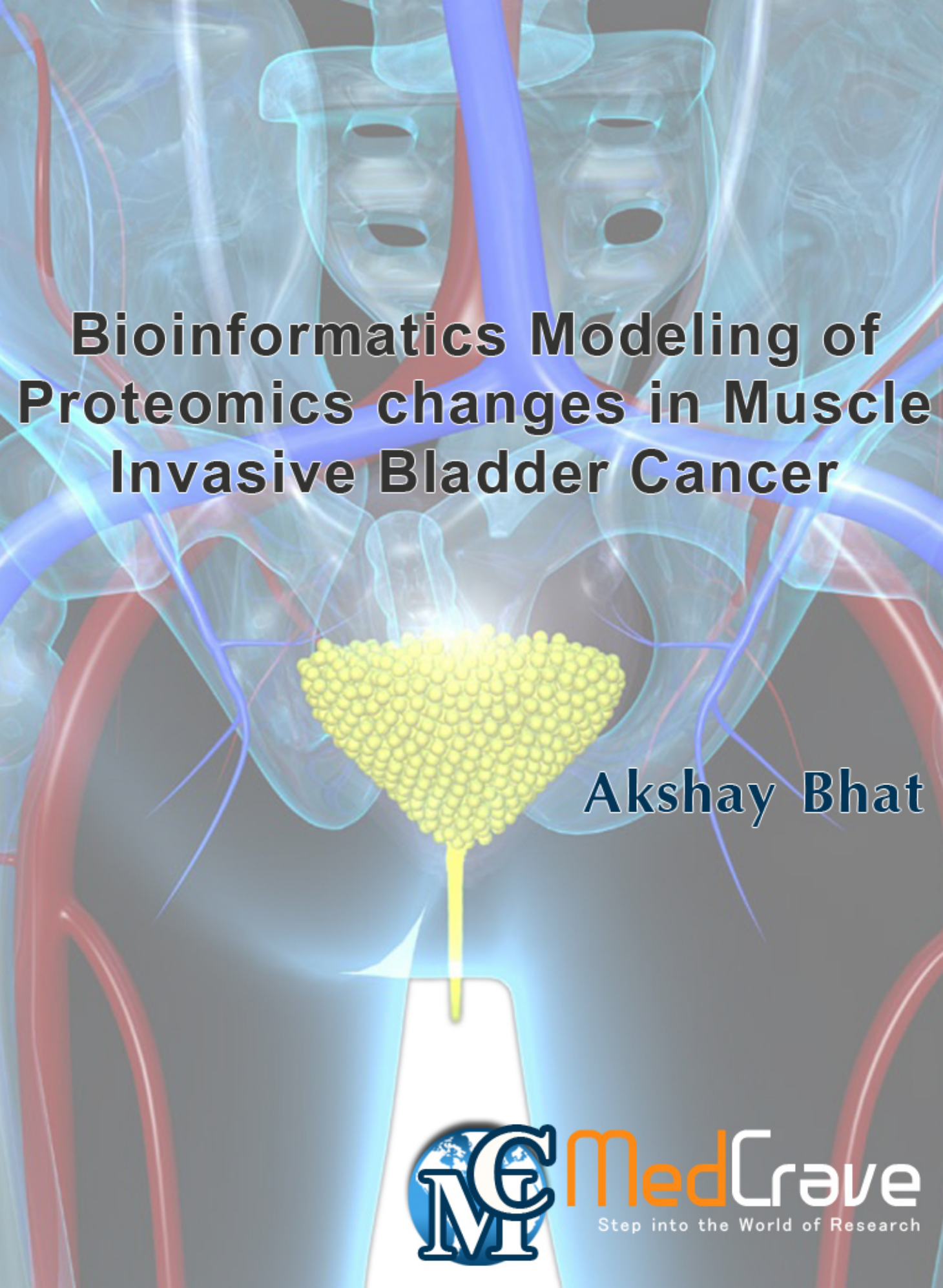

Bioinformatics Modeling of Proteomics changes in Muscle Invasive Bladder Cancer

Akshay Bhat

MedCrave
Step into the World of Research

Bioinformatics Modeling of Proteomics changes in Muscle Invasive Bladder Cancer

By

Akshay Bhat

Charite University Berlin - Marie Curie Fellow

Germany

Published By:

MedCrave Group LLC

August 24, 2016

Contents

Abbreviations

BC	Bladder Cancer
BCG	Bacillus Calmette-Guerin
CIS	Carcinoma in-situ
GEO	Gene Expression Omnibus
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
MIBC	Muscle Invasive Bladder Cancer
OMIM	Online Mendelian Inheritance in Man
PE	Pathway Enrichment
PPI	Protein-Protein Interaction
RDBMS	Relational Database Management System
ROR	Ruby on Rails
TCGA	The Cancer Genome Atlas
TNM	Tumor-Node-Metastasis

Abstract

Background

Bladder Cancer (BC) has two distinct phenotypes. Non-muscle invasive BC has good prognosis and is treated by tumor resection and intravesical therapy, whereas muscle-invasive BC has poor prognosis and requires radical cystectomy combined with cisplatin-based chemotherapy. High-throughput sequencing technologies allow identification of individual molecular signatures to characterize the invasive phenotype.

Objective

Based on this background the objective of this thesis comprises of three parts. The first aim of the thesis was to characterize muscle-invasive BC on a molecular level by incorporating signatures from literature and omics profiles. The second aim was to evaluate the performance of pathway-enrichment obtained from two bioinformatics tools ImPAla and ClueGO. The third aim shows the use of bioinformatics in order to identify altered pathways relevant to ageing.

Materials

Public domain -omics signatures and molecular features associated to muscle-invasive BC were derived from literature mining to provide protein-coding genes. These features were integrated in a protein-interaction network to obtain functional pathways relevant to the phenotype. Pathway-enrichment was performed using ClueGO and ImPAla tools. The resulting pathway terms were filtered according to criterion: multiple comparison corrected p-value <0.05.

Results

In the first part of the thesis, the protein-interactions and pathway-enrichment yielded 14 significant pathway terms. Three pathway terms were not previously reported in muscle-invasive BC. The novel disease-associated pathways were regulation of actin-cytoskeleton, neurotrophin-signalling pathway and endocytosis. In the second part, 292 pathways were obtained from ClueGO and 471 pathways from ImPAla software. Comparison of the results obtained by the two applications yielded 152 pathway-terms with the same pathway name. 137 ClueGO pathway-terms were similar to 251 ImPAla pathways. In the last part, the results from a bioinformatics analysis of urinary-peptidomics data discovered a pathway-term “degradation of insulin-like growth factor-binding proteins” that was unique in the context of pathological ageing.

Conclusion

The results of this thesis suggest that there is a complex interplay between pathways characterizing the muscle-invasive phenotype of BC. Further experimental validation of the three novel pathways with respect to progression and treatment response is indicated. In addition, the comparison of two prominent pathway enrichment tools ClueGO and ImPAla showed that ClueGO has better performance than ImPAla in pathway-enrichment analysis since the output is less redundant and contains all the biologically significant information. Lastly, molecular pathways enriched in normal and pathological ageing demonstrate that with the help of appropriate peptidomics technologies, urine could be used as a useful source of information in ageing research.

Introduction

Urothelial bladder carcinoma is a common malignancy of the urinary tract system and comprises of two distinct clinical phenotypes, non muscle-invasive bladder cancer and muscle-invasive tumors. Muscle-invasive bladder tumors are treated with pre-operative (neoadjuvant) cis-platin-based chemotherapy followed by removal of the bladder, named as "radical cystectomy" [1]. However, since a significant number of patients do not respond to chemotherapy treatment, a detailed investigation of the tumor molecular signature is required to select responsive patients for bladder cancer treatment [2]. A better understanding of muscle invasive bladder cancer might be achieved by combining information obtained from individual biomarkers measured at the DNA, RNA and/or protein levels [3]. Along this background, the comprehensive characterization of altered molecular pathways provides significant clinical relevance in order to choose optimum diagnosis and treatment regimens for bladder cancer patients.

Variable for individual bladder cancer patients, initial symptoms include hematuria and flank pain [4,5]. Cystoscopy is the gold standard diagnostic procedure with a reported sensitivity of 62-84% and specificity of 43-98%. This wide variability in sensitivity and specificity indicates a significant inter-operator variability [6]. Non muscle-invasive bladder cancer comprises of distinct forms [7-9]:

a. Ta stage - the cancer is just in the innermost layer of the bladder lining.

b. T1 stage - the cancer has started to grow into the connective tissue beneath the bladder lining

c. Carcinoma in-situ (CIS) - very early, high grade, cancer cells are only in the innermost layer of the bladder lining [9].

Papillary tumors that are confined to the mucosa and that invade the lamina propria of the bladder are classified as stages Ta and T1 according to the Tumor-Node-Metastasis (TNM) classification system [10]. The papillary carcinoma (Ta and T1) phenotype has a tendency to recur locally and it rarely invades the bladder basement membrane or metastasizes to adjacent organs. However, the flat carcinoma in situ (CIS) is often multifocal and is a dangerous lesion with a high tendency for bladder muscle invasion and metastasis [11]. Treatment of non-muscle invasive bladder carcinoma (Ta, T1 and CIS) involves endoscopic transurethral resection of visible tumors followed by adjuvant treatment with intravesical instillation therapy (Mitomycin/Epirubicin or Bacillus Calmette-Guerin (BCG)) depending on the estimated risk for progression. Irrespective of aggressive treatment and vigorous follow-up, 70% of these tumors recur, and 25% of high-grade non-muscle invasive cancers (CIS) progress into invasive phenotypes [12,13].

Muscle invasive bladder cancers are advanced cancer stages and are classified as "T2- T4" according to the Tumor-Node-Metastasis classification system.

a. T2 stage - is when the cancer has spread into the muscle layer of the bladder

b. T3 stage - is when the tumor has grown through the muscle layer

c. T4 stage - is when the cancer has spread into the prostrate, uterus or vagina, or into the wall of the pelvis or abdomen

Furthermore, muscle invasive bladder tumors are also distinguished into three distinct molecular subtypes that have widely variable clinical outcomes and responses to conventional chemotherapy treatments:

a. Basal subtypes

b. Luminal subtypes

c. "p53-like" tumors [14]

The basal muscle invasive bladder cancer subtypes are susceptible to have more invasive and metastatic disease at initial diagnosis and are associated with shorter disease-specific and overall survival. The biomarkers for the basal muscle invasive bladder cancer subtype include CD44 antigen (CD44), Keratin, type II cytoskeletal 5 (KRT5), Keratin, type II cytoskeletal 72 (KRT72), Keratin, type I cytoskeletal 14 (KRT14) and Cadherin-3 (CDH3). The luminal muscle invasive bladder cancer subtypes are enriched with activating fibroblast growth factor receptor 3 (FGFR3) and human epidermal growth factor receptor 3 (ERBB3) mutations and Receptor tyrosine-protein kinase erbB-2 (ERBB2) amplifications, and the gene expression profiles are controlled by peroxisome proliferator activator receptor γ (PPARγ) and estrogen receptor activation. The wild-type p53 is required for DNA damage induced apoptosis and is a central tenet in cancer biology [15]. Therefore, it is interesting that the de-novo and induced chemoresistance in muscle invasive bladder cancers was associated with wild-type p53 gene expression signatures. Nevertheless, TP53 mutation frequencies are similar in all three subtypes of muscle invasive bladder cancers (basal, luminal and p53-like), indicating that wild-type p53 was not responsible for the baseline and chemotherapy induced p53-like gene expression. Hence, it is proposed that "p53-like tumors" as measured by mRNA expression would be a more accurate predictor of de-novo and induced muscle invasive bladder cancers chemoresistance than would analyses of TP53 mutational status [16]. The determination of the molecular basis of these p53-like signatures is not yet defined and that could overcome de novo and/ or prevent acquired chemoresistance [14].

The most important point for the planning of radical cystectomy in bladder cancer tumors is the depth of invasion or muscular involvement (T category, P stage) [17]. Various treatments have proved useful for disease control in some patients with regional bladder cancer but the most

common procedure for this situation is radical cystectomy [18]. A recent report highlights significant errors in clinical staging of patients with bladder cancer that underwent radical cystectomy [19]. This staging error of cystectomy in the bladder tumors may cause very important mistakes on the decision for radical surgery treatment. Hence, obtaining an accurate staging diagnosis is particularly crucial in patient selection for surgical treatments (i.e. cystectomy) and the choice of chemotherapy. Due to the invasive procedure of cystoscopy and in order to improve accuracy in the phenotype detection, blood or urine biomarkers could support clinical assessment [20].

High-throughput experimental platform technologies range from genomic sequencing to epigenomic, transcriptomic, proteomic and metabolomic profiling in order to characterize the molecular aspects of individual clinical phenotypes [21-28]. Genomic sequencing comprises of applying recombinant DNA, DNA sequencing methods, and bioinformatics to sequence, assemble, and analyze the function and structure of "genomes" (i.e. the complete set of DNA within a single cell of an organism). Epigenomic sequencing is the determination of key functional elements that regulate gene expression in a cell - Epigenomes provide information about the patterns in which structures such as methyl groups tag DNA and histones (the proteins around which DNA is packaged to form the chromatin), and about interactions between distant sections of chromatin). Transcriptomics is the sequencing and quantification of transcripts - mRNA and microRNA or miRNA. Proteomics is the sequencing and quantification of the proteome and peptidome, and metabolomics is the technique of identifying and quantifying of metabolites. These techniques provide datasets that comprise of DNA-mutations, DNA-methylations, mRNAs, miRNA, proteins, peptides and metabolites. The advent of these approaches that generate a comprehensive view of the molecular landscape for a biological sample has introduced a paradigm shift in the way diseases are perceived [21,22,29].

A variety of datasets for such molecular characterizations have become available that are stored in public databases, for e.g. in Array Express [30] or Gene Expression Omnibus (GEO) [31], which is a database that stores mRNA and miRNA datasets from transcriptomics experiments, Human Proteinpedia is a public repository that provides information on proteomics datasets [32,33], Human Protein Atlas (http://www.proteinatlas.org/), which is an online portal that contains information on immunohistochemically validated proteins, or large data consolidation resources such as GeneCards [34] that provides information for genomic, proteomic, transcriptomic, genetic and functional information on all known and predicted human genes. This database aims to provide a quick overview of the current available biomedical information about the searched gene, including the human genes, the encoded proteins, and the relevant diseases.

In regard to disease specific omics data, valuable general sources in oncology include The Cancer Genome Atlas (TCGA) (http://cancergenome.nih.gov/), Oncomine [35], and Online Mendelian Inheritance in Man (OMIM) [36]. The Cancer Genome Atlas oncology portal currently lists single nucleotide polymorphism, methylation data, mutations, mRNAs, miRNAs and proteins relevant to bladder cancer. A recent report presents a systems biology approach for the analysis of the muscle invasive bladder cancer dataset contained in The Cancer Genome Atlas [37]. Another database for bladder cancer that provides molecular features in regard to miRNAs identified in literature is also available [38,39]. In addition, a user-friendly analysis tool is also available and allows the evaluation of gene expression profiles determined by microarray studies across bladder cancer patients [40].

Though omics profiling has provided an abundance of data, technical boundaries involving incompleteness of the individual molecular datasets together with the static representation of cellular activity limit the insights on molecular processes and their interaction dynamics [41-43].

A large number of biological pathway analysis tools are available, including KEGG [44], PANTHER [45], REACTOME [46] and AmiGO [47] described in PathGuide (http://www.pathguide.org/), and allow detection of significant metabolic and signaling pathways. Albeit there are several well curated and reliable pathway database resources [48], significant efforts have been taken to expand biological pathway coverage beyond any single pathway data source. This is frequently carried out by integrating different pathway sources to build high quality integrative pathway models. However, biological data integration from heterogeneous sources has been challenging due to variability at the syntactic and semantic level. Syntactic variability is due to heterogeneity of molecular feature and pathway data formats, representation schemas and retrieval methods. Semantic variability is due to incompatible pathway names, signaling event representations and molecular identifiers. For example, different pathway databases may choose to provide information on post-translation modifications, interacting proteins within a complex, or cellular location. Hence all these limitations have inhibited the growth of high quality integrative pathway models [49-51].

Previous omics studies report biomarkers associated with bladder cancer, and therapeutic targets that could allow development of personalized therapies [52-56]. However, the information gathered from these large number of omics experiments is not fully exploited, as the datasets generated are either scattered in many publications and databases or held in supplementary data files.

Therefore, the aim of this thesis was primarily to characterize muscle invasive bladder carcinoma on a molecular level by incorporating scientific literature and omics data. In addition, the objective was to evaluate the performance of pathway enrichment analysis obtained from two bioinformatics tools ImPAla and ClueGO. Thirdly, the goal was to use bioinformatics and systems biology approaches in order to identify significant molecular pathways in age-associated diseases.

Methods

Data sources for characterizing bladder cancer pathophysiology

In order to retrieve molecular features associated with muscle invasive bladder cancer, "National Center for Biotechnology Information" (NCBI) PubMed, Web of Science, Google Scholar and the omics repositories Gene Expression Omnibus (GEO) [31] and ArrayExpress [30] were manually queried. Since the study involved the molecular characterization of muscle invasive bladder cancer, the criteria for selecting keywords depended specifically in regard to the muscle invasive phenotype. The keywords for the literature search included "bladder OR urothelial OR transitional cell" AND "neoplasm OR tumor OR carcinoma" AND "muscle" AND "invas* OR aggress* OR progress* OR inflammation" (Database version of June, 2015). The list of publications relevant to muscle invasion in bladder cancer was isolated from the complete list of retrieved papers. Publications were further screened for adequacy in sample size (at least 50 samples included in study design), magnitude of differential abundance (>2-fold change for proteomics, transcriptomics, metabolomics and miRNAs), False Discovery Rate <0.1 for mutations, p-values<0.05 for methylation and -omics studies, in addition to the specific phenotypic conditions; T2a/b, T3a/b, T4a/b. The muscle invasive bladder cancer specific molecular features retrieved from the publications comprised of various sources such as DNA-mutations, DNA-methylation, mRNAs, miRNAs, proteins (immunohistochemistry validations and proteomics) and metabolites. The features were then combined for further systems biology analysis.

Protein-protein interactions

In order to retrieve protein-protein interaction information for the muscle invasive bladder cancer associated proteins, protein-protein interaction databases including IntAct [57], BioGRID [58], STRING [59] and Reactome [46] were queried. By downloading the protein interaction information contained in each database, an integrated database was developed in order to contain all available non-redundant human protein-protein interaction information. This unique human protein list along with the protein-protein interactions information were then downloaded into the Cytoscape [60-61] software to yield the human interactome based on experimental evidence. The proteins relative to muscle invasive bladder cancer were then selected from this human interactome and were put on a separate list. Muscle invasive bladder cancer proteins that had at least one binding partner in the list of muscle invasive bladder cancer specific proteins were retained in order to generate the muscle invasive bladder cancer specific interactome.

Pathway enrichment

To retrieve molecular pathway information for muscle invasive bladder cancer, proteins from the muscle invasive bladder cancer interactome were subjected to pathway enrichment analysis. This analysis used two additional sub-applications from Cytoscape; ClueGO [62] and CluePedia [63]. The statistical criterion used in generating molecular pathways included a two-sided hypergeometry test. Information from pathway databases such as Kyoto Encyclopedia of Genes and Genomes [44]and Reactome [46] databases was used in retrieving significant pathways associated to muscle invasive bladder cancer with a Bonferroni corrected p-value<0.05. In addition, the list of pathways was inspected manually and redundant pathway-terms were combined hereby. The filtered list of pathway-terms was then divided into previously known pathways and novel findings in the context of muscle invasive bladder carcinoma.

Comparison of pathway enrichment tools ClueGO and ImPAla

To evaluate the performance of pathway enrichment, ClueGO and ImPAla enrichment tools were compared. ClueGO provides an advantage to perform cluster comparisons for pathway enrichment and allows the option to separately input up and down regulated molecules in the software. In addition, ClueGO provides an optional redundancy reduction feature ("Fusion") to assess Gene Ontology (GO) terms that share similar associated features in a parent-child relation. This option was selected in the ClueGO pathway enrichment analysis to eliminate the redundant pathway terms. In contrast, ImPAla does not provide an option of redundancy reduction for pathway terms. The pathway databases selected for enrichment were KEGG. The statistical selection criterion taken into account for the enrichment analysis was the corrected for multiple comparisons p-value<0.05. The overlap assessment between the pathway outputs was performed manually.

Results

Publication 1: protein interactome for muscle invasive bladder cancer

In the present study the bioinformatics model of proteomic changes in bladder cancer involved integrating available public domain data sets from PubMed, Google scholar and Web of science in the context of bladder muscle-invasive carcinoma (Figure 1).

In the first step of this analysis, the data collected was filtered using statistical measurements to include fold-change values, p-values and sample size for the specific phenotype of muscle invasive bladder cancer. The molecular features were then incorporated into systems biology tools to model protein-protein interaction networks, and further mapping them to biological molecular pathways.

The results revealed fifteen pathways as being affected in the progressive disease. Eleven from these pathways were reported previously and four pathways were novel findings in the context of muscle invasive bladder cancer (Figure 2). The fact that the majority of pathways identified by our analysis are involved in muscle invasive bladder cancer

supports the validity of our approach. Moreover, the four novel pathways revealed by our analysis could be validated experimentally and offer new targets for biomarker discovery or therapy of muscle invasive bladder cancer (Table 1).

Publication 2: Comparison of ClueGO and ImPAla for integrated pathway enrichment analysis

In this study, the total number of Kyoto Encyclopaedia of Genes and Genomes (KEGG) and Reactome pathway terms obtained from ClueGO was 292. ImPAla produced 471 pathways (Table 2).

By comparing the pathway results, 152 pathway terms exactly overlapped in ClueGO and ImPAla. 137 pathway terms from ClueGO were highly similar to 251 ImPAla pathway terms. Therefore, the total calculated overlap of pathways between the two tools equalled to 289 ClueGO pathways that correspond to 403 ImPAla enriched pathways. In addition, the software also produced unique pathway terms. There were 3 unique pathways from the total 292 ClueGO pathway terms whereas 68 pathways were unique from ImPAla. Both the enrichment tools yielded redundancy in the output results, however results from ImPAla were characterized by higher redundancies in pathway terms (for e.g. the pathway terms "DNA replication", "synthesis of DNA"). Moreover, from the unique set of 68 ImPAla pathway terms, 12 pathway terms were not related in the context of bladder cancer. Some of these pathways include alcoholism, amphetamine addiction, inflammatory bowel disease (IBD), malaria, viral myocarditis and prion diseases. On the contrary, the 3 unique pathways obtained by ClueGO were relevant to bladder cancer. It was also noted that the overlapping pathway terms from ImPAla and ClueGO contained pathway names that are not relevant in the context of bladder cancer. These common terms totalled to 34 ImPAla and 30 ClueGO pathway terms. The common pathway terms included oocyte meiosis, tuberculosis, type II diabetes mellitus, circadian clock and shigellosis. The comparison of significant overlapping pathways obtained from ClueGO and ImPAla is represented as a Venn diagram in (Figure 3).

Publication 3: Identification of urinary age-specific peptides in a healthy population

In the last section of the thesis, the bioinformatics approach used in the molecular characterization of bladder cancer muscle invasion was applied to a peptidomics dataset relevant to ageing associated disorders.

Ageing is a complex systemic process and "omics" approaches aiming at the study of multiple features simultaneously have been applied to unravel novel underlying molecular processes [64]. Proteomics studies confirmed that oxidative stress occurs ubiquitously during ageing [65]. However, a shortcoming in most of these studies was the use of animal models. The scarcity of human subjects can be largely attributed to the inability in obtaining appropriate tissue samples. Thus, a way forward in ageing research could be the investigation of readily available body fluids.

In this study, a small-scale urinary peptidome of 324 healthy individuals was investigated. The patients aged between 2 to 73 years and showed the feasibility to obtain high-resolution molecular information readily available from body fluids such as urine [66].

Subsequently, the urinary peptidome profiles of 11,560 individuals were investigated in an attempt to identify specific ageing-associated alterations and to elucidate pathological derailment in normal ageing (Table 3). The results obtained showed perturbations mainly in collagen homeostasis, trafficking of toll-like receptors and endosomal pathways that were significantly associated to the healthy ageing group. Moreover, degradation of insulin-like growth factor-binding proteins was a unique identification deregulated in pathological ageing cohorts (Figures 4a & 4b).

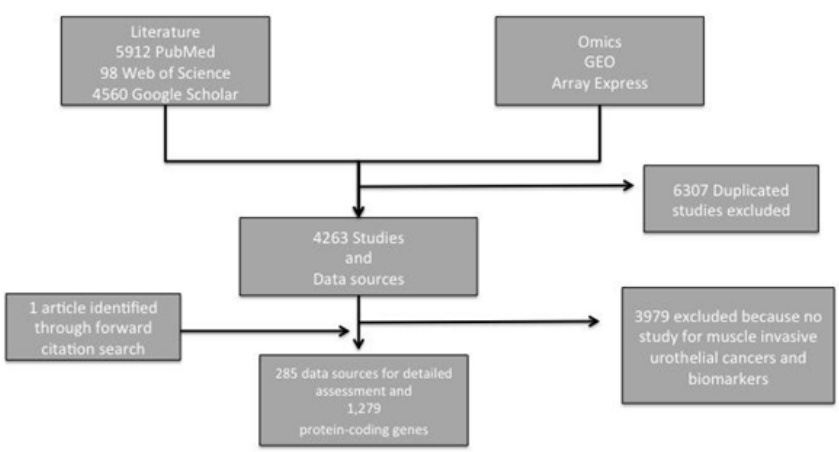

Figure 1: Data assembly workflow.

PubMed, Google Scholar and Web of Science literature analysis and Omics data source screening for the systems based analysis in muscle invasive bladder cancer.

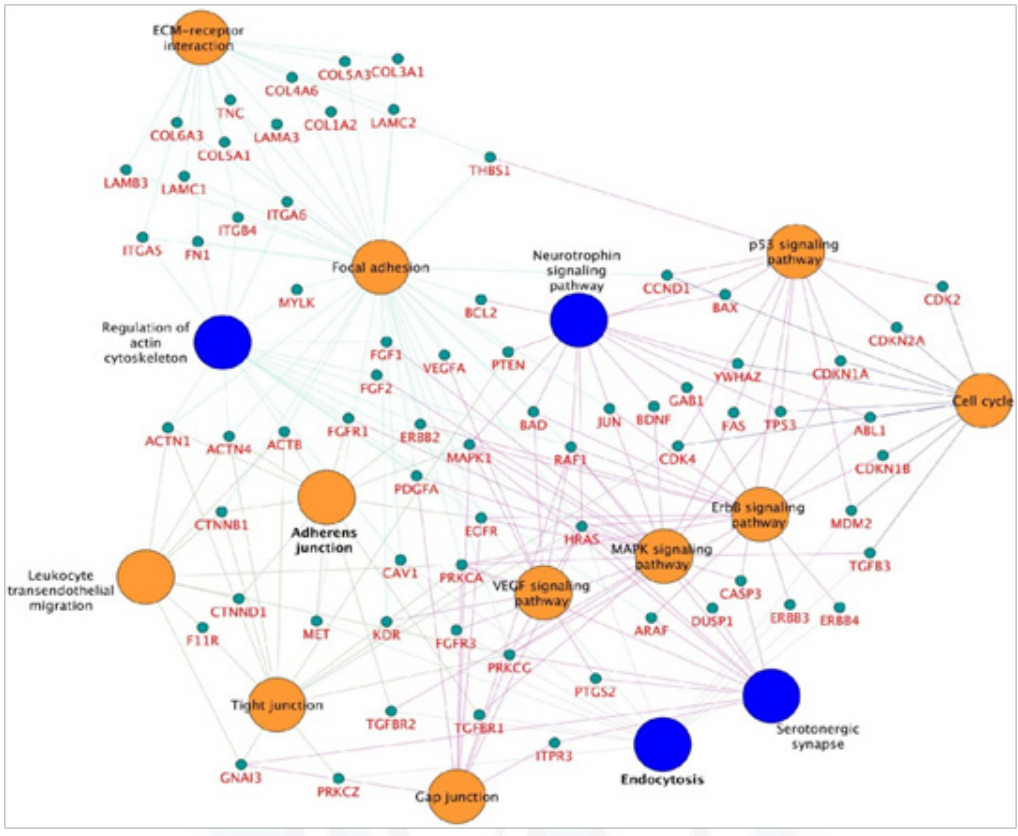

Figure 2: Muscle Invasive Bladder carcinoma interactome.

Nodes (circles) in orange denote pathways identified as relevant in both literature and enrichment analysis, nodes in blue depicts pathways of relevance according to enrichment analysis.

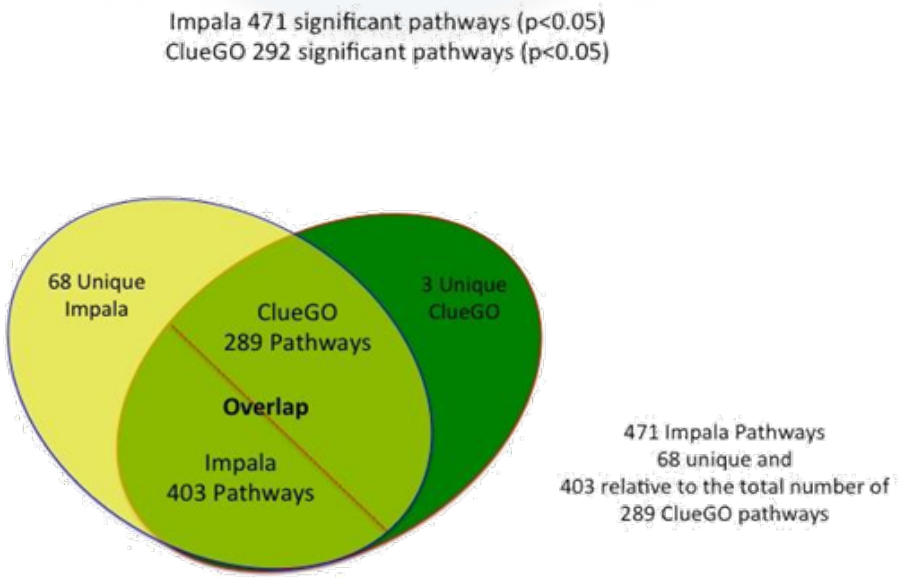

Figure 3: Venn Diagram represents the overlap of pathway terms between ClueGO and ImPAla software. All pathways enriched are selected based on p-value < 0.05

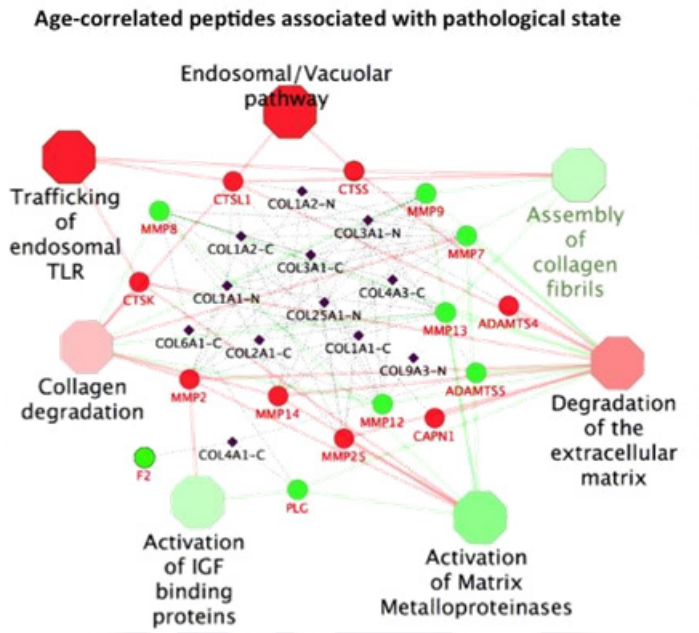

Figure 4a: Molecular pathways associated with pathological ageing.

The network represents each pathway as individual octagonal node, while the circled nodes represent the predicted proteases that were targeted from the identified urinary peptides denoted in purple diamond nodes. The edges (links) between pathways denote an approximation of biological interaction between the pathways based on the cross-pathway feature overlap.

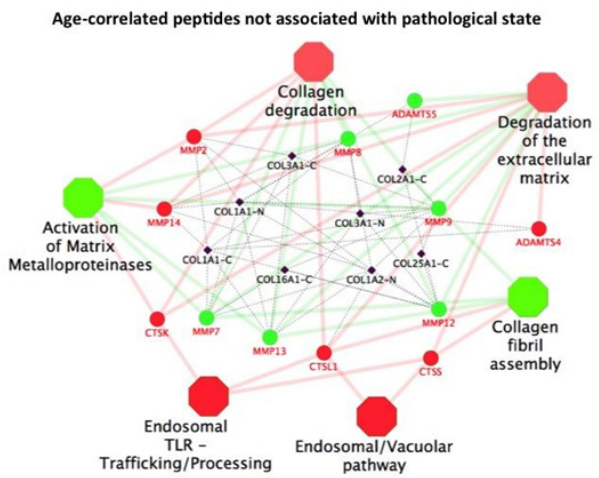

Figure 4b: Molecular pathways associated with normal ageing.

Legends for the diamond nodes with a suffix of "-C/N" represent the peptide's cleavage site; i.e. "-C" for C-terminus and "-N" for the N-terminus.

Table 1: Abbreviations for protein coding genes described in Figure 2.

LAMB3 – Laminin subunit beta-3	LAMA3 - Laminin subunit alpha-3
COL6A3 – Collagen alpha-3(VI)	COL1A2 - Collagen alpha-2(I)
chain TNC - Tenascin	chain LAMC2 - Laminin
COL4A6 - Collagen alpha-6(IV) chain	gamma-2 ITGA5 - Integrin
COL5A3 - Collagen alpha-3(V) chain	FN1 - Fibronectin
COL3A1 - Collagen alpha-1(III) chain	ITGB4 - Integrin beta-4
LAMC1 - Laminin subunit gamma-1	ITGA6 - Integrin alpha
COL5A1 - Collagen alpha-1(V) chain	MYLK - Myosin light chain kinase, smooth
THBS1 - Thrombospondin-1	muscle ACTN1 - Alpha-actinin-
ACTN4 - Alpha-actinin-4	ACTB - Actin, cytoplasmic 1 FGF2
FGFR1-Fibroblast growth factor receptor 1	- Fibroblast growth factor 2
FGF1 - Fibroblast growth factor 1	VEGFA - Vascular endothelial growth factor
BCL2 - Apoptosis regulator Bcl-2	PTEN - Phosphatidylinositol 3,4,5-trisphosphate 3-phosphatase and dual-specificity protein phosphatase PTEN
CCND1 - G1/S-specific cyclin-D1	BAX - Apoptosis regulator BAX
CTNNB1 - Catenin beta-1	ERBB2 - Receptor tyrosine-protein kinase erbB-2
PDGFA - Platelet-derived growth factor subunit A	MAPK1 - Mitogen-activated protein kinase 1
BAD - Bcl2-associated agonist of cell death	CTNND1 - Catenin delta-1
F11R - Junctional adhesion molecule A	MET - Hepatocyte growth factor receptor
KDR - Vascular endothelial growth factor receptor 2	CAV1 - Caveolin-1
JUN - Transcription factor AP-1	BDNF - Brain-derived neurotrophic factor
GAB1 - GRB2-associated-binding protein 1	YWHAZ - 14-3-3 protein zeta/del
CDKN1A - Cyclin-dependent kinase inhibitor 1	CDKN2A - Cyclin-dependent kinase inhibitor 2A
CDK2 - Cyclin-dependent kinase 2	GNAI3 - Guanine nucleotide-binding protein G(k) subunit alpha
TGFBR2 - TGF-beta receptor type-2	FGFR3 - Fibroblast growth factor receptor 3
PRKCA - Protein kinase C alpha type	HRAS - GTPase Hras
CDK4 - Cyclin-dependent kinase 4	FAS - Tumor necrosis factor receptor superfamily member 6
TP53 - Cellular tumor antigen p53	ABL1 - Tyrosine-protein kinase ABL1
CDKN1B - Cyclin-dependent kinase inhibitor 1B	MDM2 - E3 ubiquitin-protein ligase Mdm2
TGFB3 - Transforming growth factor beta-3	ERBB4 - Receptor tyrosine-protein kinase erbB-4
ERBB3 - Receptor tyrosine-protein kinase erbB-3	CASP3 - Caspase-3
DUSP1 - Dual specificity protein phosphatase 1	ARAF - Serine/threonine-protein kinase A-Raf
PTGS2 - Prostaglandin G/H synthase 2	PRKCG - Protein kinase C gamma type
TGFBR1 - TGF-beta receptor type-1	ITPR3 - Inositol 1,4,5-trisphosphate receptor type 3
PRKCZ - Protein kinase C zeta type	TGFBR2 - TGF-beta receptor type-2

Table 2: General information for the results obtained from the pathway enrichment analysis.

Software	Availability	User Input	p-value Correction Method	Total Pathway Output
ClueGO	Cytoscape plugin	435 entries	Bonferroni	292
ImPAla	Web-based	435 entries	Benjamini Hochberg	471

Table 3: Abbreviations for protease and peptide names in figure 4a and 4b.

Protease	Peptides
MMP2 - 72 kDa type IV collagenase	COL1A2-N - Collagen alpha-2(I) chain
MMP14 - Matrix metalloproteinase-14	COL3A1-N - Collagen alpha-1(III) chain
MMP8 - Neutrophil collagenase	COL1A2-C - Collagen alpha-2(I) chain
ADAMTS5 - A disintegrin and metalloproteinase with thrombospondin motifs 5	COL3A1-C - Collagen alpha-1(III) chain
MMP9 - Matrix metalloproteinase-9	COL4A3-C - Collagen alpha-3(IV) chain
CTSK - Cathepsin K	COL1A1-N - Collagen alpha-1(I) chain
MMP7 - Matrilysin	COL25A1-N - Collagen alpha-1(XXV) chain
MMP13 - Collagenase 3	COL6A1-C - Collagen alpha-1(VI) chain
CTSL1 - Cathepsin L1	COL2A1-C - Collagen alpha-1(II) chain
CTSS - Cathepsin S	COL1A1-C - Collagen alpha-1(I) chain
MMP12 - Macrophage metalloelastase	COL9A3-N - Collagen alpha-3(IX) chain
ADAMTS4 - A disintegrin and metalloproteinase with thrombospondin motifs 4	COL4A1-C - Collagen alpha-1(IV) chain
PLC - 1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase gamma-1	COL25A1-C - Collagen alpha-1(XXV) chain
F2 - Prothrombin	COL16A1
MMP25 - Matrix metalloproteinase-25 CAPN1 - Calpain-1 catalytic subunit	

Discussion

For an early diagnosis and successful targeted treatment, molecular characterization of individual disease phenotypes and prediction of novel biomarkers is essential. By the use of network biology approaches such as bioinformatics analysis tools, pathway databases, and statistical criteria, a comprehensive understanding of the complex molecular mechanisms in genetic disorders can be achieved. This would be advantageous in better prognosis and early clinical intervention of the individual phenotype. Based on this background, the aim of this thesis was to characterize muscle invasive bladder carcinoma on a molecular level by incorporating signatures from scientific literature screening and omics profiling. The characterization was achieved by integrating collected data to perform protein-protein interactions and pathway enrichment analysis.

In first part, of this study, automated data retrieval from the literature resulted in a first collection of molecular features associated with muscle invasive bladder cancer, and, combination with omics profiling data, allowed the creation of a mechanistic (pathway) map linked to muscle invasive bladder cancer. By deriving bladder cancer-associated protein coding genes on the basis of such pathway maps provides a systematic foundation for experimental analysis regarding association with development of muscle-invasive disease.

In the second part of the thesis, the performance of pathway enrichment was compared for significant pathway outputs yielded from ClueGO and ImPAla in the context to bladder cancer. Only two widely used and up-to-date pathway database resources, KEGG and Reactome were selected. Adding more pathway databases in the analysis would introduce higher redundancy in pathway outputs. In regard to manually updating database sources, the ClueGO application allows users to update individual pathway database source in order to obtain latest data whereas ImPAla is an omics-integration focusing towards metabolomics integration and pathway enrichment application that contains the latest update of January 2015 [67]. In addition, ImPAla also allows the incorporation of differential expression information for molecules such as magnitude of differentially expressed fold changes and

multiple comparison corrected p-values. Nevertheless, ImPAIa does not offer the option to input separately up and down regulated genes and does not predict activation/deactivation of an affected pathway in contrast to ClueGO. In addition, ClueGO provides users to analyze different omics datasets such as genes, mRNAs, proteins, single nucleotide polymorphisms, metabolites and miRNAs. This gives the advantage in using one analysis and visualization tool for all high-throughput sequencing and profiling experiments. Furthermore, having a single analysis tool also helps to prevent errors due to compatibility when transferring data between different software applications. Therefore, ClueGO is preferable to ImPAIa for pathway enrichment and in the comprehensive characterization of molecular diseases.

In the last section of the thesis, the analysis of the urinary peptidome of ageing-associated peptides was detected. Differentially expressed age-associated peptides were identified using capillary electrophoresis coupled with mass spectrometry (CE-MS). The Proteasix software was then used in order to predict proteases that cleaved the identified urinary age associated peptides [68]. The generated data) were then subjected to systems biology and bioinformatics approaches such as pathway enrichment analysis in order to characterize molecular pathways that were associated with normal and pathological ageing. Findings demonstrated that with the help of appropriate peptidomics technologies, urine could be used as a powerful biological fluid in ageing research.

Conclusion

In conclusion, it is demonstrated in this thesis that by using bioinformatics and systems biology methodologies a better understanding of complex molecular mechanisms such as tumor invasion in bladder cancer is possible. In addition, ClueGO pathway enrichment tool has better performance than ImPAIa in pathway enrichment analysis since the pathway output is less redundant and contains all the biologically significant information. The combination of a systems biology approach and individual proteins biochemical features offers a thorough molecular description of muscle invasive bladder cancer.

References

1. Grossman HB, Natale RB, Tangen CM, Speights VO, Vogelzang NJ, et al. (2003) Neoadjuvant chemotherapy plus cystectomy compared with cystectomy alone for locally advanced bladder cancer. N Engl J Med 349(9): 859-866.
2. Kiss B, Skuginna V, Fleischmann A, Bell RH3, Collins C, et al. (2015) Bcl-2 predicts response to neoadjuvant chemotherapy and is overexpressed in lymph node metastases of urothelial cancer of the bladder. Urol Oncol 33(4): 166.e1-8.
3. Ru Y, Dancik GM, Theodorescu D (2011) Biomarkers for prognosis and treatment selection in advanced bladder cancer patients. Curr Opin Urol 21(5): 420-427.
4. Henning A, Wehrberger M, Madersbacher S, Pycha A, Martini T, et al. (2013) Do differences in clinical symptoms and referral patterns contribute to the gender gap in bladder cancer? BJU Int 112(1): 68-73.
5. Raman JD, Shariat SF, Karakiewicz PI, Lotan Y, Sagalowsky AI, et al. (2011) Does preoperative symptom classification impact prognosis in patients with clinically localized upper-tract urothelial carcinoma managed by radical nephroureterectomy? Urol Oncol 29(6): 716-723.
6. Cauberg Evelyne CC, de la Rosette JJMCH, de Reijke TM (2011) Emerging optical techniques in advanced cystoscopy for bladder cancer diagnosis: A review of the current literature. Indian J Urol 27(2): 245-251.
7. Niegisch G, Lorch A, Droller MJ, Lavery HJ, Stensland KD, et al. (2013) Neoadjuvant chemotherapy in patients with muscle-invasive bladder cancer: which patients benefit? Eur Urol 64(3): 355-357.
8. Sternberg CN, Apolo AB (2014) Everything old is new again! Neoadjuvant chemotherapy in the treatment of muscle-invasive bladder cancer. J Clin Oncol 32(18): 1868-1870.
9. Knowles MA, Hurst CD (2014) Molecular biology of bladder cancer: new insights into pathogenesis and clinical diversity. Nat Rev Cancer 15(1): 25-41.
10. Anastasiadis A, de Reijke TM (2012) Best practice in the treatment of nonmuscle invasive bladder cancer. Ther Adv Urol 4(1): 13-32.
11. Mitra AP, Cote RJ (2009) Molecular pathogenesis and diagnostics of bladder cancer. Annu Rev Pathol 4: 251-285.
12. Vishnu P, Mathew J, Tan WW (2011) Current therapeutic strategies for invasive and metastatic bladder cancer. Onco Targets Ther 4: 97-113.
13. Mitra AP, Datar RH, Cote RJ (2005) Molecular staging of bladder cancer. BJU Int 96(1): 7-12.
14. Choi W, Porten S, Kim S, Willis D1, Plimack ER, et al. (2014) Identification of distinct basal and luminal subtypes of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy. Cancer Cell 25(2): 152-165.
15. Lowe SW, Ruley HE, Jacks T, Housman DE (1993) p53-dependent apoptosis modulates the cytotoxicity of anticancer agents. Cell 74(6): 957-967.
16. McConkey DJ, Lee S, Choi W, Tran M, Majewski T, et al. (2010) Molecular genetics of bladder cancer: Emerging mechanisms of tumor initiation and progression. Urol Oncol 28(4): 429-440.
17. Bayraktar Z, Gurbuz G, Taşci AI, Sevin G (2001) Staging error in the bladder tumor: the correlation between stage of TUR and cystectomy. Int Urol Nephrol 33(4): 627- 629.
18. Lerner SP, Skinner E, Skinner DG (1992) Radical cystectomy in regionally advanced bladder cancer. Urol Clin North Am 19(4): 713-723.
19. Kim PH, Cha EK, Sfakianos JP, Iyer G2, Zabor EC, et al. (2015) Genomic predictors of survival in patients with high-grade urothelial carcinoma of the bladder. Eur Urol 67(2):198-201.
20. Emmert-Streib F, Abogunrin F, de Matos Simoes R, Duggan B, Ruddock MW, et al. (2013) Collectives of diagnostic biomarkers identify high-risk subpopulations of hematuria patients: exploiting heterogeneity in large-scale biomarker data. BMC Med 11: 12.
21. Mayer P, Mayer B, Mayer G (2012) Systems biology: building

a useful model from multiple markers and profiles. Nephrol Dial Transplant 27(11): 3995-4002.

22. Kreeger PK, Lauffenburger DA (2010) Cancer systems biology: a network modeling perspective. Carcinogenesis 31(1): 2-8.

23. Chihara Y, Kanai Y, Fujimoto H, Sugano K, Kawashima K, et al. (2013) Diagnostic markers of urothelial cancer based on DNA methylation analysis. BMC Cancer 13: 275.

24. Weiss GJ, Liang WS, Demeure MJ, Kiefer JA, Hostetter G, et al. (2013) A pilot study using next-generation sequencing in advanced cancers: feasibility and challenges. PLoS One 8(10): e76438.

25. Mischak H, Schanstra JP (2011) CE-MS in biomarker discovery, validation, and clinical application. Proteomics Clin Appl 5(1-2): 9-23.

26. Alberice JV, Amaral AFS, Armitage EG, Lorente JA, Algaba F, et al. (2013) Searching for urine biomarkers of bladder cancer recurrence using a liquid chromatography-mass spectrometry and capillary electrophoresis-mass spectrometry metabolomics approach. J Chromatogr A 1318: 163-170.

27. Pignot G, Cizeron-Clairac G, Vacher S, Susini A, Tozlu S, et al. (2013) microRNA expression profile in a large series of bladder tumors: Identification of a 3-miRNA signature associated with aggressiveness of muscle-invasive bladder cancer. Int J Cancer 132(11): 2479-2491.

28. Liu Y, Noon AP, Aguiar Cabeza E, Shen J, Kuk C, et al. (2014) Next-generation RNA sequencing of archival formalin-fixed paraffin-embedded urothelial bladder cancer. Eur Urol 66(6): 982-986.

29. Hanash S (2003) Disease proteomics. Nature 422(6928): 226-232.

30. Rustici G, Kolesnikov N, Brandizi M, Burdett T, Dylag M, et al. (2013) ArrayExpress update--trends in database growth and links to data analysis tools. Nucleic Acids Res 41(Database issue): D987-D990.

31. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, et al. (2013) NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res 41(Database issue): D991-D995.

32. Muthusamy B, Thomas JK, Prasad TSK, Pandey A (2013) Access guide to human proteinpedia. Curr Protoc Bioinformatics. Chapter 1: Unit 1.21.

33. Prasad TSK, Kandasamy K, Pandey A (2009) Human Protein Reference Database and Human Proteinpedia as discovery tools for systems biology. Methods Mol Biol 577: 67-79.

34. Stelzer G, Dalah I, Stein TI, Satanower Y, Rosen N, et al. (2011) In-silico human genomics with GeneCards. Hum Genomics 5(6): 709-717.

35. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, et al. (2004) ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia 6(1): 1-6.

36. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res 33: D514-D517.

37. Weinstein JN, Akbani R, Broom BM, et al. (2014) Comprehensive molecular characterization of urothelial bladder carcinoma. Nature 507(7492): 315-322.

38. Zabolotneva AA, Zhavoronkov AA, Shegay P V, Gaifullin NM, Alekseev BY, et al. (2013) A systematic experimental evaluation of microRNA markers of human bladder cancer. Front Genet 4: 247.

39. Zabolotneva AA, Zhavoronkov A, Garazha A V, Roumiantsev SA, Buzdin AA (2012) Characteristic patterns of microRNA expression in human bladder cancer. Front Genet 3: 310.

40. Dancik GM (2015) An online tool for evaluating diagnostic and prognostic gene expression biomarkers in bladder cancer. BMC Urol 15(1): 59.

41. Mayer G, Heinze G, Mischak H, Hellemons ME, Heerspink HJ, et al. (2011) Omics-bioinformatics in the context of clinical data. Methods Mol Biol 719: 479-497.

42. Abu-Asab MS, Chaouchi M, Alesci S, Galli S, Laassri M, et al. (2011) Biomarkers in the age of omics: time for a systems biology approach. OMICS 15(3): 105-112.

43. De Las Rivas J, Fontanillo C (2010) Protein-protein interactions essentials: key concepts to building and analyzing interactome networks. PLoS Comput Biol 6(6): e1000807.

44. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M (2004) The KEGG resource for deciphering the genome. Nucleic Acids Res 32(Database issue): D277-D280.

45. Mi H, Muruganujan A, Thomas PD (2013) PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res 41(Database issue): D377-D386.

46. Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, et al. (2011) Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res 39(Database issue): D691-D697.

47. Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, et al. (2009) AmiGO: online access to ontology and annotation data. Bioinformatics 25(2): 288-289.

48. Frantzi M, Bhat A, Latosinska A (2014) Clinical proteomic biomarkers: relevant issues on study design & technical considerations in biomarker development. Clin Transl Med 3(1): 7.

49. Krishnamurthy L, Nadeau J, Ozsoyoglu G, Ozsoyoglu M, Schaeffer G, et al. (2003) Pathways database system: an integrated system for biological pathways. Bioinformatics 19(8): 930-937.

50. Stein LD (2003) Integrating biological databases. Nat Rev Genet 4(5): 337-345.

51. Philippi S, Köhler J (2006) Addressing the problems with life-science databases for traditional uses and systems biology. Nat Rev Genet 7(6): 482-488.

52. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, et al. (2011) COSMIC: mining complete cancer genomes in the

Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res 39(Database issue): D945-D950.

53. Gui Y, Guo G, Huang Y, Hu X, Tang A, et al. (2011) Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. Nat Genet 43(9): 875-878.

54. Hurst CD, Platt FM, Taylor CF, Knowles MA (2012) Novel tumor subgroups of urothelial carcinoma of the bladder defined by integrated genomic analysis. Clin Cancer Res 18(21): 5865-5877.

55. Lindgren D, Sjödahl G, Lauss M, Staaf J, Chebil G, et al. (2012) Integrated genomic and gene expression profiling identifies two major genomic circuits in urothelial carcinoma. PLoS One 7(6): e38863.

56. Williams S V, Hurst CD, Knowles MA (2013) Oncogenic FGFR3 gene fusions in bladder cancer. Hum Mol Genet 22(4): 795-803.

57. Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, et al. () The IntAct molecular interaction database in 2012. Nucleic Acids Res 40(Database issue): D841-D846.

58. Stark C, Breitkreutz B-J, Chatr-Aryamontri A, Boucher L, Oughtred R, et al. (2011) The BioGRID Interaction Database: 2011 update. Nucleic Acids Res 39: D698-D704.

59. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, et al. (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 41(Database issue): D808-D815.

60. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13(11): 2498-2504.

61. Saito R, Smoot ME, Ono K, Ruscheinski J, Wang PL, et al. (2012) A travel guide to Cytoscape plugins. Nat Methods 9(11): 1069-1076.

62. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, et al. (2009) ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics 25(8): 1091-1093.

63. Bindea G, Galon J, Mlecnik B (2013) CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data. Bioinformatics 29(5): 661-663.

64. Calabrese V, Dattilo S, Petralia A, Parenti R, Pennisi M, et al. (2015) Analytical approaches to the diagnosis and treatment of aging and aging-related disease: redox status and proteomics. Free Radic Res 49(5): 511-524.

65. Nkuipou-Kenfack E, Koeck T, Mischak H, Pich A4, Schanstra JP, et al. (2014) Proteome analysis in the assessment of ageing. Ageing Res Rev 18: 74-85.

66. Zürbig P, Decramer S, Dakna M, Jantos J, Good DM, et al. (2009) The human urinary proteome reveals high similarity between kidney aging and chronic kidney disease. Proteomics 9(8): 2108-2117.

67. Kamburov A, Cavill R, Ebbels TMD, Herwig R, Keun HC (2011) Integrated pathway-level analysis of transcriptomics and metabolomics data with IMPaLA. Bioinformatics 27(20): 2917-2918.

68. Klein J, Eales J, Zürbig P, Vlahou A, Mischak H, et al. (2013) Proteasix: A tool for automated and large-scale prediction of proteases involved in naturally occurring peptide generation. Proteomics 13(7): 1077-1082.