INSILICO MODELING OF BLADDER CANCER ASSOCIATED PROTEIN

Sayed zafar*, Surya Chandra Rejinthala

Sreenidhi Institute of Science & Technology, Yamnampet, Ghatkesar Hyderabad - 501 301 Telangana, India.

Correspondence to Author:
Sayed zafar
Sreenidhi Institute of Science & Technology, Yamnampet, Ghatkesar Hyderabad - 501 301 Telangana, India.

E-mail: sayed.zafar768@gmail**.com**

Abstract— Bladder cancer is one of several types of cancer arising from the tissues of the urinary bladder. It is a disease in which cells grow abnormally and have the potential to spread to other parts of the body[1] . It has been suggested that mutations at BLCAP BC10 is associated with BLCAP (bladder cancer associated protein) over expression induces S phase arrest and apoptosis in the tissue which cause urothelial carcinoma squamous cell carcinoma and adenocarcinoma thus we can say Bladder cancer-associated protein, a potential prognostic biomarker in human bladder cancer**.** Deletion of GSTM1 gene has a modest increase in risk of bladder cancer. GSTM1 gene product glutathione S-transferase M1 (GSTM1) participates in the detoxification process of carcinogens such as polycyclic aromatic hydrocarbons found in cigarette smoke[2]. More than 60,000 men and 16,000 women are diagnosed with bladder cancer each year. Smoking can only partially explain this higher incidence in men. One other reason is that the androgen receptor, which is much more active in men than in women, plays a major part in the development of the cancer. The present project aims to predict the structure of BLCAP using comparative modeling method using bioinformatics software tools

Keywords:— urinary bladder , carcinoma , GSTM1 , prognostic , BLCAP , urothelial , glutathione

1 INTRODUCTION

Bladder cancer is the ninth most common cause of cancer worldwide for both sexes combined and the second most common malignancy of the genitourinary tract. Bladder neo-plasias account for about 5% of all diagnosed cancers, affecting one in 4,000 people. An estimated 357,000 bladder cancer cases occurred in 2002 with more than 145,000 deaths reported in the same year[3]

Malignancy of the urinary bladder comprises a large variety of histologically hetero-geneous tumor types arising predominantly in the epithelial lining of the urinary bladder (urothelium) and the ureters. Tu-mor types of the urothelium include urothelial carcinomas (UCs),squamous cell carcinomas, and adenocarcinomas as well as other less frequent lesions

UCs account for more than 90% of bladder carcinomas and comprise a wide spec-trum of lesions ranging in clinical severity from superficial bladder cancer to muscle-invasive and metastatic disease with the latter having a poor prognosis. Clinical management of the superficial form of the disease is currently done by endoscopic resection of the tumor supplemented with instil-lations of cytotoxic/cytostatic agents. Although intravesical instillations are widely used to avoid recurrences and even progression, up to 80% of patients with superficial bladder cancer lesions will recur, and of these, 25% will progress to invasive disease

Current prognostic parameters such as grade and stage, multifocality of carcinomas, and lymph node status cannot predict with certainty the long term outcome of bladder cancer, and as a result, there is a pressing need to identify markers that may be associated with bladder cancer progression and predict tumor behavior.

Recently, it was identified bladder cancer-associated protein (BLCAP) as a novel potential biomarker based on a limited study,Loss of BLCAP mRNA expression correlates with the invasive potential of UCs, and subsequent studies by others in several cancer types, such as cervical and renal as well as human tongue carcinoma and osteosarcoma , have all shown differential expression of this protein in cancer. Moreover, overexpression of BLCAP in the human cervical cancer HeLa cell line and tongue carcinoma Tca8113 cell line has been shown to inhibit cell growth and to induce apoptosis, suggesting a role for this protein in the regulation of these cellular processes. BLCAP is a small (87-amino acid), evolutionary conserved protein with no homology to any know protein, and its cellular function is largely unknown[4]

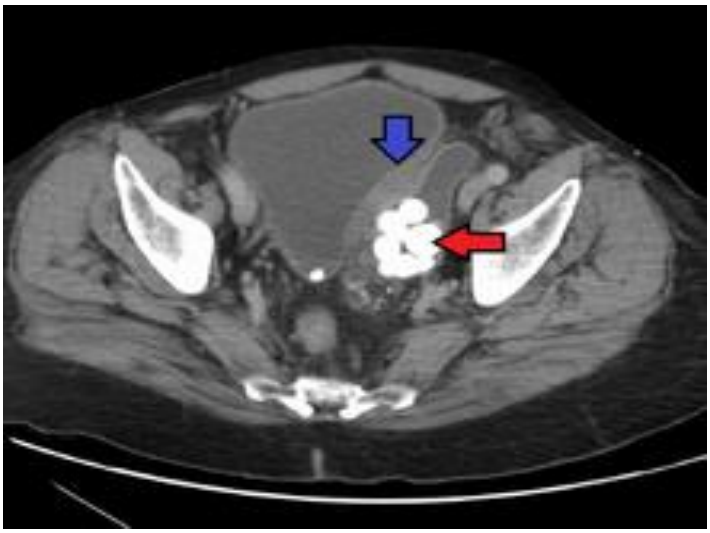


Fig. 1 Transitional cell carcinoma of the bladder. The white in the bladder is contrast.

TYPE

A to I RNA editing is catalyzed by a family of adenosine deaminases acting on RNA (ADARs) that specifically recognize adenosines within double-stranded regions of pre-mRNAs and deaminate them to inosine. Inosines are recognised as guanosine by the cells translational machinery. There are three members of the ADAR family ADARs 1-3 with ADAR 1 and ADAR 2 being the only enzymatically active members.ADAR3 is thought to have a regulatory role in the brain. ADAR1 and ADAR 2 are widely expressed in tissues while ADAR 3 is restricted to the brain. The double stranded regions of RNA are formed by base-pairing between residues in the close to region of the editing site with residues usually in a neighboring intron but can be an exonic sequence. The region that base pairs with the editing region is known as an Editing Complementary Sequence (ECS). [5]

LOCATION

The editing sites are all concentrated together between the last 150 nucleotides of intron 1 and the beginning of exon 2. There are 17 identified editing sites in total in the pre-mRNA of this protein.[6] Of these, 11 are found within the intronic sequence (1-11), 3 are in the 5'UTR region (5a,5b,5c) while 3 are found within the coding sequence(Y/C site, Q/R site, K/R site). Some of these editing sites occur in the highly conserved amino terminal of the protein.

The Y/C editing site is located at amino acid 2 of the final protein. The codon change introduces a tyrosine (UAU) to a (UGU) cysteine substitution.[7]

The Q/R site is a second coding region found at amino acid 5 of the final protein. Here the glutamine (Q_) is codon is converted to an arginine (R). [8]

The third K/R editing site within the coding sequence is found at amino acid position 15 of the final protein where a Lysine is converted to an Arginine. [8]

The ECS is predicted to be found in the intron with the double stranded structure formed containing all 17 of the editing sites. It is likely since all the editing sites fall within the duplex region that editing occurs in exonic and intronic sequences at the same time. There is a high level of conservation of the last 150 nucleotides of the intronic region and the start of exon 2. [8]

1. 1 REGULATION

The BLCAP protein is expressed in a wide range of tissues not just those associated with the nervous system.This indicates that editing may involve ADAR 1 enzyme.[9] However ADAR1 and ADAR2 have been demonstrated to cooperate to edit BLCAP transcript. The pre-mRNA of this protein is edited in many tissues like heart, bladder, lymphocytes, fibroblast, epithelial cells and brain but the frequency of editing varies in different tissues. There is an overall decrease in BLCAP-editing level in Astrocytomas, Bladder cancer and Colorectal cancer when compared with the relevant normal tissues. HEK 293t cells transfected with either EGFP-ADAR1, EGFP-ADAR2 or untransfected HEK293 cells were used to determine which ADAR enzyme is involved in editing at specific sites in 5'UTR and coding region. The editing level at the Y/C site was 16% while in tumour cells was an average of 21% in brain.It has been shown that ADAR1 does not edit the sites in 5' UTR but ADAR2 edits 5b and 5c sites.Y/c is edited by both and edits the Q/R and K/R sites at higher levels than ADAR1. Low levels of editing are also detected in untransfected vectors. These results indicate that ADAR1 and ADAR2 can edited all sites with ADAR2 being more efficient at the majority of sites.[8]

1.2 EPIDEMIOLOGY

Globally, in 2010, bladder cancer resulted in 170,000 deaths up from 114,000 in 1990. This is an increase of 19.4%, adjusted for increase in total world population. In the United States, bladder cancer is the fourth most common type of cancer in men and the ninth most common cancer in women. More than 60,000 men and 16,000 women are diagnosed with bladder cancer each year. Smoking can only partially explain this higher incidence in men. One other reason is that the androgen receptor, which is much more active in men than in women, plays a major part in the development of the cancer. Bladder cancer is the tenth most common cancer in the UK (10 083 people were diagnosed with the disease in 2016), and (around 5383 people died in 2016).[10]

1.3 SIGNS AND SYMPTOMS

Bladder cancer characteristically causes blood in the urine (hematuria), which may be visible (gross/macroscopic hematuria) or detectable only by microscope (microscopic hematuria). Blood in the urine is the most common symptom in bladder cancer, and is painless. Visible blood in the urine may be of only short duration, and a urine test may be required to confirm non visible blood. Between 80–90% of people with bladder cancer initially presented with visible blood.[11] Blood in the urine may also be caused by other conditions, such as bladder or ureteric stones, infection, kidney disease, kidney cancers or vascular malformations, though these conditions

(except kidney cancers) would typically be painful.

Other possible symptoms include pain during urination (dysuria), frequent urination, or feeling the need to urinate without being able to do so. These signs and symptoms are not specific to bladder cancer, and may also be caused by non-cancerous conditions, including prostate infections, overactive bladder or cystitis.

Patients with advanced disease refer pelvic or bony pain, lower-extremity swelling, or flank pain. Rarely, a palpable mass can be detected on physical examination.[12]

1.4 CAUSES

Tobacco smoking is the main known contributor to urinary bladder cancer; in most populations, smoking is associated with over half of bladder cancer cases in men and one-third of cases among women,[13] however these proportions have reduced over recent years since there are fewer smokers in Europe and North America.[14]

Infection with Schistosoma haematobium has been show to cause bladder cancer, especially of the squamous cell type. Schistosoma eggs induces a chronic inflammatory state in the bladder wall resulting in tissue fibrosis.[15] Higher levels of N-nitroso compounds(nitrate) has been detected in urine samples of people with schistosomiasis.[16] N-Nitroso compounds have been implicated in the pathogenesis of schistosomiasis related bladder cancer. They are know to cause alkylation DNA damage, specially Guanine to Adenine transition mutations in the H-ras and p53 tumor suppressor gene.[17] Mutations of P53 are detected in 73% of the tumors, BCL-2 mutations accounting for 32% and the combination of the two accounting for 13%. [18]

It has been suggested that mutations at HRAS,PIK3CA, TERT, KRAS2, RB1, TSC1 and FGFR3 may be associated in some cases.[19][20] Deletions of parts or whole of chromosome 9 is common in bladder cancer.[21] Low grade cancer are known to harbor mutations in RAS pathway (15%) and the fibroblast growth factor receptor 3 (FGFR3) gene (60%), both of which play a role in MAPK pathway. p53 and RB gene mutations are implicated in high-grade muscle invasive tumors.[22] 89% of muscle invasive cancers have shown mutations in chromatin-remodelling and histone modifying genes.[23] Deletion of both copies of the GSTM1 gene has a modest increase in risk of bladder cancer. GSTM1 gene product glutathione S-transferase M1 (GSTM1) participates in the detoxification process of carcinogens such as polycyclic aromatic hydrocarbons found in cigarette smoke.[24]

1.5 DIAGNOSIS & TREATMENT

Tests and procedures used to diagnose bladder cancer may include:

- **Cystoscopy.** To perform cystoscopy, your doctor inserts a small, narrow tube (cystoscope) through the urethra. The cystoscope has a lens that allows your doctor to see the inside of your urethra and bladder, to examine these structures for signs of disease.
- **Biopsy.** During cystoscopy, your doctor may pass a special tool through the scope and into your bladder to collect a cell sample (biopsy) for testing. This procedure is sometimes called transurethral resection of bladder tumor (TURBT). TURBT can also be used to treat bladder cancer.
- **Urine cytology.** A sample of your urine is analyzed under a microscope to check for cancer cells in a procedure called urine cytology.
- **Imaging tests.** Imaging tests, such as computerized tomography (CT) urogram or retrograde pyelogram, allow your doctor to examine the structures of your urinary tract.[25]

During a CT urogram, a contrast dye injected into a vein in your hand eventually flows into your kidneys, ureters and bladder. X-ray images taken during the test provide a detailed view of your urinary tract and help your doctor identify any areas that might be cancer.

Retrograde pyelogram is an X-ray exam used to get a detailed look at the upper urinary tract. During this test, your doctor threads a thin tube (catheter) through your urethra and into your bladder to inject contrast dye into your ureters. The dye then flows into your kidneys while X-ray images are captured.

Determining the extent of the cancer

After confirming that you have bladder cancer, your doctor may recommend additional tests to determine whether your cancer has spread to your lymph nodes or to other areas of your body.

Tests may include:

- CT **scan**
- **Magnetic** resonance imaging (MRI)
- **Bone** scan
- Chest X-ray

Your doctor uses information from these procedures to assign your cancer a stage. The stages of bladder cancer are indicated by Roman numerals ranging from 0 to IV. The lowest stages indicate a cancer that's confined to the inner layers of the bladder and that hasn't grown to affect the muscular bladder wall. The highest stage — stage IV — indicates cancer has spread to lymph nodes or organs in distant areas of the body.

The cancer staging system continues to evolve and is becoming more complex as doctors improve cancer diagnosis and treatment.[25] Your doctor uses your cancer stage to select the treatments that are right for you.

Bladder cancer grade

Bladder cancer tumors are further classified based on how the cancer cells appear when viewed through a microscope. This is known as tumor grade, and your doctor may describe bladder cancer as either low grade or high grade:

- **Low-grade bladder tumor.** This type of tumor has cells that are closer in **appearance** and organization to normal cells (well-differentiated). A low-grade tumor usually grows more slowly and is less likely to invade the muscular wall of the bladder than is a high-grade

tumor.

- **High-grade bladder tumor.** This type of tumor has cells that are abnormal-looking and that lack any resemblance to normal-appearing tissues (poorly differentiated). A high-grade tumor tends to grow more aggressively than a low-grade tumor and may be more likely to spread to the muscular wall of the bladder and other tissues and organs.

Treatment options for bladder cancer depend on a number of factors, including the type of cancer, grade of the cancer and stage of the cancer, which are taken into consideration along with your overall health and your treatment preferences.

Bladder cancer treatment may include:

- **Surgery,** to remove **cancerous** tissue
- **Chemotherapy in the bladder (intravesical chemotherapy),** to treat tumors that are confined to the lining of the bladder but have a high risk of recurrence or progression to a higher stage
- **Reconstruction,** to **create** a new way for urine to exit the body after bladder removal
- **Chemotherapy for the whole body (systemic chemotherapy),** to increase the chance for a cure in a person having surgery to remove the bladder, or as a primary treatment in cases where surgery isn't an option
- **Radiation therapy,** to destroy cancer cells, often as a primary treatment in **cases** where surgery isn't an option or isn't desired

Immunotherapy, to trigger the body's immune system to fight cancer cells, either in the bladder or throughout the body

2 MATERIALS AND METHODS

2.1 DESIGN OF EXPERIMENT:

The present experiment aims to retrieve the protein sequence for teritiary structure prediction by GOR method (Garnier-Osguthorpe-Robson) and to analyze the functional residues of the BLCAP. The bladder cancer associated protein in humans was studied. I-TASSER (Iterative Threading ASSEmbly Refinement), an online tool was used to predict three-dimensional structure by a technique called fold recognition or threading for the project.

2.2 DATABASES USED:

Since the approach is computational and highly data intensive, the project heavily relied on public database available online.

- *3.2.1 UniProtKB(Protein knowledgebase)*

The UniProt Knowledgebase (UniProtKB) is the central hub for the collection of functional information on proteins, with accurate, consistent and rich annotation. In addition to capturing the core data mandatory for each UniProtKB entry (mainly, the amino acid sequence, protein name or description, taxonomic data and citation information), as much annotation information as possible is added. UniProt Knowledgebase (UniProtKB) is a protein database partially curated by experts, consisting of two sections: UniProtKB/Swiss-Prot (containing reviewed, manually annotated entries) and UniProtKB/TrEMBL (containing unreviewed, automatically annotated entries).[26] As of 19 March 2014, release "2014_03" of UniProtKB/Swiss-Prot contains 542,782 sequence entries (comprising 193,019,802 amino acids abstracted from 226,896 references) and release "2014_03" of UniProtKB/TrEMBL contains 54,247,468 sequence entries (comprising 17,207,833,179 amino acids)

SIGNIFICANCE

UniProtKB/Swiss-Prot

UniProtKB/Swiss-Prot is a manually annotated, non-redundant protein sequence database. It combines information extracted from scientific literature and biocurator-evaluated computational analysis. The aim of UniProtKB/Swiss-Prot is to provide all known relevant information about a particular protein. Annotation is regularly reviewed to keep up with current scientific findings. The manual annotation of an entry involves detailed analysis of the protein sequence and of the scientific literature.

Sequences from the same gene and the same species are merged into the same database entry. Differences between sequences are identified, and their cause documented (for example alternative splicing, natural variation, incorrect initiation sites, incorrect exon boundaries, frameshifts, unidentified conflicts). A range of sequence analysis tools is used in the annotation of UniProtKB/Swiss-Prot entries. Computer-predictions are manually evaluated, and relevant results selected for inclusion in the entry. These predictions include post-translational modifications, transmembrane domains and topology, signal peptides, domain identification, and protein family classification.

Relevant publications are identified by searching databases such as PubMed. The full text of each paper is read, and information is extracted and added to the entry. Annotation arising from the scientific literature includes, but is not limited to:

- Protein and gene names
- Function
- Enzyme-specific information such as catalytic activity, cofactors and catalytic residues
- Subcellular location
- Protein-protein interactions
- Pattern of expression
- Locations and roles of significant domains and sites
- Ion-, substrate- and cofactor-binding sites
- Protein variant forms produced by natural genetic variation, RNA editing, alternative splicing, proteolytic processing, and post-translational modification

Annotated entries undergo quality assurance before inclusion into UniProtKB/Swiss-Prot. When new data becomes available, entries are updated.

UniProtKB/TrEMBL

UniProtKB/TrEMBL contains high-quality computationally analyzed records, which are enriched with automatic annotation. It was introduced in response to increased dataflow resulting from genome projects, as the time- and labour-consuming manual annotation process of UniProtKB/Swiss-Prot could not be broadened to include all available protein sequences. The translations of annotated coding sequences in the EMBL-Bank/GenBank/DDBJ nucleotide sequence database are automatically processed and entered in UniProtKB/TrEMBL. UniProtKB/TrEMBL also contains sequences from PDB, and from gene prediction, including Ensembl, RefSeq and CCDS.

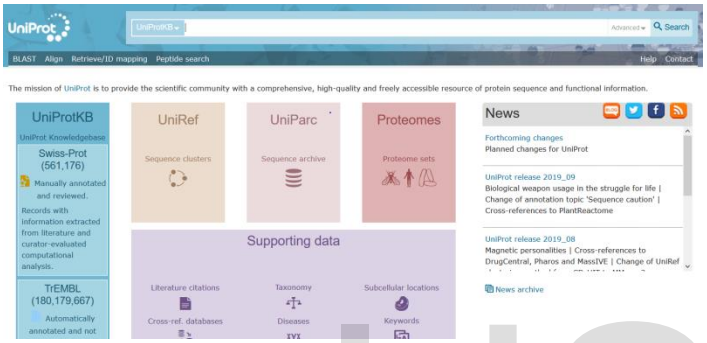


Fig.2 Homepage – uniprot

SEARCHING AND DISPLAYING RECORDS IN UNIPROT:

- The protein structure was retrived from the UNIPROT database and the hit with 87 residues with the highest score and annotated was selected.
- The sequence in FASTA format was downloaded

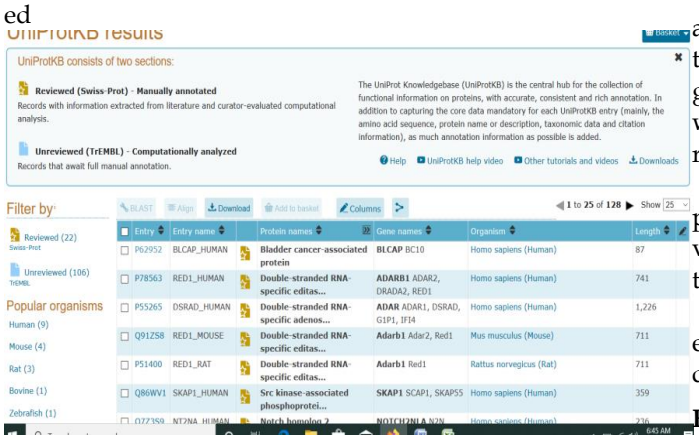


Fig. 3 The UNIPROT records page showing varioushits

Fig.4 FASTA sequence ,UNIPROT

fig.5 similar proteins ,UNIPROT

2.3 SOFTWARE USED:

2.3.1 SPDBV (Swiss-PdbViewer):

Swiss-PdbViewer (aka DeepView) is an application that provides a user friendly interface allowing to analyze several proteins at the same time. The proteins can be superimposed in order to deduce structural alignments and compare their active sites or any other relevant parts. Amino acid mutations, H-bonds, angles and distances between atoms are easy to obtain thanks to the intuitive graphic and menu interface.

Swiss-PdbViewer (aka DeepView) has been developed since 1994 by Nicolas-Guex. Swiss-PdbViewer is tightly linked to SWISS-MODEL, an automated homology modeling server developed within the Swiss Institute of Bioinformatics (SIB) at the Structural Bioinformatics Group at the Biozentrum in Basel.

Working with these two programs greatly reduces the amount of work necessary to generate models, as it is possible to thread a protein primary sequence onto a 3D template and get an immediate feedback of how well the threaded protein will be accepted by the reference structure before submitting a request to build missing loops and refine side chain packing.

Swiss-Pdb Viewer can also read electron density maps, and provides various tools to build into the density. In addition, various modeling tools are integrated and residues can be mutated.

Finally, as a special bonus, POV-Ray scenes can be generated from the current view in order to make stunning ray-traced quality images.

Features of SPDBV:

The Toolbar contains the tool buttons and menus of the program:

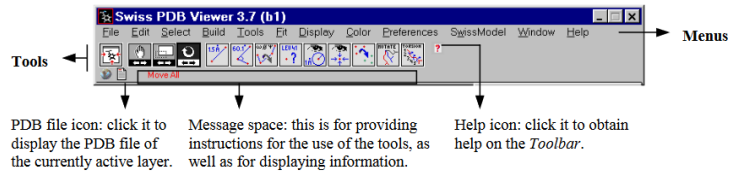


Fig.6 contains the menus and tools of the program.

Superposing and aligning all homologous templates

If several templates were selected, theyfirst of all need to be superposed by doing one of the following:•clickFit>Magic Fit and Fit>Generate Structural Alignment, or•clickFit>Iterative Magic Fit (the structural alignment will be automatically done).

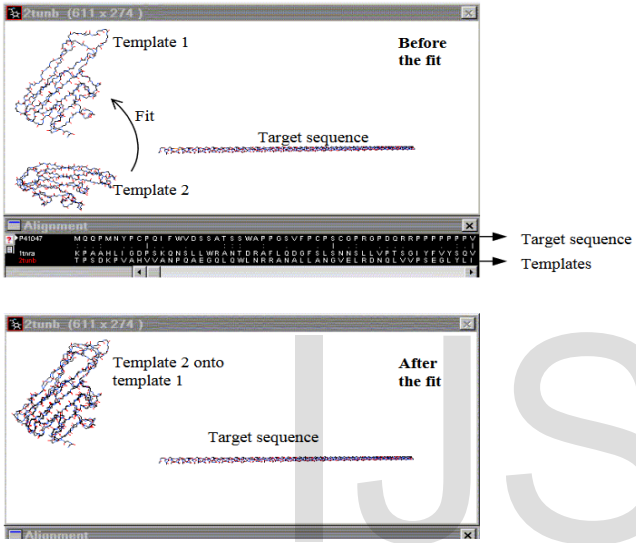


Fig.7 Templates alignment

2.3.2 ONLINE TOOLS USED:

GOR4 (Garnier-Osguthorpe-Robson):

The GOR method (Garnier-Osguthorpe-Robson) is an information theory-based method for the prediction of secondary structures in proteins. It was developed in the late 1970s shortly after the simpler Chou-Fasman method. Like Chou-Fasman, the GOR method is based on probability parameters derived from empirical studies of known protein tertiary structures solved by X-ray crystallography. However, unlike Chou-Fasman, the GOR method takes into account not only the propensities of individual amino acids to form particular secondary structures, but also the conditional probability of the amino acid to form a secondary structure given that its immediate neighbors have already formed that structure. The method is therefore essentially Bayesian in its analysis

Method

The GOR method analyzes sequences to predict alpha helix, beta sheet, turn, or random coil secondary structure at each position based on 17-amino-acid sequence windows. The original description of the method included four scoring matrices of size 17×20, where the columns correspond to the log-odds score, which reflects the probability of finding a given amino acid at each position in the 17-residue sequence. The four matrices reflect the probabilities of the central, ninth amino acid being in a helical, sheet, turn, or coil conformation. In subsequent revisions to the method, the turn matrix was eliminated due to the high variability of sequences in turn regions (particularly over such a large window). The method was considered as best requiring at least four contiguous residues to score as alpha helices to classify the region as helical, and at least two contiguous residues for a beta sheet.

Algorithm

The mathematics and algorithm of the GOR method were based on an earlier series of studies by Robson and colleagues reported mainly in the *Journal of Molecular Biology* and The Biochemical Journal . The latter describes the information theoretic expansions in terms of conditional information measures. The use of the word "simple" in the title of the GOR paper reflected the fact that the above earlier methods provided proofs and techniques somewhat daunting by being rather unfamiliar in protein science in the early 1970s; even Bayes methods were then unfamiliar and controversial. An important feature of these early studies, which survived in the GOR method, was the treatment of the sparse protein sequence data of the early 1970s by expected information measures. That is, expectations on a Bayesian basis considering the distribution of plausible information measure values given the actual frequencies (numbers of observations). The expectation measures resulting from integration over this and similar distributions may now be seen as composed of "incomplete" or extended zeta functions, e.g. z(s,observed frequency) − z(s,expected frequency) with incomplete zeta function $z(s, n) = 1 + (1/2)^s + (1/3)^s + (1/4)^s + \ldots. + (1/n)^s$. The GOR method used s=1. Also, in the GOR method and the earlier methods, the measure for the contrary state to e.g. helix H, i.e. ~H, was subtracted from that for H, and similarly for beta sheet, turns, and coil or loop. Thus the method can be seen as employing a zeta function estimate of log predictive odds. An adjustable decision constant could also be applied, which thus also implies a decision theory approach; the GOR method allowed the option to use decision constants to optimize predictions for different classes of protein. The expected information measure used as a basis for the information expansion was less important by the time of publication of the GOR method because protein sequence data became more plentiful, at least for the terms considered at that time. Then, for s=1, the expression z(s,observed frequency) − z(s,expected frequency) approaches the natural logarithm of (observed frequency / expected frequency) as frequencies increase. However, this measure (including use of other values of s) remains important in later more general applications with high-dimensional data, where data for more complex terms in the information expansion are inevitably sparse

I-TASSER(Iterative Threading ASSEmbly Refinement):

I-TASSER (**I**terative **T**hreading **ASSE**mbly **R**efinement) is a bioinformatics method for predicting three-dimensional structure model of protein molecules from amino acid sequences. It

detects structure templates from the Protein Data Bank by a technique called fold recognition (or threading). The full-length structure models are constructed by reassembling structural fragments from threading templates using replica exchange Monte Carlo simulations. I-TASSER is one of the most successful protein structure prediction methods in the community-wide CASP experiments.
I-TASSER has been extended for structure-based protein function predictions, which provides annotations on ligand binding site, gene ontology and enzyme commission by structurally matching structural models of the target protein to the known proteins in protein function databases. It has an on-line server built in the Yang Zhang Lab at the University of Michigan, Ann Arbor, allowing users to submit sequences and obtain structure and function predictions. A standalone package of I-TASSER is available for download at the I-TASSER website

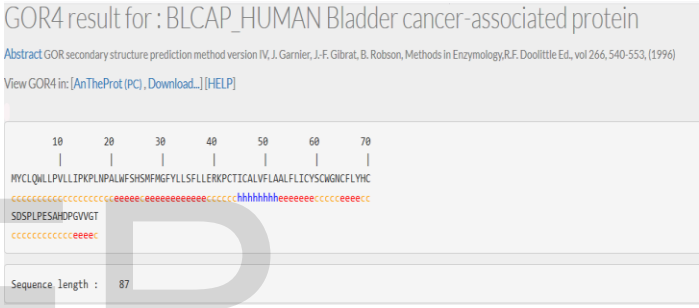
Method

I-TASSER is a template-based method for protein structure and function prediction. The pipeline consists of six consecutive steps:

- Secondary structure prediction by PSSpred
- Template detection by LOMETS
- Fragment structure assembly using replica-exchange Monte Carlo simulation
- Model selection by clustering structure decoys using SPICKER
- Atomic-level structure refinement by fragment guided molecular dynamics simulation (FG-MD) or ModRefiner
- Structure-based biology function annotation by COACH

The I-TASSER server allows users to generate automatically protein structure and function predictions.

- Input
    - Mandatory:
        - Amino acid sequence with length from 10 to 1,500 residues
    - Optional (user can provide optionally restraints and templates to assist I-TASSER modeling):
        - Contact restraints
        - Distance maps
        - Inclusion of special templates
        - Exclusion of special templates
        - Secondary structures
- Output
    - Structure prediction:
        - Secondary structure prediction
        - Solvent accessibility prediction
        - Top 10 threading alignment from LOMETS
        - Top 5 full-length atomic models (ranked based on cluster density)
        - Top 10 proteins in PDB which are structurally closest to the predicted models
        - Estimated accuracy of the predicted models (including a confidence score of all models, predicted TM-score and RMSD for the first model, and per-residue error of all models)
        - B-factor estimation
    - Function prediction:
        - Enzyme Classification (EC) and the confidence score
        - Gene Ontology (GO) terms and the confidence score
        - Ligand-binding sites and the confidence score
        - An image of the predicted ligand-binding sites



2.4 PROTOCOL:

Step I: Selection of the protein :

Over expression of BLCAP (previously was known as BC10) to inhibit cell growth and to induce apoptosis, suggesting a role for this protein in the regulation of these cellular processes. Recently, it was identified bladder cancer-associated protein (BLCAP) as a novel potential biomarker ,Loss of BLCAP mRNA expression correlates with the invasive potential of UCs, and subsequent studies by others in several cancer types, such as cervical and renal as well as human tongue carcinoma and osteosarcoma , have all shown differential expression of this protein in cancer. BLCAP is a small (87-amino acid), evolutionary conserved protein with no homology to any known-protein, and its cellular function is largely unknown .thus we have selected this protein.

Step II: Retrieval of data from uniprot:

1. BLCAP CP 10 sequence, was retrived from uniprotKB for variant search and 128 hits were obtained for the search. The records usually contain entry, entry name, Protein names, Gene names, Organism, Length.

2. The significance is also mentioned in the records.

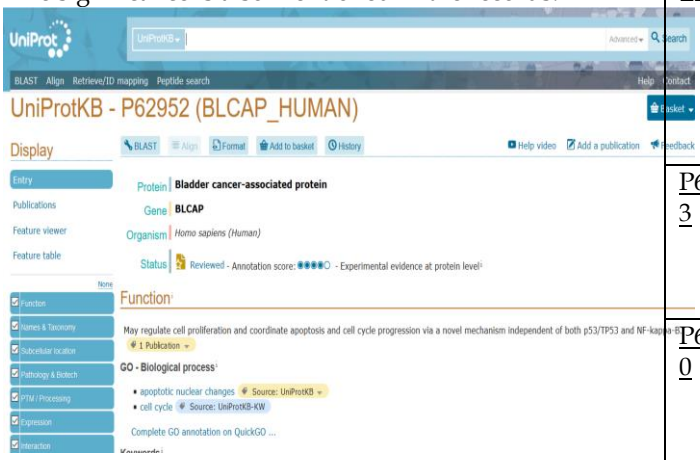


Fig.8 BLCAP data,UNIPROT

3. FASTA sequence of the BLCAP without any structure was selected with 87 residues.

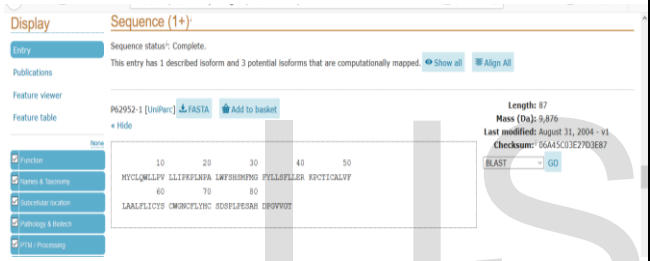


Fig.9 FASTA Sequence

Fig.10 Screen capture representing the results from GOR4.

The record of all the similar protein was collected and tabulated as follows.

TABLE.1 Screen capture representing the records of all similar protein

Entry	Entry name	Protein names	Gene names	Organism	Length
P62953	BLCAP_FELCA	**Bladder cancer-associated protein**	**BLCAP** BC10	Felis catus (Cat) (Felis silvestris catus)	87
P62950	BLCAP_RAT	**Bladder cancer-associated protein**	**Blcap**	Rattus norvegicus (Rat)	87
Q5R692	BLCAP_PONAB	**Bladder cancer-associated protein**	**BLCAP**	Pongo abelii (Sumatran orangutan) (Pongo pygmaeus abelii)	87
P62954	BLCAP_BOVIN	**Bladder cancer-associated protein**	**BLCAP**	Bos taurus (Bovine)	87
P62951	BLCAP_MOUSE	**Bladder cancer-associated protein**	**Blcap** Bc10	Mus musculus (Mouse)	87
A0A2K5XLV4	A0A2K5XLV4_MANLE	Uncharacterized protein	**BLCAP**	Mandrillus leucophaeus (Drill) (Papio leucophaeus)	87
A0A2K5BZC7	A0A2K5BZC7_AOTNA	Uncharacterized protein	**BLCAP**	Aotus nancymaae (Ma's night monkey)	87
A0A0D9SBN7	A0A0D9SBN7_CHLSB	Uncharacterized protein	**BLCAP**	Chlorocebus sabaeus (Green monkey) (Cercopithecus sabaeus)	87
A0A4X1THT8	A0A4X1THT8_PIG	Uncharacterized protein	**BLCAP**	Sus scrofa (Pig)	87

Step III: Prediction of secondary structures of proteins

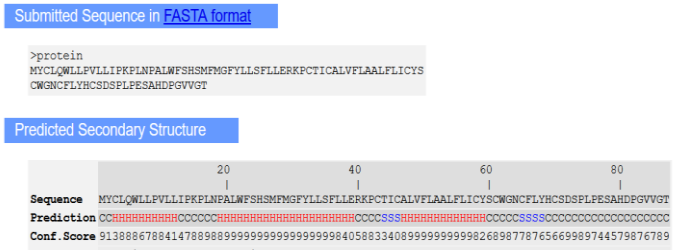


Fig.11, Screen capture representing the results from itasser.

The results of both the software were compared and based on the template used to predict the structure of BLCAP we consider the results of I-TASSER to be selected

Fig.12 representing the list of templates used,I-TASSER

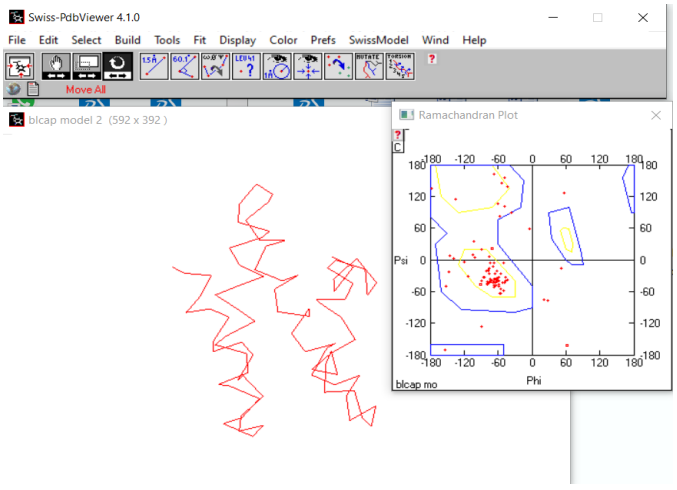
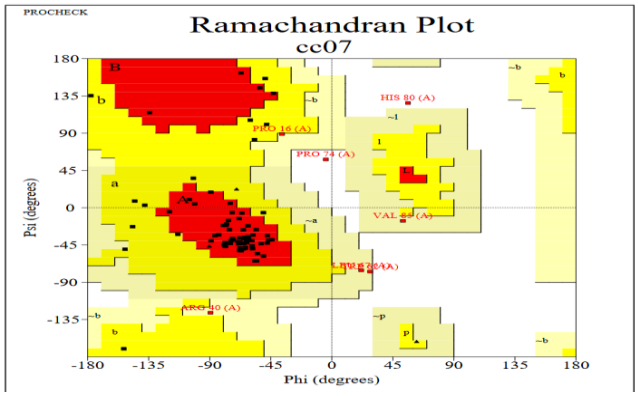


Fig.13 Representing obtained protein visualization in SPDBV

OBTAINING RAMACHANDRAN PLOT USING PDBSUM:



Plot statistics

Residues in most favoured regions [A,B,L]	56	76.7%
Residues in additional allowed regions [a,b,l,p]	12	16.4%
Residues in generously allowed regions [~a,~b,~l,~p]	4	5.5%
Residues in disallowed regions	1	1.4%
Number of non-glycine and non-proline residues	73	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	4	
Number of proline residues	8	
Total number of residues	87	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Fig.14 Representing ramachandran plot and its analysis

- The number of residues in the disallowed region was only one thus the predicted structure was loop modeled in SPDBV

2.5 Loop modeling of the protein using SPDBV:

- The residue which is obtained in the disallowed region was selected
- Then,5 residues from the above and 5 residues from the below of residue of disallowed region were also selected
- Loop was built using "build loop…" option and anchors were selected
- Loops generated were verified from the list and visualized using Ramachandran plot and the structure of the protein with zero residues was obtained

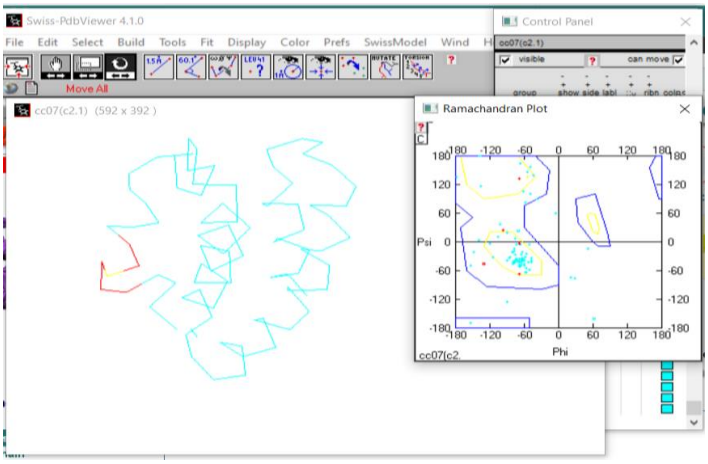
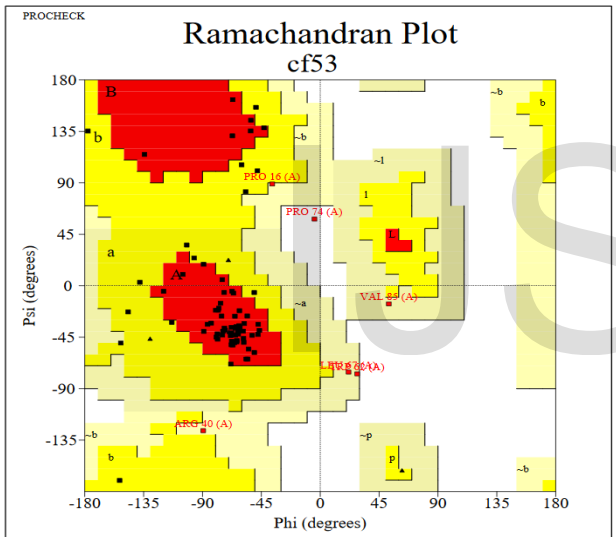


Fig.15 Representing loop modeling using SPDBV



Plot statistics

Residues in most favoured regions [A,B,L]	58	79.5%
Residues in additional allowed regions [a,b,l,p]	11	15.1%
Residues in generously allowed regions [~a,~b,~l,~p]	4	5.5%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	73	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	4	
Number of proline residues	8	
Total number of residues	87	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Fig.16 Representing Ramachandran plot of loop modeled protein and its analysis

2.6 Energy minimization using MODREFINER:

- The obtained result was sent to MODREFINER for energy minimization and to construct and refine atomic-level protein models from C-alpha traces.

Fig.17 (a)The energy values before energy minimization

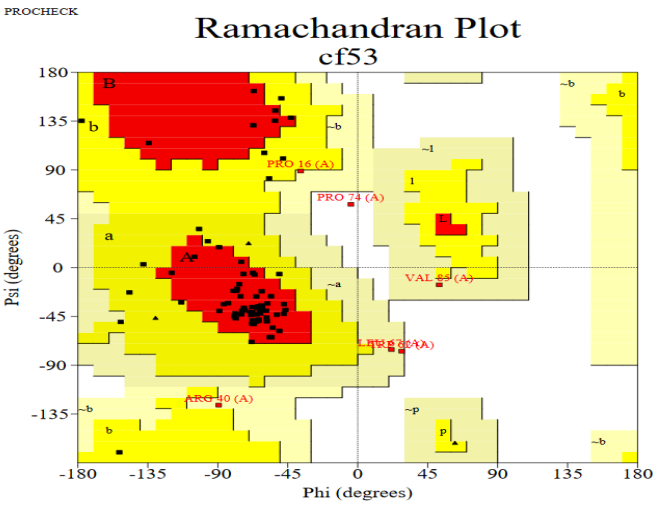
```
REMARK    File generated by Swiss-PdbViewer  4.00b0
REMARK    http://www.expasy.org/spdbv/
REMARK    File generated by Swiss-PdbViewer  4.00b0
REMARK    http://www.expasy.org/spdbv/
HEADER protein
REMARK PARENT N/A (Threading alignments from the templates are
below the Z-score cut-off)
ATOM      1  N   MET A   1      54.265  42.151  40.640  1.00
8.70
ATOM      2  CA  MET A   1      54.659  42.508  39.277  1.00
8.70
ATOM      3  C   MET A   1      55.254  43.922  39.246  1.00
8.70
ATOM      4  O   MET A   1      54.988  44.709  38.339  1.00
8.70
ATOM      5  CB  MET A   1      53.494  42.294  38.292  1.00
8.70
ATOM      6  CG  MET A   1      53.019  40.835  38.249  1.00
8.70
ATOM      7  SD  MET A   1      54.262  39.659  37.646  1.00
8.70
ATOM      8  CE  MET A   1      53.354  38.106  37.863  1.00
8.70
ATOM      9  H   MET A   1      54.471  42.911  41.257  1.00
99.99
ATOM     10  H   MET A   1      53.294  41.941  40.758  1.00
99.99
ATOM     11  H   MET A   1      54.724  41.346  41.015  1.00
99.99
ATOM     12  N   TYR A   2      55.923  44.291  40.350  1.00
```

Fig.17 (b)The energy values after energy minimization

3 RESULTS AND DISCUSSION

```
REMARK    http://www.expasy.org/spdbv/
HEADER protein
REMARK PARENT N/A (Threading alignments from the templates are
below the Z-score cut-off)
ATOM      1  N   MET A   1      44.307  56.022  59.210  1.00
10.27
ATOM      2  CA  MET A   1      42.948  56.517  59.408  1.00
10.27
ATOM      3  C   MET A   1      41.930  55.672  58.625  1.00
10.27
ATOM      4  O   MET A   1      42.139  54.479  58.413  1.00
10.27
ATOM      5  CB  MET A   1      42.921  58.030  59.100  1.00
10.27
ATOM      6  CG  MET A   1      43.306  58.378  57.653  1.00
10.27
ATOM      7  SD  MET A   1      43.222  60.137  57.219  1.00
10.27
ATOM      8  CE  MET A   1      41.438  60.455  57.308  1.00
10.27
ATOM      9 1H   MET A   1      44.288  55.230  58.600  1.00
99.99
ATOM     10 2H   MET A   1      44.942  56.680  58.804  1.00
99.99
ATOM     11 3H   MET A   1      44.776  55.731  60.044  1.00
99.99
ATOM     12  N   TYR A   2      40.835  56.269  58.142  1.00
9.20
ATOM     13  CA  TYR A   2      39.982  55.649  57.118  1.00
9.20
```

The 3D model that was obtained using multiple templates (4cofA & 4roeA) by the above mentioned approach was visualized (Figure 18).The stereo chemical quality of the model was assessed using RAMACHANDRAN PLOT analysis (Figure.16), where 79.5% of the residues were in the most favored region, 15.1 % in allowed region, 5.5 % in generously allowed region and 0% of the residues lying in the disallowed regions, number of non glycine non proline residues are 73,number of end residues (excluding glycine and proline) are 2 ,number of glycine residues which are represented as triangles are four and number of proline residues are 8,Further studies using PDBSUM and Verify 3-D were also carried out to check the reliability of the model. The plot indicates that none of the residues lie in the disallowed region. The model thus obtained in this study is the best model of BCLAP.



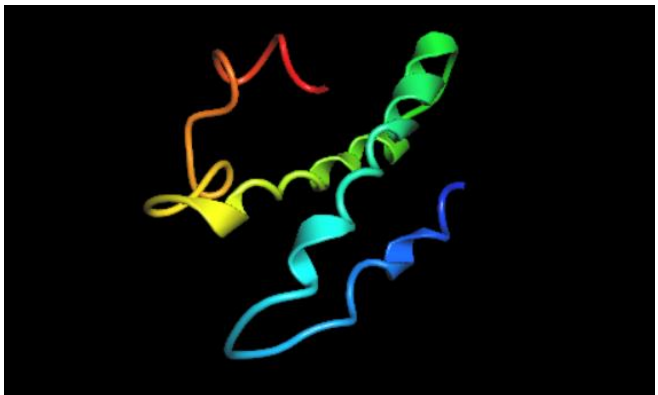


Fig.18 Final 3D model BCLAP BC10 visualized

4 REFERENCES

.
[1]"Cancer Fact sheet N°297". World Health Organization . February 2014. Archived from the original on 29 December 2010 . Retrieved 10 June 2014 .
[2] Engel, LS; Taioli, E; Pfeiffer, R; Garcia-Closas, M; Marcus, PM; Lan, Q; Boffetta, P; Vineis, P; Autrup, H; Bell, DA; Branch, RA; Brockmöller, J; Daly, AK; Heckbert, SR; Kalina, I; Kang, D; Katoh, T; Lafuente, A; Lin, HJ; Romkes, M; Taylor, JA; Rothman, N (15 July 2002). "Pooled analysis and meta-analysis of glutathione S-transferase M1 and bladder cancer: a HuGE review". American Journal of Epidemiology .
[3] https://emedicine.medscape.com/article/438262-overview.
[4] Jemal A, Siegal R, Ward E, et al. Cancer statistics. CA:Cancer J Clin 2008; 58: 71–76
[5] Hartner JC, Schmittwolf C, Kispert A, Müller AM, Higuchi M, Seeburg PH (February 2004). "Liver disintegration in the mouse embryo caused by deficiency in the RNA-editing enzyme ADAR1". J. Biol. Chem. **279** (6): 4894–902.
[6] Galeano F, Leroy A, Rossetti C, Gromova I, Gautier P, Keegan LP, Massimi L, Di Rocco C, O'Connell MA, Gallo A (July 2010). "Human BLCAP transcript: new editing events in normal and cancerous tissues". International Journal of Cancer. **127** (1): 127–37.
[7] Levanon EY, Hallegger M, Kinar Y, Shemesh R, Djinovic-Carugo K, Rechavi G, Jantsch MF, Eisenberg E (2005). "Evolutionarily conserved human targets of adenosine to inosine RNA editing". Nucleic Acids Research. **33** (4): 1162–8.
[8] Galeano F, Leroy A, Rossetti C, Gromova I, Gautier P, Keegan LP, Massimi L, Di Rocco C, O'Connell MA, Gallo A (July 2010). "Human BLCAP transcript: new editing events in normal and cancerous tissues". International Journal of Cancer. **127** (1): 127–37.
[9] Riedmann EM, Schopoff S, Hartner JC, Jantsch MF (June 2008). "Specificity of ADAR-mediated RNA editing in newly identified targets". RNA. **14** (6): 1110–8.
[10]Hemelt M, Zeegers MP (2000). "The effect of smoking on the male excess of bladder cancer: a meta-analysis and geographical analyses". Int J Cancer. **124** (2): 412–9.
[11]Avellino, Gabriella J.; Bose, Sanchita; Wang, David S. (1 June 2016). "Diagnosis and Management of Hematuria". Surgical Clinics of North America. **96** (3): 503–515.

[12] Avellino, Gabriella J.; Bose, Sanchita; Wang, David S. (1 June 2016). "Diagnosis and Management of Hematuria". Surgical Clinics of North America. **96** (3): 503–515.
[13] Zeegers MP; Tan, FE; Dorant, E; Van Den Brandt, PA (2000). "The impact of characteristics of cigarette smoking on urinary tract cancer risk: a meta-analysis of epidemiologic studies". Cancer. **89** (3): 630–9.
[14] Osch, Frits H. M. van; Jochems, Sylvia H. J.; Schooten, Frederik-Jan van; Bryan, Richard T.; Zeegers, Maurice P. (20 April 2016). "Quantified relations between exposure to tobacco smoking and bladder cancer risk: a meta-analysis of 89 observational studies". International Journal of Epidemiology . **45** (3): 857–870.
[15] Zaghloul, Mohamed S. (December 2012). "Bladder cancer and schistosomiasis". Journal of the Egyptian National Cancer Institute. **24** (4): 151–159.
[16] Mostafa, MH; Helmi, S; Badawi, AF; Tricker, AR; Spiegelhalder, B; Preussmann, R (April 1994). "Nitrate, nitrite and volatile N-nitroso compounds in the urine of Schistosoma haematobium and Schistosoma mansoni infected patients". Carcinogenesis. **15** (4): 619–25.
[17] Badawi, AF (2 August 1996). "Molecular and genetic events in schistosomiasis-associated human bladder cancer: role of oncogenes and tumor suppressor genes". Cancer Letters. 105 (2): 123–38.
[18] Chaudhary, KS; Lu, QL; Abel, PD; Khandan-Nia, N; Shoma, AM; el Baz, M; Stamp, GW; Lalani, EN (January 1997). "Expression of bcl-2 and p53 oncoproteins in schistosomiasis-associated transitional and squamous cell carcinoma of urinary bladder". British Journal of Urology. 79 (1): 78–84
[19] Online Mendelian Inheritance in Man (OMIM) 109800
[20] Zhang, X; Zhang, Y (September 2015). "Bladder Cancer and Genetic Mutations". Cell Biochemistry and Biophysics. 73(1): 65–9.
[21] Reference, Genetics Home. "Bladder cancer". Genetics Home Reference.
[22]Ahmad, I.; Sansom, O. J.; Leung, H. Y. (15 March 2012). "Exploring molecular genetics of bladder cancer: lessons learned from mouse models". Disease Models & Mechanisms. **5** (3): 323–332.
[23] Humphrey, PA; Moch, H; Cubilla, AL; Ulbright, TM; Reuter, VE (July 2016). "The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs-Part B: Prostate and Bladder Tumours" (PDF). European Urology. **70** (1): 106–119.
[24] "Bladder Cancer Report" (PDF). World Cancer Research Fund : International. Retrieved 9 November 2019.
[25]https://www.mayoclinic.org/diseases-conditions/bladder-cancer/diagnosis-treatment/drc-20356109
[26] https://www.uniprot.org/help/uniprotkb