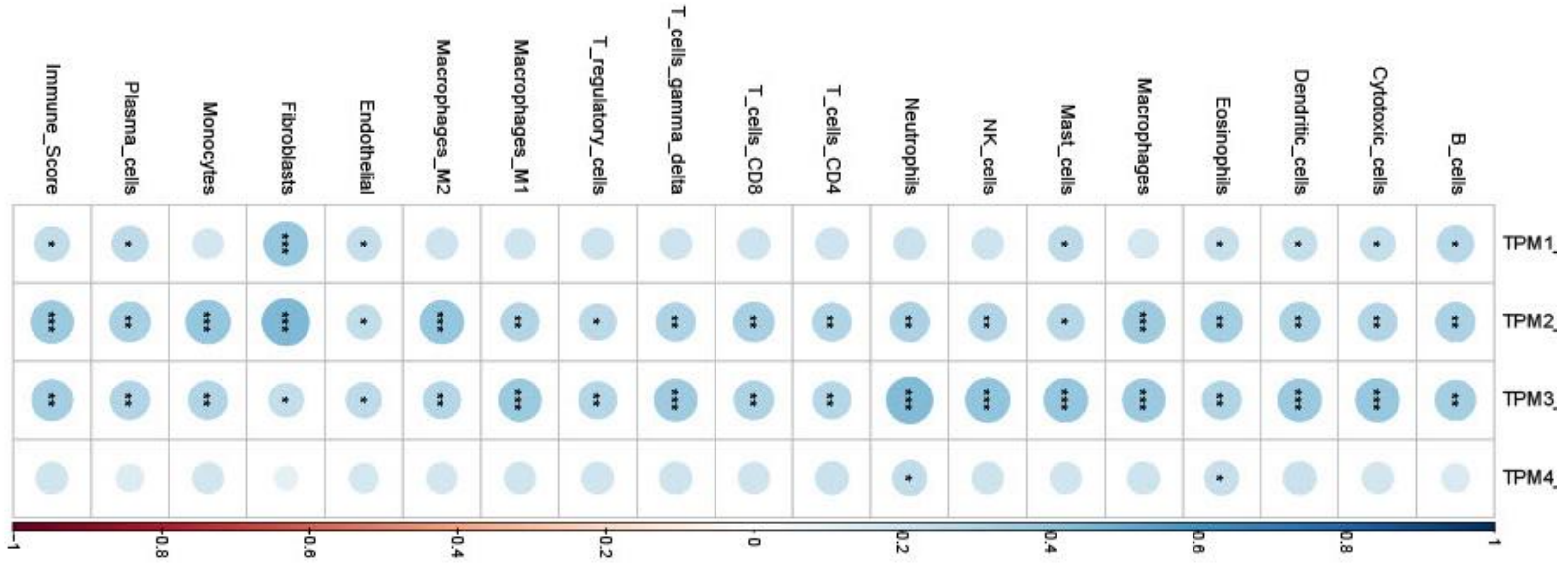
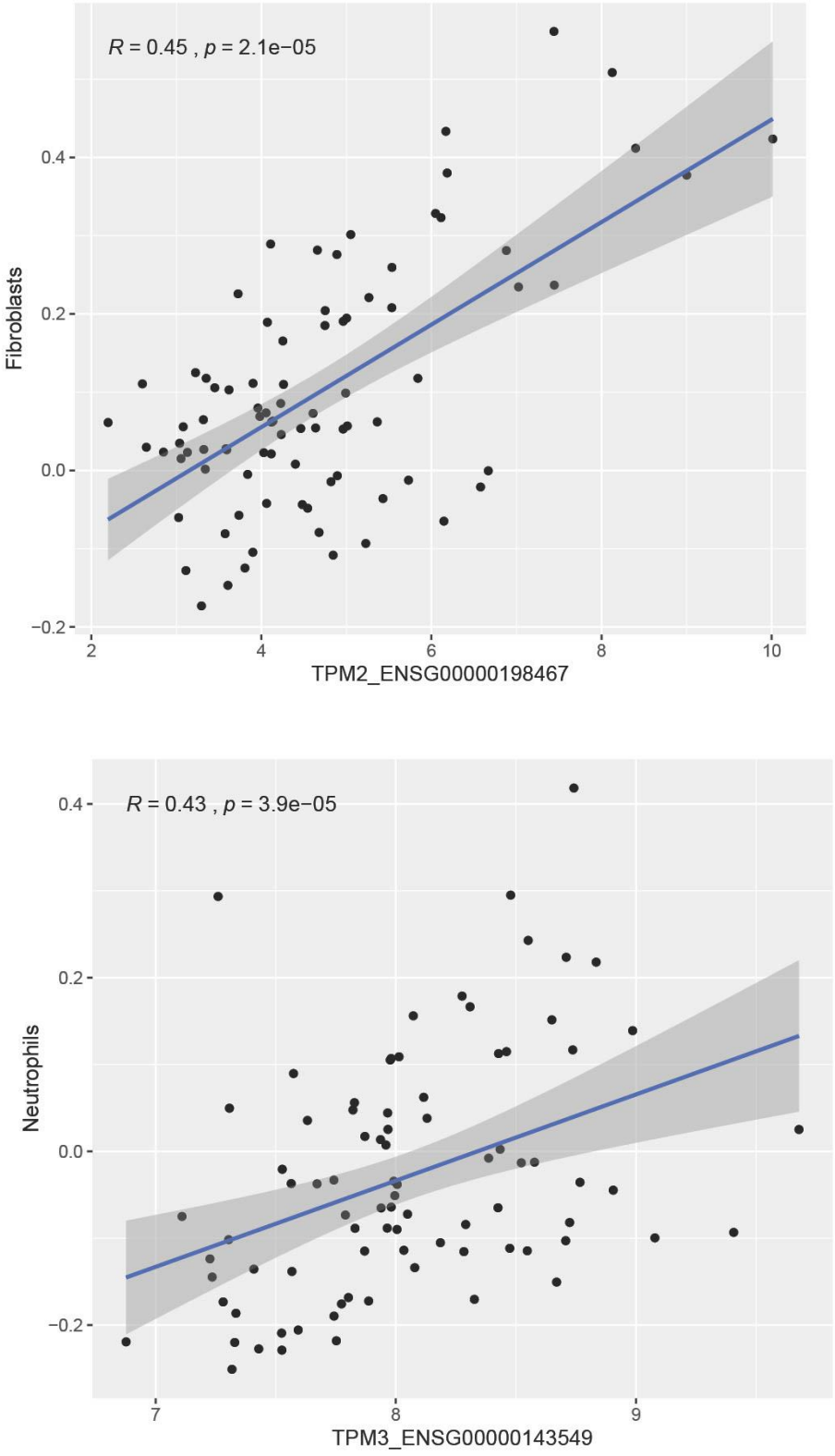


Figure 4.5b TPM gene expression levels in relation to specific immune cells. *TPM2 gene expression is positively correlated with the fibroblast concentration in the tissues (top), whereas TPM3 expression is higher in neutrophil enriched tissues (bottom).*



4.6 Summary

- *Relative isoform expression of two pairs of isoforms from TPM1 and TPM4 indicate diagnostic value for the detection of UBC. Tpm1.7 and Tpm4.2 have higher expression in cancerous tissues in comparison to Tpm1.4 and Tpm4.1, which are the isoforms with high expression in normal urothelial cells.*
- *In UROMOL Class 2a, isoform usage was indicated by Tpm4.2, Tpm3.2 and Tpm1.7, whereas Class 2b had increased expression of Tpm1.4, Tpm2.1, Tpm3.2 and Tpm4.2. Classes 1 and 3 showed increased expression of Tpm4.2 with the latter also having increased Tpm4.1 usage.*
- *Tpm2.1 has a positive correlation with Fibroblast recruitment (p:2.1e-05), whereas Tpm3.2 is found in tissues that have neutrophil enrichment (p:3.9e-05).*
- *Tpm isoforms from TPM1 (Tpm1.7 and Tpm1.4) and TPM4 (Tpm4.2 and Tpm4.1) did not prove prognostic as RFS, PFS and OS were not affected and had no correlation with isoform expression, apart from high expression of Tpm1.7 which was associated with risk of progression of NMIBC to MIBC ($p<0.05$) as observed in NMIBC samples of the BCPP cohort.*

Discussion:

Observations of differential isoform expression patterns in *TPM1* and *TPM4*, in BC and benign urothelial cells led to exploration of the 4 potential isoforms as diagnostic biomarkers (Tpm1.7, Tpm1.4 and Tpm4.2, Tpm4.1). HMW Tpm1.7 and LMW Tpm4.2

have been found to be upregulated in cancer cells. When compared to their counterparts, HMW Tpm1.4 and HMW Tpm4.1, which are expressed in normal bladder urothelium, we were able to determine a ratio of expression which supports diagnostic potential. Further analyses using the single sample classifier of UROMOL 2021 subtyping revealed that Class 2a had higher expression of LMW isoforms compared to other classes. This class mostly comprised of high-risk HG NMIBC tumours, indicating the importance of LMW Tpm isoforms in cellular proliferation and mobility, as discussed previously (207, 233). As per UROMOL, Class 2b tumours have higher infiltration of stomal cells and immune cell differentiation (177). In our data, we found increased EMT markers, by fibroblast recruitment and increased expression of both HMW and LMW isoforms. Additionally, increased immune cell infiltration by neutrophils and macrophages were found in Class 2b, which was proportional to Tpm3.2 expression.

At the censor date, there was no significant difference in survival outcomes between groups with normal or high expression levels of LMW Tpm4.2. However, groups with high expression of HMW Tpm1.7 increased risk of progression of HG NMIBC tumours to MIBC ($p<0.05$).

Chapter 5

CHARACTERIZATION of TROPOMYOSIN ISOFORM EXPRESSION USING LONG-READ RNA SEQUENCING

5.1 Background

Short-read RNA sequencing is a powerful and widely used technique for transcriptomic analysis, but it has its own limitations which include the possibility of missing exon-splice junctions (that do not cover the length of the isoform), wrong quantifications due to inadequate mapping of repetitive or long sequences to ambiguous alignment (secondary to the (short) read lengths typically between 100-300 bp). Sequence similarities may exist within regions of the genome (particularly within isoforms) making it challenging to assign reads correctly. As mentioned previously, adequate read depth is necessary for lowly-expressed isoforms, where SR-seq data may fail to reconstruct the correct isoform sequence even with full coverage (242). Long-read RNA sequencing (LR-seq) techniques allow for end-to-end sequencing without the need for an assembly step. The read lengths can span over 10kb per single read covering the entire length of the gene. LR-seq is advantageous to SR-seq in challenging genomes that have very similar or repetitive sequences shared across the gene family (as in the case of Tropomyosins).

The efficiency of LR-seq is seen with reduced mapping and alignment errors, ability to sequence long tandem repeats (repetitive sequences of two or more bases), detection of structural variants and suitability for de novo sequencing. However, one may argue that these methods might not have enough sequencing depth compared to short-read sequencing and therefore lack power in quantification.

Two prominent technologies exist that provide the LR-seq platform: Pacific Biosciences (PacBio) and Oxford-Nanopore technologies (Nanopore/ ONT). Whereas PacBio utilizes single molecule real-time (SMRT) sequencing, ONT is dependent on the passage of nucleotides through a (nanopore) membrane and their detection by change in the electric current.

The differences and similarities between the two are highlighted in the table below.

Table 5.1 | Comparison of Third generation sequencing platforms: Pacific Biosciences and Oxford Nanopore Technologies (265)

	PacBio Sequel II SMRT cell	**ONT**
Principle of operation	SMRT sequencing: real-time monitoring of DNA polymerase activity to incorporate fluorescently labelled nucleotides to reconstruct DNA strands	Nanopore membrane: identification of nucleotide bases through changes in electrical activity as they pass through a nanopore membrane
Sample prep	Amplification of cDNA to create a SMRTbell library requiring high quality RNA	Simpler technique with use of cDNA
Read lengths	Very long reads 1kb- 15kb	Variable long reads, 100 – 10kb
Throughput	Low	High
Accuracy	Relatively higher >99%	High 92%
Error rate correction	Corrected with circular consensus sequencing	Self-correction tools (266, 267)

5.2 Aims and objectives

Having discovered the potential isoforms of interest from *IsoTPM* applied to *in silico* data from SR-seq, we proceeded to validation of *TPM* isoform expression in cell lines and patient samples. This chapter focuses on the characterization of tropomyosin isoforms in bladder cancer tissues and cell lines using long-read RNA sequencing.

Objectives:

- Design and optimization of long-read RNA sequencing method for *TPM* isoform characterization in BC cell lines and patient samples.

- Comparison of isoform expression in cell lines, tumour tissues and benign bladder urothelial tissues.

- Correlation of SR-seq with LR-seq results to verify target isoform expression

5.3 Long-read sequencing sample selection

To interrogate the isoform landscape of *TPM* genes in BC, we performed long-read RNA sequencing (LR-seq) in parallel on matched samples from BC tissue specimens and cell lines for comparison with *in silico* analyses of SR-seq. LR-seq is not commonly used for quantitative analysis but as a qualitative check to identify isoforms of interest. With the use of targeted isoform sequencing using primer pairs at 5' and 3' ends to cover the entire length of the isoforms, we were able to capture full-length reads of isoforms. Targeted isoform quantification was performed with SMRT-sequencing using the PacBio platform to provide full-length isoform circular consensus sequence (CCS) reads (discussed in next section). CCS reads were demultiplexed using the exact barcode matches on the primer sequences and aligned to GRCh38 human reference genome using Minimap2 (268); any unmatched reads or reads with duplication of barcodes were excluded. The library comprised of the following samples: 12 benign bladder tissues, 6 paracancerous, 18 LG-NMIBC (G1/G2 Ta), 37 HG-NMIBC (G3 Ta/T1), 17 MIBC (G3T2+) tissues, and 5 BC cell lines.

5.3.1 PacBio SMRTBell sequencing and the Iso-Seq pipeline

PacBio developed an isoform sequencing (Iso-Seq) pipeline which is a multistep protocol for transcriptomic sequencing incorporating *(1)* library construction *(2)* sequence data collection and *(3)* data processing (269). It involves initial conversion of high-quality RNA into cDNA prior to PCR amplification. The double-stranded PCR products are then ligated with hairpin adapters at both ends to create a closed-loop circular construct of 'SMRTbell'. This is fed into a SMRTcell housing nanoscale

observation chambers called Zero Mode Waveguides (ZMW). At the base of each ZMW is a DNA polymerase that attaches to the hairpin adapter and incorporates fluorescently labelled nucleotides (fed into the ZMW chamber) to replicate the DNA. The light waves emitted in the process are recorded in real-time and signify base detection for construction of sequences or a subread. Multiple subreads from a single molecule in a ZMW are collated to create a circular consensus sequence (CCS).

Iso-seq does not require a reference genome for construction of sequencing reads and can allow direct sequencing of transcripts of up to 10kb in length. Demultiplexing of the CCS reads requires a multistep process (270). HiFi CCS reads were filtered, with removal of primers and identification of barcodes using limma (241, 271). Full length-reads were refined, clustered and polished based on per base quality values (QVs) for transcript consensus sequences.

5.3.2 SMRTBell library preparation

RNA extraction and cDNA synthesis from BCPP tissues was performed as described in Section 2.3. The cDNA library was constructed from 112 archived fresh-frozen sections of bladder tissues from BCPP (section 2.2.3.1), 22 fresh-frozen sections from Wrexham (section 2.2.3.2), 5 BC cell lines (Ward Lab), and 32 archived urine supernatants from HBRC (section 3.2.3.3). The SMRTBell library preparation protocol provided by PacBio was followed to create the amplicons in a two-step PCR process with barcoded M13 target primers for Multiplex SMRT sequencing (272). The following QC checks were performed at each step:

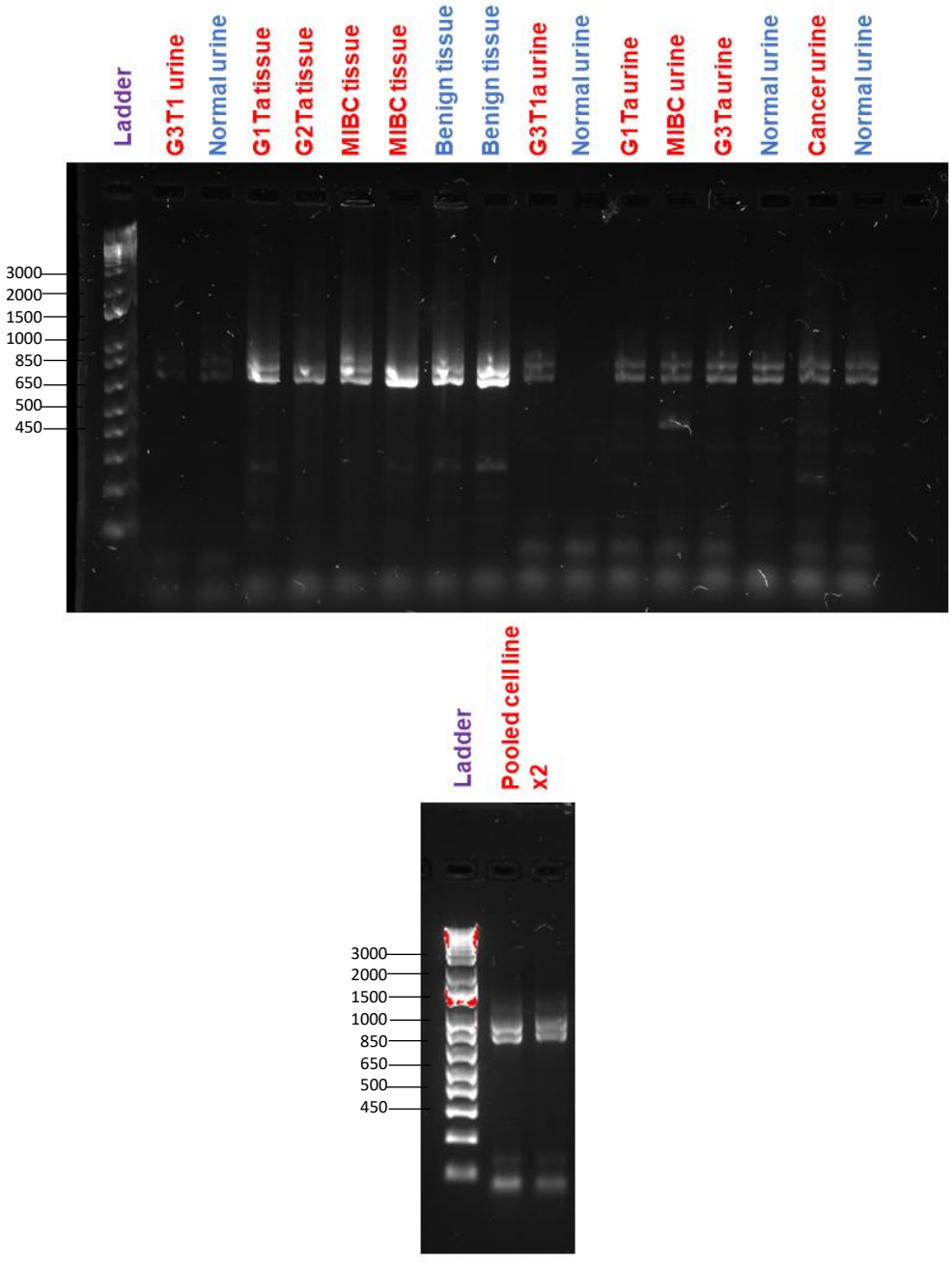
QC1- Initial standard RT-PCR was performed using target specific primers (supplied by Sigma Aldrich), and generation of *TPM* amplicons was checked by agarose GE which showed amplicon size between 650-850kb (Figure 5.1, top).

Following validation of adequate band intensity, a second set of primers with M13 adapters (supplied by integrated DNA Technologies, iDT) was validated using pooled cell line cDNA (Figure 5.1, bottom). 2µl of pooled cell line cDNA was added to a reaction mixture and PCR1 was performed using the protocol as described previously. Addition of adapters showed lengthening of amplicons to increase size to fall between 850-1000kb.

Figure 5.1 Agarose GE of PCR1 products: RT-PCR with target specific primers.

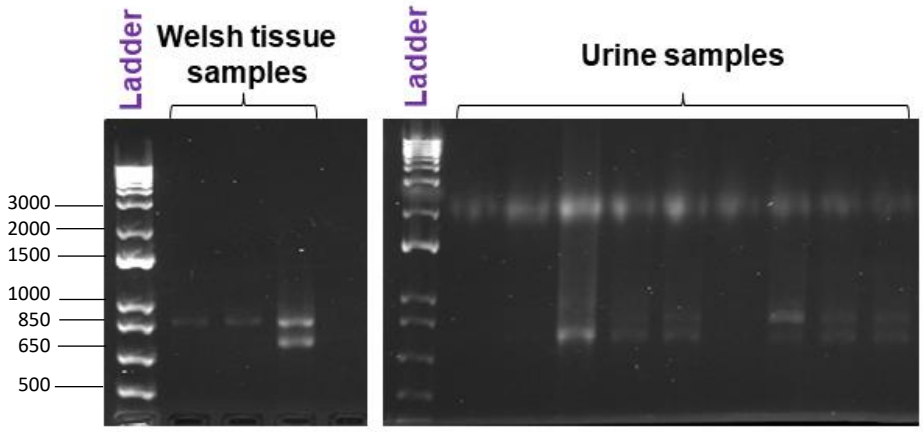
(Top) Randomly selected HBRC urine and BCPP tissue samples were checked on Agarose GE following RT-PCR to verify target TPM isoform amplification, size 650-850kb.

(Bottom) Target specific primers with M13 adapters were verified for amplification using pooled cell line (in duplicates), size 850-1000kb. The first lanes are 1kb DNA ladders followed by the samples.



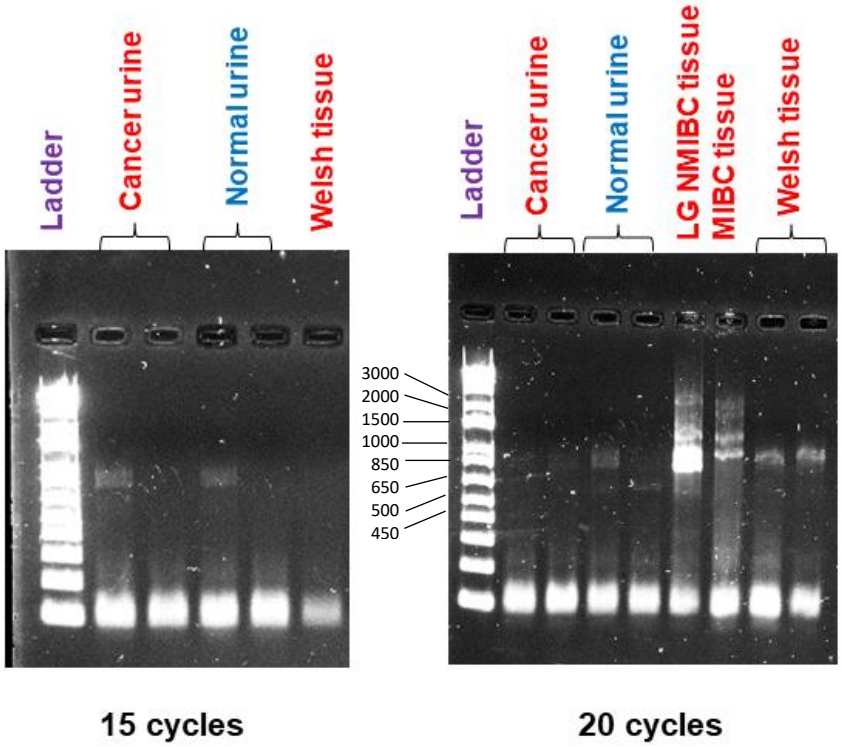
QC2- PCR1 with multiplexed target specific M13 primers was performed for all samples to create the library. Samples that had low RNA concentration (urine and BC tissues from Wrexham) were tested on 1.5% agarose GE for adequate band intensity (Figure 5.2), to ensure Tpm amplicons were present.

Figure 5.2 QC check- Agarose GE of PCR1 products. *Products from PCR1 reaction of samples that had low RNA concentration, ≤5ng/µl (HBRC urine and Wrexham (Welsh) tissues) were selected to evaluate TPM expression.*



Before proceeding to PCR2, randomly selected samples were run at 15-cyles and 20-cycles with barcoded primers and checked on 1.5% agarose GE to determine optimal cycle length for amplification (Figure 5.3).

Figure 5.3 QC check- Agarose GE with PCR2 products. *Randomly selected samples were tested for optimization of PCR2 cycle length. 5 samples were run at 15-cycles (left) and the same 5 samples + 2 additional BCPP tissue samples were run at 20-cycles (right) to test barcoded primers.*



QC3- Due to low RNA concentration in the urine samples, these samples were cleaned using AMpure beads individually, between the two PCR cycles (PCR1 and PCR2) to remove any impurities and decrease the probability of primer-dimer formation. Archived urine supernatants had limited availability of RNA in comparison to fresh urine pelleted cells, see limitations Section 8.5.

QC4- All the samples from each of the groups (urine/ BC tissues from Wrexham/ BCPP tissues) were pooled into 3 groups following PCR2. The final purified pooled sample was checked for amplicon size and variability using agarose GE (Figure 5.4) and the Bioanalyzer (Figure 5.5).

Figure 5.4 QC check- Agarose GE of pooled PCR2 products. *Pooled samples of PCR2 products from each of the 3 groups were analysed on agarose GE pre- and post-clean up with AMpure beads.*

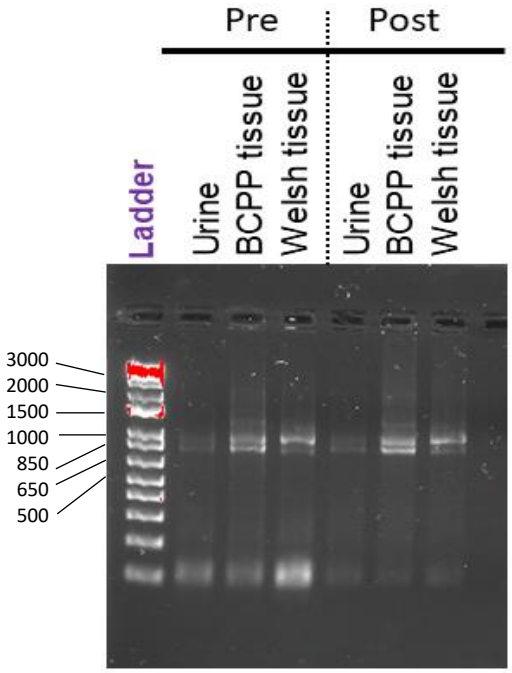
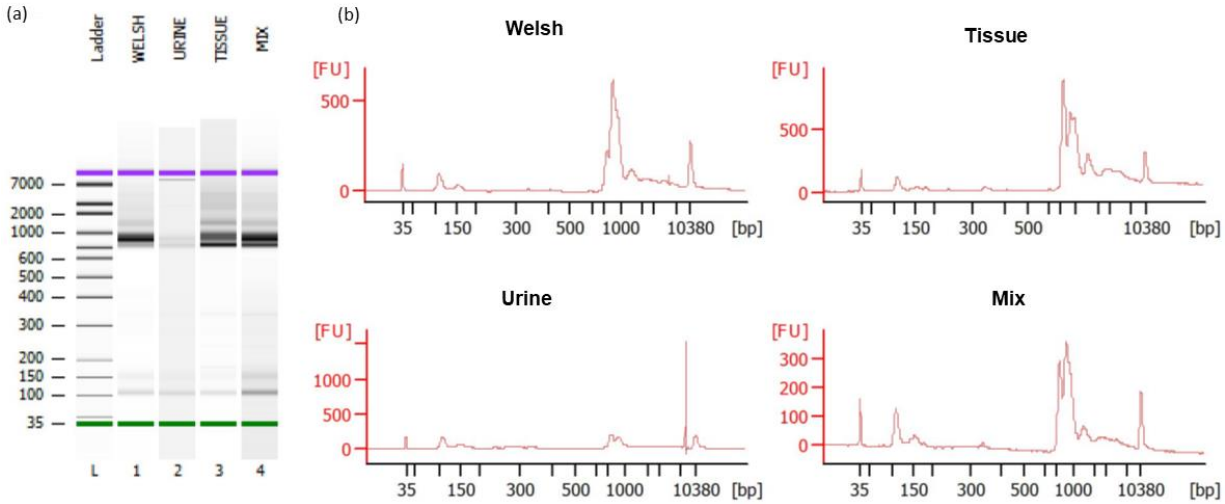


Figure 5.5 Electropherogram traces from the Bioanalyzer of the 3 pooled samples and the final pool (Mix). *(a) The first lane represents the ladder with subsequent multiplexed pools of Welsh tissues, urine, BCPP tissues and the final pool (Mix). (b) The traces show the composition of the individual library groups and the final (Mix) pool. Library sizes (in bp) are on the x-axis with sample intensity on the y-axis. The two peaks at each end of the spectrum indicate upper and lower control markers. Welsh tissues have a size of 862bp, urine = 749bp, BCPP tissues = 764bp, and the final Mix has a double peak of 758 and 880bp.*



5.4 Long-read RNA-seq data analysis

We obtained a total of 211,526 CCS reads which were classified further into known or novel isoforms based on their splice site match (full-splice match and incomplete splice match, novel in catalogue and novel not in catalogue, respectively). PSI for isoform expression was generated using the Arnold Lab bioinformatics pipeline. We assessed the similarity coefficient in the samples profiled by LR-seq using Pearson correlation with corresponding PSI from SR-seq (from *IsoTPM*) on isoforms that were full-splice matched (202,596 CCS reads, 95.78%).

TPM4 contributed to the highest number of reads (34.34%) followed by *TPM1* (28.42%) and *TPM3* (21.74%), whereas *TPM2* made up only 11.28% of the total number of reads.

HMW Tpm4.1 had higher usage in benign and paracancerous tissues accounting for more than 50% of *TPM4* expression (PSI ± SD: 0.591 ± 0.39, 0.577 ± 0.43, respectively). However, for bladder cancer tissues and cell lines, isoform usage was contributed mostly by LMW Tpm4.2 (PSI ± SD: LG NMIBC- 0.615 ± 0.32, HG NMIBC- 0.863 ± 0.24, MIBC- 0.618 ± 0.28, cell lines- 0.876 ± 0.18). Additionally, a very small proportion was represented by a novel isoform in benign tissues (0.044), MIBC (0.037) and LG-NMIBC (0.014) (Figure 5.6a (i)).

When compared with the SR-seq analyses, Tpm4.1 and Tpm4.2 were closely matched with Pearson correlation of R=0.7173 (p-value: 1.19e-13) and R= 0.3204 (p-value:4.137e-05), respectively (Figure 5.6a(ii)).

Figure 5.6a(i) TPM4 isoform distribution in samples from LR-seq. *x-axis represent the sample type and y-axis shows the distribution of isoforms (inclusion, PSI ± SD) that contributed to overall gene expression. TPM4 gene expression is represented by HMW Tpm4.1 in benign and paracancerous subtype, whereas LMW Tpm4.2 is found in cancer samples.*

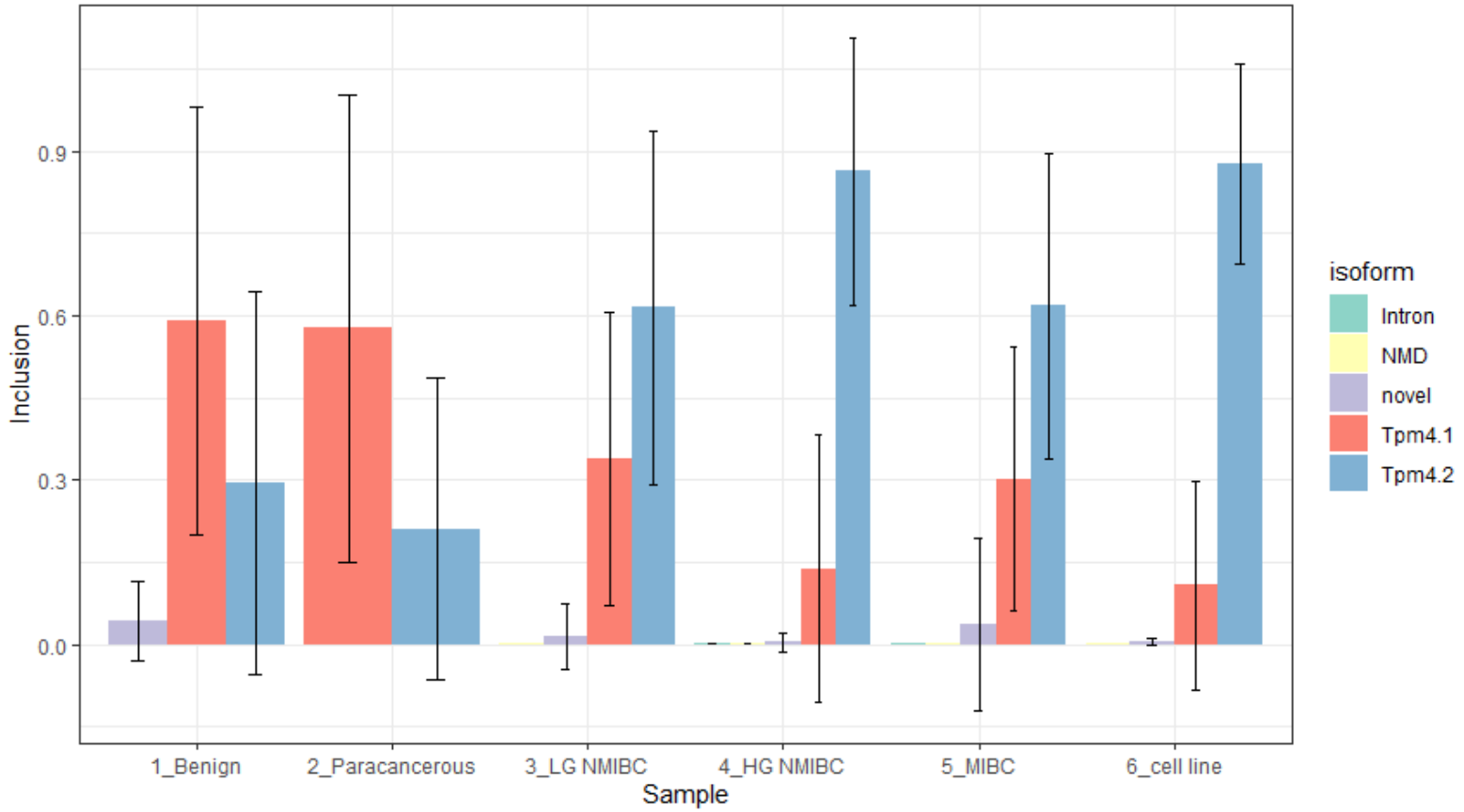
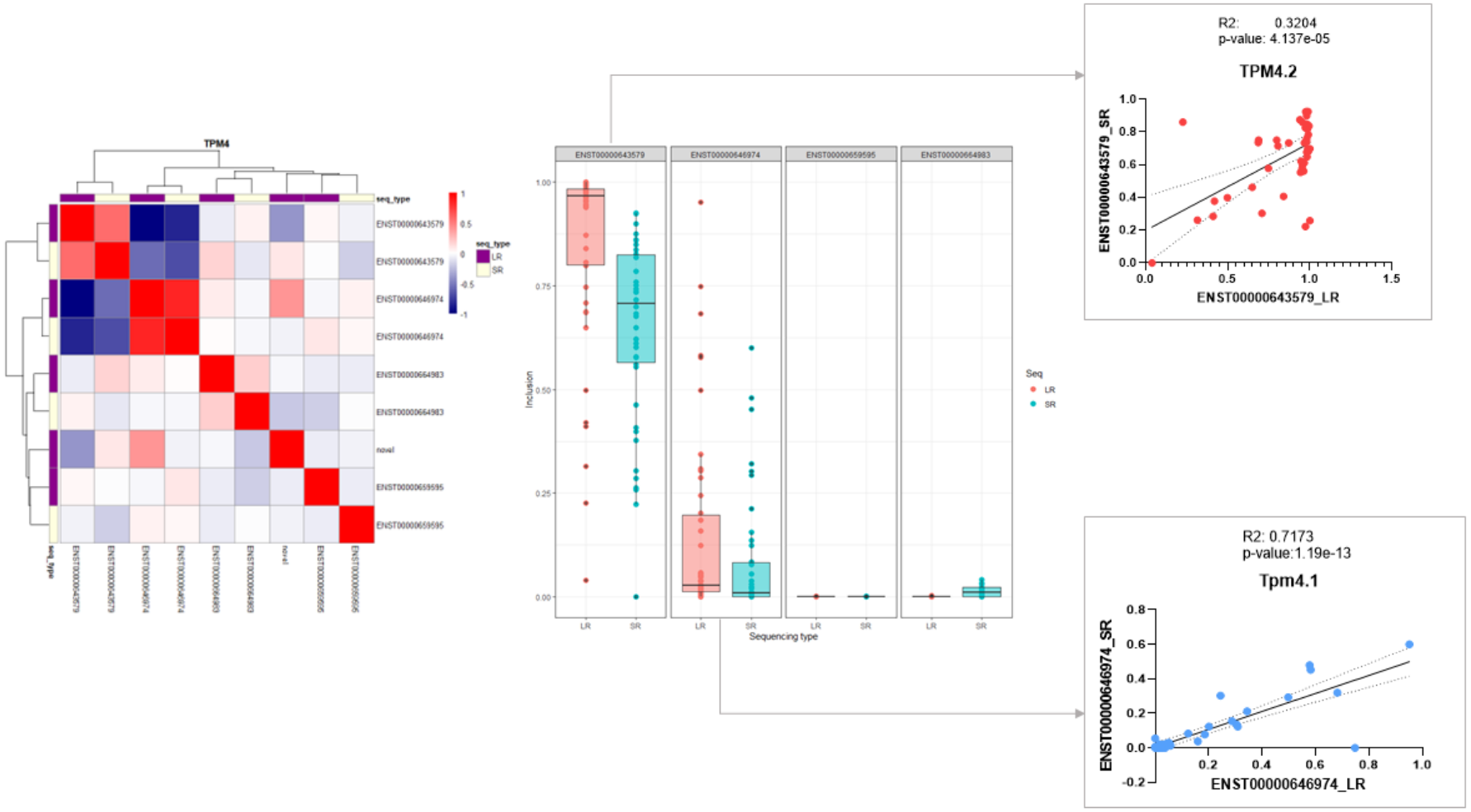


Figure 5.6a(ii) *The heatmap shows the* ***correlation between SR-seq and LR-seq of the same isoforms of TPM4****. The boxplot illustrates the inclusion levels (y-axis) per isoform (grey bar) in the RNA sequencing platform (x-axis). Pearson correlation coefficient was then calculated for similarity quotient between SR-seq and LR-seq reads for Tpm4.2 (red, p: 4.137e-05) and Tpm4.1 (blue, p: 1.19e-13) for comparison and validation of consistency in isoform expression.*



TPM1 isoform usage was shared predominantly between Tpm1.4, Tpm1.7 and Tpm1.9 across all the samples except paracancerous tissues which either exhibited Tpm1.4 (PSI: 0.629 ± 0.48) or a novel isoform (PSI: 0.110 ± 0.29). With advancing disease, the proportion of Tpm1.7 usage increased more contributing to at least one-third of overall *TPM1* expression (PSI: LG NMIBC- 0.296 ± 0.33, HG NMIBC- 0.398 ± 0.26, MIBC- 0.630 ± 0.27, cell lines- 0.841 ± 0.06). (Figure 5.6b(i)). In LG-NMIBC, alternative HMW isoforms (Tpm1.15, Tpm1.16, Tpm1.3 and Tpm1.2) contributed to at least 40% of total *TPM1* gene expression (PSI: 0.021, 0.108, 0.113, 0.159, respectively).

As opposed to SR-seq, additional isoforms were identified in LR-seq for LG NMIBC. Additionally, higher number of reads for Tpm1.7 and Tpm1.9 were obtained through LR-seq. (Figure 5.6b(ii)).

Figure 5.6b(i) TPM1 isoform distribution in samples from LR-seq. *x-axis represent the samples and y-axis shows the distribution of isoforms (inclusion, PSI ± SD) that contributed to overall gene expression.*

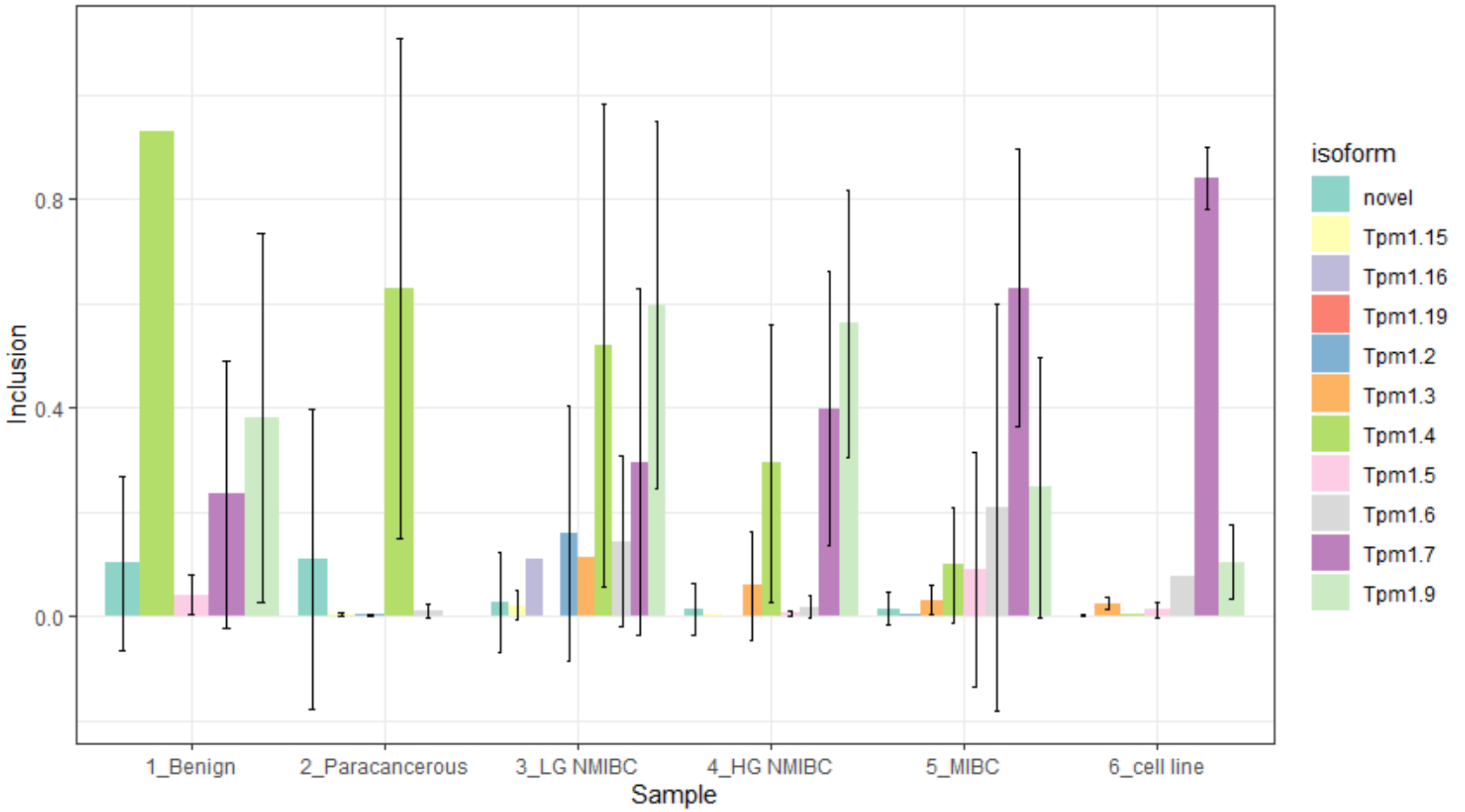
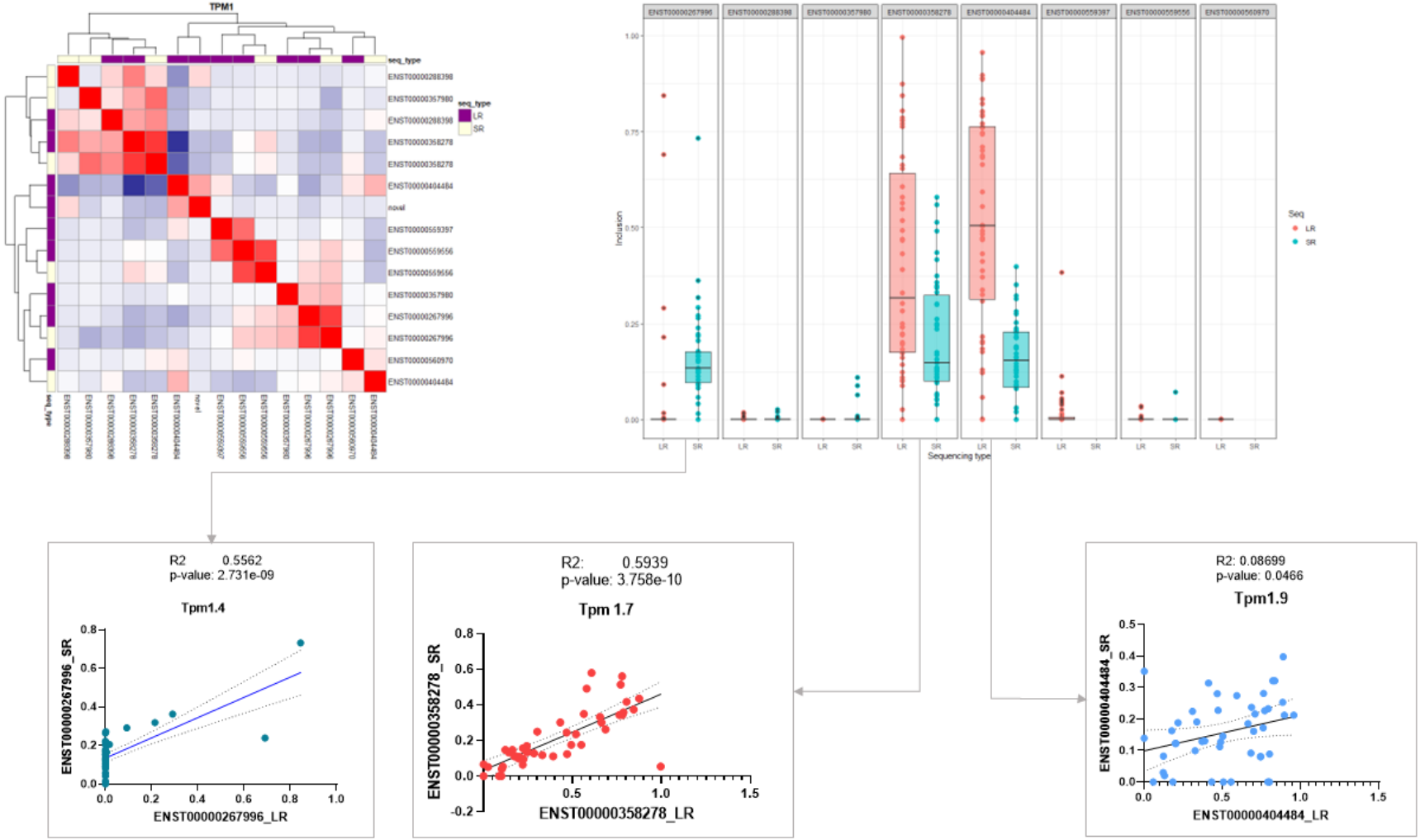


Figure 5.6b(ii) *The heatmap shows the* ***correlation between SR-seq and LR-seq of the same isoforms of TPM1.*** *The boxplot illustrates the inclusion levels (y-axis) per isoform (grey bar) in the RNA sequencing platform (x-axis). Pearson correlation coefficient was then calculated for similarity quotient between SR-seq and LR-seq reads for Tpm1.4 (teal, p: 2.73e-09), Tpm1.7 (red, p: 3.758e-10) and Tpm1.9 (blue, p: 0.0466) for comparison and validation of consistency in isoform expression.*



Whereas *TPM3* expression appeared to be characterized by Tpm3.1 in almost all the samples irrespective of tissue type (benign/ paracancerous/ tumour), it was contradictory to the SR-seq results which exhibited the dominant isoform as Tpm3.2 (Figure 5.6c(i,ii)).

To understand the discrepancy between the two sequencing results for LMW *TPM3* isoform, we explored the reads using the integrated genome viewer (iGV version 280 available on the BlueBear Linux portal) on a BCPP tissue sample that had reported the highest number of reads for Tpm3.2 (from SR-seq) and compared it to the results from LR-seq (which showed Tpm3.1 expression). On inspection the reads demonstrated origin from exon 6a (Tpm3.1) rather than 6b (Tpm3.2). The two isoforms (Tpm3.1 and Tpm3.2) appeared to have identical sequence coverage with the only difference observed in the differential usage of exon 6, which can explain why the SR-seq may have inadvertently aligned it to the incorrect isoform (Figure 5.6c(iii, iv)).

Figure 5.6c(i) TPM3 isoform distribution in samples from LR-seq. *x-axis represent the samples and y-axis shows the distribution of isoforms (inclusion, PSI ± SD) that contributed to overall gene expression.*

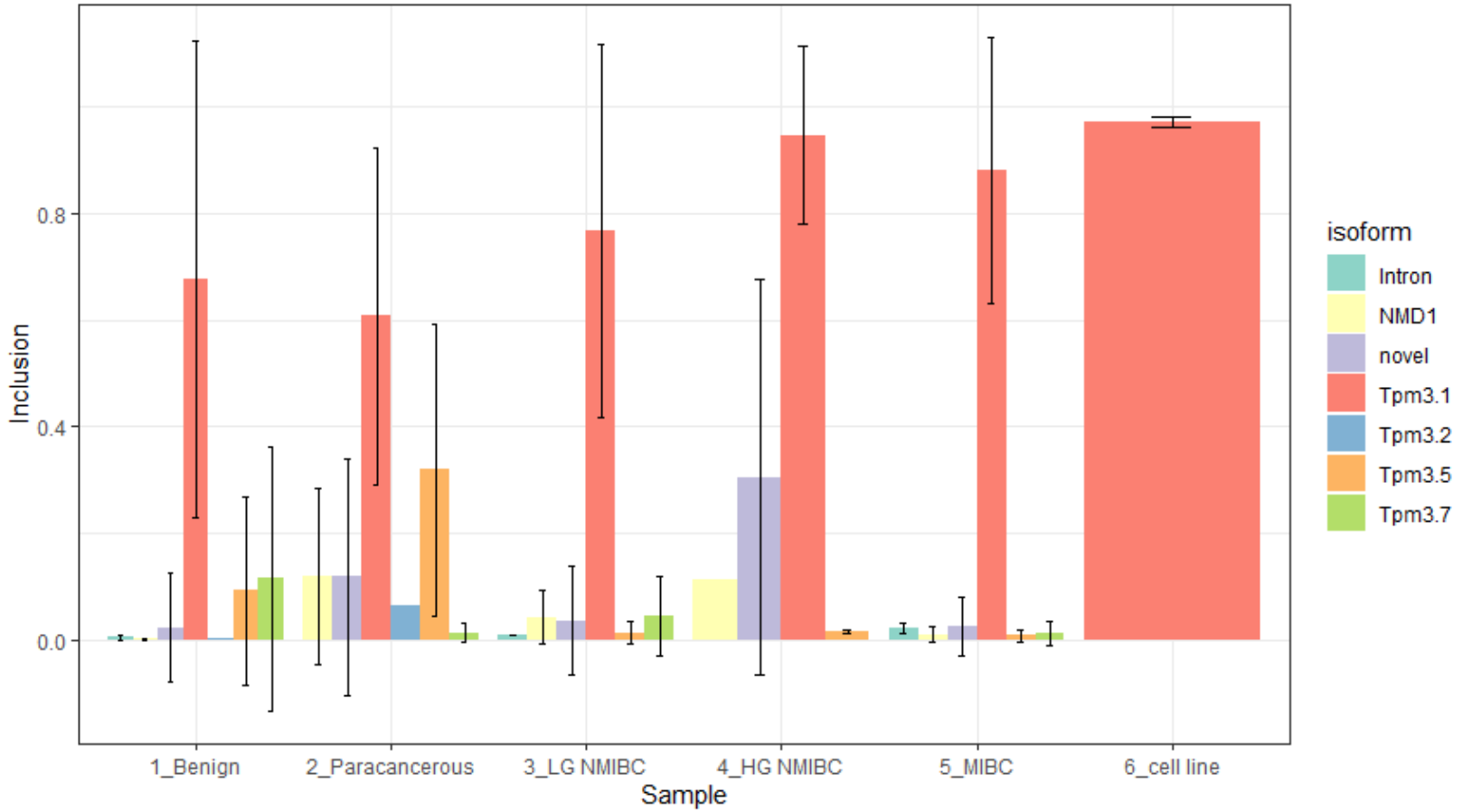


Figure 5.6c(ii) *The heatmap shows the* ***correlation between SR-seq and LR-seq of the same isoforms of TPM3****. The boxplot illustrates the inclusion levels (y-axis) per isoform (grey bar) in the RNA sequencing platform (x-axis). Pearson correlation coefficient was then calculated for similarity quotient between SR-seq and LR-seq reads for Tpm3.2 (purple, p: 0.3184) and Tpm3.1 (blue, p: 0.4227) for comparison and validation of consistency in isoform expression.*

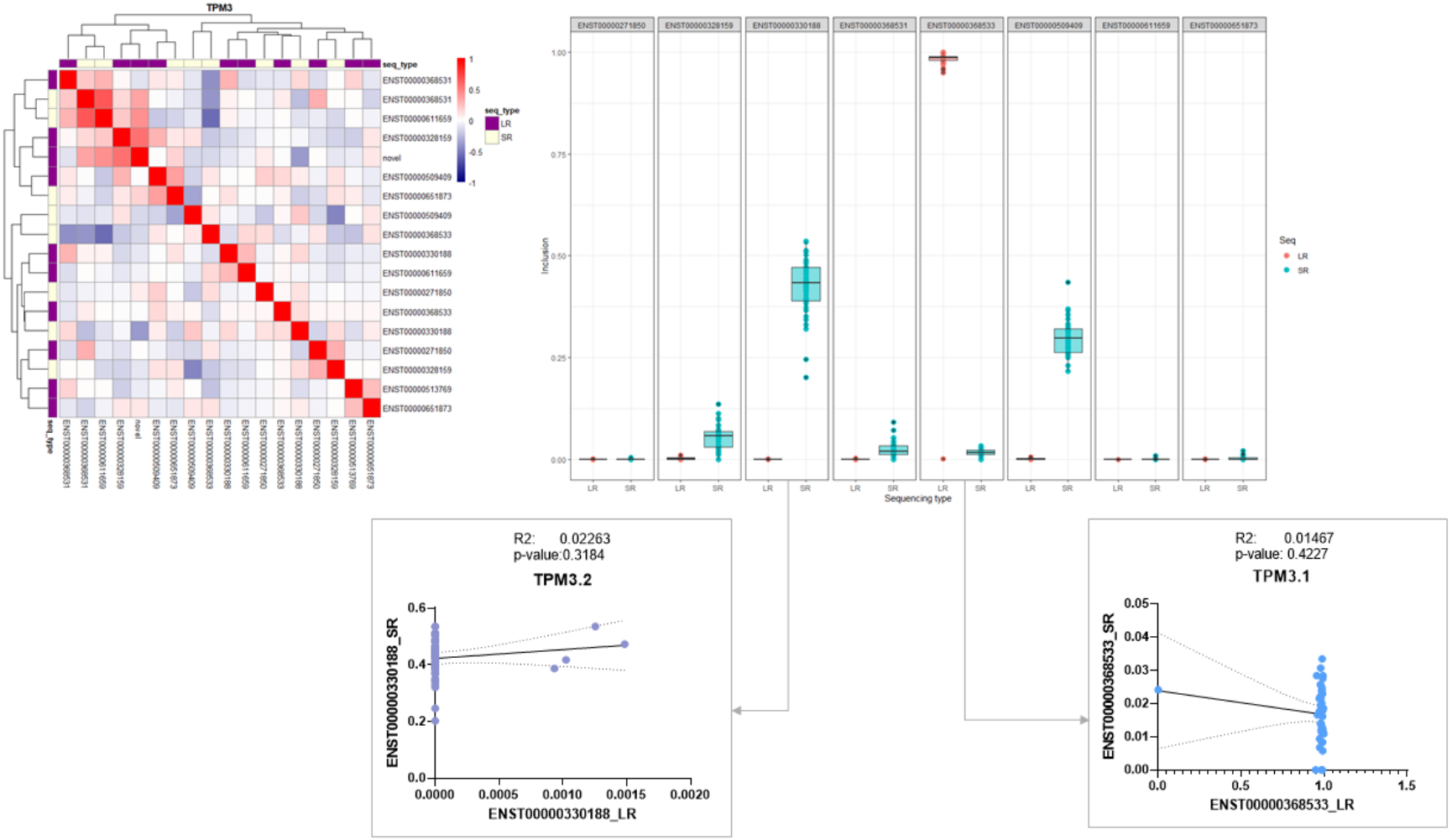


Figure 5.6c (iii) iGV snapshot comparing LR-seq and SR-seq in the same BCPP sample. *The top bar in red shows the result from LR-seq whereas the light blue bar indicates SR-seq. The vertical arrows are aligned to the schematic representation of the exons from the two isoforms:* ***Tpm3.1*** *and* ***Tpm3.2****, shown at the bottom of the figure. It is evident that there is higher usage of* ***exon 6a*** *rather than* ***exon 6b*** *indicative of isoform* ***Tpm3.1****.*

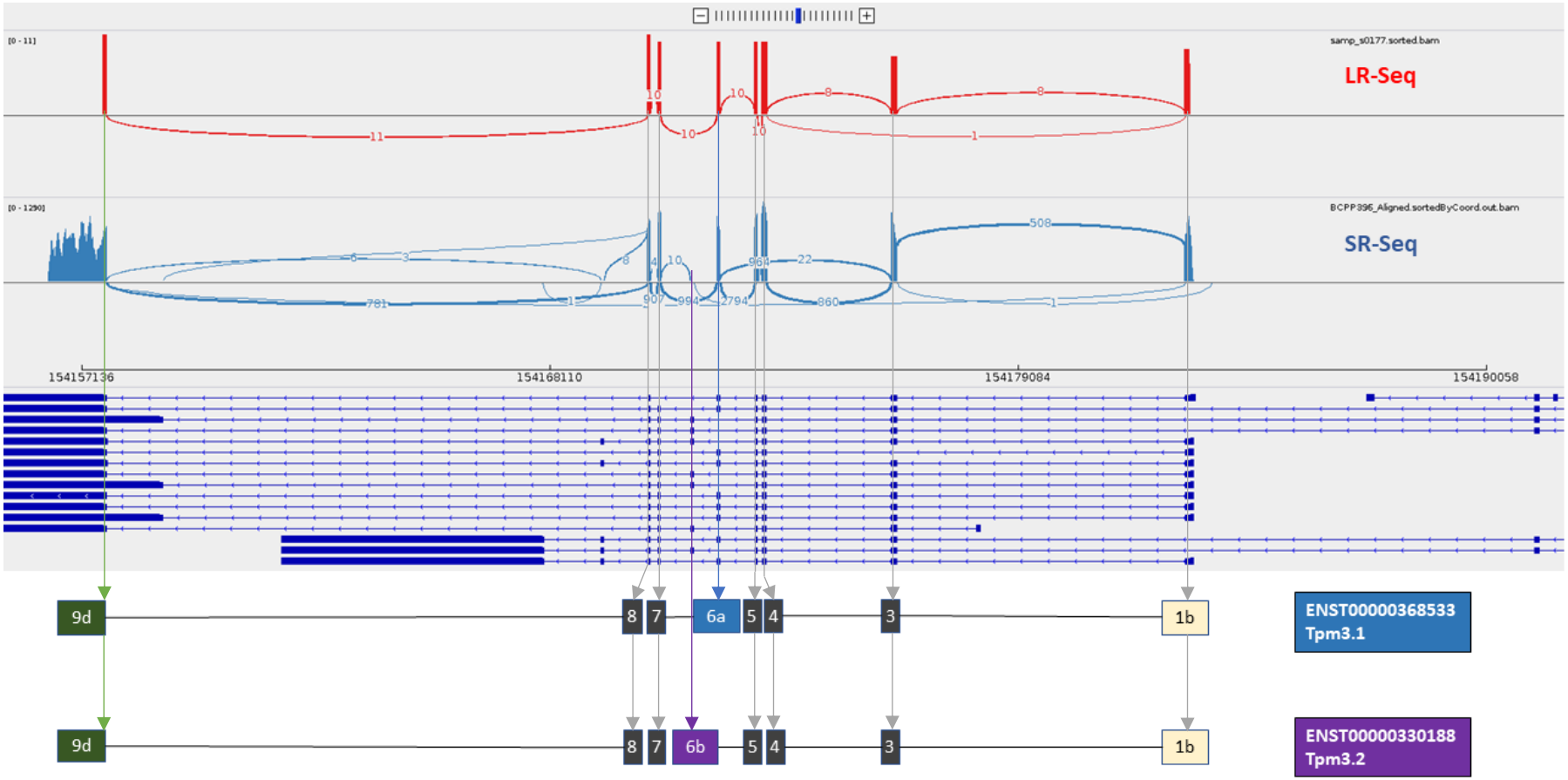


Figure 5.6c(iv) Comparison of isoform sequence of Tpm3.1 and Tpm3.2. *The red box indicates the differences in the two isoform sequences due to differential usage of exon 6. No difference is found in the remainder of the bp coverage at the 3' or 5' end.*



Tpm2.1 made up the majority of read counts representing *TPM2* expression in all sample types (Figure 5.6d(i,ii)). MIBC samples expressed an additional HMW isoform, Tpm2.4 (ENST00000647435) with PSI of 0.354 ± 0.46; this isoform was also found in benign (PSI: 0.091 ± 0.13) and paracancerous (PSI 0.061 ± 0.13) samples. A novel isoform was found for *TPM2* with highest expression in benign samples (PSI: 0.252 ± 0.28) followed by HG NMIBC (PSI: 0.120 ± 0.23).

Figure 5.6d(i) TPM2 isoform distribution in samples from LR-seq. *x-axis represent the samples and y-axis shows the distribution of isoforms (inclusion, PSI ± SD) that contributed to overall gene expression.*

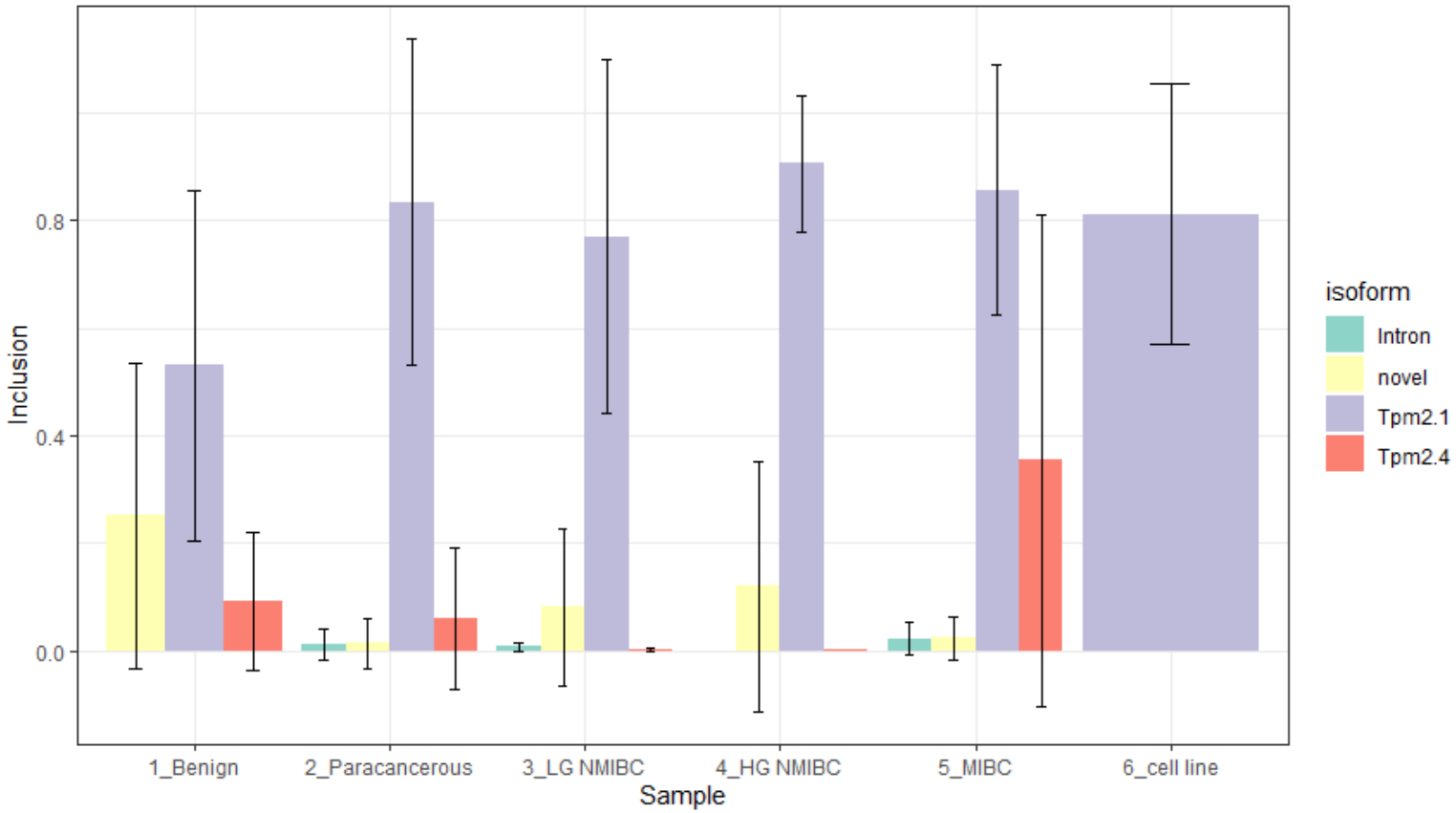
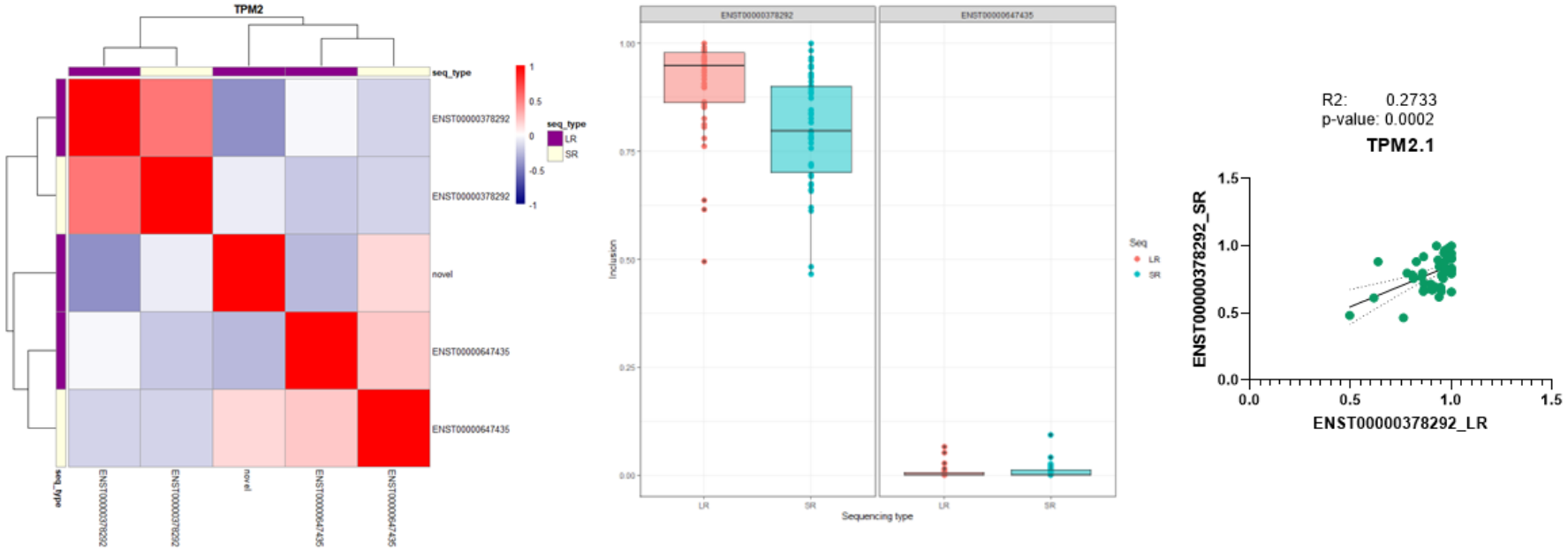


Figure 5.6d(ii) *The heatmap shows the* ***correlation between SR-seq and LR-seq of the same isoforms of TPM2****. The boxplot illustrates the inclusion levels (y-axis) per isoform (grey bar) in the RNA sequencing platform (x-axis). Pearson correlation coefficient was then calculated for similarity quotient between SR-seq and LR-seq reads for Tpm2.1 (green, p: 0.0002) for comparison and validation of consistency in isoform expression.*



5.5 Summary

- *LR- sequencing can characterize TPM isoforms from urothelial tissues and cell lines using the PacBio Isoseq pipeline.*
- *TPM4 and TPM1 made up the majority of the reads (34.34% and 28.42%, respectively) followed by TPM3 (21.74%) and TPM2 (11.28%) to account for overall TPM gene (1-4) expression.*
- *Tpm4.2 and Tpm1.7 showed patterns of expression specific to bladder cancer tissues, exhibiting increased expression (increasing PSI values) proportionate to invasiveness of disease.*
- *Tpm3.1 had similar expression across all the urothelial subtypes. However, when compared to SR-seq, Tpm3.2 was the assigned isoform to show similar expression pattern. On reviewing in iGV browser, the difference in the two isoforms was based on differential usage of exon 6.*
- *Tpm2.1 contributed to overall TPM2 expression with a novel isoform found in benign and HG NMIBC samples.*

Discussion:

LR-seq provides sequencing of full-length isoforms to overcome the limitations encountered with SR-seq, as discussed previously. We performed LR-seq on the same (matched) samples processed via SR-seq to compare the differences and similarities in isoform identification and composition. This process was facilitated by creating primers at 5’ and 3’- ends of the isoforms to allow full-length capture of the isoforms

amplified. PacBio Isoseq pipeline aided in identifying multiplexed isoforms of *TPM* genes that had very closely related base pair sequences. Tpm4.2 and Tpm1.7 made up the majority of the read counts in BC cells and cell lines, corroborating the findings from SR-seq. Concerning LMW isoform of *TPM3*, LR-seq identified Tpm3.1 as the most dominant isoform. To verify the conflicting result of SR-seq and LR-seq identification of *TPM3* isoform, the sequences of the two isoforms and reads were reviewed in the iGV browser, which concluded that SR-seq results were incorrect. Nonetheless, in literature the two isoforms Tpm3.1/3.2 are grouped together as the only difference in sequence is the differential usage of exon 6 (200, 207, 273).

It is possible that Salmon incorrectly assigned the SR-seq reads to Tpm3.2 during quantification due to the low complexity of differences within the sequences of all other exons of *TPM3*. The reads were not unique (as tested with BLAST) and therefore not adequately assigned to the corresponding isoform. When comparing the two nucleotide sequences (length of 76 bp) of Tpm3.1 and Tpm3.2, it was observed that the differences were in the 5’ end of exon 6 with an overall sequence similarity score of 61.36%.

Chapter 6

HIGH THROUGHPUT ASSAY DEVELOPMENT FOR BIOMARKER USE

6.1 Aims and objectives

The sequencing experiments described thus far have implicated Tpm1.7 and Tpm4.2 are found in bladder carcinogenesis. The target isoforms were explored for quantitative analysis in cell lines, BCPP archived tissues and urine supernatants. The relative isoform expression of *TPM1* (Tpm1.7/Tpm1.4) and *TPM4* (Tpm4.2/Tpm4.1) genes was tested. This chapter focuses on the experimental confirmation of NGS isoform predictions with subsequent development of the biomarker assay.

Objectives:

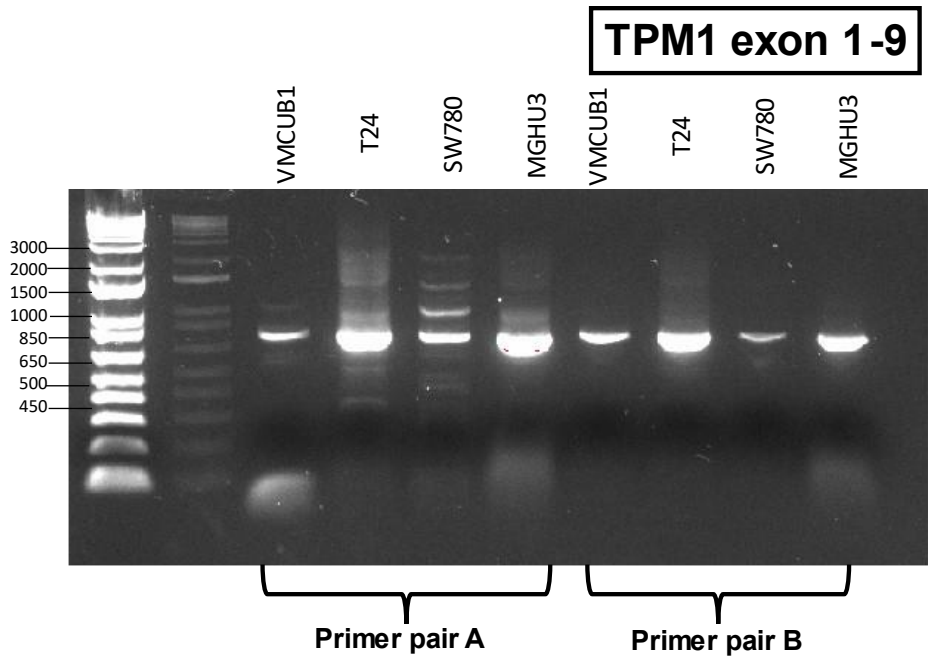
- Design and optimize a biomarker assay to assess isoform expression in BC cell lines.
- Process a pilot cohort of BCPP tissue (matched) samples to compare *TPM* isoform expression *in silico* and *in vitro.*
- Perform preliminary assay testing on urine samples to identify potential utility as POC test.
- Compare expression profiles between normal and cancerous samples.

6.2 PCR

Initial validation experiments were performed on BC cell lines to assess for presence of *TPM1* gene expression, using RT-PCR and GE (Fig 6.1). The first batch of primer pairs were designed using Primer Blast (254) for *TPM1* (forward and reverse) to target exon1 and exon9 respectively, to enable full coverage of the entire length of the isoforms (Table 2.4, Chapter 2).

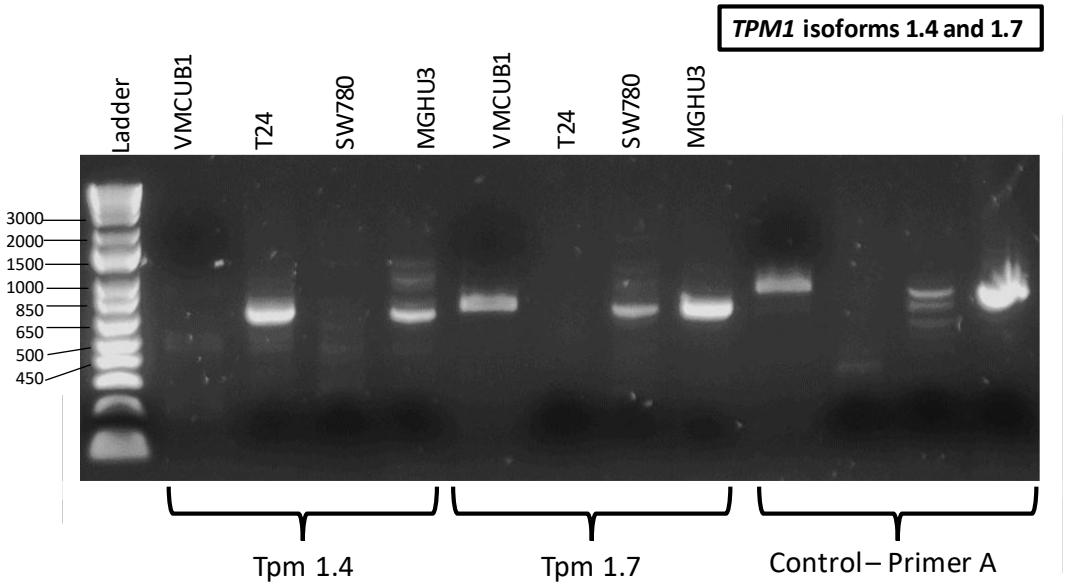
These primer pairs were amplified using a 35-cycle standard PCR and assessed for band intensity and amplicon size using 1.5% agarose gel electrophoresis (Figure 6.1).

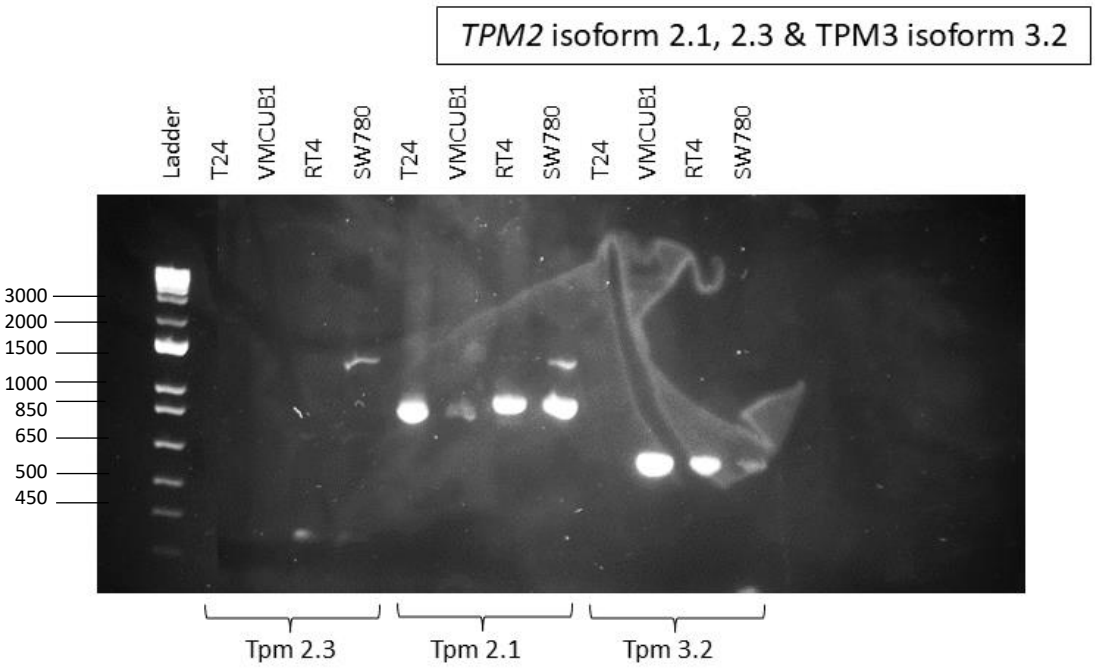
Figure 6.1 Agarose gel electrophoresis for validation of TPM expression using primer pairs targeted to exons 1 and 9 from TPM1. *The first lane is the 1kb DNA ladder. Successive lanes (from third) are bladder cancer cell lines at ~50ng/µl cDNA concentration, collectively divided into 2 sets of target-specific primer pair A and primer pair B.*

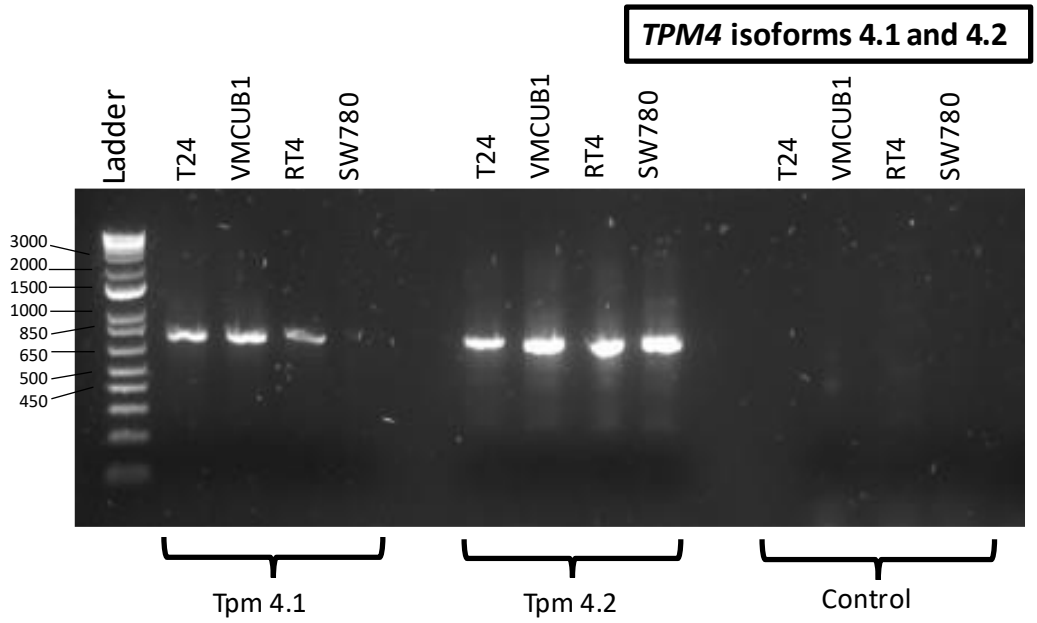


Following confirmation of adequate band intensity from expected-length amplicons (~750-850 bp), a second set of isoform-specific primers were designed using a combination of custom Python scripts and Primer blast. These isoforms were used for RT-PCR and GE to confirm the *in silico* predictions of the MDT isoforms from the 4 *TPM* genes (Figure 6.2).

Figure 6.2 Agarose gel electrophoresis for validation of isoform-specific primers of TPM1-4 genes. *The first lanes in all gels are 1kb DNA ladders. Consecutive lanes are representative of bladder cancer cell lines with ~50ng/µl cDNA concentration per PCR reaction.*







The primer pairs showed adequate band intensity on GE and target amplicons were verified by Sanger sequencing. The PCR products were excised from the gels, purified

and Sanger sequenced; subsequent results were checked for isoform matches using NCBI-BLASTn (274) to confirm the identity of the target isoforms.

Tpm1.7, Tpm4.2 and Tpm4.1 were accurately identified. Primers intended for Tpm1.4 amplified the sequences matching Tpm1.3 in the initial tests. Therefore, the primers were redesigned, and following RT-PCR and GE, were sequenced to verify the correct amplicon. Due to the complexity and significant sequence overlap shared between isoforms of *TPM3*, multiple isoforms matched the queried sequence. The table below shows the input sequences (query coverage) and the percentage matched (sequence coverage) of the PCR products.

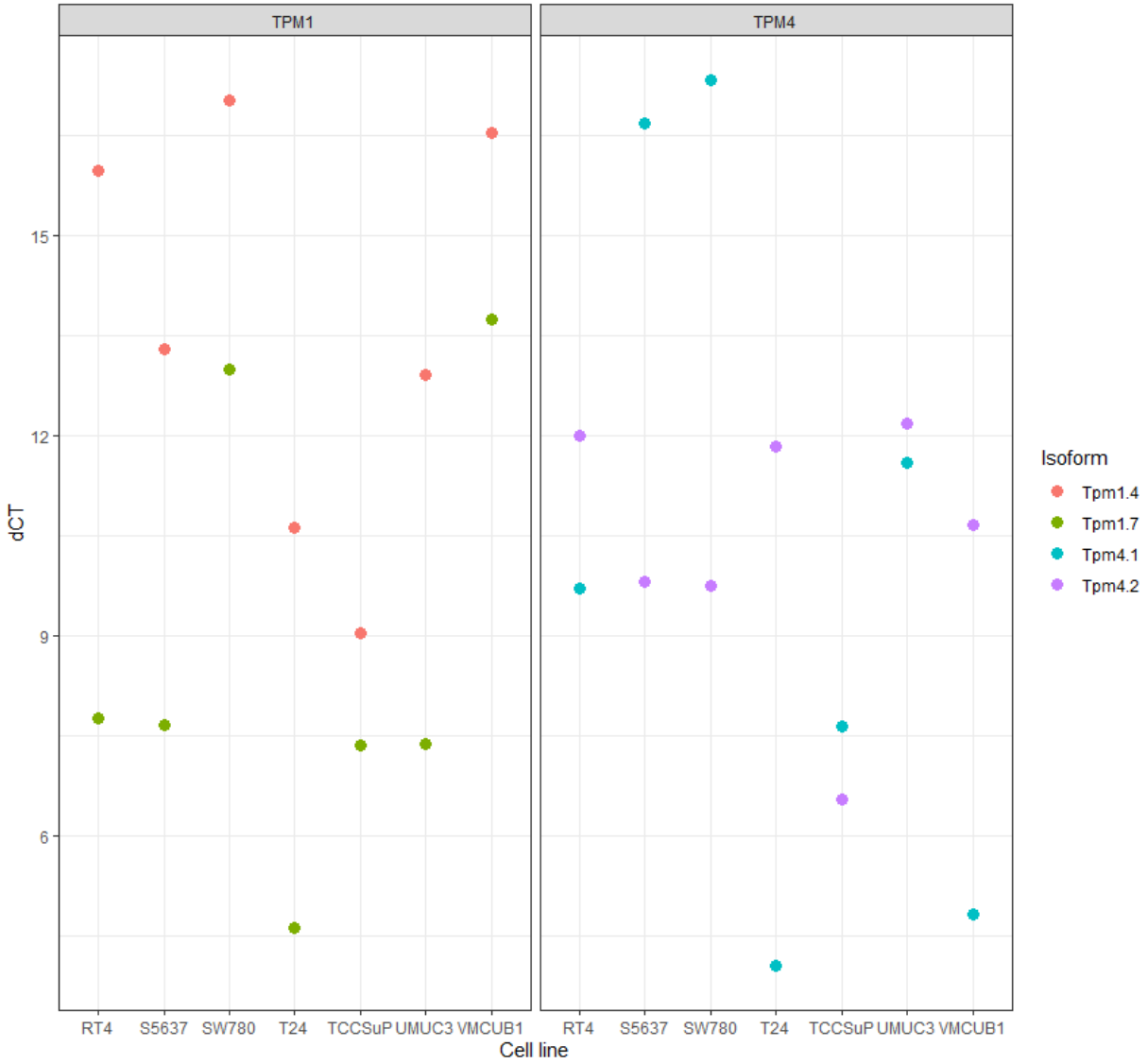
Table 6.1 | RT-PCR products that were sequenced were 'BLAST' with NCBI BLAST-n, with % query and sequence coverage matching the intended target isoforms.

PCR product sequenced (Sample name)	Query cov	Seq cov	TPM Nomenclature	RefSeq Accession ID
T201F	98	99.54	Tpm1.3	NM_001018020.2
T201R	98	99.55	Tpm1.3	NM_001018020.2
T206F	98	99.35	Tpm1.7	NM_001018006.2
T206R	98	99.52	Tpm1.7	NM_001018006.2
T218F	99	95.28	Tpm4.2	NM_003290.2
T218R	97	94.64	Tpm4.2	NM_003290.2
V221F	85	100	Tpm4.1	NM_001145160.1
TM24-IIB	98	99.73	Tpm2.1	NM_213674.1
VMCUB-III	100	99.75	Tpm3.1, 3.2, 3.4, 3.5, 3.7	
SW780-IIB	91	100	Tpm2.1	NM_213674.1
SW780-TMI (206)	98	99.37	Tpm1.7	NM_001018006.2
RT4-IIB	87	100	Tpm2.1	NM_213674.1
RT4-III	99	99.75	Tpm3.1, 3.2, 3.4, 3.5, 3.7	
RT4-TMI (206)	98	99.53	Tpm1.7	NM_001018006.2
T3-208	99	99.64	Tpm3.7, Tpm3.5	
T4-218	99	99.37	Tpm4.2	NM_003290.3
T4-221	99	99.7	Tpm4.1	NM_001145160.2
T201F_redesigned	98	95.63	Tpm1.4	NM_001018007.2
T201R_redesigned	98	95.35	Tpm1.4	NM_001018007.2

6.3 Optimization of qPCR assay

Following confirmation of intended target isoforms, the verified primer pairs were used to evaluate relative isoform expression in 7 BC cell lines using qPCR with **SYBR-Green** dye. GAPDH was applied as a house-keeping gene and technical replicates (x3) for each BC cell line per target isoform were used. The qPCR cycle endpoint was set to 35, if no amplification was found by 35 cycles, it was considered a negative result i.e. low or nil isoform expression. To assess relative isoform expression for *TPM1* and *TPM4*, average C_T was computed, then normalized to GAPDH to calculate delta C_T, and 2^-delta C_T (Figure 6.3). Tpm1.7 had higher expression than Tpm1.4 in all BC cell lines tested. Tpm4.2 had high expression in all cell lines, however, Tpm4.1 had variable expression depending on the characteristics of the cell line (grade/stage). Muscle-invasive cell lines showed high expression for Tpm4.1, whereas non-muscle-invasive/ low grade cell lines had higher expression of Tpm4.2, comparatively.

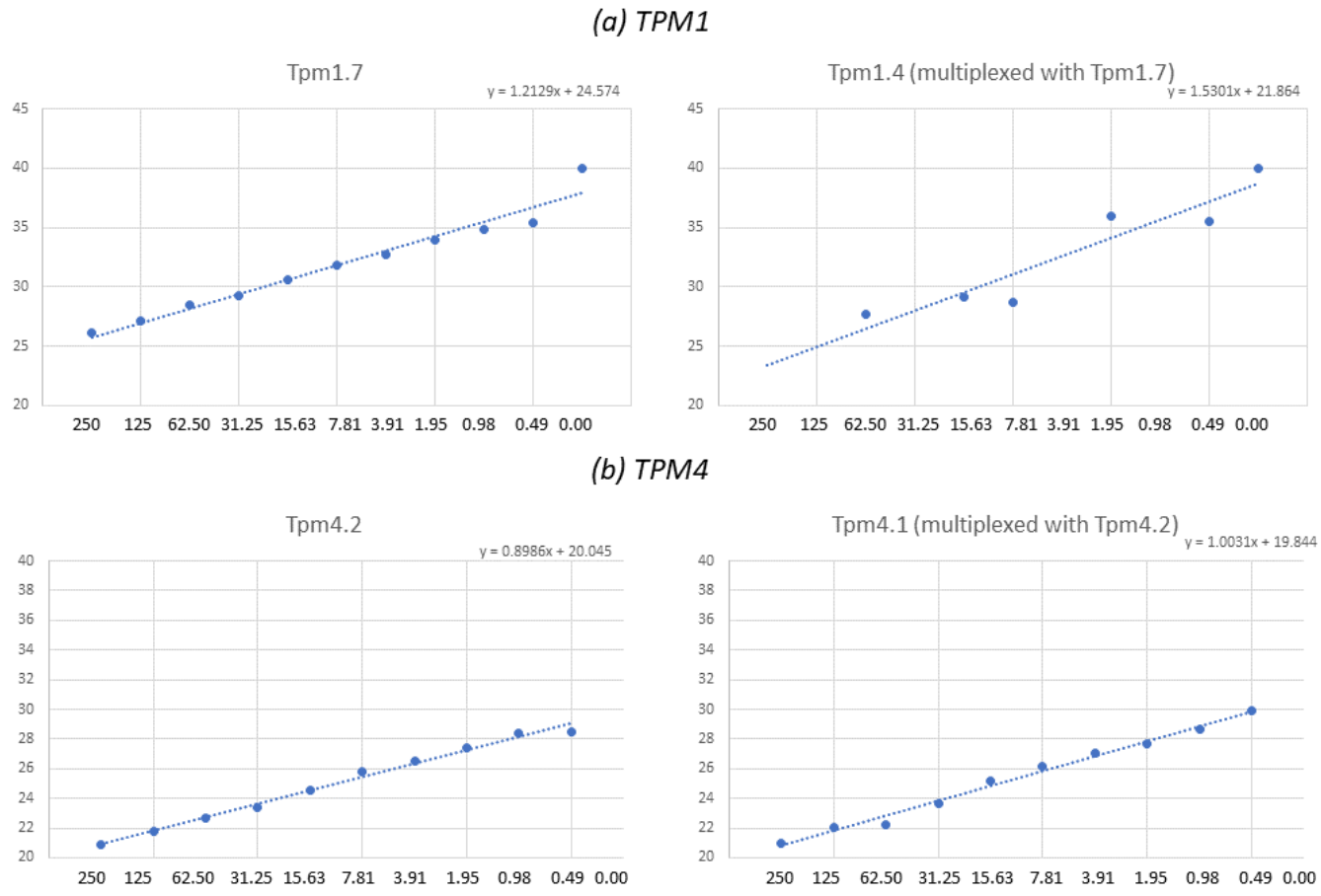
Figure 6.3 dC_T values of TPM isoform expression relative to GAPDH. *The scatter plot shows the dC_T values of target TPM isoforms relative to GAPDH. Smaller dC_T values (y-axis) indicate higher expression of the isoform in the cell line tested (x-axis). 150ng of cDNA was used at a dilution of 1:10, resulting in 30ng of cDNA per sample for each PCR reaction.*



Variability in reproducibility of results, decreased specificity with use of SYBR green and inability to multiplex several isoforms prompted redesigning of qPCR assays using **Taqman** probes. Taqman increases the specificity of amplicons and allows for multiplexing of target isoforms within one sample (275). The forward and reverse primer pairs for *TPM1* isoforms were reused with a probe targeted at exon 6. *TPM4* isoform assays were redesigned with shorter length amplicons (length 100-200 bp), with probes located at exon 2b/3 and exon 1b/3 splice-junctions *(See Table 2.5, Chapter 2 for sequences of probes and primers)*. The qPCR cycle-endpoint was set to 40. Validation experiments with the optimized qPCR assay for *TPM4* improved C_T values to fall within a range of 20-30. Therefore, the *TPM1* assay was also redesigned to create shorter-length amplicons. However, due to differential usage of exon 6 by the target isoforms, amplicon lengths could only be decreased by 5% of the original length with the 3' reverse primers positioned at exon 7.

A dilution series qPCR experiment was performed to run QC checks on the Taqman probe assays. cDNA from pooled BC cell lines (in duplicates) were added to a qPCR reaction mix containing 6µl cDNA, 10 µl of Taqman Multiplex Enzyme Mastermix, and 4 µl of assay (at a final concentration of 0.4 µM primer and 0.25 µM probe). For each 1:2 dilution of cDNA, the amplification cycle shifted by x2 to the right (Figure 6.4).

Figure 6.4 Dilution charts for qPCR Taqman assay. *QC check was performed on the multiplexed qPCR assay for TPM1(a) and TPM4(b). Initial concentration of 250ng pooled cDNA (x-axis) was used with serial dilutions to note how the C_T levels (y-axis) were affected. Each multiplexed assay had at least 2 isoforms from each gene. Decreasing the amount of cDNA concentration by half in each sample causes the C_T value to increase by 2 cycles.*



6.4 qPCR assay testing in BCPP tissue samples

Following optimization of qPCR, we tested the multiplexed probe-based assay (to quantify RNA expression exhibited by *TPM1* and *TPM4* isoforms) on BCPP tissue samples and BC cell lines. Input cDNA for qPCR testing was from the same samples that were used to generate the LR-seq PacBio library. This was done to allow for consistency across and comparison of the two methods. qPCR assay testing was performed in duplicates for each sample. Average C_T was calculated for all 4 *TPM* isoforms to assess expression in each sample (Figure 6.5).

The ratios of *TPM1* and *TPM4* isoform expression were determined with eventual calculation of fold-change expression of Tpm1.7 to Tpm1.4, and Tpm4.2 to Tpm4.1. (based on our previous hypothesis and findings that indicated upregulation of Tpm1.7 and Tpm4.2 in bladder cancer compared to Tpm1.4 and Tpm4.1, respectively) (Figure 6.6).

Tpm1.7 had higher relative expression (to Tpm1.4) in bladder cancer tissues (Kruskal-Wallis p: 0.031**), with an increase in expression to severity of disease (NMIBC to MIBC). Similarly, LMW isoform Tpm4.2 had higher relative expression (to Tpm4.1) (Kruskal-Wallis p:9.18e-06***). However, Tpm4.1 expression also increased concurrently with Tpm4.2 in MIBC tissues, causing an overall increase in gene expression for *TPM4* (indicated by lower C_T values and decrease in relative expression of the two isoforms). Potentially, this could be due to Tpm4.1 from muscle cells, signifying the ‘invasive’ nature of the tumour (discussed later).

Figure 6.5 Isoform expression in BCPP tissues and cell lines indicated by C_T values. *The box plot shows the qPCR assay (absolute) C_T values (y-axis) of the target isoforms (x-axis) from TPM1 (a) and TPM4 (b) in the order: cell lines, NMIBC low grade -G1/G2Ta, NMIBC high grade- G3Ta/ G3T1 and MIBC tumour tissue RNA. High C_T values indicate LOW expression in comparison to Low C_T values that are indicative of HIGH expression of the isoform.*

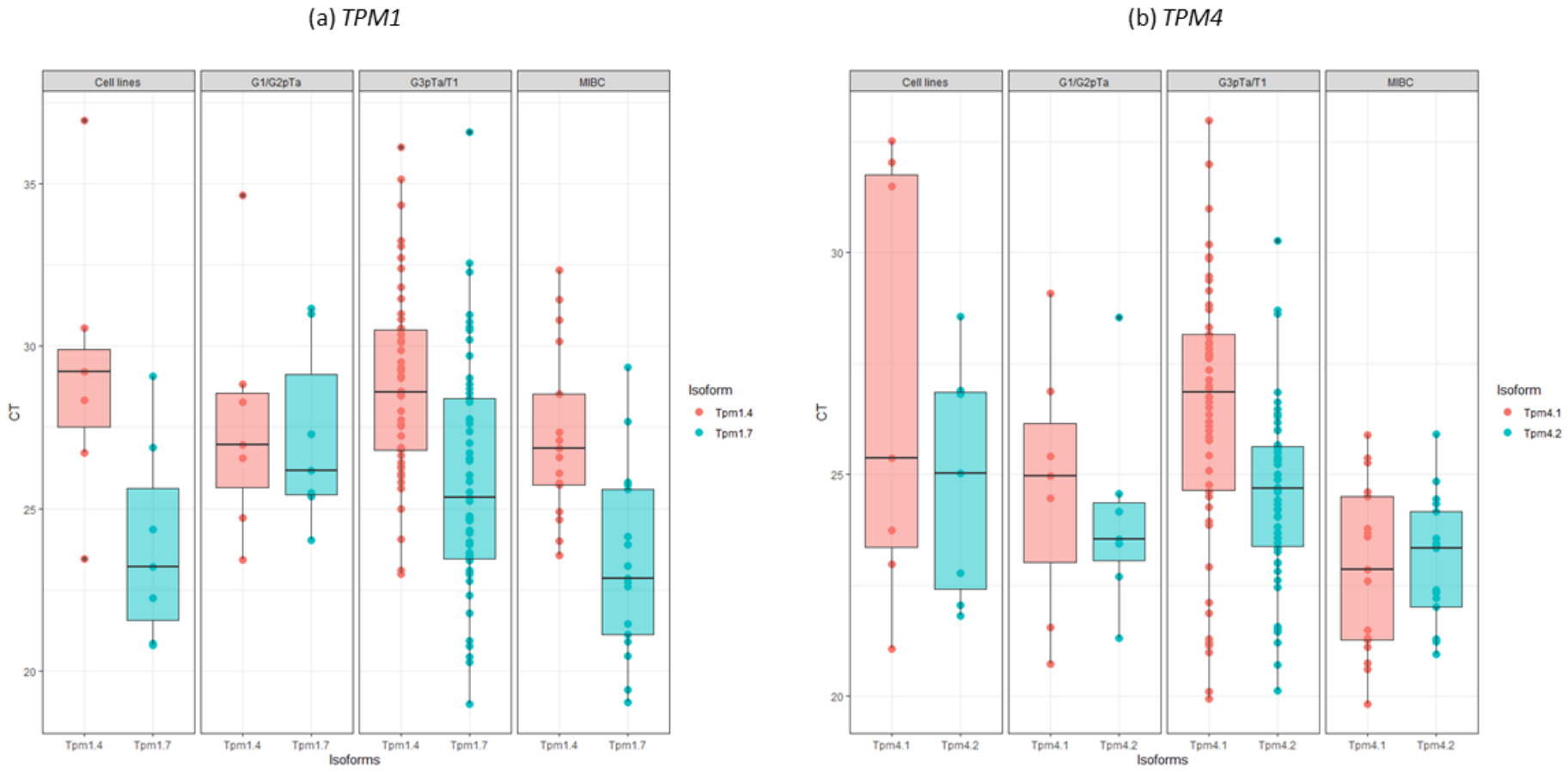
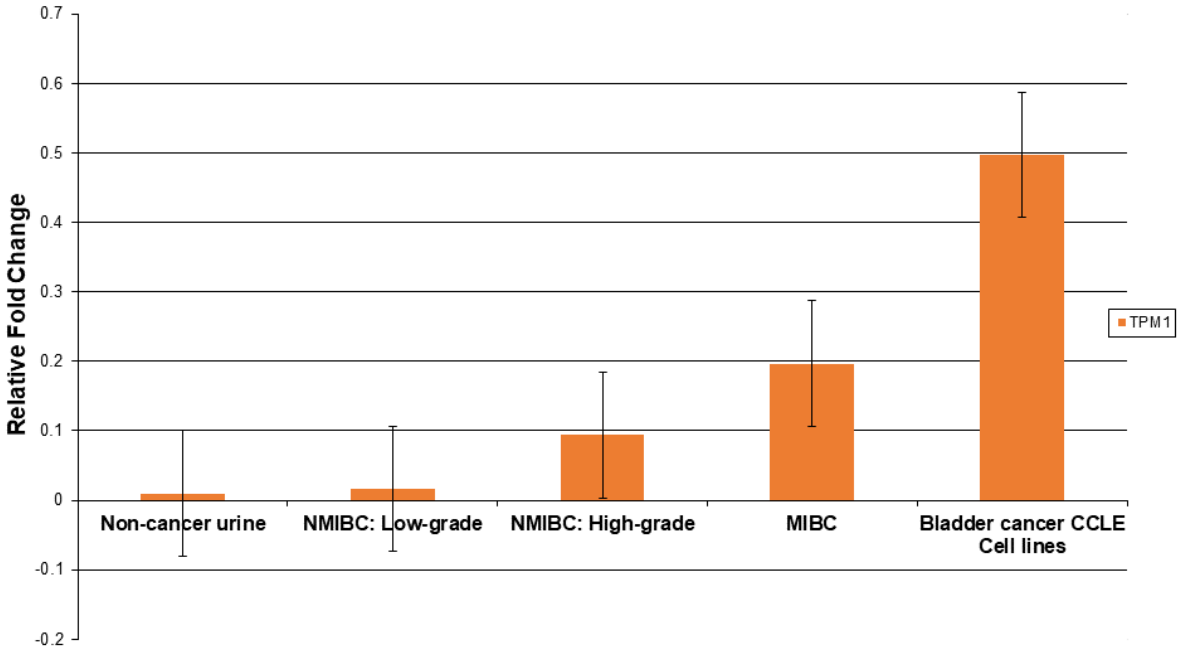
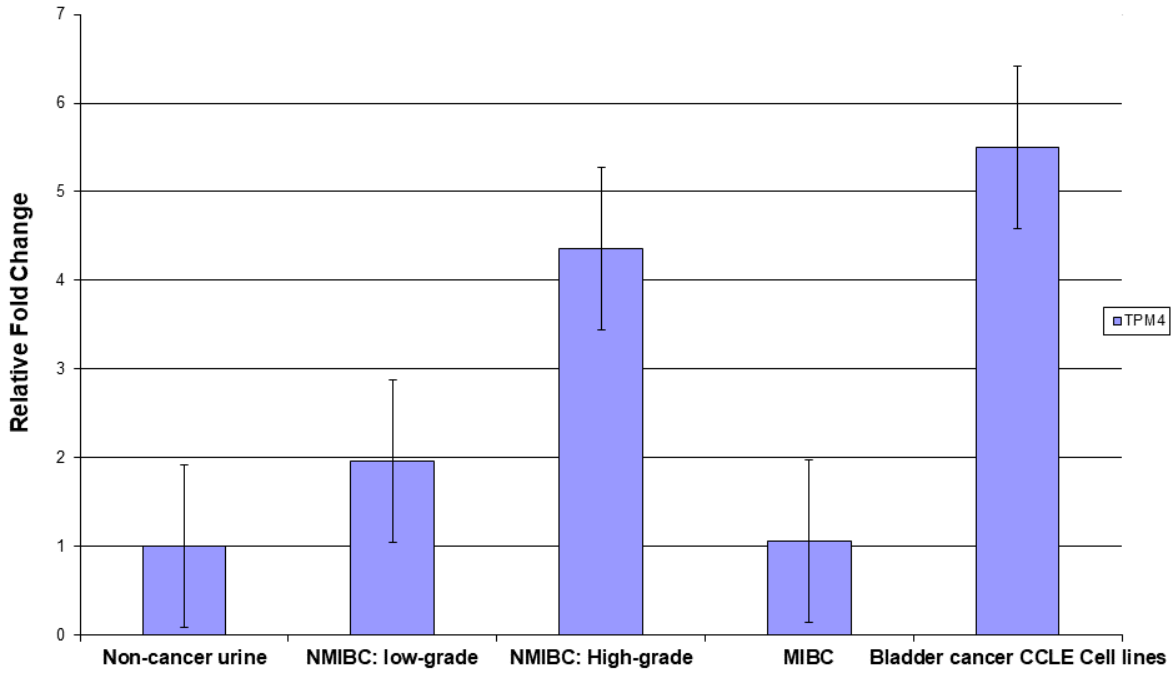


Figure 6.6 Fold-change expression of TPM1 and TPM4 isoforms. *qPCR analysis of TPM isoforms on BCPP tissues and BC cell lines. A set of 8 non-cancerous urine samples were used as control.*

(a) The graph shows that the ***relative fold change expression (y-axis) of Tpm1.7 (to Tpm1.4)*** *was higher in BCPP cancer tissues and cell lines. The increase in expression is seen as a trend with increase in severity of disease (NMIBC low-grade to MIBC).*

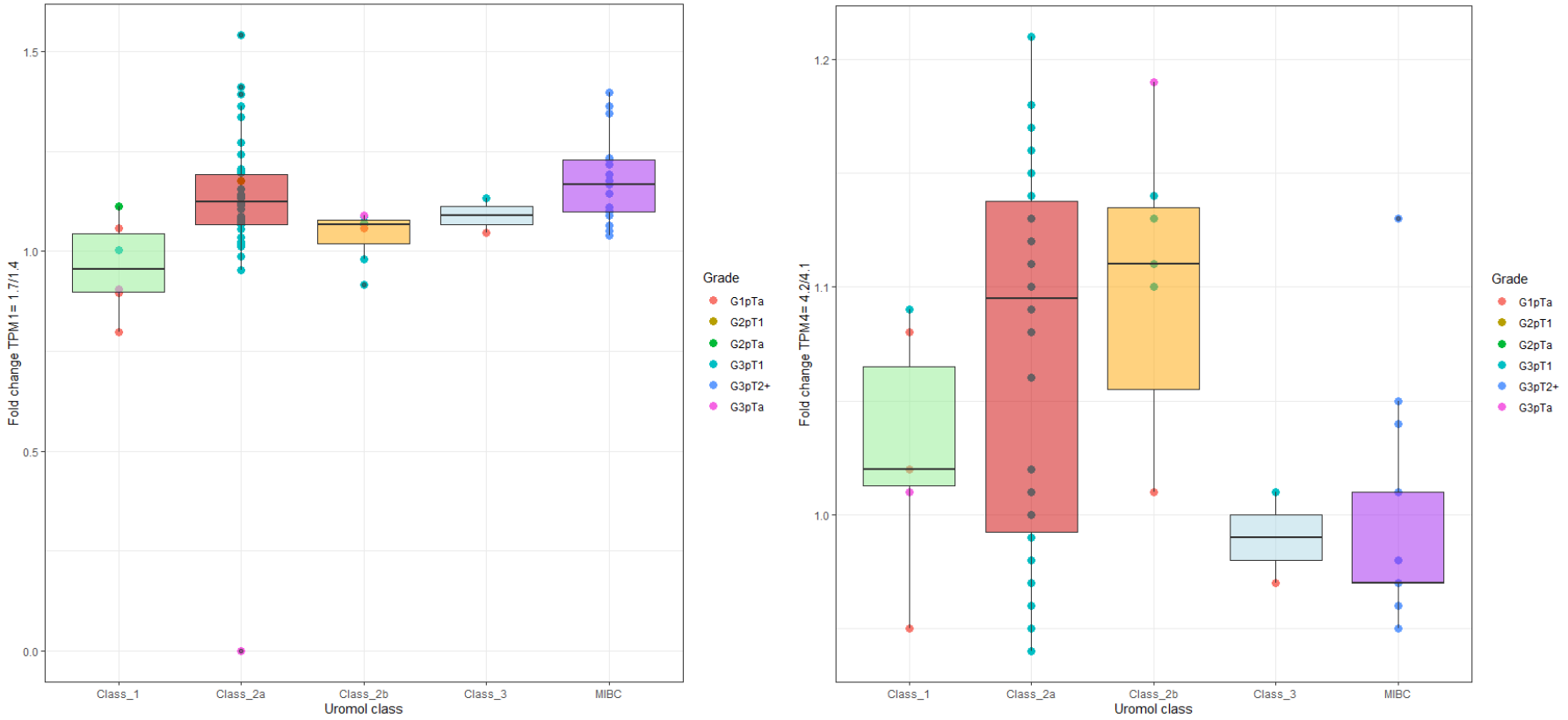


*(b) Similarly, **Tpm4.2 had higher relative expression (to Tpm4.1)** in cancer cells. Of note, Tpm4.2 has a sudden marked decrease in relative expression in MIBC. This is not due to decrease in expression of Tpm4.2 itself but an upregulation of Tpm4.1 as well, creating an overall increase in TPM4 gene expression, as noted previously.*



Furthermore, we assessed the relative expression of *TPM* isoforms in relation to UROMOL classes using the fold-change expression of Tpm1.7 to Tpm1.4 from the qPCR analysis. We found that isoform usage of Tpm1.7 is increased for Class 2a, as well as for MIBC tissues (Kruskal-Wallis p: 0.0026**). Similarly, fold-change expression of Tpm4.2 to 4.1 was evaluated and the observations show highest rates for Class 2a and 2b, especially for G3T1 tumours (Kruskal-Wallis p: 0.0022**) (Figure 6.7). Note that Class 2a was characterized by high risk G3T1 tumours.

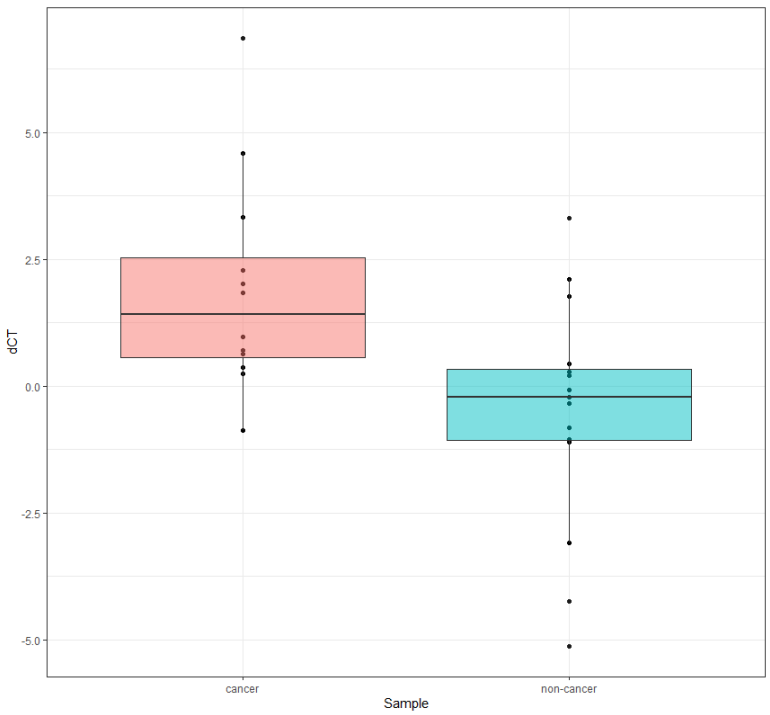
Figure 6.7 qPCR fold-change expression of TPM isoforms according to Uromol subtypes. *Fold change expression of Tpm1.7 to Tpm1.4 (left) and Tpm4.2 to Tpm4.1 (right) is shown. There is no significant difference in the classes in case of TPM1 expression. However, some variability is observed in case of TPM4 expression.*



6.5 qPCR testing in urine supernatants

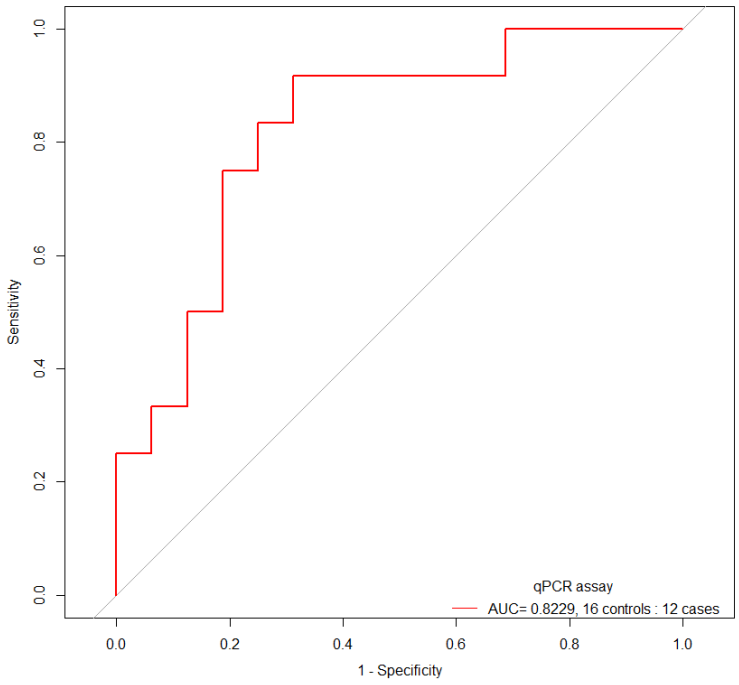
Following qPCR testing on BC tissues and cell lines, we performed preliminary biomarker assay testing (*TPM4* isoforms) with archived urine supernatants taken from patients who had BC, and others who were diagnosed with benign bladder conditions (n=12 & 16, respectively). As these samples were banked urine supernatants, RNA extraction was more challenging in comparison to fresh urine which would have provided a greater amount from pelleted cells (discussed further in section 8.5). The results indicated a dCT of 1.41 ± 2.15 (median ± SD) of relative isoform expression (Tpm4.2/4.1) in urine of bladder cancer vs a dCT of -0.216 ± 2.18 (median ± SD) in urine samples of benign bladder disease (Wilcox exact test p: 0.003122**) (Figure 6.8).

Figure 6.8 qPCR assay testing in urine samples *of bladder cancer patients compared to others with benign conditions of bladder (Wilcox exact test p: 0.003122**) using relative isoform expression of Tpm4.2/ Tpm4.1.*



Whilst accepting these data are based on a small set of samples, this assay model can effectively distinguish the presence of BC when compared to other benign bladder conditions. Performing a ROC check on *TPM4* relative isoform expression in the urine samples as a marker for disease, it was found to have an AUC value of 82.3% for diagnosis when comparing urine samples from BC patients with those harbouring no disease (Figure 6.9).

Figure 6.9 ROC curve- qPCR assay analysis using TPM4 isoform expression ratio tested in urine samples, *showing an accuracy of 82.3% as a diagnostic biomarker. 12 cases of BC were compared with 16 controls (urine samples were taken from patients that attended the haematuria clinic).*



6.6 Summary

- *Utilizing RT-PCR, TPM1-TPM4 genes were characterized in BC cell lines for isoform expression using isoform specific primers, with subsequent confirmation of PCR products using Sanger sequencing.*
- *Tpm1.7, Tpm1.4, Tpm4.2 and Tpm4.1 were correctly identified. Due to high overlap of sequences for TPM3, isoform-specific primers could not be created because at least two isoforms had identical sequences originating from the same gene.*
- *Taqman probes increased sensitivity for isoform quantification in qPCR analysis, simultaneously allowing for multiplexing isoforms of the same gene (TPM1 and TPM4).*
- *Patterns exist in TPM isoform expression between different subgroups of BC phenotype. Tpm1.7 and Tpm4.2 had highest expression in Uromol Class 2, which is characterized by high-risk G3T1 tumours.*
- *qPCR biomarker assay can successfully be used to quantify isoforms in BC tissues and cell lines with subsequent testing in urine to assess feasibility.*

Discussion:

While there is considerable heterogeneity between NMIBC and MIBC subtypes regarding isoform expression, clear distinction between BC and non-cancerous samples can be observed. Amplification of target isoforms of *TPM1* and *TPM4* was performed in NMIBC and MIBC tissue cells using the principles of Taqman qPCR to multiplex isoforms from the same gene for relative isoform expression analysis.

Other urinary biomarkers viz. CxBladder and Xpert BC Monitor (discussed in Section 1.2.4.4) have also utilized RT-qPCR for amplification and evaluation of target gene expression for diagnosing BC (114, 276, 277). From our data, the findings thus far have illustrated HMW Tpm1.7 and LMW Tpm4.2 are upregulated in cancer cells, whilst HMW Tpm1.4 and HMW Tpm4.1 have decreased expression. Higher expression of these Tpm1.7 and Tpm4.2 is noted in high-risk tumours that have an increased risk of progression. Relative isoform expression of *TPM1* and *TPM4* genes has provided a platform for development of a biomarker assay that could potentially be used as POC test (Chapter 8).

Chapter 7

TROPOMYOSIN PROTEIN ISOFORMS AS BIOMARKERS FOR BLADDER CANCER

7.1 Background

Analysis of a complex proteome (such as Tropomyosin) that has a diverse set of splice-form variants with closely matched sequences is particularly challenging. Two forms of protein identification methods have been applied in this part of the project. In the first platform, the bottom-up approach of tandem mass spectrometry (LC-MS/MS) is used for quantifying (and identifying) abundance of (unique) peptides following enzymatic digestion with trypsin prior to proteomic analysis.

Commercially available antibodies for Tropomyosin proteins provide minimal information about the isoform specificity and/ or antigen used as epitope, thereby making it difficult to test target isoforms. The antibodies used herein have been found in the literature to provide at least part of the peptide/ exon sequence that is most closely related to a specific isoform (194).

7.2 Aims and objectives

The purpose of this chapter was to investigate methods capable of identifying unique peptides that would correspond to the target Tpm isoforms. As proteins are more stable in comparison to RNA, alternative techniques of protein quantification in BC cell lines and BCPP tissues were investigated to characterize Tpm protein expression.

Additionally, we sought to identify if a Tpm protein-based immunoassay could be devised to aid in POC testing for BC diagnosis.

Objectives:

- Comparison of alternative protein quantification methods to identify the optimal technique for detecting Tpm isoforms
- Identification of peptide sequences specific to target isoforms
- Investigate available antibodies for suitability to recognize intended isoforms

7.3 Mass Spectrometry

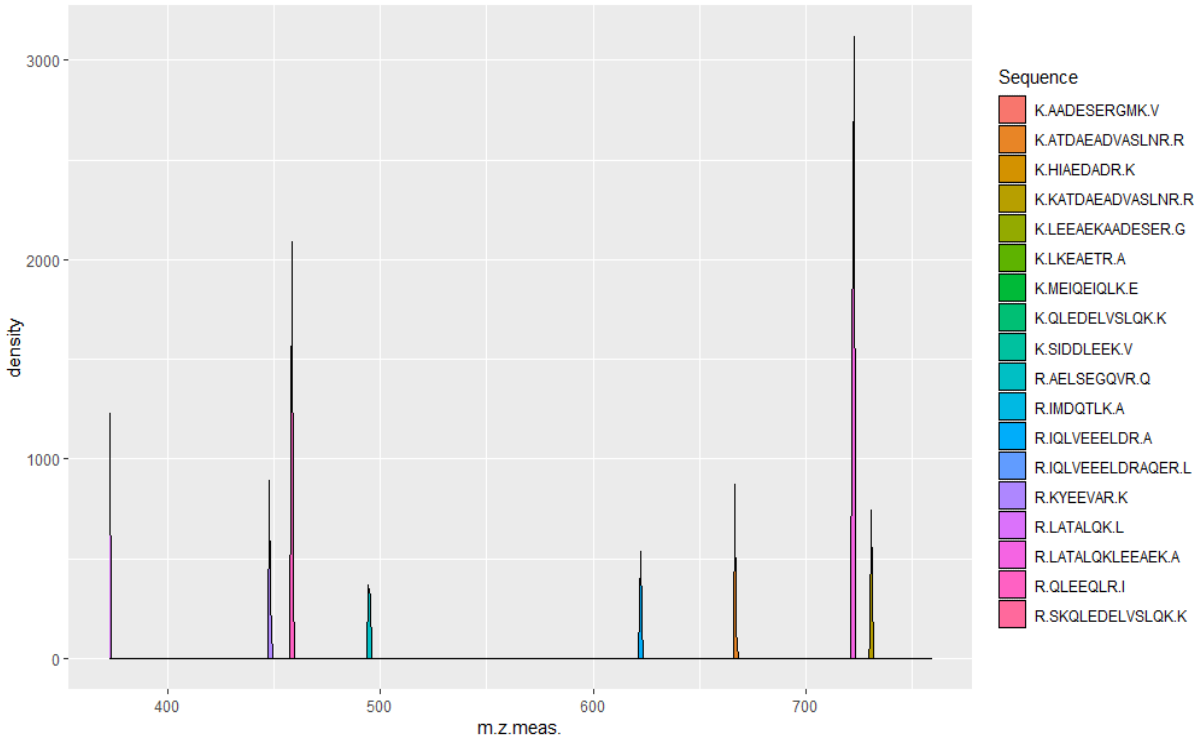
Mass spectrometry (MS) was performed on 5 BC cell lines (RT4, 5637, T24, SW780 and VMCUB1). LC-MS/MS identified 68 unique Tpm peptides (resulting from Tryptic digestion). The list of peptides was fed into a custom Python script to identify their isoform origin and corresponding exon ID. The outcome revealed that no isoform-specific peptides could be found due to high sequence similarity, with at least two isoforms sharing one peptide sequence each (Table 7.1, Figure 7.1).

Table 7.1 | List of peptide sequences from bladder cancer cell lines. Peptides identified by MS with isoform and exon alignment, in BC cell lines. Data for peptide sequences that were shared in >10 isoforms is in Appendix. Only peptides that belonged to protein-coding isoforms (found through other applications discussed before) are named.

TPM gene	Exon	Peptide sequence	Isoforms sharing peptide sequence	Isoform
1	6a/7	KALMAAEDKYSQKE	7	Tpm 1.7
1	1a/2b	KQLEDELVSLQKK	7	Tpm1.7
1	1b/3	RETAEADVASLNRR	7	Tpm 1.9
1	1b	MAGSSSLEAVRR	8	Tpm 1.9
1	5/6a	RAELSEGQVRQ	8	Tpm1.7
1	6a	RQLEEQLRI	8	Tpm 1.7
2	6b/7	KSLMASEEEYSTKE	2	
2	8/9d	KTIDDLEETLASAKE	2	
2	6a	RQLEEELRT	2	
2	4	KMELQEMQLKE	4	
2	1a/2b	KQLEEEQQALQKK	4	
2	6b/7	KSLEAQADKYSTKE	4	
2	5	KLVILEGELERS	5	
2	5	RKLVILEGELERS	5	
3	2b/3	KAADAEAEVASLNRR	4	
3	2b/3	KAADAEAEVASLNRRI	4	
3	6a/7	KCLSAAEEKY	5	
3	9d	KCTKEEHLCTQRM	5	
3	9d	KEEHLCTQRM	5	
3	8/9d	KTIDDLEDKLKC	5	

3	5/6a	RAELAESRC	7	
3	6a	RCREMDEQIRL	7	
3	6a	REMDEQIRL	7	
3	6a	RLMDQNLKC	7	
3	1b/3	RAREQAEAEVASLNRR	9	
3	1b/3	REQAEAEVASLNRRI	9	
3	1b	KIQVLQQQADDAEERAE RL	10	
3	1b	MAGITTIEAVKRK	10	
3	1b	RKIQVLQQQADDAEERA	10	
4	2b/3	KASDAEGDVAALNRR	3	Tpm 4.1
4	1a/2b	KQVEEELTHLQKK	3	Tpm 4.1
4	1b/3	REKAEGDVAALNRR	5	Tpm 4.2
4	1b	KIQALQQQADEAEDRA	6	Tpm 4.2
4	8/9d	KTIDDLEEKL	6	Tpm 4.1 and 4.2
4	1b	MAGLNSLEAVKRK	6	Tpm 4.2
4	1b	RKIQALQQQADEAEDRA	6	Tpm 4.2
4	6b/7	KSLEAASEKY	10	Tpm 4.1 and 4.2
4	8	RAEFAERT	10	Tpm 4.1 and 4.2
2; 4	1a	KLDKENAIDRA	6	Tpm 2.1 and Tpm 4.1
3; 4	8/9a	KTIDDLEDELYAQKL	8	

Figure 7.1 MS data for Tpm1.7. *The plots show the peptide signal intensity (density, y-axis) detected through MS and their mass to charge ratio (x-axis), with the peptide sequences shown on the side. These peptides make up the sequence for Tpm1.7. Detailed below is the entire sequence for Tpm1.7, indicating which exons they belong to. They have been coloured according to the density plot for easy recognition.*



MDAIKKKMQMLKLDKENALDRAEQAEADKKAAEDRSKQ	exon 1a
LEDELVSLQKKLKGTEDELDKYSEALKDAQEKLELAEKKATD	exon 2b
AEADVASLNRRIQLVEEELDRAQERLATALQKLEEAEKAADESER	exon 3
GMKVIESRAQKDEEKMEIQEIQLKEAKHIAEDADRKYEE	exon 4
VARKLVIIESDLERAEERAELSEG	exon 5
QVRQLEEQLRIMDQTLKALMAAEDK	exon 6a
YSQKEDRYEEEIKVLSDKLKE	exon 7
AETRAEFAERSVTKLEKSIDDLE	exon 8
EKVAHAKEENLSMHQMLDQTLLELNNM	exon 9d

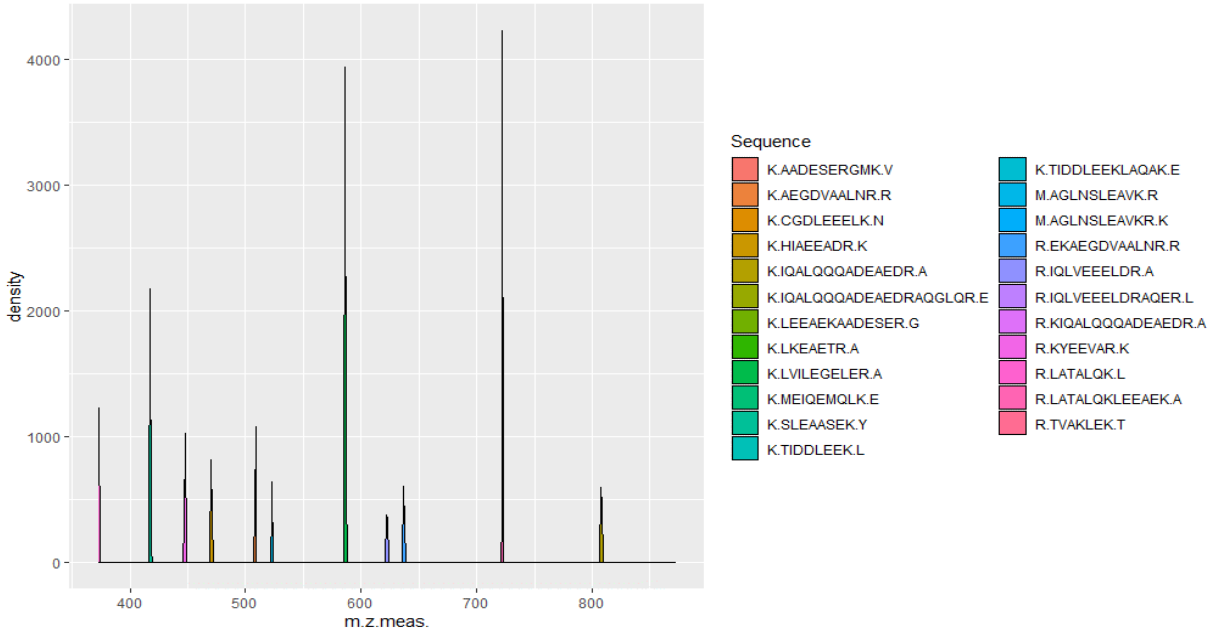
Three BCPP tissue samples were also researched for discovery of unique peptides using LC-MS/MS, which identified a total of 45 peptides with **one** unique peptide assigned to an exon-splice junction of an isoform from *TPM1* (KAEDSLLAAEEAAAKL) (Table 7.2, Figure 7.2).

Table 7.2 | List of peptide sequences from BCPP tissues. Peptides identified by MS with isoform and exon alignment in BCPP BC tissues. Data for peptide sequences that were shared in >10 isoforms is in Appendix. Only peptides that belonged to protein-coding isoforms (found through other applications discussed before) are named.

TPM gene	Exon	Peptide sequence	Isoforms sharing peptide sequence	Isoform
1	2a	KAEDSLLAAEEAAAKL	1	Tpm1.15
1	2a	RVLEELHKA	4	Tpm 1.4
1	1a/2b	KQLEDELVSLQKK	7	Tpm 1.7
1	5/6a	RAELSEGQVRQ	8	Tpm 1.7
1	6a	RQLEEQLRI	8	Tpm 1.7
2	6a/7	KSLMASEEEYSTKE	2	Tpm 2.1 and Tpm 2.4
2	8/9d	KTIDDLEETLASAKE	2	Tpm 2.2 and Tpm 2.3
2	6a	RQLEEELRT	2	Tpm 2.1 and Tpm 2.4
2	4	KMELQEMQLKE	4	
2	1a/2b	KQLEEEQQALQKK	4	
2	6b/7	KSLEAQADKYSTKE	4	
2	5	KLVILEGELERS	5	
2	5	RKLVILEGELERS	5	
3	7/8	KTIDDLEDKLKC	5	
3	8	RMLDQTLLDLNEM	5	
3	5	RLMDQNLKC	7	
3	1--2	REQAEAEVASLNRR	9	
3	1	MAGITTIEAVKR	10	
3	1b	MAGITTIEAVKRK	10	
3	1b	RKIQVLQQQADDAEERA	10	
4	1b/3	REKAEGDVAALNRR	5	Tpm 4.2
4	1b	KIQALQQQADEAEDRA	6	Tpm 4.2
4	7/8	KTIDDLEEKL	6	Tpm 4.2 and Tpm 4.1
4	1b	MAGLNSLEAVKR	6	Tpm 4.2

4	1b	MAGLNSLEAVKRK	6	Tpm 4.2
4	1b	RKIQALQQQADEAEDRA	6	Tpm 4.2
2; 4	1a	KLDKENAIDRA	6	Tpm 2.1 and Tpm 4.1

Figure 7.2 MS data for Tpm4.2. *The plots show the peptide signal intensity (density, y-axis) detected through MS and their mass to charge ratio (x-axis), with the peptide sequences shown on the side. These peptides make up the sequence for Tpm4.2. Detailed below is the entire sequence for Tpm4.2, indicating which exons they belong to. They have been coloured according to the density plot for easy recognition.*



MAGLNSLEAVKRKIQALQQQADEAEDRAQGLQRELDGERERREK exon 1b

AEGDVAALNRRIQLVEEELDRAQERLATALQKLEEAEKAADESER exon 3

GMKVIENRAMKDEEKMEIQEMQLKEAKHIAEEADRKYEE exon 4

VARKLVILEGELERAEERAEVSEL exon 5

KCGDLEEELKNVTNNLKSLEAASEK exon 6b

YSEKEDKYEEEIKLLSDKLKE exon 7

AETRAEFAERTVAKLEKTIDDLEE exon 8

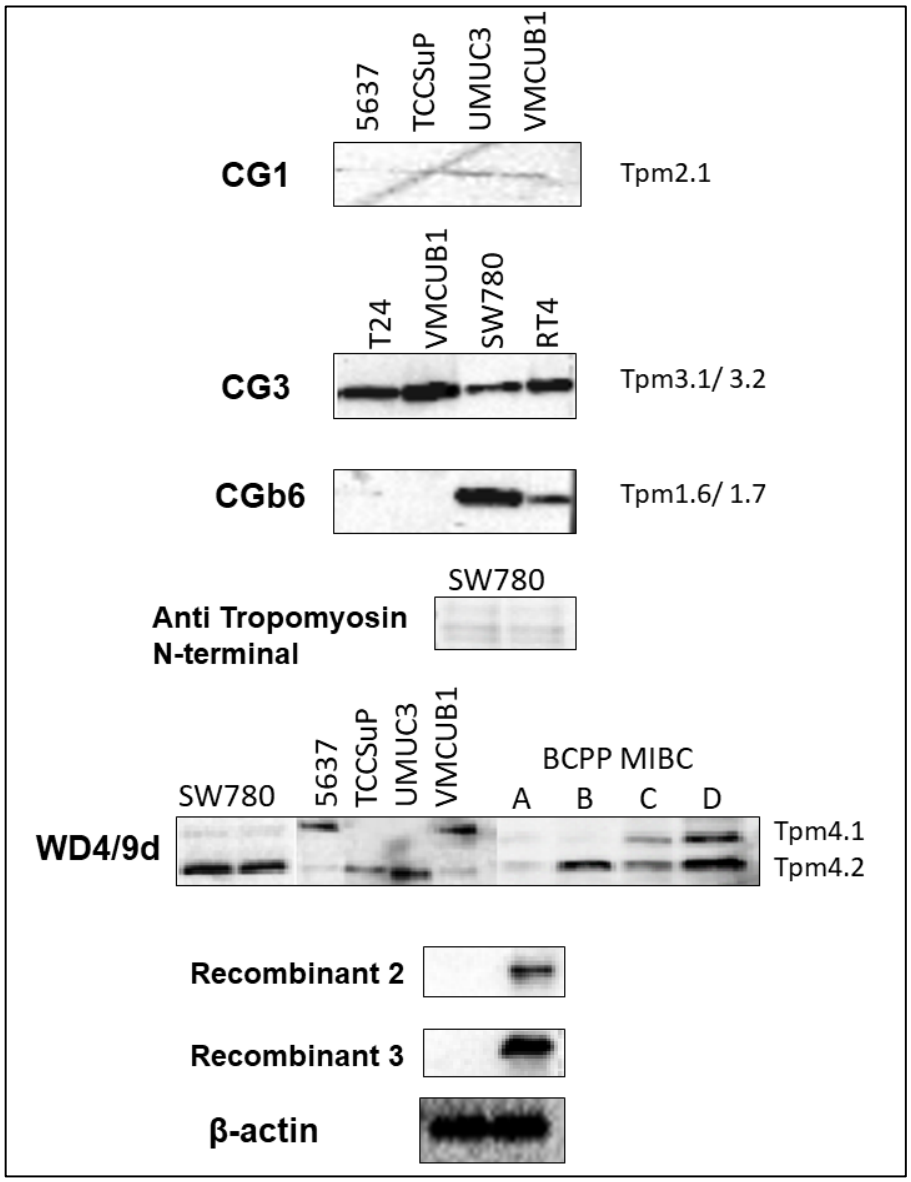
KLAQAKEENVGLHQTLDQTLNELNCI exon 9d

7.4 Western blotting

Tpm proteins were investigated in BC cell lines and BCPP tissues using a panel of antibodies for *TPM1-4* by Western blotting. Total protein was obtained from the tissues or cell lines after processing in urea buffer as described in 2.8.2.

Very weak staining of bands was detected with CG1 antibody in BC cell lines, which identifies Tpm2.1, suggesting a decrease or absence of the isoform, relative to other Tpm isoforms. The CGb6 antibody is non-specific for *TPM1* and identifies both, Tpm1.6 and Tpm1.7; 2 out of 4 cell lines had bands detected for this antibody. Most of the cell lines detected bands for CG3 antibody at LMW of 32kDa signifying the target epitope of exon 1b from *TPM3*. WD4/9d is specific for the C-terminus (exon 9d) of *TPM4*. Therefore, this antibody generated 2 bands for each of the cell lines, one for HMW Tpm4.1 and the other for LMW Tpm4.2. Additionally, 4 samples of BCPP MIBC tissues were tested with this antibody to check if the isoform usage could be identified by intensity of the bands. For both isoforms of *TPM4*, one sample showed very weak bands while another had intense bands. The remaining samples had moderate intensity for Tpm4.2, or both Tpm4.1 and Tpm4.2. Anti-tropomyosin *TPM4* antibody designed to target the N-terminus of *TPM4* showed very light bands for the SW780 cell line. It can be assumed that the N-terminus epitope is targeted at exon 1a, as there were very light bands detected for Tpm4.1 when the same cell line was tested with WD4/9d. β-actin and recombinant protein 2 and 3 were used as controls. Figure 7.3.

Figure 7.3 Antibody detection of Tpm protein isoforms. *Representative western blots of pelleted cell lines and/ or BCPP tissues were probed with the antibodies as listed (left,* ***in bold****), to detect the Tpm isoforms (right). The bottom 3 blots are controls, Recombinant Tpm 2 protein, Recombinant Tpm 3 protein and β-actin.*



7.5 Summary

- *Tandem MS detected multiple peptide fragments of Tpm isoforms, with most peptides detected for Tpm1.7 and Tpm4.2.*
- *Due to overlap of sequence coverage, peptides covering exons 3, 4, 5, 7 and 8 were similar in isoforms from the same gene and across genes.*
- *Only one peptide fragment was identified as being unique, which belonged to TPM1 isoform (Tpm1.15), which is not a commonly expressed isoform in BCPP samples.*
- *Immunoblotting with commercially available Tpm antibodies were non-specific for Tpm isoforms. However, WD4/9d detected LMW Tpm4.2 and HMW Tpm4.1 in BC cell lines and MIBC BCPP samples. CG3 antibody indicated expression for Tpm3.1/3.2 in BC cell lines. Very weak bands for Tpm2.1 were evident indicating evidence of downregulation, as detected by CG1. Expression of Tpm1.6/Tpm1.7 was found in 2/4 BC cell lines by CGb6, which is a non-specific Ab for TPM1.*

Discussion:

Protein isoform quantification proved more challenging due to the sequence similarity of amino acids of the isoforms of the same gene. Nonetheless, tandem MS was used to quantify peptide fragments to evaluate isoform origin and expression in cell lines and 4 BC tumour tissues. Consistent with results from sequencing and RT-qPCR, MS results indicated expression of Tpm1.7 and Tpm4.2.

Antibodies have been tested to detect skeletal isoforms of Tpms with limited data on non-muscle cell Tpms (194). Exon 9d specific Ab was used to identify isoform expression from *TPM4* gene that indicated presence of LMW Tpm4.2 in BC cell lines and tissues, and Tpm4.1 in high-grade BC cell lines and MIBC specimens.

Chapter 8

DISCUSSION

8.1 Bioinformatics pipeline development and comparison with others

In order to characterize isoforms resulting from AS events, RNA-sequencing data can be sourced via the 'methodology of the year 2023: LR-seq' (278) or through SR-seq (which is the more commonly used platform); both methods have limitations. Whilst LR-seq includes high cost and incomplete capture of all reads, SR-seq can erroneously assign reads to an alternative isoform, have positional bias, or even fall below the threshold of adequate reads for mapping.

At the time of conception of the *IsoTPM* pipeline, available bioinformatics algorithms were ill-suited for analysing large sets of heterogeneous data without high false-discovery-rate to quantify isoforms, or conversely, missing out pertinent reads due to (short) inadequate lengths. To overcome these difficulties, *IsoTPM* was devised to incorporate both SR-seq and LR-seq technologies and to provide a suitable comparison between the two modalities for characterization of *TPM* isoforms in BC.

In this project, large data were used incorporating multiple sets of Illumina-based sequencing with consequent increased splicing variability and heterogeneity within short-read RNA sequenced datasets. Exploration of such extensively diverse datasets pose challenges with processing, computational analysis, and visualization. *IsoTPM* is based on indexing previously annotated transcripts (with known splice junctions) for abundance quantification with subsequent normalization of the varied datasets and comparison by use of 'inclusion/ PSI' (explained in Chapter 3). Additionally,

representation of complex data was simplified by developing a series of algorithms utilizing non-parametric based statistical tests for comparative analyses.

Inference of *TPM* isoform expression curated from SR-seq assisted in verification by analysis of LR-seq data, with subsequent validation *in vitro* by applying experimental methods integrating RNA- and protein-based quantification.

In our lab, continuing bioinformatics research on wide isoform networks within bladder cancer is ongoing using the principles applied in the development of *IsoTPM*. The pipeline is not restricted to isoform usage of Tropomyosins alone, but can be expanded to identify large isoform networks involved in urothelial carcinogenesis. The rather complicated isoform family of Tropomyosins serves as a testbed for the bioinformatics pipeline, recognizing common pitfalls and allowing for improvements to accommodate future experimental splicing studies.

IsoTPM was applied to two other projects as outlined below:

- *MRE11* isoform expression was analysed in TCGA RNA seq SNP Array data (that contained rs1805363) to identify correlations with the expression of isoform 2 in response to ionizing radiation in the context of MIBC treatment. Total *MRE11A* expression was quantified using *IsoTPM* and analysed in variant samples (genotype G vs. A) versus wild type in TCGA. We found 10 isoform variants of *MRE11A*, of which 5 were expressed in TCGA. This work was performed in collaboration with the University of Oxford (279).
- The *IsoTPM* pipeline was also used to investigate non-coding mutation isoform expression in wild type and mutant BCs across 5 hotspots: *GPR126*, *PLEKHS1*, *TBC1D12*, *LEPROTL1* and *WDR74*. No differential expression of the isoforms of these genes was identified, and the frequency of mutations was similar in all grades

and stages of BC. However, mutations occurred concurrently and in association with *APOBEC* signature (280).

8.2 Tropomyosin isoform expression in urothelial carcinoma of bladder

8.2.1 Overview

In this project the expression of Tropomyosin genes and their isoforms in urothelial bladder cancer have been comprehensively explored. Tpms are the key regulators of cellular structural configuration (in conjunction with actin). In BC cells, it has been observed that a marked decrease in the expression of HMW isoforms (particularly Tpm2.1) with upregulation of LMW isoforms (Tpm3.1/3.2, Tpm4.2) is required to promote cell proliferation and aid in cell motility (216). These findings were observed in our analysis of RNA-seq datasets, which illustrated overall downregulation of HMW isoforms of *TPM2* and upregulation of LMW isoforms of *TPM4* and *TPM3.* Nascent urothelial cells have organized actin filaments requiring HWM Tpms (Tpm2.1, Tpm1.4) to stabilize and strengthen cellular architectural integrity. The turnover of HMW isoforms is maintained in a slow homeostatic environment to strengthen actin filament organization (as bundled or cross-linked fibres). Furthermore, our results indicated that an increase in the expression of Tpm1.7 and Tpm4.2 isoforms, as seen in bladder cancer cells, is associated with the changing morphology and reorganization of cytoskeletal elements to enhance cytokinesis and cellular turnover, leading to increased tumour aggressiveness. This phenomenon is illustrated by an upward trend in expression of these isoforms from benign urothelium to NMIBC and/or MIBC. Interestingly, Tpm1.7 is a HMW isoform of *TPM1* which is expressed by alternative

usage of exon 6(a), resulting in *mis*expression, leading to fracture of the stress fibres and alterations in cellular morphology and a carcinomatous phenotype (281). LMW isoforms are unable to bind strongly to actin filaments, thereby increasing their vulnerability to severing by ADF/ cofilin leading to branching of networks. The branched networks/ membrane ruffles are found in lamellipodia and filopodia of the malignant urothelial cell to assist in migration, and breach and escape through the lamina propria. Tpm4.2, when bound to the F-actin filament, recruits myosin IIA which is implicated in the formation of the cleavage furrow to aid in cytokinesis (282). We demonstrated an increase in the HMW isoform Tpm4.1 in muscle-invasive BC was observed, potentially attributable to smooth (detrusor) muscle cells origin to signify the 'invasive' characteristics of the tumour.

8.2.2 Co-expression of *TPM4* isoforms in muscle-invasive bladder cancer

Increased expression of Tpm4.1 in muscle-invasive bladder cancer is a novel finding, and it would be noteworthy if the upregulation of Tpm4.1 is a consequence of, rather than an instigator of events signifying breach of the basement membrane and invasion into muscle: when bound to F-actin, HMW isoform Tpm4.1 facilitates cell-cell adhesion, whereas Tpm4.2 promotes cellular proliferation (283, 284). Increased co-expression of these two isoforms propagates tumour cells to grow and advance beyond the lamina propria in MIBC with an overall increase in cellular/ tumour density (284). It is postulated that the initial phase of MIBC tumour expansion selectively upregulates Tpm4.2. However, as tumour cells invade, their stability is maintained by increased expression of Tpm4.1 alongside Tpm4.2. ECM stiffness, assisted by recruitment of

fibroblasts and epithelial cells, attempts to anchor cells via focal adhesion molecules (285, 286). This is counteracted by increased tyrosine kinase activity (signalling via *MAPK/ERK* pathways) to modulate adhesion, thereby increasing cell proliferation secondary to altered *TPM4* isoform- actin dynamics.

It would be interesting to further investigate the concentration of HMW Tpm4.1 (bound to actin) in MIBC tumour tissues and its co-expression with LMW Tpm4.2 as a quantitative marker comparing protein and/ or RNA expression. This novel finding could potentially assist in determining aggressiveness of the BC type and aid in predicting disease progression based on concomitant expression of Tpm4.1 and Tpm4.2. If both isoforms are found to be higher in BC cells compared to normal urothelium (even if relative isoform expression is similar to that of urothelium), it would indicate presence of BC. The basis of these findings would prove to be of clinical relevance when translating into a quantitative point-of-care test (discussed in more detail in section 8.3).

8.2.3 Immune correlation with *TPM* isoform expression

The tumour microenvironment (TME) serves an important role in selection and recruitment of immune cells to affect tumour development, propagation, and eventual prognosis (287). It is particularly leveraged in bladder cancer as intravesical immunotherapy regimens are a standard of care for NMIBC and, with the upcoming advances in immune checkpoint inhibitors, immunotherapy has also dramatically changed treatment strategies for MIBC e.g. Enfortumab Vedotin +/- Pembrolizumab (147). Polarized tumour-associated macrophages can influence treatment response to

immunotherapy and play an active role in prognosis (288). In addition to immune cells (T-cells, B-cells, tumour-associated macrophages, dendritic cells, and a small proportion of natural killer (NK) cells), the bladder cancer TME includes the tumour cells, inflammatory chemokines and cytokines, stromal epithelial cells and fibroblasts (285).

In our analyses, there were associations of *TPM* isoforms with immune cells and direct correlations with fibroblast recruitment and macrophages. Specifically, *TPM2* gene expression was found to be directly correlated to stromal enhancement with fibroblasts, whereas *TPM3* expression had a linear correlation with neutrophils and tumour-associated macrophages. At the isoform level, Tpm2.1 and Tpm3.1/ 3.2 were found to be upregulated in these instances. Additionally, the UROMOL single-sample classifier grouped these tumours under class 2b and 2a, respectively. Previous analyses by Goel et al in the BCPP samples using whole exome sequencing had also shown increased immune and EMT markers for class 2b (289). It is also observed that increased expression of Tpm2.1 indirectly dampens expression of other *TPM* genes since it appears to be responsible for anoikis (programmed cell death) in urothelial cells (290). Conversely, increased expression of *TPM2* has been proposed to be a fibroblast-specific marker and a poor prognostic indicator in colorectal cancer (291). It is postulated that the enhanced ECM and fibroblast enrichment observed in our data is associated with cancer-associated fibroblasts that alter the TME to provide support to the BC cells to continue growth and assist in cellular adhesion. Notably, cancer-associated fibroblasts (due to epigenetic modifications) have been associated with BC tumour progression, EMT transition and effects on the immune system (292, 293).

Neutrophils, deemed as the ‘first responders’ of the immune system, are found to be elevated (or within high normal range) in the presence of cancer (294). In the early phases of carcinomatous change, the TME manipulates the immune cells to create favourable phenotypic polarization states to promote tumour growth and invasion. The correlation between neutrophils and Tpm3.1 implicates this LMW isoform as a candidate oncogene, strongly associated with carcinogenesis and progression (216, 295).

8.3 Clinical significance and relevance

8.3.1 qPCR assay as a biomarker tool for detection of urothelial carcinoma of bladder

In our work, two pairs of isoforms of *TPM1* and *TPM4* genes (Tpm1.7/ Tpm1.4 and Tpm4.2/ Tpm4.1) have shown strong potential as diagnostic biomarkers of BC. We have elicited the foundations of an RNA-based biomarker test based on the observation of differential isoform usage in BC vs. benign bladder conditions. Preliminary qPCR assay testing with urine supernatants showed a dCT ratio of 1.41 between relative isoform expression in cancer vs benign conditions (p-value: 0.00312, n=12 and 16, respectively). As elicited in urine samples (following protocol refinement in tumour tissues), using a relative isoform expression level has an overall higher AUC value (82.3%) for diagnosis compared to individual isoform expression.

Overall *TPM* gene expression does not appreciate the difference in isoform usage – it is a consequence of relative isoform usage that enables detection of BC from other benign conditions of the bladder. This highlights the potential importance of AS and

the significance of bespoke bioinformatics pipelines such as *IsoTPM* in identifying these characteristics, that influence fundamental cell morphology and behaviour.

8.3.2 Detection in urine

Tropomyosins are very stable and can easily be extracted from tissue and/ or urine samples. Due to the stability of Tpms and their resistance to temperature changes and/ or degradation by proteases, their longer bioavailability in urine could prove advantageous for use as a diagnostic biomarker.

The caveats with using urine as a biomarker source are the quality of the sample, its storage, transport, and processing for analysis. Optimal procedural design is required to prevent contamination.

As noted previously, currently available biomarkers may provide high sensitivity but not equal that of flexible cystoscopy (sensitivity: 85%, specificity: 87%)(61), with the inability to differentiate between BC and other (inflammatory) conditions that may affect the urothelium. A possible solution is to utilize a combinatorial score of biomarker readouts with other *risk factors for BC* as well as non-invasive imaging results, with the aim of improving the sensitivity and specificity closer to 100%.

With the use of qPCR, such an assay (following refinement), could serve as an inexpensive test available in the primary care setting to triage patients who require further investigations (and subsequent intervention) for haematuria. In addition to the advantage of having a quick turnaround time for results, the other benefits may include cost-effectiveness (by reducing the need for avoidable referral to secondary care), combined with significant reduction in numbers of cystoscopies and a resultant

decreased burden to the healthcare systems (e.g. the NHS in the UK). Due to the simplicity of the assay design, it would eliminate the need to send urine samples to specialized laboratories, decreasing administrative burden on primary care. Such a streamlined workflow would allow for improved operational efficiency in parallel with optimal patient care. From a patient perspective, timely diagnosis can be achieved with rapid access to referral and treatment where needed, and improved outcome. Additionally, a non-invasive *rapid result* test can significantly reduce anxiety, waiting times and provide enhanced patient satisfaction.

Nonetheless, the RNA-based assay can provide numerous benefits as a diagnostic test for primary care use, it is essential to ensure that appropriate utilization and interpretation of results is maintained via proper training, quality assurance and adherence to contemporary guidelines on managing haematuria patients. Therefore, the biomarker assay must undergo validation studies to assess its feasibility and accuracy, such that it should exemplify high sensitivity (with NPV). Such data (with transparency) would ultimately allow increased uptake and implementation of non-invasive diagnostic tests in the clinical setting.

8.4 Our findings and the *IsoTPM* pipeline are of relevance to other cancers[(1)]

Tropomyosins have been studied in other cancers, as well as in some neurodegenerative diseases. In this section, only those closely related to the urothelium i.e. having an epithelial/ mucosal morphology or associated with the urinary tract have been discussed.

8.4.1 Prostate cancer (CaP)

Alternative splicing events give rise to several isoforms in normal cells and may not be directly involved in carcinogenesis, as discussed. However, some isoforms do stimulate cellular expansion and the growth of malignant cells, providing a “tumour microenvironment”. This phenomenon is observed in prostate cancer where mutagenic pathways cause overexpression of the *TPM1* isoform, providing a platform for distant metastases (262). Thorsen et at found alternative expression of the mutually exclusive exon 6 and compared it between normal prostatic tissue and CaP. The Tpm isoform variant found in localized prostate cancer (as well as metastatic CaP) expressed exon 6a (Tpm1.7). The exon array data analyses were also undertaken for bladder urothelium, and the results indicated exclusive usage of exon 6b (Tpm1.6) in normal samples whereas bladder cancer had both 6a and 6b (262).

These findings have been corroborated in this project, where Tpm1.7 was found to have the highest expression in BC tissues and cell lines.

8.4.2 Malignant Breast Epithelial cells

The molecular classification for MIBC that includes the luminal- and basal- subtypes are highly reminiscent of the molecular subtypes and heterogeneity noted in breast cancer (167). Previous studies have demonstrated commonalities between the two cancers with regards to gene expression signatures, transcription factor alterations, and cellular pathways (167, 296, 297). HMW tropomyosin isoforms of *TPM1* and *TPM2* are expressed in normal breast epithelial cells along with the expression of LMW isoforms of *TPM1* and *TPM4* (akin to differential isoform expression in normal urothelial

cells). Expression of these isoforms appears to be downregulated in malignant breast epithelial cells. An additional four novel isoforms of the *TPM1* gene have been identified in nine malignant breast epithelial cell lines, isoforms that and are not otherwise expressed in normal adult breast tissue (298). The *TPM2* isoform has been found markedly upregulated in aggressive forms of breast cancer with lymph node metastases (298, 299), and could potentially serve as a prognostic biomarker.

8.4.3 Renal cell cancer

TPM1 gene expression has been closely linked to miRNA (as previously discussed in the context of BC, where *miR-96* contradicts normal *TPM1* expression). In renal cell carcinoma, *miR-21* has been observed to counteract *TPM1* activity by promoting cell cycle progression, growth and inhibiting apoptosis (which is otherwise amenable to increased *TPM1* expression) (264, 300).

8.4.4 Oesophageal cancer

Ras effector regulatory pathways *MEK/ERK* and *PI3K/AKT* are compromised in oesophageal squamous cell carcinoma cell lines, leading to downregulation and decreased expression of *TPM1* and *TPM2* (263, 301). A similar pattern is observed in BC, where downregulation of HMW Tpm isoforms leads to higher grade and stage (224).

Given the fundamental cellular roles of Tpms, and the considerable differences observed between protein isoforms, it is feasible to consider that Tpm isoforms and AS

events represent promising diagnostic and prognostic biomarkers across a range of cancers and other diseases. Hence, *IsoTPM* represents an important tool for biomedical research.

8.5 Limitations

This project was undertaken amidst the global pandemic, thereby requiring adjustments for several challenges, particularly in sample collection, delay in transfer and results acquisition. Furthermore, this is a PhD project incorporating multidisciplinary approaches: bioinformatics analyses of transcriptomes from large datasets (using short-read and long-read RNA sequencing platforms), subsequent curation of significant isoforms, and the eventual development of a biomarker assay. The exploitation of numerous methods of quantification required multistep QC checks to ensure continuity and consistency in evaluation and interpretation of data.

Initial work package design for validation and quantification of tropomyosin isoforms *in vivo* included multiple testing in urine and tissue samples from treatment-naïve patients undergoing their first 'index' TURBT. Due to the pandemic, sample collection and transfer from our collaborators at Wrexham Academic Centre was difficult, leading to unforeseen delays and reduced numbers of samples procured. Adjustments were made to utilize available resources whilst still allowing for meaningful endpoints for data validation. The available alternatives to fresh (whole) urine samples were archived urine supernatants. However, extracting RNA from urine supernatant is much more challenging than RNA extraction from pelleted urine cells. Nonetheless, we were able

to quantify and test for the expression of LMW Tpm isoforms in the urine samples using qPCR.

The bladder cancer cell lines cultured in-house may differ from the original cell line composition due to cell culture media and local environmental conditions, as their line origins and validity were confirmed prior to cell line experiments. This could explain the differences in fold-change expression of isoforms in qPCR when compared to RNA-seq data. Notably, cell lines are not representative of real tumours and primary cultures of real BC tumours may be another avenue to explore.

Tropomyosins have a diverse number of isoforms that encompass significant sequence overlap which made it challenging to specifically identify target isoforms. Multiple checks with optimization of designs and protocols were time-consuming and thus prolonged the development of an optimal biomarker assay.

Regarding long-read RNA sequencing, forward and reverse primers were designed at exon 1 and 9 to facilitate high capture quantification of isoforms. Strategic placement of primers at the 3’ and 5’ end meant full-length reads were captured effectively. Although advantageous, any shorter length isoforms not encased within the specified targets may have been missed. During the demultiplexing step we found that the barcode sequence tags used as adapters for sample recognition were truncated prematurely leading to loss of the original sample pool (and ZMW reads). The built-in criteria by lima undergoes various filtering steps to clip barcode positions to generate a score which determines whether a read has ‘passed’ or ‘failed’. ZMW reads that did not fulfil the required length were deemed as ‘failed’ and were removed from the analyses, meaning that not all our samples were processed to generate reads.

8.6 Future experiments

8.6.1 Antibody development

As noted earlier, the commercially available antibodies that we tested do not provide isoform specificity. Whilst some have exon-specific epitopes (as in the case of WD9/4d which identifies exon 9d from *TPM4*), others can identify multiple isoforms from the same gene (CGb6) and even across other *TPM* genes (Anti-tropomysosin N-terminal Ab).

For the purpose of isoform specificity and to enable detection of LMW and HMW target isoforms from *TPM1* and *TPM4*, it is proposed that development of an enhanced peptide-targeted antibody prototype incorporating ELISA, or phage display (or a combination of both) may prove valuable.

With the high sensitivity (and specificity) provided by ELISA to quantitatively detect interaction of antigen/ protein with monoclonal (and/ or polyclonal) antibodies, it can be used to identify exact sequences of the target Tpm isoforms in BC samples for high throughput diagnostics. However, there may be a risk of cross-reactivity with other Tpm isoforms which may prove challenging for designing the ‘perfect’ antibody (302).

Alternatively, phage display may be a viable option to generate specific Tpm-isoform-targeted antibodies using recombinant tagged bacterio’*phage*’ proteins. Phage libraries are constructed by genetically fusing antibody fragments that coat the surface of a filamentous bacteriophage particle (usually M13) followed by selection of the most suitable high density, high antigenicity, high affinity clone through biopanning (a technique which allows careful selection of peptides that bind to a target). Apart from high costs, this technique requires rigorous validation, confirmation of the correct target

affinity and specificity as there is a risk of selection bias during *in vitro* library construction (303).

8.6.2 Validation studies for qPCR biomarker assay

Our data thus far has revealed relative expression of *TPM* isoforms can be used as a biomarker assay to detect BC. Preliminary testing was conducted on a small cohort of urine samples taken from patients that attended the haematuria clinic and were either found to have BC (cases) or had other benign diseases (controls). Further validation studies in urine samples are needed to exploit the potential of the test to confirm its diagnostic performance for translation into clinical practice.

8.6.3 Whole gene transcriptomic profiling and co-expression analyses with *TPM*

At present, *IsoTPM* is based on quantification of known (predicted) isoforms using the Ensembl annotation as index transcriptome. Future modifications to the existing pipeline could include identification of novel isoforms based on local splicing events (exon skipping, or novel donor/acceptor sites) using splice-junction aware software tools (RMATs, MAJIQ) (304, 305). New bioinformatics pipelines have also come to the fore recently (305), which are capable of analysing similar data structures as used in this project. It would be pertinent to investigate if our results are consistent irrespective of quantification algorithms. Furthermore, similarities in co-expression of *TPM* isoforms with early and/ or late cell cycle regulons in order to investigate the downstream effects

on cytoskeletal proteins may shed further insight into the functional capabilities of tropomyosins and the consequences of isoform switching.

CONCLUSION

Whilst tropomyosin isoforms have been investigated for decades, their true potential as cancer biomarkers had not been comprehensively evaluated to date in urothelial carcinoma of the bladder. Given their complexity of isoforms, as well as tissue- and cell-specific functions, multimodal characterization through use of transcriptomic profiling, sequencing, and expression analyses have provided insight into their significance in bladder carcinogenesis. They play an integral role in combination with the actin cytoskeleton in the maintenance of cell morphology, and alterations in gene expression may mediate EMT, leading to more aggressive disease and worse survival outcomes. Notwithstanding. AS and isoform switching are characteristics that could lead to promising and cost-effective throughput diagnostic biomarkers. With ongoing technological advances, the investigation of tropomyosin isoforms could mediate further advances in BC diagnosis and risk stratification. *IsoTPM* represents the foundations of such investigations. Further studies are warranted to validate our findings and to discover novel isoforms genome-wide that may play a role in promoting and maintaining cancer cell survival.

Nonetheless the findings from this study have led to development of an RNA-based biomarker assay that may enable the simple and cost-effective detection of bladder cancer in urine samples.

REFERENCES

1. Humayun-Zakaria N, Arnold R, Goel A, Ward D, Savill S, Bryan RT. Tropomyosins: Potential Biomarkers for Urothelial Bladder Cancer. Int J Mol Sci. 2019;20(5).
2. Humayun-Zakaria N, Ward DG, Arnold R, Bryan RT. Trends in urine biomarker discovery for urothelial bladder cancer: DNA, RNA, or protein? Transl Androl Urol. 2021;10(6):2787-808.
3. My diagnosis counts. Available from: https://fightbladdercancer.co.uk/sites/default/files/downloads/My-diagnosis-counts.pdf.
4. UK CR. 2021 [Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bladder-cancer#heading-One.
5. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians. 2021;71(3):209-49.
6. Saginala K, Barsouk A, Aluru JS, Rawla P, Padala SA, Barsouk A. Epidemiology of Bladder Cancer. Med Sci (Basel). 2020;8(1).
7. Uhlig A, Strauss A, Seif Amir Hosseini A, Lotz J, Trojan L, Schmid M, et al. Gender-specific Differences in Recurrence of Non–muscle-invasive Bladder Cancer: A Systematic Review and Meta-analysis. European Urology Focus. 2018;4(6):924-36.
8. Marks P, Soave A, Shariat SF, Fajkovic H, Fisch M, Rink M. Female with bladder cancer: what and why is there a difference? Translational Andrology and Urology. 2016;5(5):668-82.
9. Dobruch J, Daneshmand S, Fisch M, Lotan Y, Noon AP, Resnick MJ, et al. Gender and Bladder Cancer: A Collaborative Review of Etiology, Biology, and Outcomes. Eur Urol. 2016;69(2):300-10.
10. Lucca I, Fajkovic H, Klatte T. Sex steroids and gender differences in nonmuscle invasive bladder cancer. Curr Opin Urol. 2014;24(5):500-5.
11. Koti M, Ingersoll MA, Gupta S, Lam CM, Li X, Kamat AM, et al. Sex Differences in Bladder Cancer Immunobiology and Outcomes: A Collaborative Review with Implications for Treatment. European Urology Oncology. 2020;3(5):622-30.
12. Ball KS, Hounsome L, Verne J, Kockelbergh R. Non-transitional cell carcinoma only partly explains adverse survival outcomes in females with T1–T4 bladder cancer: A summary of UK epidemiological data. Journal of Clinical Urology. 2017;10(1_suppl):14-8.
13. Mushtaq J, Thurairaja R, Nair R. Bladder cancer. Surgery (Oxford). 2019;37(9):529-37.
14. Martin JW, Carballido EM, Ahmed A, Farhan B, Dutta R, Smith C, et al. Squamous cell carcinoma of the urinary bladder: Systematic review of clinical characteristics and therapeutic approaches. Arab J Urol. 2016;14(3):183-91.
15. Shokeir AA. Squamous cell carcinoma of the bladder: pathology, diagnosis and treatment. BJU International. 2004;93(2):216-20.
16. Ismaili N. A rare bladder cancer - small cell carcinoma: review and update. Orphanet Journal of Rare Diseases. 2011;6(1):75.
17. Boström CE, Gerde P, Hanberg A, Jernström B, Johansson C, Kyrklund T, et al. Cancer risk assessment, indicators, and guidelines for polycyclic aromatic hydrocarbons in the ambient air. Environ Health Perspect. 2002;110 Suppl 3(Suppl 3):451-88.
18. Vineis P, Pirastu R. Aromatic amines and cancer. Cancer Causes Control. 1997;8(3):346-55.
19. Cancer IAfRo. IARC Monographs on the identification of carcinogenic hazards to humans [Available from: https://monographs.iarc.who.int/agents-classified-by-the-iarc/.
20. Teoh JY-C, Huang J, Ko WY-K, Lok V, Choi P, Ng C-F, et al. Global Trends of Bladder Cancer Incidence and Mortality, and Their Associations with Tobacco Use and Gross Domestic Product Per Capita. European Urology. 2020;78(6):893-906.

21. Rink M, Furberg H, Zabor EC, Xylinas E, Babjuk M, Pycha A, et al. Impact of smoking and smoking cessation on oncologic outcomes in primary non-muscle-invasive bladder cancer. Eur Urol. 2013;63(4):724-32.
22. Samanic C, Kogevinas M, Dosemeci M, Malats Nr, Real FX, Garcia-Closas M, et al. Smoking and Bladder Cancer in Spain: Effects of Tobacco Type, Timing, Environmental Tobacco Smoke, and Gender. Cancer Epidemiology, Biomarkers & Prevention. 2006;15(7):1348-54.
23. Vineis P, Caporaso N, Tannenbaum SR, Skipper PL, Glogowski J, Bartsch H, et al. Acetylation phenotype, carcinogen-hemoglobin adducts, and cigarette smoking. Cancer Res. 1990;50(10):3002-4.
24. Lee HW, Park SH, Weng MW, Wang HT, Huang WC, Lepor H, et al. E-cigarette smoke damages DNA and reduces repair activity in mouse lung, heart, and bladder as well as in human lung and bladder cells. Proc Natl Acad Sci U S A. 2018;115(7):E1560-e9.
25. Bjurlin MA, Matulewicz RS, Roberts TR, Dearing BA, Schatz D, Sherman S, et al. Carcinogen Biomarkers in the Urine of Electronic Cigarette Users and Implications for the Development of Bladder Cancer: A Systematic Review. European Urology Oncology. 2021;4(5):766-83.
26. van Osch FH, Jochems SH, van Schooten F-J, Bryan RT, Zeegers MP. Quantified relations between exposure to tobacco smoking and bladder cancer risk: a meta-analysis of 89 observational studies. International Journal of Epidemiology. 2016;45(3):857-70.
27. Lammers RJM, Witjes WPJ, Hendricksen K, Caris CTM, Janzing-Pastors MHC, Witjes JA. Smoking Status Is a Risk Factor for Recurrence After Transurethral Resection of Non–Muscle-Invasive Bladder Cancer. European Urology. 2011;60(4):713-20.
28. Boström PJ, Alkhateeb S, Trottier G, Athanasopoulos PZ, Mirtti T, Kortekangas H, et al. Sex differences in bladder cancer outcomes among smokers with advanced bladder cancer. BJU International. 2012;109(1):70-6.
29. Safiri S, Kolahi AA, Naghavi M. Global, regional and national burden of bladder cancer and its attributable risk factors in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease study 2019. BMJ Glob Health. 2021;6(11).
30. Brown KF, Rumgay H, Dunlop C, Ryan M, Quartly F, Cox A, et al. The fraction of cancer attributable to modifiable risk factors in England, Wales, Scotland, Northern Ireland, and the United Kingdom in 2015. British Journal of Cancer. 2018;118(8):1130-41.
31. Dietrich H, Dietrich, B. Ludwig Rehn (1849-1930) - Pioneering findings on the aetiology of bladder tumours. World J Urol. 2001;19:151-3.
32. Jalilian H, Ziaei M, Weiderpass E, Rueegg CS, Khosravi Y, Kjaerheim K. Cancer incidence and mortality among firefighters. International Journal of Cancer. 2019;145(10):2639-46.
33. Hadkhale K, Martinsen JI, Weiderpass E, Kjærheim K, Sparén P, Tryggvadóttir L, et al. Occupational variation in bladder cancer in Nordic males adjusted with approximated smoking prevalence. Acta Oncologica. 2019;58(1):29-37.
34. Johns LE, Houlston RS. N-acetyl transferase-2 and bladder cancer risk: a meta-analysis. Environ Mol Mutagen. 2000;36(3):221-7.
35. Engel LS, Taioli E, Pfeiffer R, Garcia-Closas M, Marcus PM, Lan Q, et al. Pooled analysis and meta-analysis of glutathione S-transferase M1 and bladder cancer: a HuGE review. Am J Epidemiol. 2002;156(2):95-109.
36. Marcus PM, Hayes RB, Vineis P, Garcia-Closas M, Caporaso NE, Autrup H, et al. Cigarette smoking, N-acetyltransferase 2 acetylation status, and bladder cancer risk: a case-series meta-analysis of a gene-environment interaction. Cancer Epidemiol Biomarkers Prev. 2000;9(5):461-7.
37. Pesch B, Gawrych K, Rabstein S, Weiss T, Casjens S, Rihs H-P, et al. N-acetyltransferase 2 Phenotype, Occupation, and Bladder Cancer Risk: Results from the EPIC Cohort. Cancer Epidemiology, Biomarkers & Prevention. 2013;22(11):2055-65.
38. Aktas D, Ozen H, Atsu N, Tekin A, Sozen S, Tuncbilek E. Glutathione S-transferase M1 gene polymorphism in bladder cancer patients: a marker for invasive bladder cancer? Cancer Genetics and Cytogenetics. 2001;125(1):1-4.

39. Joost P, Therkildsen C, Dominguez-Valentin M, Jönsson M, Nilbert M. Urinary Tract Cancer in Lynch Syndrome; Increased Risk in Carriers of MSH2 Mutations. Urology. 2015;86(6):1212-7.
40. van der Post RS, Kiemeney LA, Ligtenberg MJ, Witjes JA, Hulsbergen-van de Kaa CA, Bodmer D, et al. Risk of urothelial bladder cancer in Lynch syndrome is increased, in particular among MSH2 mutation carriers. J Med Genet. 2010;47(7):464-70.
41. Vasen HF, Stormorken A, Menko FH, Nagengast FM, Kleibeuker JH, Griffioen G, et al. MSH2 mutation carriers are at higher risk of cancer than MLH1 mutation carriers: a study of hereditary nonpolyposis colorectal cancer families. J Clin Oncol. 2001;19(20):4074-80.
42. Shi R, Wang X, Wu Y, Xu B, Zhao T, Trapp C, et al. APOBEC-mediated mutagenesis is a favorable predictor of prognosis and immunotherapy for bladder cancer patients: evidence from pan-cancer analysis and multiple databases. Theranostics. 2022;12(9):4181-99.
43. Lada AG, Dhar A, Boissy RJ, Hirano M, Rubel AA, Rogozin IB, et al. AID/APOBEC cytosine deaminase induces genome-wide kataegis. Biol Direct. 2012;7:47; discussion
44. Shi MJ, Meng XY, Lamy P, Banday AR, Yang J, Moreno-Vega A, et al. APOBEC-mediated Mutagenesis as a Likely Cause of FGFR3 S249C Mutation Over-representation in Bladder Cancer. Eur Urol. 2019;76(1):9-13.
45. Kripnerova M, Parmar HS, Pesta M, Kohoutova M, Kuncova J, Drbal K, et al. Urothelial Cancer Stem Cell Heterogeneity. Adv Exp Med Biol. 2019;1139:127-51.
46. Bolla SR ON, Jetti R. Histology, Bladder: Treasure Island (FL): StatPearls Publishing; 2022.
47. HICKS RM. THE MAMMALIAN URINARY BLADDERAN ACCOMMODATING ORGAN. Biological Reviews. 1975;50(2):215-46.
48. Wang C, Ross WT, Mysorekar IU. Urothelial generation and regeneration in development, injury, and cancer. Developmental Dynamics. 2017;246(4):336-43.
49. Mostofi FK, Sobin LH, Torloni H, World Health O. Histological typing of urinary bladder tumours / F. K. Mostofi, in collaboration with L. H. Sobin, H. Torloni and pathologists in fourteen countries. Geneva: World Health Organization; 1973.
50. van der Kwast T, Liedberg F, Black PC, Kamat A, van Rhijn BWG, Algaba F, et al. International Society of Urological Pathology Expert Opinion on Grading of Urothelial Carcinoma. Eur Urol Focus. 2022;8(2):438-46.
51. Sylvester RJ, Rodríguez O, Hernández V, Turturica D, Bauerová L, Bruins HM, et al. European Association of Urology (EAU) Prognostic Factor Risk Groups for Non–muscle-invasive Bladder Cancer (NMIBC) Incorporating the WHO 2004/2016 and WHO 1973 Classification Systems for Grade: An Update from the EAU NMIBC Guidelines Panel. European Urology. 2021;79(4):480-8.
52. Jahnson S, Jancke G, Olsson H, Aljabery F. Bladder cancer grading using the four-tier combination of the World Health Organization (WHO) 1973 and WHO 2004 classifications. BJU Int. 2023;132(6):656-63.
53. Simerville JA, Maxted WC, Pahira JJ. Urinalysis: a comprehensive review. Am Fam Physician. 2005;71(6):1153-62.
54. Price SJ, Shephard EA, Stapley SA, Barraclough K, Hamilton WT. Non-visible versus visible haematuria and bladder cancer risk: a study of electronic records in primary care. Br J Gen Pract. 2014;64(626):e584-9.
55. EDWARDS TJ, DICKINSON AJ, NATALE S, GOSLING J, MCGRATH JS. A prospective analysis of the diagnostic yield resulting from the attendance of 4020 patients at a protocol-driven haematuria clinic. BJU International. 2006;97(2):301-5.
56. Trinh TW, Glazer DI, Sadow CA, Sahni VA, Geller NL, Silverman SG. Bladder cancer diagnosis with CT urography: test characteristics and reasons for false-positive and false-negative results. Abdom Radiol (NY). 2018;43(3):663-71.
57. Palou J, Rodríguez-Rubio F, Huguet J, Segarra J, Ribal MJ, Alcaraz A, et al. Multivariate analysis of clinical parameters of synchronous primary superficial bladder cancer and upper urinary tract tumor. J Urol. 2005;174(3):859-61; discussion 61.

58. Hilton S, Jones LP. Recent advances in imaging cancer of the kidney and urinary tract. Surg Oncol Clin N Am. 2014;23(4):863-910.
59. Panebianco V, Narumi Y, Altun E, Bochner BH, Efstathiou JA, Hafeez S, et al. Multiparametric Magnetic Resonance Imaging for Bladder Cancer: Development of VI-RADS (Vesical Imaging-Reporting And Data System). Eur Urol. 2018;74(3):294-306.
60. Woo S, Panebianco V, Narumi Y, Del Giudice F, Muglia VF, Takeuchi M, et al. Diagnostic Performance of Vesical Imaging Reporting and Data System for the Prediction of Muscle-invasive Bladder Cancer: A Systematic Review and Meta-analysis. European Urology Oncology. 2020;3(3):306-15.
61. Zheng C, Lv Y, Zhong Q, Wang R, Jiang Q. Narrow band imaging diagnosis of bladder cancer: systematic review and meta-analysis. BJU Int. 2012;110(11 Pt B):E680-7.
62. Bryan RT, Billingham LJ, Wallace DM. Narrow-band imaging flexible cystoscopy in the detection of recurrent urothelial cancer of the bladder. BJU Int. 2008;101(6):702-5; discussion 5-6.
63. Tschirdewahn S, Harke NN, Hirner L, Stagge E, Hadaschik B, Eisenhardt A. Narrow-band imaging assisted cystoscopy in the follow-up of patients with transitional cell carcinoma of the bladder: a randomized study in comparison with white light cystoscopy. World J Urol. 2020;38(6):1509-15.
64. Hsueh TY, Chiu AW. Narrow band imaging for bladder cancer. Asian J Urol. 2016;3(3):126-9.
65. Lotan Y, Bivalacqua TJ, Downs T, Huang W, Jones J, Kamat AM, et al. Blue light flexible cystoscopy with hexaminolevulinate in non-muscle-invasive bladder cancer: review of the clinical evidence and consensus statement on optimal use in the USA — update 2018. Nature Reviews Urology. 2019;16(6):377-86.
66. Oude Elferink P, Witjes JA. Blue-light cystoscopy in the evaluation of non-muscle-invasive bladder cancer. Ther Adv Urol. 2014;6(1):25-33.
67. Russo GI, Sholklapper TN, Cocci A, Broggi G, Caltabiano R, Smith AB, et al. Performance of Narrow Band Imaging (NBI) and Photodynamic Diagnosis (PDD) Fluorescence Imaging Compared to White Light Cystoscopy (WLC) in Detecting Non-Muscle Invasive Bladder Cancer: A Systematic Review and Lesion-Level Diagnostic Meta-Analysis. Cancers (Basel). 2021;13(17).
68. Burke DM, Shackley DC, O'Reilly PH. The community-based morbidity of flexible cystoscopy. BJU Int. 2002;89(4):347-9.
69. Almallah YZ, Rennie CD, Stone J, Lancashire MJ. Urinary tract infection and patient satisfaction after flexible cystoscopy and urodynamic evaluation. Urology. 2000;56(1):37-9.
70. Cusumano JA, Hermenau M, Gaitanis M, Travis M, LaPlante KL, Tran TY, et al. Evaluation of post-flexible cystoscopy urinary tract infection rates. Am J Health Syst Pharm. 2020;77(22):1852-8.
71. Yafi FA, Brimo F, Steinberg J, Aprikian AG, Tanguay S, Kassouf W. Prospective analysis of sensitivity and specificity of urinary cytology and other urinary biomarkers for bladder cancer. Urol Oncol. 2015;33(2):66.e25-31.
72. Têtu B, Tiguert R, Harel F, Fradet Y. ImmunoCyt/uCyt+ improves the sensitivity of urine cytology in patients followed for urothelial carcinoma. Mod Pathol. 2005;18(1):83-9.
73. Reid MD, Osunkoya AO, Siddiqui MT, Looney SW. Accuracy of grading of urothelial carcinoma on urine cytology: an analysis of interobserver and intraobserver agreement. Int J Clin Exp Pathol. 2012;5(9):882-91.
74. Babjuk M, Burger M, Compérat EM, Gontero P, Mostafid AH, Palou J, et al. European Association of Urology Guidelines on Non-muscle-invasive Bladder Cancer (TaT1 and Carcinoma In Situ) - 2019 Update. Eur Urol. 2019;76(5):639-57.
75. Witjes JA. Follow-up in non-muscle invasive bladder cancer: facts and future. World J Urol. 2020.
76. Devlies W, de Jong JJ, Hofmann F, Bruins HM, Zuiverloon TCM, Smith EJ, et al. The Diagnostic Accuracy of Cystoscopy for Detecting Bladder Cancer in Adults Presenting with Haematuria: A

Systematic Review from the European Association of Urology Guidelines Office. European Urology Focus.
77. EAU Guidelines. Amsterdam: EAU Guidelines Office, Arnhem, The Netherlands. http://uroweb.org/guidelines/compilations-of-all-guidelines/; 2022.
78. Scotland N. Scottish Referral Guidelines for Suspected Cancer 2/9/20222019-2023 [Available from: https://www.cancerreferral.scot.nhs.uk/urological-cancers/?alttemplate=guideline.
79. Khadhouri S, Gallagher KM, MacKenzie K, Shah TT, Gao C, Moore S, et al. IDENTIFY: The investigation and detection of urological neoplasia in patients referred with suspected urinary tract cancer: A multicentre cohort study. Int J Surg Protoc. 2020;21:8-12.
80. Khadhouri S, Gallagher KM, MacKenzie KR, Shah TT, Gao C, Moore S, et al. The IDENTIFY study: the investigation and detection of urological neoplasia in patients referred with suspected urinary tract cancer – a multicentre observational study. BJU International. 2021;128(4):440-50.
81. Khadhouri S, Gallagher KM, MacKenzie KR, Shah TT, Gao C, Moore S, et al. Developing a Diagnostic Multivariable Prediction Model for Urinary Tract Cancer in Patients Referred with Haematuria: Results from the IDENTIFY Collaborative Study. Eur Urol Focus. 2022;8(6):1673-82.
82. Babjuk M, Soukup V, Pesl M, Kostírová M, Drncová E, Smolová H, et al. Urinary cytology and quantitative BTA and UBC tests in surveillance of patients with pTapT1 bladder urothelial carcinoma. Urology. 2008;71(4):718-22.
83. Kibar Y, Goktas S, Kilic S, Yaman H, Onguru O, Peker AF. Prognostic value of cytology, nuclear matrix protein 22 (NMP22) test, and urinary bladder cancer II (UBC II) test in early recurrent transitional cell carcinoma of the bladder. Ann Clin Lab Sci. 2006;36(1):31-8.
84. Richter P, Junker K, Franz M, Berndt A, Geyer C, Gajda M, et al. IIICS de novo glycosylated fibronectin as a marker for invasiveness in urothelial carcinoma of the urinary bladder (UBC). J Cancer Res Clin Oncol. 2008;134(10):1059-65.
85. Landman J, Chang Y, Kavaler E, Droller MJ, Liu BC. Sensitivity and specificity of NMP-22, telomerase, and BTA in the detection of human bladder cancer. Urology. 1998;52(3):398-402.
86. Kinders R, Jones T, Root R, Bruce C, Murchison H, Corey M, et al. Complement factor H or a related protein is a marker for transitional cell cancer of the bladder. Clin Cancer Res. 1998;4(10):2511-20.
87. Babjuk M, Kostírová M, Mudra K, Pecher S, Smolová H, Pecen L, et al. Qualitative and quantitative detection of urinary human complement factor H-related protein (BTA stat and BTA TRAK) and fragments of cytokeratins 8, 18 (UBC rapid and UBC IRMA) as markers for transitional cell carcinoma of the bladder. Eur Urol. 2002;41(1):34-9.
88. Ecke TH, Arndt C, Stephan C, Hallmann S, Lux O, Otto T, et al. Preliminary Results of a Multicentre Study of the UBC Rapid Test for Detection of Urinary Bladder Cancer. Anticancer Res. 2015;35(5):2651-5.
89. Ecke TH, Weiß S, Stephan C, Hallmann S, Arndt C, Barski D, et al. UBC(®) Rapid Test-A Urinary Point-of-Care (POC) Assay for Diagnosis of Bladder Cancer with a focus on Non-Muscle Invasive High-Grade Tumors: Results of a Multicenter-Study. Int J Mol Sci. 2018;19(12).
90. Ecke TH, Weiß S, Stephan C, Hallmann S, Barski D, Otto T, et al. UBC(®) Rapid Test for detection of carcinoma in situ for bladder cancer. Tumour Biol. 2017;39(5):1010428317701624.
91. Eissa S, Swellam M, Sadek M, Mourad MS, El Ahmady O, Khalifa A. Comparative evaluation of the nuclear matrix protein, fibronectin, urinary bladder cancer antigen and voided urine cytology in the detection of bladder tumors. J Urol. 2002;168(2):465-9.
92. Giannopoulos A, Manousakas T, Gounari A, Constantinides C, Choremi-Papadopoulou H, Dimopoulos C. Comparative evaluation of the diagnostic performance of the BTA stat test, NMP22 and urinary bladder cancer antigen for primary and recurrent bladder tumors. J Urol. 2001;166(2):470-5.

93. Gkialas I, Papadopoulos G, Iordanidou L, Stathouros G, Tzavara C, Gregorakis A, et al. Evaluation of urine tumor-associated trypsin inhibitor, CYFRA 21-1, and urinary bladder cancer antigen for detection of high-grade bladder carcinoma. Urology. 2008;72(5):1159-63.
94. Gleichenhagen J, Arndt C, Casjens S, Meinig C, Gerullis H, Raiko I, et al. Evaluation of a New Survivin ELISA and UBC(®) Rapid for the Detection of Bladder Cancer in Urine. Int J Mol Sci. 2018;19(1).
95. Heicappell R, Schostak M, Müller M, Miller K. Evaluation of urinary bladder cancer antigen as a marker for diagnosis of transitional cell carcinoma of the urinary bladder. Scand J Clin Lab Invest. 2000;60(4):275-82.
96. Maffezzini M, Campodonico F, Capponi G, Canepa G, Casazza S, Bandelloni R, et al. Prognostic significance of fluorescent in situ hybridisation in the follow-up of non-muscle-invasive bladder cancer. Anticancer Res. 2010;30(11):4761-5.
97. Dimashkieh H, Wolff DJ, Smith TM, Houser PM, Nietert PJ, Yang J. Evaluation of urovysion and cytology for bladder cancer detection: a study of 1835 paired urine samples with clinical and histologic correlation. Cancer Cytopathol. 2013;121(10):591-7.
98. Friedrich MG, Hellstern A, Toma MI, Hammerer P, Huland H. Are false-positive urine markers for the detection of bladder carcinoma really wrong or do they predict tumor recurrence? Eur Urol. 2003;43(2):146-50; discussion 50-1.
99. Mian C, Lodde M, Haitel A, Egarter Vigl E, Marberger M, Pycha A. Comparison of two qualitative assays, the UBC rapid test and the BTA stat test, in the diagnosis of urothelial cell carcinoma of the bladder. Urology. 2000;56(2):228-31.
100. Sánchez-Carbayo M, Herrero E, Megías J, Mira A, Soria F. Initial evaluation of the new urinary bladder cancer rapid test in the detection of transitional cell carcinoma of the bladder. Urology. 1999;54(4):656-61.
101. Vasdev N, Hampson A, Agarwal S, Swamy R, Chilvers M, Hampson A, et al. The role of URO17™ biomarker to enhance diagnosis of urothelial cancer in new hematuria patients-First European Data. BJUI Compass. 2021;2(1):46-52.
102. van Kessel KEM, Van Neste L, Lurkin I, Zwarthoff EC, Van Criekinge W. Evaluation of an Epigenetic Profile for the Detection of Bladder Cancer in Patients with Hematuria. The Journal of Urology. 2016;195(3):601-7.
103. Beukers W, Kandimalla R, Masius RG, Vermeij M, Kranse R, van Leenders GJ, et al. Stratification based on methylation of TBX2 and TBX3 into three molecular grades predicts progression in patients with pTa-bladder cancer. Mod Pathol. 2015;28(4):515-22.
104. Kandimalla R, van Tilborg AA, Kompier LC, Stumpel DJ, Stam RW, Bangma CH, et al. Genome-wide analysis of CpG island methylation in bladder cancer identified TBX2, TBX3, GATA2, and ZIC4 as pTa-specific prognostic markers. Eur Urol. 2012;61(6):1245-56.
105. Roperch JP, Grandchamp B, Desgrandchamps F, Mongiat-Artus P, Ravery V, Ouzaid I, et al. Promoter hypermethylation of HS3ST2, SEPTIN9 and SLIT2 combined with FGFR3 mutations as a sensitive/specific urinary assay for diagnosis and surveillance in patients with low or high-risk non-muscle-invasive bladder cancer. BMC Cancer. 2016;16(1):704.
106. Sieverink CA, Batista RPM, Prazeres HJM, Vinagre J, Sampaio C, Leão RR, et al. Clinical Validation of a Urine Test (Uromonitor-V2(®)) for the Surveillance of Non-Muscle-Invasive Bladder Cancer Patients. Diagnostics (Basel). 2020;10(10).
107. Batista R, Vinagre J, Prazeres H, Sampaio C, Peralta P, Conceição P, et al. Validation of a Novel, Sensitive, and Specific Urine-Based Test for Recurrence Surveillance of Patients With Non-Muscle-Invasive Bladder Cancer in a Comprehensive Multicenter Study. Front Genet. 2019;10:1237.
108. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29(7):644-52.

109. Martin JA, Wang Z. Next-generation transcriptome assembly. Nature Reviews Genetics. 2011;12(10):671-82.
110. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10(1):57-63.
111. Nagalakshmi U, Waern K, Snyder M. RNA-Seq: a method for comprehensive transcriptome analysis. Curr Protoc Mol Biol. 2010;Chapter 4:Unit 4.11.1-3.
112. Garneau NL, Wilusz J, Wilusz CJ. The highways and byways of mRNA decay. Nat Rev Mol Cell Biol. 2007;8(2):113-26.
113. Yang E, van Nimwegen E, Zavolan M, Rajewsky N, Schroeder M, Magnasco M, et al. Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes. Genome Res. 2003;13(8):1863-72.
114. Laukhtina E, Shim SR, Mori K, D'Andrea D, Soria F, Rajwa P, et al. Diagnostic Accuracy of Novel Urinary Biomarker Tests in Non-muscle-invasive Bladder Cancer: A Systematic Review and Network Meta-analysis. Eur Urol Oncol. 2021;4(6):927-42.
115. Grunewald CM, Niegisch G, Albers P. Using Circulating Tumor DNA To Guide Adjuvant Therapy in Bladder Cancer: IMvigor010 and IMvigor011. Eur Urol Focus. 2022;8(3):646-7.
116. Birkenkamp-Demtröder K, Christensen E, Nordentoft I, Knudsen M, Taber A, Høyer S, et al. Monitoring Treatment Response and Metastatic Relapse in Advanced Bladder Cancer by Liquid Biopsy Analysis. Eur Urol. 2018;73(4):535-40.
117. Crupi E, de Padua TC, Marandino L, Raggi D, Dyrskjøt L, Spiess PE, et al. Circulating tumor DNA as a Predictive and Prognostic Biomarker in the Perioperative Treatment of Muscle-invasive Bladder Cancer: A Systematic Review. European Urology Oncology. 2023.
118. Powles T, Bellmunt J, Comperat E, De Santis M, Huddart R, Loriot Y, et al. Bladder cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up[☆]. Annals of Oncology. 2022;33(3):244-58.
119. Bryan RT, Collins SI, Daykin MC, Zeegers MP, Cheng KK, Wallace DM, et al. Mechanisms of recurrence of Ta/T1 bladder cancer. Ann R Coll Surg Engl. 2010;92(6):519-24.
120. Sylvester RJ, Oosterlinck W, van der Meijden AP. A single immediate postoperative instillation of chemotherapy decreases the risk of recurrence in patients with stage Ta T1 bladder cancer: a meta-analysis of published results of randomized clinical trials. J Urol. 2004;171(6 Pt 1):2186-90, quiz 435.
121. Friedrich MG, Pichlmeier U, Schwaibold H, Conrad S, Huland H. Long-term intravesical adjuvant chemotherapy further reduces recurrence rate compared with short-term intravesical chemotherapy and short-term therapy with Bacillus Calmette-Guérin (BCG) in patients with non-muscle-invasive bladder carcinoma. Eur Urol. 2007;52(4):1123-29.
122. Shelley MD, Kynaston H, Court J, Wilt TJ, Coles B, Burgon K, et al. A systematic review of intravesical bacillus Calmette-Guérin plus transurethral resection vs transurethral resection alone in Ta and T1 bladder cancer. BJU Int. 2001;88(3):209-16.
123. Shelley MD, Wilt TJ, Court J, Coles B, Kynaston H, Mason MD. Intravesical bacillus Calmette-Guérin is superior to mitomycin C in reducing tumour recurrence in high-risk superficial bladder cancer: a meta-analysis of randomized trials. BJU Int. 2004;93(4):485-90.
124. Böhle A, Jocham D, Bock PR. Intravesical bacillus Calmette-Guerin versus mitomycin C for superficial bladder cancer: a formal meta-analysis of comparative studies on recurrence and toxicity. J Urol. 2003;169(1):90-5.
125. Grimm MO, van der Heijden AG, Colombel M, Muilwijk T, Martínez-Piñeiro L, Babjuk MM, et al. Treatment of High-grade Non-muscle-invasive Bladder Carcinoma by Standard Number and Dose of BCG Instillations Versus Reduced Number and Standard Dose of BCG Instillations: Results of the European Association of Urology Research Foundation Randomised Phase III Clinical Trial "NIMBUS". Eur Urol. 2020;78(5):690-8.

126. Koch GE, Smelser WW, Chang SS. Side Effects of Intravesical BCG and Chemotherapy for Bladder Cancer: What They Are and How to Manage Them. Urology. 2021;149:11-20.
127. Kuperus JM, Busman RD, Kuipers SK, Broekhuizen HT, Noyes SL, Brede CM, et al. Comparison of Side Effects and Tolerability Between Intravesical Bacillus Calmette-Guerin, Reduced-Dose BCG and Gemcitabine for Non-Muscle Invasive Bladder Cancer. Urology. 2021;156:191-8.
128. van der Meijden AP, Sylvester RJ, Oosterlinck W, Hoeltl W, Bono AV. Maintenance Bacillus Calmette-Guerin for Ta T1 bladder tumors is not associated with increased toxicity: results from a European Organisation for Research and Treatment of Cancer Genito-Urinary Group Phase III Trial. Eur Urol. 2003;44(4):429-34.
129. McElree IM, Steinberg RL, Mott SL, O'Donnell MA, Packiam VT. Comparison of Sequential Intravesical Gemcitabine and Docetaxel vs Bacillus Calmette-Guérin for the Treatment of Patients With High-Risk Non-Muscle-Invasive Bladder Cancer. JAMA Netw Open. 2023;6(2):e230849.
130. Steinberg RL, Packiam VT, Thomas LJ, Brooks N, Vitale A, Mott SL, et al. Intravesical sequential gemcitabine and docetaxel versus bacillus calmette-guerin (BCG) plus interferon in patients with recurrent non-muscle invasive bladder cancer following a single induction course of BCG. Urol Oncol. 2022;40(1):9.e1-9.e7.
131. Stein JP, Lieskovsky G, Cote R, Groshen S, Feng AC, Boyd S, et al. Radical Cystectomy in the Treatment of Invasive Bladder Cancer: Long-Term Results in 1,054 Patients. J Clin Oncol. 2023;41(22):3772-81.
132. Johnson SB, Yu JB. Bladder Preserving Trimodality Therapy for Muscle-Invasive Bladder Cancer. Current Oncology Reports. 2018;20(9):66.
133. Mak RH, Hunt D, Shipley WU, Efstathiou JA, Tester WJ, Hagan MP, et al. Long-term outcomes in patients with muscle-invasive bladder cancer after selective bladder-preserving combined-modality therapy: a pooled analysis of Radiation Therapy Oncology Group protocols 8802, 8903, 9506, 9706, 9906, and 0233. J Clin Oncol. 2014;32(34):3801-9.
134. James ND, Hussain SA, Hall E, Jenkins P, Tremlett J, Rawlings C, et al. Radiotherapy with or without Chemotherapy in Muscle-Invasive Bladder Cancer. New England Journal of Medicine. 2012;366(16):1477-88.
135. Hall E, Hussain SA, Porta N, Lewis R, Crundwell M, Jenkins P, et al. Chemoradiotherapy in Muscle-invasive Bladder Cancer: 10-yr Follow-up of the Phase 3 Randomised Controlled BC2001 Trial. Eur Urol. 2022;82(3):273-9.
136. Zlotta AR, Ballas LK, Niemierko A, Lajkosz K, Kuk C, Miranda G, et al. Radical cystectomy versus trimodality therapy for muscle-invasive bladder cancer: a multi-institutional propensity score matched and weighted analysis. Lancet Oncol. 2023;24(6):669-81.
137. Alfred Witjes J, Max Bruins H, Carrión A, Cathomas R, Compérat E, Efstathiou JA, et al. European Association of Urology Guidelines on Muscle-invasive and Metastatic Bladder Cancer: Summary of the 2023 Guidelines. Eur Urol. 2024;85(1):17-31.
138. Excellence NIfHaC. Bladder cancer: diagnosis and management 2015 [Available from: https://www.nice.org.uk/guidance/ng2/chapter/Recommendations#treating-muscle-invasive-bladder-cancer-2.
139. Neoadjuvant chemotherapy in invasive bladder cancer: update of a systematic review and meta-analysis of individual patient data advanced bladder cancer (ABC) meta-analysis collaboration. Eur Urol. 2005;48(2):202-5; discussion 5-6.
140. Griffiths G, Hall R, Sylvester R, Raghavan D, Parmar MK. International phase III trial assessing neoadjuvant cisplatin, methotrexate, and vinblastine chemotherapy for muscle-invasive bladder cancer: long-term results of the BA06 30894 trial. J Clin Oncol. 2011;29(16):2171-7.
141. Powles T, Kockx M, Rodriguez-Vida A, Duran I, Crabb SJ, Van Der Heijden MS, et al. Clinical efficacy and biomarker analysis of neoadjuvant atezolizumab in operable urothelial carcinoma in the ABACUS trial. Nat Med. 2019;25(11):1706-14.

142. Szabados B, Kockx M, Assaf ZJ, van Dam PJ, Rodriguez-Vida A, Duran I, et al. Final Results of Neoadjuvant Atezolizumab in Cisplatin-ineligible Patients with Muscle-invasive Urothelial Cancer of the Bladder. Eur Urol. 2022;82(2):212-22.
143. Necchi A, Anichini A, Raggi D, Briganti A, Massa S, Lucianò R, et al. Pembrolizumab as Neoadjuvant Therapy Before Radical Cystectomy in Patients With Muscle-Invasive Urothelial Bladder Carcinoma (PURE-01): An Open-Label, Single-Arm, Phase II Study. Journal of Clinical Oncology. 2018;36(34):3353-60.
144. Powles T, Park SH, Voog E, Caserta C, Valderrama BP, Gurney H, et al. Avelumab Maintenance Therapy for Advanced or Metastatic Urothelial Carcinoma. N Engl J Med. 2020;383(13):1218-30.
145. Grivas P, Park SH, Voog E, Caserta C, Gurney H, Bellmunt J, et al. Avelumab First-line Maintenance Therapy for Advanced Urothelial Carcinoma: Comprehensive Clinical Subgroup Analyses from the JAVELIN Bladder 100 Phase 3 Trial. Eur Urol. 2023;84(1):95-108.
146. Balar AV, Castellano D, O'Donnell PH, Grivas P, Vuky J, Powles T, et al. First-line pembrolizumab in cisplatin-ineligible patients with locally advanced and unresectable or metastatic urothelial cancer (KEYNOTE-052): a multicentre, single-arm, phase 2 study. Lancet Oncol. 2017;18(11):1483-92.
147. O'Donnell PH, Milowsky MI, Petrylak DP, Hoimes CJ, Flaig TW, Mar N, et al. Enfortumab Vedotin With or Without Pembrolizumab in Cisplatin-Ineligible Patients With Previously Untreated Locally Advanced or Metastatic Urothelial Cancer. J Clin Oncol. 2023:Jco2202887.
148. Kamoun A, de Reyniès A, Allory Y, Sjödahl G, Robertson AG, Seiler R, et al. A Consensus Molecular Classification of Muscle-invasive Bladder Cancer. Eur Urol. 2020;77(4):420-33.
149. Hedegaard J, Lamy P, Nordentoft I, Algaba F, Høyer S, Ulhøi BP, et al. Comprehensive Transcriptional Analysis of Early-Stage Urothelial Carcinoma. Cancer Cell. 2016;30(1):27-42.
150. Cambier S, Sylvester RJ, Collette L, Gontero P, Brausi MA, van Andel G, et al. EORTC Nomograms and Risk Groups for Predicting Recurrence, Progression, and Disease-specific and Overall Survival in Non–Muscle-invasive Stage Ta–T1 Urothelial Bladder Cancer Patients Treated with 1–3 Years of Maintenance Bacillus Calmette-Guérin. European Urology. 2016;69(1):60-9.
151. Sylvester RJ, van der Meijden APM, Oosterlinck W, Witjes JA, Bouffioux C, Denis L, et al. Predicting Recurrence and Progression in Individual Patients with Stage Ta T1 Bladder Cancer Using EORTC Risk Tables: A Combined Analysis of 2596 Patients from Seven EORTC Trials. European Urology. 2006;49(3):466-77.
152. Droller MJ. Bladder cancer. Current Problems in Surgery. 1981;18(4):205-79.
153. Lee R, Droller MJ. The natural history of bladder cancer. Implications for therapy. Urol Clin North Am. 2000;27(1):1-13, vii.
154. Cairns P, Shaw ME, Knowles MA. Initiation of bladder cancer may involve deletion of a tumour-suppressor gene on chromosome 9. Oncogene. 1993;8(4):1083-5.
155. Baud E, Catilina P, Boiteux JP, Bignon YJ. Human bladder cancers and normal bladder mucosa present the same hot spot of heterozygous chromosome-9 deletion. Int J Cancer. 1998;77(6):821-4.
156. Robertson AG, Kim J, Al-Ahmadie H, Bellmunt J, Guo G, Cherniack AD, et al. Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer. Cell. 2017;171(3):540-56.e25.
157. Hurst CD, Cheng G, Platt FM, Castro MAA, Marzouka NS, Eriksson P, et al. Stage-stratified molecular profiling of non-muscle-invasive bladder cancer enhances biological, clinical, and therapeutic insight. Cell Rep Med. 2021;2(12):100472.
158. Jebar AH, Hurst CD, Tomlinson DC, Johnston C, Taylor CF, Knowles MA. FGFR3 and Ras gene mutations are mutually exclusive genetic events in urothelial cell carcinoma. Oncogene. 2005;24(33):5218-25.

159. Hernández S, López-Knowles E, Lloreta J, Kogevinas M, Amorós A, Tardón A, et al. Prospective Study of FGFR3 Mutations As a Prognostic Factor in Nonmuscle Invasive Urothelial Bladder Carcinomas. Journal of Clinical Oncology. 2006;24(22):3664-71.
160. Kompier LC, Lurkin I, van der Aa MN, van Rhijn BW, van der Kwast TH, Zwarthoff EC. FGFR3, HRAS, KRAS, NRAS and PIK3CA mutations in bladder cancer and their potential as biomarkers for surveillance and therapy. PLoS One. 2010;5(11):e13821.
161. Comprehensive molecular characterization of urothelial bladder carcinoma. Nature. 2014;507(7492):315-22.
162. Wikman FP, Lu ML, Thykjaer T, Olesen SH, Andersen LD, Cordon-Cardo C, et al. Evaluation of the performance of a p53 sequencing microarray chip using 140 previously sequenced bladder tumor samples. Clin Chem. 2000;46(10):1555-61.
163. Ecke TH, Lenk SV, Schlechte HH, Loening SA. Tissue polypeptide antigen (TPA) in comparison with mutations of tumour suppressor gene P53 (TP53) in patients with bladder cancer. Anticancer Res. 2003;23(2a):957-62.
164. Ecke TH, Schlechte HH, Schulze G, Lenk SV, Loening SA. Four tumour markers for urinary bladder cancer--tissue polypeptide antigen (TPA), HER-2/neu (ERB B2), urokinase-type plasminogen activator receptor (uPAR) and TP53 mutation. Anticancer Res. 2005;25(1b):635-41.
165. Sherr CJ, McCormick F. The RB and p53 pathways in cancer. Cancer cell. 2002;2(2):103-12.
166. Lindgren D, Frigyesi A, Gudjonsson S, Sjödahl G, Hallden C, Chebil G, et al. Combined Gene Expression and Genomic Profiling Define Two Intrinsic Molecular Subtypes of Urothelial Carcinoma and Gene Signatures for Molecular Grading and Outcome. Cancer Research. 2010;70(9):3463-72.
167. Damrauer JS, Hoadley KA, Chism DD, Fan C, Tiganelli CJ, Wobker SE, et al. Intrinsic subtypes of high-grade bladder cancer reflect the hallmarks of breast cancer biology. Proceedings of the National Academy of Sciences. 2014;111(8):3110-5.
168. Sjödahl G, Lauss M, Lövgren K, Chebil G, Gudjonsson S, Veerla S, et al. A Molecular Taxonomy for Urothelial Carcinoma. Clinical Cancer Research. 2012;18(12):3377-86.
169. Marzouka N-A-D, Eriksson P, Bernardo C, Hurst CD, Knowles MA, Sjödahl G, et al. The Lund Molecular Taxonomy Applied to Non–Muscle-Invasive Urothelial Carcinoma. The Journal of Molecular Diagnostics. 2022;24(9):992-1008.
170. Eriksson P, Aine M, Veerla S, Liedberg F, Sjödahl G, Höglund M. Molecular subtypes of urothelial carcinoma are defined by specific gene regulatory systems. BMC Med Genomics. 2015;8:25.
171. Seiler R, Ashab HAD, Erho N, van Rhijn BWG, Winters B, Douglas J, et al. Impact of Molecular Subtypes in Muscle-invasive Bladder Cancer on Predicting Response and Survival after Neoadjuvant Chemotherapy. European Urology. 2017;72(4):544-54.
172. Rosenberg JE, Hoffman-Censits J, Powles T, van der Heijden MS, Balar AV, Necchi A, et al. Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. Lancet. 2016;387(10031):1909-20.
173. Sharma P, Callahan MK, Bono P, Kim J, Spiliopoulou P, Calvo E, et al. Nivolumab monotherapy in recurrent metastatic urothelial carcinoma (CheckMate 032): a multicentre, open-label, two-stage, multi-arm, phase 1/2 trial. The Lancet Oncology. 2016;17(11):1590-8.
174. Bryan RT, Tselepis C. Cadherin switching and bladder cancer. J Urol. 2010;184(2):423-31.
175. Bryan RT, Atherfold PA, Yeo Y, Jones LJ, Harrison RF, Wallace DM, et al. Cadherin switching dictates the biology of transitional cell carcinoma of the bladder: ex vivo and in vitro studies. J Pathol. 2008;215(2):184-94.
176. Lo H-W, Hsu S-C, Xia W, Cao X, Shih J-Y, Wei Y, et al. Epidermal growth factor receptor cooperates with signal transducer and activator of transcription 3 to induce epithelial-mesenchymal transition in cancer cells via up-regulation of TWIST gene expression. Cancer research. 2007;67(19):9066-76.

177. Lindskrog SV, Prip F, Lamy P, Taber A, Groeneveld CS, Birkenkamp-Demtröder K, et al. An integrated multi-omics analysis identifies prognostic molecular subtypes of non-muscle-invasive bladder cancer. Nature Communications. 2021;12(1):2301.
178. McKee AE, Silver PA. Systems perspectives on mRNA processing. Cell Research. 2007;17(7):581-90.
179. Nilsen TW, Graveley BR. Expansion of the eukaryotic proteome by alternative splicing. Nature. 2010;463(7280):457-63.
180. Liu Y, Gonzàlez-Porta M, Santos S, Brazma A, Marioni JC, Aebersold R, et al. Impact of Alternative Splicing on the Human Proteome. Cell Rep. 2017;20(5):1229-41.
181. Wahl MC, Will CL, Lührmann R. The spliceosome: design principles of a dynamic RNP machine. Cell. 2009;136(4):701-18.
182. Pajares MJ, Ezponda T, Catena R, Calvo A, Pio R, Montuenga LM. Alternative splicing: an emerging topic in molecular and clinical oncology. The Lancet Oncology. 2007;8(4):349-57.
183. Dueñas M, Pérez-Figueroa A, Oliveira C, Suárez-Cabrera C, Sousa A, Oliveira P, et al. Gene Expression Analyses in Non Muscle Invasive Bladder Cancer Reveals a Role for Alternative Splicing and Tp53 Status. Scientific Reports. 2019;9(1):10362.
184. Wang BD, Lee NH. Aberrant RNA Splicing in Cancer and Drug Resistance. Cancers (Basel). 2018;10(11).
185. Climente-González H, Porta-Pardo E, Godzik A, Eyras E. The Functional Impact of Alternative Splicing in Cancer. Cell Reports. 2017;20(9):2215-26.
186. Varey AHR, Rennel ES, Qiu Y, Bevan HS, Perrin RM, Raffy S, et al. VEGF 165 b, an antiangiogenic VEGF-A isoform, binds and inhibits bevacizumab treatment in experimental colorectal carcinoma: balance of pro- and antiangiogenic VEGF-A isoforms has implications for therapy. British journal of cancer. 2008;98(8):1366-79.
187. Pawlak G, McGarvey TW, Nguyen TB, Tomaszewski JE, Puthiyaveettil R, Malkowicz SB, et al. Alterations in tropomyosin isoform expression in human transitional cell carcinoma of the urinary bladder. Int J Cancer. 2004;110(3):368-73.
188. Bailey K. Tropomyosin: a new asymmetric protein component of the muscle fibril. Biochem J. 1948;43(2):271-9.
189. Lazarides E. Tropomyosin antibody: the specific localization of tropomyosin in nonmuscle cells. Journal of Cell Biology. 1975;65(3):549-61.
190. Gunning P, O'neill G, Hardeman E. Tropomyosin-Based Regulation of the Actin Cytoskeleton in Time and Space. Physiological Reviews. 2008;88(1):1-35.
191. Meiring JCM, Bryce NS, Wang Y, Taft MH, Manstein DJ, Liu Lau S, et al. Co-polymers of Actin and Tropomyosin Account for a Major Fraction of the Human Actin Cytoskeleton. Current Biology. 2018;28(14):2331-7.e5.
192. Lees JG, Bach CTT, O'Neill GM. Interior decoration. Cell Adhesion & Migration. 2011;5(2):181-6.
193. Li X, Tobacman LS, Mun JY, Craig R, Fischer S, Lehman W. Tropomyosin Position on F-Actin Revealed by EM Reconstruction and Computational Chemistry. Biophysical Journal. 2011;100(4):1005-13.
194. Schevzov G, Whittaker SP, Fath T, Lin JJC, Gunning PW. Tropomyosin isoforms and reagents. BioArchitecture. 2011;1(4):135-64.
195. Gunning PW, Hardeman EC, Lappalainen P, Mulvihill DP. Tropomyosin – master regulator of actin filament function in the cytoskeleton. Journal of Cell Science. 2015;128(16):2965-74.
196. Tojkander S, Gateva G, Lappalainen P. Actin stress fibers--assembly, dynamics and biological roles. J Cell Sci. 2012;125(Pt 8):1855-64.
197. NAUMANEN P, LAPPALAINEN P, HOTULAINEN P. Mechanisms of actin stress fibre assembly. Journal of Microscopy. 2008;231(3):446-54.

198. Brayford S, Bryce NS, Schevzov G, Haynes EM, Bear JE, Hardeman EC, et al. Tropomyosin Promotes Lamellipodial Persistence by Collaborating with Arp2/3 at the Leading Edge. Current Biology. 2016;26(10):1312-8.
199. Colpan M, Moroz NA, Kostyukova AS. Tropomodulins and tropomyosins: working as a team. J Muscle Res Cell Motil. 2013;34(3-4):247-60.
200. Robaszkiewicz K, Ostrowska Z, Marchlewicz K, Moraczewska J. Tropomyosin isoforms differentially modulate the regulation of actin filament polymerization and depolymerization by cofilins. The FEBS Journal. 2016;283(4):723-37.
201. Ishikawa R, Yamashiro S, Matsumura F. Differential modulation of actin-severing activity of gelsolin by multiple isoforms of cultured rat cell tropomyosin. Potentiation of protective ability of tropomyosins by 83-kDa nonmuscle caldesmon. J Biol Chem. 1989;264(13):7490-7.
202. Burtnick LD, Urosev D, Irobi E, Narayan K, Robinson RC. Structure of the N-terminal half of gelsolin bound to actin: roles in severing, apoptosis and FAF. Embo j. 2004;23(14):2713-22.
203. Creed SJ, Desouza M, Bamburg JR, Gunning P, Stehn J. Tropomyosin isoform 3 promotes the formation of filopodia by regulating the recruitment of actin-binding proteins to actin filaments. Exp Cell Res. 2011;317(3):249-61.
204. Lamb MC, Tootle TL. Fascin in Cell Migration: More Than an Actin Bundling Protein. Biology (Basel). 2020;9(11).
205. Schevzov G, Vrhovski B, Bryce NS, Elmir S, Qiu MR, O'Neill GM, et al. Tissue-specific Tropomyosin Isoform Composition. Journal of Histochemistry & Cytochemistry. 2005;53(5):557-70.
206. Geeves MA, Hitchcock-DeGregori SE, Gunning PW. A systematic nomenclature for mammalian tropomyosin isoforms. J Muscle Res Cell Motil. 2015;36(2):147-53.
207. Janco M, Bonello TT, Byun A, Coster ACF, Lebhar H, Dedova I, et al. The impact of tropomyosins on actin filament assembly is isoform specific. Bioarchitecture. 2016;6(4):61-75.
208. Manstein D, Meiring J, Hardeman E, Gunning P. Actin–tropomyosin distribution in non-muscle cells. Journal of Muscle Research and Cell Motility. 2020;41:1-12.
209. Zhang W, Liu HT. MAPK signal pathways in the regulation of cell proliferation in mammalian cells. Cell Research. 2002;12(1):9-18.
210. Sinkala M, Nkhoma P, Mulder N, Martin DP. Integrated molecular characterisation of the MAPK pathways in human cancers reveals pharmacologically vulnerable mutations and gene dependencies. Communications Biology. 2021;4(1):9.
211. Grieco L, Calzone L, Bernard-Pierrot I, Radvanyi F, Kahn-Perlès B, Thieffry D. Integrative Modelling of the Influence of MAPK Network on Cancer Cell Fate Decision. PLOS Computational Biology. 2013;9(10):e1003286.
212. Nakopoulou L, Zervas A, Constantinides C, Deliveliotis C, Stefanaki K, Dimopoulos C. Epithelial differentiation antigens and epidermal growth factor receptors in transitional cell bladder carcinoma: correlation with prognosis. Urol Int. 1995;54(4):191-7.
213. Snell KIE, Ward DG, Gordon NS, Goldsmith JC, Sutton AJ, Patel P, et al. Exploring the roles of urinary HAI-1, EpCAM & EGFR in bladder cancer prognosis & risk stratification. Oncotarget. 2018;9(38):25244-53.
214. Sanguedolce F, Zanelli M, Palicelli A, Bisagni A, Zizzo M, Ascani S, et al. HER2 Expression in Bladder Cancer: A Focused View on Its Diagnostic, Prognostic, and Predictive Role. Int J Mol Sci. 2023;24(4).
215. Santarpia L, Lippman SM, El-Naggar AK. Targeting the MAPK–RAS–RAF signaling pathway in cancer therapy. Expert Opinion on Therapeutic Targets. 2012;16(1):103-19.
216. Schevzov G, Kee AJ, Wang B, Sequeira VB, Hook J, Coombes JD, et al. Regulation of cell proliferation by ERK and signal-dependent nuclear translocation of ERK is dependent on Tm5NM1-containing actin filaments. Molecular Biology of the Cell. 2015;26(13):2475-90.
217. Syed V. TGF-β Signaling in Cancer. J Cell Biochem. 2016;117(6):1279-87.

218. Fabregat I, Fernando J, Mainez J, Sancho P. TGF-beta signaling in cancer treatment. Curr Pharm Des. 2014;20(17):2934-47.
219. Bakin AV, Safina A, Rinehart C, Daroqui C, Darbary H, Helfman DM. A critical role of tropomyosins in TGF-beta regulation of the actin cytoskeleton and cell motility in epithelial cells. Mol Biol Cell. 2004;15(10):4682-94.
220. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100(1):57-70.
221. Hanahan D. Hallmarks of Cancer: New Dimensions. Cancer Discovery. 2022;12(1):31-46.
222. Gunning PW, Hardeman EC. Tropomyosins. Current Biology. 2017;27(1):R8-R13.
223. Desouza-Armstrong M, Gunning PW, Stehn JR. Tumor suppressor tropomyosin Tpm2.1 regulates sensitivity to apoptosis beyond anoikis characterized by changes in the levels of intrinsic apoptosis proteins. Cytoskeleton. 2017;74(6):233-48.
224. Liu G, Zhao X, Zhou J, Cheng X, Ye Z, Ji Z. Long non-coding RNA MEG3 suppresses the development of bladder urothelial carcinoma by regulating miR-96 and TPM1. Cancer Biol Ther. 2018;19(11):1039-56.
225. He C, Zhang Q, Gu R, Lou Y, Liu W. miR-96 regulates migration and invasion of bladder cancer through epithelial-mesenchymal transition in response to transforming growth factor-β1. Journal of Cellular Biochemistry. 2018;119(9):7807-17.
226. Bryan RT, Hussain SA, James ND, Jankowski JA, Wallace DM. Molecular pathways in bladder cancer: part 1. BJU Int. 2005;95(4):485-90.
227. Bryan RT, Hussain SA, James ND, Jankowski JA, Wallace DMA. Molecular pathways in bladder cancer: Part 2. BJU International. 2005;95(4):491-6.
228. Chen S, Shen Z, Gao L, Yu S, Zhang P, Han Z, et al. TPM3 mediates epithelial-mesenchymal transition in esophageal cancer via MMP2/MMP9. Ann Transl Med. 2021;9(16):1338.
229. Xu X, Wang Y, Bryce NS, Tang K, Meagher NS, Kang EY, et al. Targeting the actin/tropomyosin cytoskeleton in epithelial ovarian cancer reveals multiple mechanisms of synergy with anti-microtubule agents. Br J Cancer. 2021;125(2):265-76.
230. Kee AJ, Yang L, Lucas CA, Greenberg MJ, Martel N, Leong GM, et al. An actin filament population defined by the tropomyosin Tpm3.1 regulates glucose uptake. Traffic. 2015;16(7):691-711.
231. Bryce NS, Schevzov G, Ferguson V, Percival JM, Lin JJ, Matsumura F, et al. Specification of actin filament function and molecular composition by tropomyosin isoforms. Mol Biol Cell. 2003;14(3):1002-16.
232. Janco M, Bonello TT, Byun A, Coster AC, Lebhar H, Dedova I, et al. The impact of tropomyosins on actin filament assembly is isoform specific. Bioarchitecture. 2016;6(4):61-75.
233. Savill SA, Leitch HF, Harvey JN, Thomas TH. Functional structure of the promoter regions for the predominant low molecular weight isoforms of tropomyosin in human kidney cells. J Cell Biochem. 2012;113(11):3576-86.
234. Lerner SP, McConkey DJ, Hoadley KA, Chan KS, Kim WY, Radvanyi F, et al. Bladder Cancer Molecular Taxonomy: Summary from a Consensus Meeting. Bladder Cancer. 2016;2(1):37-47.
235. Pradhan AK, Emdad L, Das SK, Sarkar D, Fisher PB. Chapter Two - The Enigma of miRNA Regulation in Cancer. In: Croce CM, Fisher PB, editors. Advances in Cancer Research. 135: Academic Press; 2017. p. 25-52.
236. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21.
237. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nature Methods. 2015;12(4):357-60.
238. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology. 2013;14(4):R36.

239. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139-40.
240. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014;15(12):550.
241. Law CW, Alhamdoosh M, Su S, Dong X, Tian L, Smyth GK, et al. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. F1000Res. 2016;5.
242. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33(3):290-5.
243. Ghandi M, Huang FW, Jané-Valbuena J, Kryukov GV, Lo CC, McDonald ER, et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. Nature. 2019;569(7757):503-8.
244. Jiménez-Sánchez A, Cast O, Miller ML. Comprehensive Benchmarking and Integration of Tumor Microenvironment Cell Estimation Methods. Cancer Research. 2019;79(24):6238-46.
245. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nature Methods. 2017;14(4):417-9.
246. Frankish A, Uszczynska B, Ritchie GRS, Gonzalez JM, Pervouchine D, Petryszak R, et al. Comparison of GENCODE and RefSeq gene annotation and the impact of reference geneset on variant effect prediction. BMC Genomics. 2015;16(8):S2.
247. Srivastava A, Malik L, Sarkar H, Zakeri M, Almodaresi F, Soneson C, et al. Alignment and mapping methodology influence transcript abundance estimation. Genome Biology. 2020;21(1):239.
248. Fogh J, Fogh JM, Orfeo T. One Hundred and Twenty-Seven Cultured Human Tumor Cell Lines Producing Tumors in Nude Mice23. JNCI: Journal of the National Cancer Institute. 1977;59(1):221-6.
249. Kyriazis AA, Kyriazis AP, McCombs WB, 3rd, Peterson WD, Jr. Morphological, biological, and biochemical characteristics of human bladder transitional cell carcinomas grown in tissue culture and in nude mice. Cancer Res. 1984;44(9):3997-4005.
250. Bubeník J, Barešová M, Viklický V, Jakoubková J, Sainerová H, Donner J. Established cell line of urinary bladder carcinoma (T24) containing tumour-specific antigen. International Journal of Cancer. 1973;11(3):765-73.
251. Nayak SK, O'Toole C, Price ZH. A cell line from an anaplastic transitional cell carcinoma of human urinary bladder. British Journal of Cancer. 1977;35(2):142-51.
252. Grossman HB, Wedemeyer G, Ren L, Wilson GN, Cox B. Improved Growth of Human Urothelial Carcinoma Cell Cultures. The Journal of Urology. 1986;136(4):953-9.
253. Zeegers MP, Bryan RT, Langford C, Billingham L, Murray P, Deshmukh NS, et al. The West Midlands Bladder Cancer Prognosis Programme: rationale and design. BJU Int. 2010;105(6):784-8.
254. NCBI. Primer Blast [Available from: https://www.ncbi.nlm.nih.gov/tools/primer-blast/.
255. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics. 2012;13:134.
256. Technologies A. Agilent High Sensitivity DNA Kit Guide- Bioanalyzer.
257. Zhang C, Zhang B, Lin L-L, Zhao S. Evaluation and comparison of computational tools for RNA-seq isoform quantification. BMC Genomics. 2017;18(1):583.
258. Williams CR, Baccarella A, Parrish JZ, Kim CC. Empirical assessment of analysis workflows for differential expression analysis of human samples using RNA-Seq. BMC Bioinformatics. 2017;18(1):38.
259. Teng M, Love MI, Davis CA, Djebali S, Dobin A, Graveley BR, et al. A benchmark for RNA-seq quantification pipelines. Genome Biology. 2016;17(1):74.
260. Seyednasrollah F, Laiho A, Elo LL. Comparison of software packages for detecting differential expression in RNA-seq studies. Briefings in Bioinformatics. 2013;16(1):59-70.
261. Engström PG, Steijger T, Sipos B, Grant GR, Kahles A, Alioto T, et al. Systematic evaluation of spliced alignment programs for RNA-seq data. Nature Methods. 2013;10(12):1185-91.

262. Thorsen K, Sørensen KD, Brems-Eskildsen AS, Modin C, Gaustadnes M, Hein AM, et al. Alternative splicing in colon, bladder, and prostate cancer identified by exon array analysis. Mol Cell Proteomics. 2008;7(7):1214-24.
263. Zare M, Jazii FR, Soheili ZS, Moghanibashi MM. Downregulation of tropomyosin-1 in squamous cell carcinoma of esophagus, the role of Ras signaling and methylation. Mol Carcinog. 2012;51(10):796-806.
264. Tang C, Wang J, Wei Q, Du YP, Qiu HP, Yang C, et al. Tropomyosin-1 promotes cancer cell apoptosis via the p53-mediated mitochondrial pathway in renal cell carcinoma. Oncol Lett. 2018;15(5):7060-8.
265. Midha MK, Wu M, Chiu K-P. Long-read sequencing in deciphering human genetics to a greater depth. Human Genetics. 2019;138(11):1201-15.
266. Wyman D, Mortazavi A. TranscriptClean: variant-aware correction of indels, mismatches and splice junctions in long-read transcripts. Bioinformatics. 2019;35(2):340-2.
267. Xiao CL, Chen Y, Xie SQ, Chen KN, Wang Y, Han Y, et al. MECAT: fast mapping, error correction, and de novo assembly for single-molecule sequencing reads. Nat Methods. 2017;14(11):1072-4.
268. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094-100.
269. Biosciences P. Iso-Seq 2018 [Available from: https://www.pacb.com/wp-content/uploads/2018-10-NA-UGM-Iso-Seq-Method.pdf.
270. Biosciences P. IsoSeq [Available from: https://github.com/PacificBiosciences/IsoSeq/blob/master/isoseq-clustering.md.
271. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research. 2015;43(7):e47-e.
272. Biosciences P. Preparing SMRTbell ® Libraries using PacBio Barcoded M13 Primers for Multiplex SMRT® Sequencing [Available from: https://www.pacb.com/wp-content/uploads/Procedure-checklist-Preparing-SMRTbell-libraries-using-PacBio-barcoded-M13-primers-for-multiplex-SMRT-sequencing.pdf.
273. Manstein DJ, Meiring JCM, Hardeman EC, Gunning PW. Actin–tropomyosin distribution in non-muscle cells. Journal of Muscle Research and Cell Motility. 2020;41(1):11-22.
274. NCBI. BLASTn [Available from: https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&BLAST_SPEC=GeoBlast&PAGE_TYPE=BlastSearch.
275. Blomquist TM, Crawford EL, Lovett JL, Yeo J, Stanoszek LM, Levin A, et al. Targeted RNA-sequencing with competitive multiplex-PCR amplicon libraries. PLoS One. 2013;8(11):e79120.
276. O'Sullivan P, Sharples K, Dalphin M, Davidson P, Gilling P, Cambridge L, et al. A multigene urine test for the detection and stratification of bladder cancer in patients presenting with hematuria. J Urol. 2012;188(3):741-7.
277. Pichler R, Fritz J, Tulchiner G, Klinglmair G, Soleiman A, Horninger W, et al. Increased accuracy of a novel mRNA-based urine test for bladder cancer surveillance. BJU Int. 2018;121(1):29-37.
278. Marx V. Method of the year: long-read sequencing. Nature Methods. 2023;20(1):6-11.
279. Walker AK, Na J, Browning L, Humayun-Zakaria N, Zeegers MP, Cheng KK, et al. MRE11A Isoform Expression Associated with Outcome Following Radiotherapy in Muscle-Invasive Bladder Cancer does not Alter Cell Survival and DNA Double-Strand Break Repair Following Ionising Radiation. Bladder cancer (Amsterdam, Netherlands). 2019;5(2):147-57.
280. Jeeta RR, Gordon NS, Baxter L, Goel A, Noyvert B, Ott S, et al. Non-Coding Mutations in Urothelial Bladder Cancer: Biological and Clinical Relevance and Potential Utility as Biomarkers. Bladder Cancer. 2019;5:263-72.

281. Bharadwaj S, Hitchcock-DeGregori S, Thorburn A, Prasad GL. N terminus is essential for tropomyosin functions: N-terminal modification disrupts stress fiber organization and abolishes anti-oncogenic effects of tropomyosin-1. J Biol Chem. 2004;279(14):14039-48.
282. Maciver SK. Myosin II function in non-muscle cells. Bioessays. 1996;18(3):179-82.
283. Jeong S, Lim S, Schevzov G, Gunning PW, Helfman DM. Loss of Tpm4.1 leads to disruption of cell-cell adhesions and invasive behavior in breast epithelial cells via increased Rac1 signaling. Oncotarget. 2017;8(20):33544-59.
284. Hundt N, Steffen W, Pathan-Chhatbar S, Taft MH, Manstein DJ. Load-dependent modulation of non-muscle myosin-2A function by tropomyosin 4.2. Sci Rep. 2016;6:20554.
285. Hatogai K, Sweis RF. The Tumor Microenvironment of Bladder Cancer. Adv Exp Med Biol. 2020;1296:275-90.
286. Kang HW, Kim W-J, Yun SJ. The role of the tumor microenvironment in bladder cancer development and progression. Translational Cancer Research. 2017:S744-S58.
287. Lee YC, Lam HM, Rosser C, Theodorescu D, Parks WC, Chan KS. The dynamic roles of the bladder tumour microenvironment. Nat Rev Urol. 2022;19(9):515-33.
288. Boström MM, Irjala H, Mirtti T, Taimen P, Kauko T, Ålgars A, et al. Tumor-Associated Macrophages Provide Significant Prognostic Information in Urothelial Bladder Cancer. PLoS One. 2015;10(7):e0133552.
289. Goel A, Ward DG, Noyvert B, Yu M, Gordon NS, Abbotts B, et al. Combined exome and transcriptome sequencing of non-muscle-invasive bladder cancer: associations between genomic changes, expression subtypes, and clinical outcomes. Genome Medicine. 2022;14(1):59.
290. Raval GN, Bharadwaj S, Levine EA, Willingham MC, Geary RL, Kute T, et al. Loss of expression of tropomyosin-1, a novel class II tumor suppressor that induces anoikis, in primary breast tumors. Oncogene. 2003;22(40):6194-203.
291. Zhou Y, Bian S, Zhou X, Cui Y, Wang W, Wen L, et al. Single-Cell Multiomics Sequencing Reveals Prevalent Genomic Alterations in Tumor Stromal Cells of Human Colorectal Cancer. Cancer Cell. 2020;38(6):818-28.e5.
292. Burley A, Rullan A, Wilkins A. A review of the biology and therapeutic implications of cancer-associated fibroblasts (CAFs) in muscle-invasive bladder cancer. Front Oncol. 2022;12:1000888.
293. Caramelo B, Zagorac S, Corral S, Marqués M, Real FX. Cancer-associated Fibroblasts in Bladder Cancer: Origin, Biology, and Therapeutic Opportunities. Eur Urol Oncol. 2023;6(4):366-75.
294. Coffelt SB, Wellenstein MD, de Visser KE. Neutrophils in cancer: neutral no more. Nat Rev Cancer. 2016;16(7):431-46.
295. Meiring JCM, Bryce NS, Wang Y, Taft MH, Manstein DJ, Liu Lau S, et al. Co-polymers of Actin and Tropomyosin Account for a Major Fraction of the Human Actin Cytoskeleton. Curr Biol. 2018;28(14):2331-7.e5.
296. Cheung KJ, Gabrielson E, Werb Z, Ewald AJ. Collective invasion in breast cancer requires a conserved basal epithelial program. Cell. 2013;155(7):1639-51.
297. Carroll DK, Carroll JS, Leong CO, Cheng F, Brown M, Mills AA, et al. p63 regulates an adhesion programme and cell survival in epithelial cells. Nat Cell Biol. 2006;8(6):551-61.
298. Dube S, Yalamanchili S, Lachant J, Abbott L, Benz P, Mitschow C, et al. Expression of Tropomyosin 1 Gene Isoforms in Human Breast Cancer Cell Lines. Int J Breast Cancer. 2015;2015:859427.
299. Dube S, Thomas A, Abbott L, Benz P, Mitschow C, Dube DK, et al. Expression of tropomyosin 2 gene isoforms in human breast cancer cell lines. Oncol Rep. 2016;35(6):3143-50.
300. Lv L, Huang F, Mao H, Li M, Li X, Yang M, et al. MicroRNA-21 is overexpressed in renal cell carcinoma. Int J Biol Markers. 2013;28(2):201-7.
301. Zare M, Hadi F, Alivand MR. Considering the downregulation of Tpm1.6 and Tpm1.7 in squamous cell carcinoma of esophagus as a potent biomarker. Personalized Medicine. 2018;15(5):361-70.

302. Alhajj M, Zubair M, Farhana A. Enzyme Linked Immunosorbent Assay. StatPearls. Treasure Island (FL): StatPearls Publishing

Copyright © 2024, StatPearls Publishing LLC.; 2024.

303. Nagano K, Tsutsumi Y. Phage Display Technology as a Powerful Platform for Antibody Drug Discovery. Viruses. 2021;13(2).
304. Shen S, Park JW, Huang J, Dittmar KA, Lu ZX, Zhou Q, et al. MATS: a Bayesian framework for flexible detection of differential alternative splicing from RNA-Seq data. Nucleic Acids Res. 2012;40(8):e61.
305. Vaquero-Garcia J, Aicher JK, Jewell S, Gazzara MR, Radens CM, Jha A, et al. RNA splicing analysis using heterogeneous and large RNA-seq datasets. Nature Communications. 2023;14(1):1230.

Appendix 1- all bioinformatics scripts can be found at:

https://github.com/Nadahzak/TropICana-B

Appendix 2- Annotation

TPM1

TPM2

TPM3

TPM4

TPM1- New updates: 38 transcripts instead of previous 32 transcripts

	canonical
	change in bp/ aa length; sequence revision
	change in Biotype
	deleted transcript in update
	new transcripts added
	MANE select transcript i.e. highest level of accuracy

Transcript Support Level	TSL1	all splice junctions of the transcript are supported by at least one non-suspect mRNA
	TSL2	the best supporting mRNA is flagged as suspect or the support is from multiple ESTs
	TSL3	the only support is from a single EST
	TSL4	the best supporting EST is flagged as suspect
	TSL5	no single transcript supports the model structure
	TSL n/a	not analyzed

		GRCh38							GRCh37				
Name	**Tpm nomenclature**	**Nucleotide ID**	**Transcript ID**	**bp**	**weight**	**Protein**	**Biotype**	**Flags**	**Name**	**Transcript ID**	**bp**	**Protein**	**Biotype**
TPM1-201	**1.4**	NM_001018007.2	ENST00000267996.11	1670	HMW	284aa	Protein coding	TSL:1, GENCODE	**TPM1-005**	ENST00000267996.7	1670	284aa	Protein coding
TPM1-202	**1.5**	NM_000366.6	ENST00000288398.10	1295	HMW	284aa	Protein coding	TSL:1, GENCODE	**TPM1-001**	ENST00000288398.6	1295	284aa	Protein coding
TPM1-203	**transcript variant 12**		ENST00000317516.12	1160	LMW	245aa	Protein coding	TSL:1, GENCODE	**TPM1-016**	ENST00000317516.7	738	245aa	Protein coding
TPM1-204	**1.12br / v6**		ENST00000334895.10	1564	LMW	245aa	Protein coding	TSL:1, GENCODE	**TPM1-011**	ENST00000334895.5	1637	245aa	Protein coding
TPM1-205	**1.15**	NM_001365778.1	ENST00000357980.9	1807	HMW	326aa	Protein coding	TSL:2, GENCODE	**TPM1-008**	ENST00000357980.4	1800	326aa	Protein coding
TPM1-206	**1.7**	NM_001018006.2	ENST00000358278.7	1717	HMW	284aa	Protein coding	TSL:1, GENCODE	**TPM1-018**	ENST00000358278.3	1717	284aa	Protein coding
TPM1-207	**1.1**	NM_001018005.2	ENST00000403994.9	1217	HMW	284aa	Protein coding	TSL:1, GENCODE	**TPM1-002**	ENST00000403994.3	1049	284aa	Protein coding
TPM1-208	**1.9**	NM_001330344.2	ENST00000404484.9	943	LMW	248aa	Protein coding	TSL:1, GENCODE	**TPM1-012**	ENST00000404484.4	1568	248aa	Protein coding
TPM1-209			ENST00000558072.5	525		No protein	Retained intron	TSL:2,	**TPM1-025**	ENST00000558072.1	525	No protein	Retained intron
TPM1-210			ENST00000558264.5	2687		No protein	Retained intron	TSL:2,	**TPM1-017**	ENST00000558264.1	2687	92aa	Nonsense mediated decay
TPM1-211			ENST00000558314.5	1031		No protein	Retained intron	TSL:1,	**TPM1-006**	ENST00000558314.1	1031	No protein	Retained intron
TPM1-212			ENST00000558347.5	1525		157aa	Nonsense mediated decay	TSL:5,	**TPM1-009**	ENST00000558347.1	1525	157aa	Nonsense mediated decay
TPM1-213			ENST00000558400.1	658		No protein	Retained intron	TSL:2,	**TPM1-030**	ENST00000558400.1	658	No protein	Retained intron
TPM1-214			ENST00000558544.5	1399		49aa	Nonsense mediated decay	TSL:5,	**TPM1-013**	ENST00000558544.1	1399	49aa	Nonsense mediated decay
TPM1-215			ENST00000558868.5	653		No protein	Retained intron	TSL:2,	**TPM1-021**	ENST00000558868.1	653	No protein	Retained intron
TPM1-216	**1.19**		ENST00000558910.3	1968	LMW	248aa	Protein coding	TSL:2, GENCODE	**TPM1-023**	ENST00000558910.1	782	33aa	Protein coding
TPM1-217			ENST00000559108.1	437		No protein	Retained intron	TSL:2,	**TPM1-031**	ENST00000559108.1	437	No protein	Retained intron
TPM1-218	**1.13**		ENST00000559281.6	941	LMW	248aa	Protein coding	TSL:5, GENCODE	**TPM1-015**	ENST00000559281.1	851	248aa	Protein coding
TPM1-219	**1.3**	NM_001018020.2	ENST00000559397.6	933	HMW	284aa	Protein coding	TSL:2, GENCODE	**TPM1-010**	ENST00000559397.1	1655	284aa	Protein coding
TPM1-220	**1.6**		ENST00000559556.5	977	HMW	284aa	Protein coding	TSL:2, GENCODE	**TPM1-028**	ENST00000559556.1	977	284aa	Protein coding
TPM1-221			ENST00000559831.5	573		142aa	Nonsense mediated decay	CDS 5' incompl	**TPM1-029**	ENST00000559831.1	573	142aa	Nonsense mediated decay
TPM1-222			ENST00000560131.1	537		No protein	Retained intron	TSL:2,	**TPM1-026**	ENST00000560131.1	537	No protein	Retained intron
									TPM1-019	ENST00000560445.1	485	70aa	Protein coding
TPM1-223			ENST00000560615.5	570	LMW	108aa	Protein coding	CDS 3' incompl	**TPM1-022**	ENST00000560615.1	570	108aa	Protein coding
TPM1-224			ENST00000560679.1	292		No protein	Processed transcript	TSL:3,	**TPM1-024**	ENST00000560679.1	292	8aa	Protein coding
TPM1-225			ENST00000560959.5	925	HMW	275aa	Protein coding	TSL:2, GENCODE	**TPM1-014**	ENST00000560959.1	925	275aa	Protein coding
TPM1-226	**1.16**		ENST00000560970.6	1097	HMW	284aa	Protein coding	TSL:1, GENCODE	**TPM1-003**	ENST00000560970.1	1039	265aa	Protein coding
TPM1-227			ENST00000560975.5	2855		No protein	Retained intron	TSL:5,	**TPM1-007**	ENST00000560975.1	2855	No protein	Retained intron
TPM1-228			ENST00000561242.1	673		No protein	Retained intron	TSL:2,	**TPM1-032**	ENST00000561242.1	673	No protein	Retained intron
TPM1-229	**1.2/ v8**		ENST00000561266.6	1100	HMW	284aa	Protein coding	TSL:3, GENCODE	**TPM1-027**	ENST00000561266.1	671	223aa	Protein coding
TPM1-230			ENST00000561395.5	673		68aa	Nonsense mediated decay	CDS 5' incompl	**TPM1-033**	ENST00000561395.1	673	68aa	Nonsense mediated decay
TPM1-231			ENST00000561425.5	582		40aa	Nonsense mediated decay	TSL:5,	**TPM1-020**	ENST00000561425.1	582	40aa	Nonsense mediated decay
TPM1-232			ENST00000610733.1	3957		106aa	Protein coding	TSL:2, GENCODE basic,					
TPM1-233			ENST00000644204.1	2704		No protein	Retained intron	-					
TPM1-234	**1.10br**		ENST00000651344.1	3316		No protein	Retained intron	-					
TPM1-235			ENST00000651577.1	2010		155aa	Protein coding	GENCODE basic,					
TPM1-236	**1.11br**		ENST00000651590.1	1382		251aa	Nonsense mediated decay	-					
TPM1-237			ENST00000651622.1	1163		62aa	Nonsense mediated decay		CDS 5' incomplete,				
TPM1-238			ENST00000651704.1	1780		155aa	Protein coding		GENCODE basic,				

TPM2- New updates: 10 transcripts instead of previous 8 transcripts; 2 deleted and 4 added		
	canonical	
	change in bp/ aa length; sequence revision	
	change in Biotype	
	deleted transcript in update	
	new transcripts added	
	MANE select transcript i.e. highest level of accuracy	
Transcript	TSL1	all splice junctions of the transcript are supported by at least one non-suspect mRNA
	TSL2	the best supporting mRNA is flagged as suspect or the support is from multiple ESTs
	TSL3	the only support is from a single EST
	TSL4	the best supporting EST is flagged as suspect
	TSL5	no single transcript supports the model structure
	TSL n/a	not analyzed

GRCh38						GRCh37					
Name	**TPM nomenclature**	**Transcript ID**	**bp**	**Protein**	**Biotype**	**Flags**	**Name**	**Transcript ID**	**bp**	**Protein**	**Biotype**
TPM2-201	**2.3**	ENST00000329305.6	1018	284aa	Protein coding	TSL:2, GENCODE	**TPM2-006**	ENST00000329305.2	1018	284aa	Protein coding
							TPM2-002	ENST00000360958.2	1189	284aa	Protein coding
TPM2-202	**2.1**	ENST00000378292.9	1182	284aa	Protein coding	TSL:1, GENCODE	**TPM2-001**	ENST00000378292.3	2147	284aa	Protein coding
							TPM2-005	ENST00000378300.5	1771	322aa	Protein coding
TPM2-203		ENST00000471212.5	1363	No protein	Retained intron	TSL:2,	**TPM2-003**	ENST00000471212.1	1363	No protein	Retained intron
TPM2-204		ENST00000486018.1	699	No protein	Retained intron	TSL:2,	**TPM2-004**	ENST00000486018.1	699	No protein	Retained intron
TPM2-205		ENST00000604975.1	573	No protein	Retained intron	TSL:5,	**TPM2-007**	ENST00000604975.1	573	No protein	Retained intron
TPM2-206		ENST00000607559.1	518	111aa	Protein coding	CDS 5' incompl	**TPM2-008**	ENST00000607559.1	518	111aa	Protein coding
TPM2-207		ENST00000643485.1	1580	No protein	Processed transcript	-					
TPM2-208		ENST00000644325.1	1117	133aa	Protein coding	CDS 5' incomplete,					
TPM2-209	**2.2**	ENST00000645482.3	1190	284aa	Protein coding	GENCODE basic, APPRIS ALT1, MANE Select v0.91,					
TPM2-210	**2.4**	ENST00000647435.1	1185	284aa	Protein coding	GENCODE basic, APPRIS P3,					

TPM3- New updates: 38 transcripts instead of previous 32 transcripts		
	canonical	
	change in bp/ aa length; sequence revision	
	change in Biotype	
	deleted transcript in update	
	new transcripts added	
	MANE select transcript i.e. highest level of accuracy	
Transcript	TSL1	all splice junctions of the transcript are supported by at least one non-suspect mRNA
	TSL2	the best supporting mRNA is flagged as suspect or the support is from multiple ESTs
	TSL3	the only support is from a single EST
	TSL4	the best supporting EST is flagged as suspect
	TSL5	no single transcript supports the model structure
	TSL n/a	not analyzed

GRCh38								GRCh37				
Name	**TPM Nomenclature**	**Nucleotide ID**	**Transcript ID**	**bp**	**Protein**	**Biotype**	**Flags**	**Name**	**Transcript ID**	**bp**	**Protein**	**Biotype**
TPM3-201			ENST00000271850.11	1219	285aa	Protein coding	TSL:5, GENCODE basic,	**TPM3-015**	ENST00000271850.7	1219	285aa	Protein coding
TPM3-202			ENST00000302206.9	741	158aa	Protein coding	TSL:3, GENCODE basic,	**TPM3-011**	ENST00000302206.5	741	158aa	Protein coding
TPM3-203			ENST00000312970.13	1968	No protein	Processed transcript	TSL:5,	**TPM3-022**	ENST00000312970.8	750	No protein	Processed transcript
TPM3-204	**3.4**	**NM_001043353.2**	ENST00000323144.12	1185	247aa	Protein coding	TSL:2, GENCODE basic,	**TPM3-010**	ENST00000323144.7	1088	247aa	Protein coding
TPM3-205	**3.5**	**NM_001278189.2**	ENST00000328159.9	1582	248aa	Protein coding	TSL:2, GENCODE basic,	**TPM3-006**	ENST00000328159.4	1771	248aa	Protein coding
TPM3-206	**3.2**	NM_001043351.2	ENST00000330188.13	2108	248aa	Protein coding	TSL:1, GENCODE basic,	**TPM3-009**	ENST00000330188.9	2108	248aa	Protein coding
TPM3-207			ENST00000341372.8	1933	49aa	Nonsense mediated decay	TSL:5,	**TPM3-005**	ENST00000341372.3	2007	223aa	Protein coding
TPM3-208			ENST00000341485.10	2241	247aa	Nonsense mediated decay	TSL:1,	**TPM3-003**	ENST00000341485.5	2063	232aa	Protein coding
TPM3-209			ENST00000368527.7	569	No protein	Processed transcript	TSL:4,	**TPM3-019**	ENST00000368527.3	569	No protein	Processed transcript
TPM3-210			ENST00000368530.7	1111	285aa	Protein coding	TSL:1, GENCODE basic,	**TPM3-001**	ENST00000368530.2	1523	285aa	Protein coding
TPM3-211	**3.7**	**NM_001043352.2**	ENST00000368531.6	1115	247aa	Protein coding	TSL:1, GENCODE basic,	**TPM3-026**	ENST00000368531.2	1115	247aa	Protein coding
TPM3-212	**3.1**	NM_153649.4	ENST00000368533.8	3148	248aa	Protein coding	TSL:1, GENCODE basic,	**TPM3-004**	ENST00000368533.3	3131	248aa	Protein coding
TPM3-213			ENST00000368545.7	2147	No protein	Retained intron	TSL:1,	**TPM3-014**	ENST00000368545.3	2147	No protein	Retained intron
TPM3-214			ENST00000466010.1	461	No protein	Processed transcript	TSL:2,	**TPM3-013**	ENST00000466010.1	461	No protein	Processed transcript
TPM3-215			ENST00000469717.5	4041	No protein	Retained intron	TSL:2,	**TPM3-008**	ENST00000469717.1	4041	No protein	Processed transcript
TPM3-216			ENST00000473036.2	746	No protein	Retained intron	TSL:2,	**TPM3-021**	ENST00000473036.2	746	No protein	Retained intron
TPM3-217			ENST00000504663.1	870	No protein	Retained intron	TSL:2,	**TPM3-025**	ENST00000504663.1	870	No protein	Retained intron
TPM3-218			ENST00000505010.5	540	No protein	Retained intron	TSL:2,	**TPM3-023**	ENST00000505010.1	540	No protein	Retained intron
TPM3-219			ENST00000509409.5	1248	49aa	Nonsense mediated decay	TSL:2,	**TPM3-016**	ENST00000509409.1	1248	49aa	Nonsense mediated decay
TPM3-220			ENST00000509601.1	543	95aa	Nonsense mediated decay	TSL:4,	**TPM3-020**	ENST00000509601.1	543	95aa	Nonsense mediated decay
TPM3-221			ENST00000513769.5	539	No protein	Retained intron	TSL:3,	**TPM3-024**	ENST00000513769.1	539	No protein	Retained intron
TPM3-222			ENST00000515609.1	587	75aa	Protein coding	CDS 3' incomplete, TSL:3,	**TPM3-018**	ENST00000515609.1	587	75aa	Protein coding
TPM3-223	**transcript variant 8**	**NM_001278190.2**	ENST00000611659.4	3149	227aa	Protein coding	TSL:3, GENCODE basic,					
TPM3-224	**3.12**		ENST00000651641.1	7064	285aa	Protein coding	GENCODE basic, APPRIS P1, MANE Select v0.91,					
TPM3-225			ENST00000651644.1	810	214aa	Nonsense mediated decay	-					
TPM3-226			ENST00000651731.1	2796	97aa	Nonsense mediated decay	CDS 5' incomplete,					
TPM3-227			ENST00000651873.1	1217	87aa	Nonsense mediated decay	CDS 5' incomplete,					

TPM4- New updates: 35 transcripts instead of previous 19 transcripts, 21 added and 2 deleted

	change in bp/ aa length; sequence revision
	change in Biotype
	deleted transcript in update
	new transcripts added
	MANE select transcript i.e. highest level of accuracy
Transcript	TSL1 all splice junctions of the transcript are supported by at least one non-suspect mRNA
	TSL2 the best supporting mRNA is flagged as suspect or the support is from multiple ESTs
	TSL3 the only support is from a single EST
	TSL4 the best supporting EST is flagged as suspect
	TSL5 no single transcript supports the model structure
	TSL n/a not analyzed

GRCh38								GRCh37				
Name	**isoform**	**RefSeq**	**Transcript ID**	**bp**	**Protein**	**Biotype**	**Flags**	**Name**	**Transcript ID**	**bp**	**Protein**	**Biotype**
TPM4-201			ENST00000300933.9	2330	83aa	Nonsense med	CDS 5' incomplete, TSL:1,	**TPM4-002**	ENST00000300	2542	248aa	Protein coding
TPM4-202			ENST00000344824.11	2359	158aa	Protein coding	TSL:5, GENCODE basic,	**TPM4-001**	ENST00000344	1417	284aa	Protein coding
								TPM4-201	ENST00000538	1542	284aa	Protein coding
TPM4-203			ENST00000586193.6	3369	No protein	Retained intro	TSL:3,	**TPM4-017**	ENST00000586	545	No protein	Retained intron
TPM4-204			ENST00000586499.6	2638	74aa	Nonsense med	CDS 5' incomplete, TSL:4,	**TPM4-008**	ENST00000586	708	170aa	Protein coding
TPM4-205	**tv5**	**NM_001367838.1**	ENST00000586833.7	2491	245aa	Protein coding	TSL:5, GENCODE basic,	**TPM4-009**	ENST00000586	558	179aa	Protein coding
TPM4-206			ENST00000587201.6	1879	No protein	Retained intro	TSL:2,	**TPM4-011**	ENST00000587	877	No protein	Retained intron
TPM4-207			ENST00000588032.6	707	110aa	Nonsense med	CDS 5' incomplete, TSL:3,	**TPM4-016**	ENST00000588	712	110aa	Nonsense mediated decay
TPM4-208			ENST00000588410.2	1283	No protein	Retained intro	TSL:4,	**TPM4-012**	ENST00000588	563	138aa	Protein coding
TPM4-209			ENST00000588483.2	2867	No protein	Processed tran	TSL:5,	**TPM4-014**	ENST00000588	801	16aa	Protein coding
								TPM4-010	ENST00000588	367	96aa	Protein coding
TPM4-210			ENST00000589897.1	586	69aa	Protein coding	CDS 3' incomplete, TSL:4,	**TPM4-007**	ENST00000589	586	69aa	Protein coding
TPM4-211			ENST00000590180.3	2364	No protein	Processed tran	TSL:3,	**TPM4-013**	ENST00000590	521	No protein	Processed transcript
TPM4-212			ENST00000591226.2	2675	No protein	Retained intro	TSL:NA,	**TPM4-006**	ENST00000591	2672	No protein	Retained intron
TPM4-213			ENST00000591645.3	2167	No protein	Processed tran	TSL:3,	**TPM4-018**	ENST00000591	653	No protein	Processed transcript
TPM4-214			ENST00000592138.6	1396	No protein	Retained intro	TSL:2,	**TPM4-015**	ENST00000592	635	102aa	Protein coding
								TPM4-019	ENST00000592	262	59aa	Protein coding
TPM4-215			ENST00000642221.1	2567	226aa	Protein coding	GENCODE basic,					
TPM4-216			ENST00000642789.1	2431	173aa	Nonsense med	CDS 5' incomplete,					
TPM4-217			ENST00000643494.1	2504	183aa	Nonsense med	-					
TPM4-218	**4.2**	**NM_003290.3**	ENST00000643579.2	2474	248aa	Protein coding	GENCODE basic, MANE Select v0.91,					
TPM4-219			ENST00000645471.1	3106	248aa	Nonsense med	-					
TPM4-220	**Tv4**	**NM_001367837.2**	ENST00000646575.1	2355	220aa	Protein coding	CDS 5' incomplete,					
TPM4-221	**4.1**	**NM_001145160.2**	ENST00000646974.2	2598	284aa	Protein coding	GENCODE basic, APPRIS P1,					
TPM4-222			ENST00000647037.1	2927	248aa	Nonsense med	-					
TPM4-223			ENST00000647464.2	2370	248aa	Protein coding	CDS 5' incomplete,					
TPM4-224			ENST00000653961.1	1879	No protein	Processed tran	-					
TPM4-225			ENST00000653979.1	2163	40aa	Nonsense med	CDS 5' incomplete,					
TPM4-226			ENST00000654417.1	1011	No protein	Retained intro	-					
TPM4-227			ENST00000655004.1	2447	No protein	Retained intro	-					
TPM4-228			ENST00000657915.1	2805	No protein	Processed tran	-					
TPM4-229			ENST00000658224.1	4894	No protein	Retained intro	-					
TPM4-230			ENST00000658489.1	1974	No protein	Processed tran	-					
TPM4-231			ENST00000659595.1	2330	56aa	Nonsense med	-					
TPM4-232			ENST00000663894.1	2266	No protein	Retained intro	-					
TPM4-233			ENST00000664983.1	2582	No protein	Retained intro	-					
TPM4-234			ENST00000668959.1	2540	No protein	Processed tran	-					
TPM4-235			ENST00000670382.1	2551	44aa	Nonsense med	-					

Appendix 3 - qPCR data on cell lines using GAPDH

Cell line	Tropomyosin gene	Transcript/ Isoform	qPCR avg CT	inv_CT	qPCR delta CT	2^-delta CT
RT4	TPM1	Tpm1.4	36.96	0.03	15.96	0.000016
S5637	TPM1	Tpm1.4	28.29	0.04	13.29	0.0001
SW780	TPM1	Tpm1.4	35.55	0.03	17.03	0.000007
T24	TPM1	Tpm1.4	24.63	0.04	10.63	0.00063
TCCSuP	TPM1	Tpm1.4	28.17	0.04	9.05	0.00189
UMUC3	TPM1	Tpm1.4	32.42	0.03	12.92	0.000129
VMCUB1	TPM1	Tpm1.4	33.93	0.03	16.53	0.000011
RT4	TPM1	Tpm1.7	28.77	0.03	7.77	0.004569
S5637	TPM1	Tpm1.7	22.67	0.04	7.67	0.004902
SW780	TPM1	Tpm1.7	31.51	0.03	12.99	0.000123
T24	TPM1	Tpm1.7	18.63	0.05	4.63	0.040319
TCCSuP	TPM1	Tpm1.7	26.49	0.04	7.37	0.00604
UMUC3	TPM1	Tpm1.7	26.88	0.04	7.39	0.005977
VMCUB1	TPM1	Tpm1.7	31.14	0.03	13.74	0.000073
RT4	TPM4	Tpm4.1	30.72	0.03	9.72	0.001183
S5637	TPM4	Tpm4.1	31.68	0.03	16.68	0.000009
SW780	TPM4	Tpm4.1	35.84	0.03	17.32	0.000006
T24	TPM4	Tpm4.1	18.06	0.06	4.06	0.059978
TCCSuP	TPM4	Tpm4.1	26.77	0.04	7.64	0.004999
UMUC3	TPM4	Tpm4.1	31.08	0.03	11.59	0.000324
VMCUB1	TPM4	Tpm4.1	22.24	0.04	4.84	0.034899
RT4	TPM4	Tpm4.2	33	0.03	12	0.000245
S5637	TPM4	Tpm4.2	24.81	0.04	9.81	0.001113
SW780	TPM4	Tpm4.2	28.28	0.04	9.76	0.001152
T24	TPM4	Tpm4.2	25.83	0.04	11.83	0.000276
TCCSuP	TPM4	Tpm4.2	28.26	0.04	6.56	0.010609
UMUC3	TPM4	Tpm4.2	31.68	0.03	12.19	0.000214
VMCUB1	TPM4	Tpm4.2	28.07	0.04	10.67	0.000613

Appendix 3 – qPCR data of all samples using Taqman probes

Grade	Samples	Avg CT Tpm4.1	Avg CT Tpm4.2	Ratio	Fold-change TPM4	dCT TPM4	Avg CT Tpm1.4	Avg CT Tpm1.7	Ratio	Fold-change TPM1	dCT TPM1
cell line	VMCUB1	23.73	25.02	1.05	0.95	1.29	30.57	26.88	0.87926	1.137325	-3.69
cell line	SW780	32.51	28.56	0.88	1.14	-3.96	29.23	24.37	0.83386	1.199248	-4.86
cell line	T24	31.49	26.81	0.85	1.17	-4.68	29.23	23.21	0.79398	1.259476	-6.02
cell line	5637	32.03	26.90	0.84	1.19	-5.13	36.95	29.09	0.78726	1.270223	-7.86
cell line	MGUH	21.07	21.81	1.04	0.97	0.75	28.32	22.25	0.78564	1.272844	-6.07
cell line	RT4	25.37	22.78	0.90	1.11	-2.59	26.72	20.85	0.78017	1.281767	-5.87
cell line	TCCSUP	22.98	22.05	0.96	1.04	-0.93	23.45	20.78	0.88624	1.128365	-2.67
G- NA	s1107	20.18	22.07	1.09	0.91	1.89	28.32	27.55	0.97274	1.028024	-0.77
G- NA	s0896	21.66	22.80	1.05	0.95	1.14	26.72	24.90	0.93188	1.073097	-1.82
G- NA	s0277	27.71	24.25	0.88	1.14	-3.46	31.02	28.92	0.93235	1.072555	-2.10
G1pTa	s1101	20.72	21.31	1.03	0.97	0.59	26.56	25.38	0.95575	1.046295	-1.17
G1pTa	s0177	25.41	23.43	0.92	1.08	-1.98	24.73	31.00	1.25376	0.797603	6.27
G1pTa	s0658	21.54	22.69	1.05	0.95	1.15	23.43	26.17	1.11703	0.895231	2.74
G1pTa	s0713	24.47	24.15	0.99	1.01	-0.32	26.97	25.48	0.94464	1.058608	-1.49
G1pTa	s1207	29.07	28.53	0.98	1.02	-0.54	28.84	27.29	0.94614	1.056931	-1.55
G2pT1	s0185	26.87	23.55	0.88	1.14	-3.33	28.28	24.04	0.85014	1.176283	-4.24
G2pTa	s0247	24.96	24.55	0.98	1.02	-0.40	34.64	31.15	0.89940	1.11185	-3.48
G3pT1	s0044	25.85	23.34	0.90	1.11	-2.51	29.88	24.77	0.82891	1.206397	-5.11
G3pT1	s0045	20.98	21.57	1.03	0.97	0.59	26.64	24.67	0.92582	1.080132	-1.98
G3pT1	s0046	24.77	23.00	0.93	1.08	-1.77	26.00	24.26	0.93291	1.071904	-1.74
G3pT1	s0047	26.09	23.45	0.90	1.11	-2.64	26.85	23.64	0.88021	1.136096	-3.22
G3pT1	s0048	22.92	21.21	0.93	1.08	-1.72	23.10	21.78	0.94267	1.060815	-1.32
G3pT1	s0049	28.82	25.34	0.88	1.14	-3.48	31.83	30.20	0.94879	1.053974	-1.63
G3pT1	s0050	27.84	24.32	0.87	1.14	-3.52	29.03	26.04	0.89693	1.11491	-2.99

G3pT1	s0051	19.94	20.11	1.01	0.99	0.18	26.10	20.76	0.79554	1.257011	-5.34
G3pT1	s0052	26.63	26.30	0.99	1.01	-0.33	28.01	30.59	1.09200	0.915749	2.58
G3pT1	s0053	29.90	25.57	0.86	1.17	-4.32	28.54	25.24	0.88442	1.130681	-3.30
G3pT1	s0054	26.72	26.38	0.99	1.01	-0.35	33.24	30.76	0.92553	1.080466	-2.48
G3pT1	s0055	26.94	26.84	1.00	1.00	-0.09	26.87	24.67	0.91833	1.088932	-2.19
G3pT1	s0525	29.14	25.62	0.88	1.14	-3.52	31.00	27.61	0.89075	1.122644	-3.39
G3pT1	s0591	28.12	24.98	0.89	1.13	-3.14	28.57	26.72	0.93517	1.069326	-1.85
G3pT1	s0613	27.35	24.42	0.89	1.12	-2.93	29.35	23.09	0.78656	1.271358	-6.27
G3pT1	s0663	29.46	25.39	0.86	1.16	-4.07	29.34	28.82	0.98206	1.018273	-0.53
G3pT1	s1115	26.76	24.59	0.92	1.09	-2.16	24.07	23.98	0.99611	1.003908	-0.09
G3pT1	s1434	27.93	25.01	0.90	1.12	-2.92	27.56	24.33	0.88310	1.13237	-3.22
G3pT1	s1445	27.98	24.71	0.88	1.13	-3.27	26.31	24.62	0.93564	1.068785	-1.69
G3pT1	s1919	21.16	21.52	1.02	0.98	0.36	28.61	20.26	0.70824	1.41195	-8.35
G3pT1	s0028	26.99	23.83	0.88	1.13	-3.16	24.98	23.41	0.93685	1.067402	-1.58
G3pT1	s0080	25.98	23.58	0.91	1.10	-2.41	25.82	22.32	0.86468	1.156504	-3.49
G3pT1	s0237	20.10	20.70	1.03	0.97	0.60	26.79	26.46	0.98764	1.012514	-0.33
G3pT1	s0255	21.28	22.44	1.05	0.95	1.16	28.47	20.43	0.71758	1.39358	-8.04
G3pT1	s0538	27.71	23.67	0.85	1.17	-4.03	26.25	23.07	0.87898	1.13768	-3.18
G3pT1	s0550	25.08	22.82	0.91	1.10	-2.26	22.98	23.45	1.02028	0.980127	0.47
G3pT1	s0735	21.22	22.61	1.07	0.94	1.39	29.52	28.54	0.96694	1.034194	-0.98
G3pT1	s1395	24.27	21.46	0.88	1.13	-2.81	29.26	18.97	0.64846	1.542105	-10.29
G3pT1	s1508	23.93	24.07	1.01	0.99	0.13	27.25	22.78	0.83576	1.196519	-4.48
G3pT1	s0054	22.11	23.02	1.04	0.96	0.91	30.82	23.08	0.74887	1.33535	-7.74
G3pT1	s0084	27.85	25.63	0.92	1.09	-2.22	27.57	25.34	0.91909	1.088029	-2.23
G3pT1	s0201	21.87	23.25	1.06	0.94	1.39	28.53	20.93	0.73363	1.36309	-7.60
G3pT1	s0418	25.76	26.63	1.03	0.97	0.88	34.34	32.55	0.94803	1.054817	-1.78
G3pT1	s1406	25.42	25.31	1.00	1.00	-0.11	30.98	29.02	0.93658	1.06771	-1.96
G3pT1	s0166	24.59	24.05	0.98	1.02	-0.55	26.41	23.90	0.90512	1.104826	-2.51

G3pT1	s0044	28.80	26.47	0.92	1.09	-2.33	32.74	28.68	0.87609	1.14143	-4.06
G3pT1	s0078	24.50	24.67	1.01	0.99	0.17	30.56	28.40	0.92926	1.076121	-2.16
G3pT1	s0079	28.72	26.01	0.91	1.10	-2.71	30.16	27.72	0.91908	1.088041	-2.44
G3pT1	s0098	30.19	26.47	0.88	1.14	-3.72	30.13	26.56	0.88152	1.134407	-3.57
G3pT1	s0109	27.61	24.89	0.90	1.11	-2.72	27.74	25.52	0.92016	1.086764	-2.21
G3pT1	s0193	31.99	30.26	0.95	1.06	-1.73	36.13	36.59	1.01262	0.987536	0.46
G3pT1	s0237	29.39	25.48	0.87	1.15	-3.90	31.47	27.76	0.88208	1.133679	-3.71
G3pT1	s0294	23.85	23.56	0.99	1.01	-0.29	29.24	23.53	0.80483	1.242494	-5.71
G3pT1	s0295	27.67	25.63	0.93	1.08	-2.04	29.07	30.50	1.04928	0.953038	1.43
G3pT1	s0298	29.86	25.22	0.84	1.18	-4.64	30.36	27.01	0.88952	1.124205	-3.35
G3pT1	s0304	26.50	26.17	0.99	1.01	-0.33	35.14	30.98	0.88176	1.134096	-4.15
G3pT1	s0400	26.19	25.62	0.98	1.02	-0.58	26.81	23.58	0.87940	1.137142	-3.23
G3pT1	s0525	30.98	25.68	0.83	1.21	-5.31	30.50	27.37	0.89740	1.114327	-3.13
G3pT1	s0591	28.17	24.23	0.86	1.16	-3.94	27.55	25.84	0.93799	1.066116	-1.71
G3pT1	s0613	27.13	24.20	0.89	1.12	-2.93	27.61	22.97	0.83222	1.201606	-4.63
G3pT1	s0663	32.97	28.62	0.87	1.15	-4.35	33.07	32.27	0.97591	1.024683	-0.80
G3pT2+	s0900	25.88	24.84	0.96	1.04	-1.04	27.11	22.75	0.83911	1.191741	-4.36
G3pT2+	s1158	23.69	23.46	0.99	1.01	-0.23	26.85	25.82	0.96165	1.039883	-1.03
G3pT2+	s0176	20.60	21.22	1.03	0.97	0.62	24.92	20.46	0.82135	1.21751	-4.45
G3pT2+	s0191	24.61	23.33	0.95	1.05	-1.28	26.10	19.41	0.74391	1.344245	-6.68
G3pT2+	s0908	21.49	22.22	1.03	0.97	0.73	23.56	21.46	0.91057	1.09821	-2.11
G3pT2+	s1474	22.59	23.40	1.04	0.97	0.81	30.14	27.66	0.91788	1.089467	-2.48
G3pT2+	s0178	21.27	22.33	1.05	0.95	1.06	24.00	22.85	0.95225	1.050148	-1.15
G3pTa	s1439	27.73	23.35	0.84	1.19	-4.38	32.39	29.70	0.91686	1.090676	-2.69
G3pTa	s1412	26.35	26.00	0.99	1.01	-0.35	25.61	28.28	1.10403	0.90577	2.66
G3pTa	s0266	28.32	28.69	1.01	0.99	0.37					0.00
Control	N0109	31.55	31.63	1.00	1.00	0.08	36.25	32.22	0.88893	1.124951	-4.03
Control	N0700	28.22	29.21	1.04	0.97	0.99	38.54	29.71	0.77091	1.297169	-8.83

Control	N0702	34.54	32.78	0.95	1.05	-1.77					0.00
Control	N0863	27.96	29.01	1.04	0.96	1.06	37.03	30.82	0.83236	1.201399	-6.21
Control	N0105	27.49	28.60	1.04	0.96	1.11					0.00
Control	N0106	27.76	28.58	1.03	0.97	0.82	36.34	28.89	0.79499	1.257884	-7.45
Control	Urine 10	24.91	25.66	1.03	0.97	0.75	32.46	27.48	0.84655	1.181261	-4.98
Control	N0453	32.53	31.72	0.98	1.03	-0.81					0.00
MIBC	T0056	21.11	22.01	1.04	0.96	0.90	24.66	21.12	0.85632	1.167793	-3.54
MIBC	T0314	21.30	22.01	1.03	0.97	0.71	27.35	23.24	0.84971	1.176879	-4.11
MIBC	T0561	24.49	25.91	1.06	0.95	1.42	32.34	29.36	0.9077	1.101686	-2.99
MIBC	T0569	20.75	21.29	1.03	0.97	0.54	26.59	19.02	0.71543	1.39776	-7.57
MIBC	T0747	23.78	24.15	1.02	0.98	0.37	28.54	25.72	0.90113	1.109713	-2.82
MIBC	T0891	25.26	22.39	0.89	1.13	-2.87	25.74	24.15	0.93841	1.065638	-1.59
MIBC	T0986	23.60	24.33	1.03	0.97	0.73	30.81	22.61	0.73378	1.362806	-8.20
MIBC	T1401	19.82	20.94	1.06	0.95	1.12	27.34	23.90	0.87431	1.143761	-3.44
MIBC	T1402	25.36	24.45	0.96	1.04	-0.92	31.42	25.58	0.81405	1.228425	-5.84
MIBC	T2200	22.86	23.56	1.03	0.97	0.70	25.78	20.90	0.81059	1.233665	-4.88

Appendix 4- MS data, list of peptides

TPM gene	exon	Peptide sequence	Total count of isof	Isoforms										
1	2a	KAEDSLLAAEEAAAKL	1	ENST00000357980.9										
1	2a	RVLEELHKA	4	ENST00000357980.9	ENST00000267996.11	ENST00000561266.6	ENST00000559397.6							
1	1a/2b	KQLEDELVSLQKK	7	ENST00000288398.10	ENST00000358278.7	ENST00000558347.5	ENST00000560970.6	ENST00000403994.9	ENST00000559556.5	ENST00000610733.1				
1	5/6a	RAELSEGQVRQ	8	ENST00000317516.12	ENST00000288398.10	ENST00000358278.7	ENST00000404484.9	ENST00000559831.5	ENST00000560959.5	ENST00000560615.5	ENST00000559397.6			
1	6a	RQLEEQLRI	8	ENST00000317516.12	ENST00000288398.10	ENST00000358278.7	ENST00000404484.9	ENST00000559831.5	ENST00000560959.5	ENST00000560615.5	ENST00000559397.6			
2	6a/7	KSLMASEEEYSTKE	2	ENST00000378292.9	ENST00000647435.1									
2	8/9d	KTIDDLEETLASAKE	2	ENST00000378292.9	ENST00000329305.6									
2	6a	RQLEEELRT	2	ENST00000378292.9	ENST00000647435.1									
2	4	KMELQEMQLKE	4	ENST00000378292.9	ENST00000329305.6	ENST00000647435.1	ENST00000645482.3							
2	1a/2b	KQLEEEQQALQKK	4	ENST00000378292.9	ENST00000329305.6	ENST00000647435.1	ENST00000645482.3							
2	6b/7	KSLEAQADKYSTKE	4	ENST00000329305.6	ENST00000607559.1	ENST00000644325.1	ENST00000645482.3							
2	5	KLVILEGELERS	5	ENST00000378292.9	ENST00000329305.6	ENST00000607559.1	ENST00000647435.1	ENST00000645482.3						
2	5	RKLVILEGELERS	5	ENST00000378292.9	ENST00000329305.6	ENST00000607559.1	ENST00000647435.1	ENST00000645482.3						
3	7/.8	KTIDDLEDKLKC	5	ENST00000368533.8	ENST00000302206.9	ENST00000330188.13	ENST00000271850.11	ENST00000611659.4						
3	8	RMLDQTLLDLNEM	5	ENST00000368533.8	ENST00000302206.9	ENST00000330188.13	ENST00000271850.11	ENST00000611659.4						
3	5	RLMDQNLKC	7	ENST00000368533.8	ENST00000368531.6	ENST00000368530.7	ENST00000328159.9	ENST00000271850.11	ENST00000651644.1	ENST00000611659.4				
3	1--2	REQAEAEVASLNRR	9	ENST00000368533.8	ENST00000341485.10	ENST00000368531.6	ENST00000330188.13	ENST00000328159.9	ENST00000323144.12	ENST00000509601.1	ENST00000651873.1	ENST00000611659.4		
3	1	MAGITTIEAVKR	10	ENST00000368533.8	ENST00000341485.10	ENST00000341372.8	ENST00000368531.6	ENST00000330188.13	ENST00000328159.9	ENST00000323144.12	ENST00000509601.1	ENST00000509409.5	ENST00000611659.4	
3	1b	MAGITTIEAVKRK	10	ENST00000368533.8	ENST00000341485.10	ENST00000341372.8	ENST00000368531.6	ENST00000330188.13	ENST00000328159.9	ENST00000323144.12	ENST00000509601.1	ENST00000509409.5	ENST00000611659.4	
3	1b	RKIQVLQQQADDAEERA	10	ENST00000368533.8	ENST00000341485.10	ENST00000341372.8	ENST00000368531.6	ENST00000330188.13	ENST00000328159.9	ENST00000323144.12	ENST00000509601.1	ENST00000509409.5	ENST00000611659.4	
3	6	KEDKYEEEIKI	11	ENST00000368533.8	ENST00000341485.10	ENST00000302206.9	ENST00000368531.6	ENST00000330188.13	ENST00000368530.7	ENST00000328159.9	ENST00000271850.11	ENST00000323144.12	ENST00000651641.1	ENST00000651731.1
4	1b/3	REKAEGDVAALNRR	5	ENST00000643579.2	ENST00000645471.1	ENST00000643494.1	ENST00000646575.1	ENST00000647037.1						
4	1b	KIQALQQQADEAEDRA	6	ENST00000659595.1	ENST00000643579.2	ENST00000645471.1	ENST00000643494.1	ENST00000647037.1	ENST00000670382.1					
4	7/.8	KTIDDLEEKL	6	ENST00000344824.11	ENST00000586833.7	ENST00000646974.2	ENST00000643579.2	ENST00000588032.6	ENST00000642221.1					
4	1b	MAGLNSLEAVKR	6	ENST00000659595.1	ENST00000643579.2	ENST00000645471.1	ENST00000643494.1	ENST00000647037.1	ENST00000670382.1					
4	1b	MAGLNSLEAVKRK	6	ENST00000659595.1	ENST00000643579.2	ENST00000645471.1	ENST00000643494.1	ENST00000647037.1	ENST00000670382.1					
4	1b	RKIQALQQQADEAEDRA	6	ENST00000659595.1	ENST00000643579.2	ENST00000645471.1	ENST00000643494.1	ENST00000647037.1	ENST00000670382.1					
4	4	KLVILEGELERA	11	ENST00000344824.11	ENST00000586833.7	ENST00000646974.2	ENST00000647464.2	ENST00000643579.2	ENST00000645471.1	ENST00000643494.1	ENST00000646575.1	ENST00000642221.1	ENST00000647037.1	ENST00000642789.1
4	3	KMEIQEMQLKE	11	ENST00000344824.11	ENST00000586833.7	ENST00000646974.2	ENST00000647464.2	ENST00000643579.2	ENST00000645471.1	ENST00000643494.1	ENST00000646575.1	ENST00000642221.1	ENST00000647037.1	ENST00000642789.1
4	4	RKLVILEGELERA	11	ENST00000344824.11	ENST00000586833.7	ENST00000646974.2	ENST00000647464.2	ENST00000643579.2	ENST00000645471.1	ENST00000643494.1	ENST00000646575.1	ENST00000642221.1	ENST00000647037.1	ENST00000642789.1
1; 4	2--3	KKATDAEADVASLNRR	11	ENST00000357980.9	ENST00000378292.9	ENST00000288398.10	ENST00000329305.6	ENST00000358278.7	ENST00000558347.5	ENST00000560970.6	ENST00000403994.9	ENST00000559556.5	ENST00000647435.1	ENST00000645482.3
2; 4	1a	KLDKENAIDRA	6	ENST00000378292.9	ENST00000329305.6	ENST00000646974.2	ENST00000589897.1	ENST00000647435.1	ENST00000645482.3					
		KATDAEADVASLNRR	12	ENST00000357980.9	ENST00000378292.9	ENST00000288398.10	ENST00000329305.6	ENST00000358278.7	ENST00000558347.5	ENST00000560970.6	ENST00000403994.9	ENST00000559831.5	ENST00000559556.5	ENST00000647435.1
		KLVIIEGDLERT	12	ENST00000368533.8	ENST00000341485.10	ENST00000302206.9	ENST00000368531.6	ENST00000330188.13	ENST00000368530.7	ENST00000328159.9	ENST00000271850.11	ENST00000323144.12	ENST00000651641.1	ENST00000651644.1
		RKLVIIEGDLERT	12	ENST00000368533.8	ENST00000341485.10	ENST00000302206.9	ENST00000368531.6	ENST00000330188.13	ENST00000368530.7	ENST00000328159.9	ENST00000271850.11	ENST00000323144.12	ENST00000651641.1	ENST00000651644.1
		KEDKYEEEIKL	16	ENST00000378292.9	ENST00000344824.11	ENST00000329305.6	ENST00000586833.7	ENST00000646974.2	ENST00000647464.2	ENST00000607559.1	ENST00000643579.2	ENST00000644325.1	ENST00000647435.1	ENST00000645471.1
		KLDKENALDRA	17	ENST00000357980.9	ENST00000267996.11	ENST00000288398.10	ENST00000368530.7	ENST00000271850.11	ENST00000358278.7	ENST00000558347.5	ENST00000560970.6	ENST00000561266.6	ENST00000403994.9	ENST00000561425.5
		KLVIIESDLERA	20	ENST00000334895.10	ENST00000357980.9	ENST00000317516.12	ENST00000267996.11	ENST00000288398.10	ENST00000358278.7	ENST00000559281.6	ENST00000560970.6	ENST00000561266.6	ENST00000403994.9	ENST00000558910.3
		RKLVIIESDLERA	20	ENST00000334895.10	ENST00000357980.9	ENST00000317516.12	ENST00000267996.11	ENST00000288398.10	ENST00000358278.7	ENST00000559281.6	ENST00000560970.6	ENST00000561266.6	ENST00000403994.9	ENST00000558910.3
		KMEIQEIQLKE	21	ENST00000334895.10	ENST00000357980.9	ENST00000317516.12	ENST00000267996.11	ENST00000288398.10	ENST00000358278.7	ENST00000559281.6	ENST00000560970.6	ENST00000561266.6	ENST00000403994.9	ENST00000558910.3
		RIQLVEEELDRA	46	ENST00000368533.8	ENST00000341485.10	ENST00000334895.10	ENST00000368531.6	ENST00000357980.9	ENST00000330188.13	ENST00000317516.12	ENST00000267996.11	ENST00000378292.9	ENST00000288398.10	ENST00000368530.7
		RIQLVEEELDRAQERL	46	ENST00000368533.8	ENST00000341485.10	ENST00000334895.10	ENST00000368531.6	ENST00000357980.9	ENST00000330188.13	ENST00000317516.12	ENST00000267996.11	ENST00000378292.9	ENST00000288398.10	ENST00000368530.7
		RLATALQKL	46	ENST00000368533.8	ENST00000341485.10	ENST00000334895.10	ENST00000368531.6	ENST00000357980.9	ENST00000330188.13	ENST00000317516.12	ENST00000267996.11	ENST00000378292.9	ENST00000288398.10	ENST00000368530.7
		RLATALQKLEEAEKA	46	ENST00000368533.8	ENST00000341485.10	ENST00000334895.10	ENST00000368531.6	ENST00000357980.9	ENST00000330188.13	ENST00000317516.12	ENST00000267996.11	ENST00000378292.9	ENST00000288398.10	ENST00000368530.7
		RKYEEVARK	48	ENST00000368533.8	ENST00000341485.10	ENST00000334895.10	ENST00000302206.9	ENST00000368531.6	ENST00000357980.9	ENST00000330188.13	ENST00000317516.12	ENST00000267996.11	ENST00000378292.9	ENST00000288398.10