



# Journal of Biological Sciences

ISSN 1727-3048

**science**  
alert

**ANSI***net*  
an open access publisher  
<http://ansinet.com>

## Compendium of Human Oncogenic Viruses: OncVir

<sup>1</sup>Baby Joseph and <sup>1,2</sup>Vrundha M. Nair

<sup>1</sup>Interdisciplinary Research Centre, Department of Biotechnology, Malankara Catholic College, Mariagiri, Kaliakkavilai-629 153, Kanyakumari District, Tamil Nadu, India

<sup>2</sup>Department of Bioinformatics, Malankara Catholic College, Mariagiri, Kaliakkavilai-629 153, Kanyakumari District, Tamil Nadu, India

---

**Abstract:** The key significance of this study is the development of a database termed as Oncvir. The resource is developed using Hypertext Preprocessor, Hyper Text Markup Language and JavaScript as front-end and MySQL as backend. OncVir mainly in-houses information on nucleotide sequence, protein sequence and protein structure information's of human oncoviruses. Oncoviruses are those categories of viruses which assist in the immortalization of the cancer cells. The key classification factor of these viruses is its genetic material either DNA or RNA. OncVir is a database which comprehends the information of human oncogenic viruses such as Hepatitis B Virus, Human Papilloma Virus; Kaposi's Sarcoma-associated Herpes Virus, Epstein-Barr virus, Merkel cell polyomavirus, Hepatitis C virus, Human T-lymphotropic virus and Human Cytomegalovirus. This knowledgebase includes information from literature and other public available databases. The database is accessible to keyword searching via., simple and advanced options. It integrates about 215 nucleotide sequence records, 158 protein sequence records and 68 structural information associated with oncoviruses. It also facilitates additional facility to users to submit their sequence and structural data using tools such as OncSeq and OncTer, respectively. Moreover, additional information including description of viruses with illustrations, overview about the database, new events and updates are also available. In short Oncvir serves as a single access point for substantial and multitudinous information about human oncoviruses.

**Key words:** Databases, Human Papilloma Virus, *Hepatitis B Virus*, cancer, sequence

---

### INTRODUCTION

A substantial proportion of human cancers are due to viral infections. Cancer is due to the progressive acquirement of mutations which results in selective growth advantages of the altered cells through the activation of intrinsic mitogenic and pro-survival pathways. (Kasper *et al.*, 2006). Infectious agents are one among the several risk factors for cancer which includes viruses and bacteria. Tumor viruses are those viruses which causes cancer in their natural host *in vivo* or in experimental animals *in vitro* (Damania and Pipas, 2009). The presence of oncoviruses in man is confirmed when identified its world-wide distribution in other vertebrates (Barbacid *et al.*, 1980). A tumor causing virus were first discovered by Dr. Francis Peyton Rous and was awarded with Nobel prize in Physiology and Medicine in 1966. This discovery leads to a novel field Oncovirology (Gurtsevitch, 2008). The world wide cancer statistics estimates that etiologic agents like virus's causes approximately 12% of human cancers (Parkin, 2006).

Human oncogenic viruses are ubiquitous pathogens which exhibits their malignant potential only under natural conditions (Didier, 1983). However, all virus infections will not result in cancer. It was recognized that the virus-induced cancer is promoted by several other risk factors apparently resulting in the progression of cancer.

Oncogenes and Tumor suppressor genes are two classes of genes in which altered expression can lead to loss of growth control i.e., cancer (Joseph and Nair, 2012). The major reason behind the involvement of viruses in cancer is because they can carry extra genetic material in to the host cell or it could alter the expression of the existing cellular genes. If a virus takes up residence in a cell and alters the properties of that cell, the cell is said to be transformed. Cancer causing DNA tumor viruses are: *Hepatitis B Virus* (HBV), Human Papilloma Virus (HPV) (Halimi and Asl, 2011), Kaposi's Sarcoma-associated Herpes Virus (KSHV or HHV-8), Epstein-Barr virus (EBV or HHV-4), Merkel cell polyomavirus and Human Cytomegalovirus (CMV or HHV-5) (Zheng, 2010). When these viruses infects a permissive host cell, the viral

genes took control of the cellular processes and force the host cell to divide uncontrollably, transforming the cell into benign tumor and later leading to cancer. Such neoplastic transformations are achieved by DNA tumor viruses (Hames and Hooper, 2000). Whereas the RNA viruses normally termed as retroviruses first convert their genetic material-RNA to DNA through reverse transcriptase enzyme before transforming a normal cell to malignant cell (Hames and Hooper, 2000). *Hepatitis C Virus* (HCV) and Human T-lymphotropic virus (HTLV-1) falls in to the category of RNA tumor viruses (Zheng, 2010).

Maintaining electronic databases is one of the applications of the interdisciplinary field, Bioinformatics (Prajapat *et al.*, 2012). Electronic databases maintain data in a persistent, non-volatile form that allows operations to be repeated and compared with other operations, with the results communicated to other researchers and developers (Bryan, 2009). In the present study, a biological database named OncVir is created which stands as a primary access point for information about human oncoviruses. Hypertext Preprocessor (PHP), Hyper Text Markup Language (HTML) and Javascript form the front end of the database. The back end of the database was developed with MySQL. The database can be run using Apache server. This database act as an integrated resource in accessing information on nucleotide sequences, protein sequences and protein structural information of human oncoviruses.

## MATERIALS AND METHODS

**Data source:** The data required for OncVir database were collected from literature references and through public domain databases such as GenBank (Benson *et al.*, 1999), PDB (Bernstein *et al.*, 1977), SwissProt (Bairoch and Apweiler, 1999) and Uniprot (Leinonen *et al.*, 2004). The lists of oncoviruses included in OncVir are listed in Table 1.

**HTML:** HTML is the language used to create web pages. It stands for Hyper Text Markup Language. It is not a programming, but a markup language. A markup language

Table 1: Comprehensive information on oncogenic viruses

Genome	Virus	Family
DNA	Human papilloma virus	Papavaviridae
	Kaposi's sarcoma-associated herpesvirus	Herpesviridae
	Merkel cell polyomavirus	Polyomaviridae
	Epstein-barr virus	Herpesviridae
	Hepatitis B virus	Hepadnaviridae
	Human cytomegalovirus	Herpesviridae
RNA	Human T-lymphotropic virus	Retroviridae
	Hepatitis C virus	Hepadnaviridae

uses a set of markup tags. HTML uses markup tags to describe web pages. HTML can include embedded scripting language code (such as JavaScript) that can affect the behavior of Web browsers and other processors. When the web client access the website, a default or homepage in HTML format is transmitted to the user. The HTML file can display images, sounds and multimedia objects. The objects are not actually stored in the HTML document. Instead, an external reference to a picture or multimedia object is inserted in the text of HTML document.

**JAVA SCRIPT:** JavaScript is a scripting language used to enable programmatic access to objects within other applications. It is primarily used in the form of user JavaScript for the development of dynamic websites. JavaScript is characterized as a dynamic, weakly typed, prototype based language with first class functions. JavaScript is used to do user side validation. It is used to create mouse over events.

**PHP:** PHP is an open-source server side scripting language. A dynamic web page interacts with the user, so that each user visiting the page sees customized information. PHP can also be used to create dynamic web pages that are generated from information accessed from a MySQL database. We can embed PHP commands within a standard HTML page. Another feature that PHP offers is connectivity to most of the common databases. PHP also offers integration with various external libraries which allow the developer to do anything from generating PDF documents to parsing XML.

**MYSQL:** MySQL is the world's most popular open source database software, with over 100 million copies of its software downloaded or distributed throughout its history. MySQL is a key part of LAMP (Linux, Apache, PHP/Perl/Python), the fast-growing open source enterprise software stack. MySQL is a data storage area. In this storage area, there are small sections called tables. Very similar to a normal HTML table, the MySQL table consists of rows, columns and cells. When the PHP script and MySQL are on the same machine, you can use local host as the address you wish to connect to local host is a shortcut to just have the machine connects to itself.

**PHP myadmin:** PhpMyAdmin is a graphical interface program that allows you to use the functions of a MySQL. MySQL's appeal originates in its relative simplicity and ease of use which is enabled by an ecosystem of open source tools such as PhpMyAdmin.

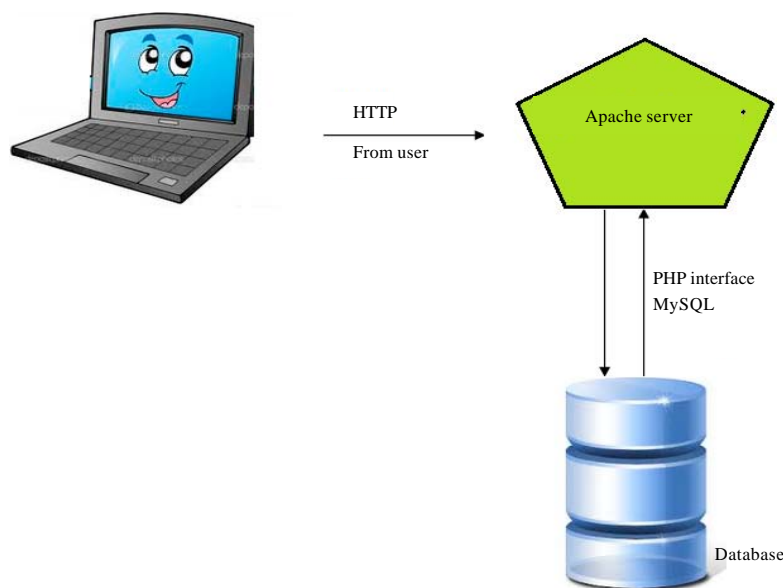


Fig. 1: Overall architecture of OncVir database systems

**Apache:** Apache is an Open-source web server originally based on NCSA server. Available on over 160 varieties of UNIX and NT. Apache can handle run-time execution of external applications (scripts). Apache support Perl, Python, Tcl and PHP. Virtual hosting allows one Apache installations to serve many different actual web sites. Apache is primarily used to serve both static content and dynamic web pages on the World Wide Web. Apache does have performance comparable to other “high-performance” web servers. The overall architecture of the database is shown in Fig. 1.

## RESULTS AND DISCUSSION

OncVir is a repository on Human Oncogenic Viruses. The database contains information on DNA and RNA oncoviruses. The aim of this database is to collect and register the available sequence (protein and nucleotide) and structural information of tumor viruses from the public domain databases and thus it supports researchers. There are approximately 185108 bases in 215 nucleotide sequence records, 48630 residues in 158 protein sequence records and 68 structural information. It is an integrated resource for human oncovirus information. The overview of OncVir is shown in Fig. 2.

OncVir was created using PHP with MySQL. This database has two faces: User part and Admin part. In user part user can enter the queries and the Admin part is accessed only by the administrator. Inserting data, updating data and deletion of desired data are done by the administrator. Using user name and password the

administrator can enter into admin homepage which contain links for Insert data, update data, delete data, change password and to post updates.

The home page (Fig. 3) of OncVir shows hyperlink to Home, overview, OncVir submission: Oncseq, Oncter and register, description, help, photo gallery, references, new events, contact us, about us and enquiry. The database facilitates two kinds of search options such as non-specific search and advanced search (Fig. 4). In simple keyword search user can use any type of keywords. In advanced search (Fig. 5) user can search the database using accession number/PDB ID, types of virus and category (protein sequence, nucleotide sequence and protein structure). If the desired keyword is used, then an intermediate page is displayed which contains the hit information on the specific oncoviruses. Further searching on hit information will show detailed features of protein sequence information, nucleotide sequence information and protein structure information.

The nucleotide sequence search with any key word results in a hit of search results. If any one of the hit is selected, the brief information such as entry No., accession number, title, nucleic acid type, definition, sequence length, date of submission, source organism, pubmed citation number, database cross reference and sequence. The search for protein sequence (Table 2) shows information's such as entry No., identification, accession number, gene name, description, organism species, organism classification, sequence length, function and sequence. The protein structure includes

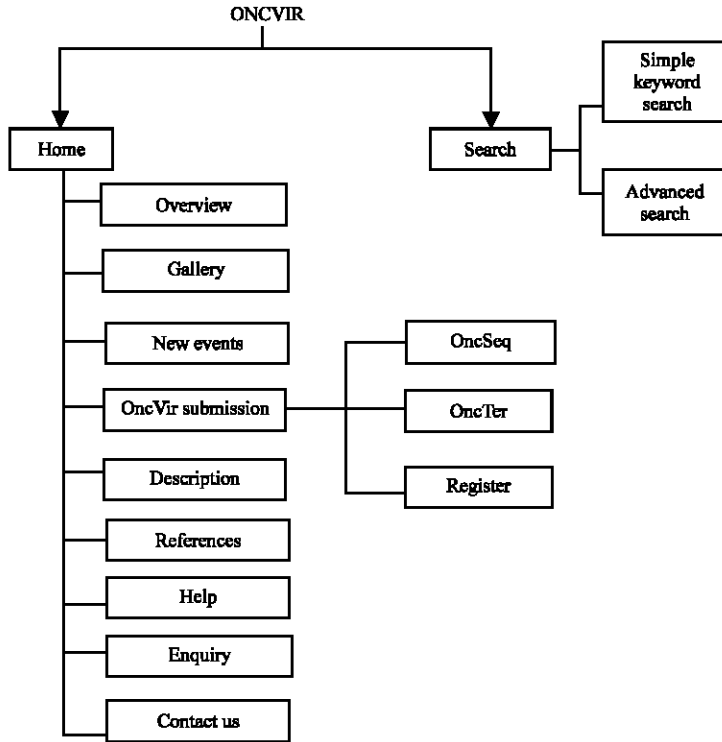


Fig. 2: Overview of OncVir

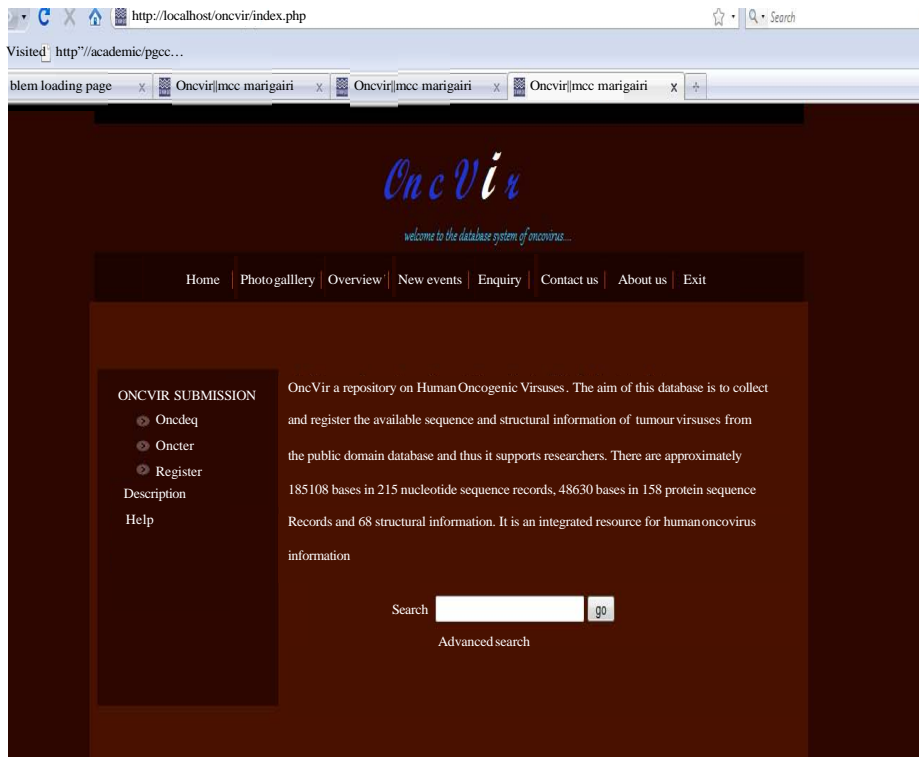


Fig. 3: Home page of OncVir

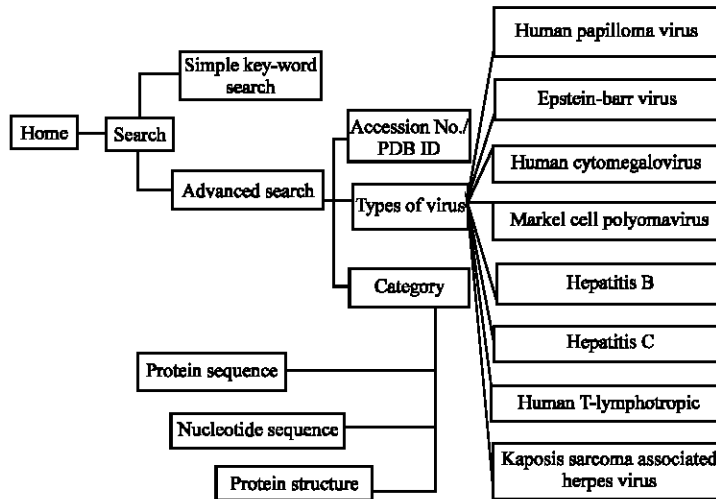


Fig. 4: Searching OncVir

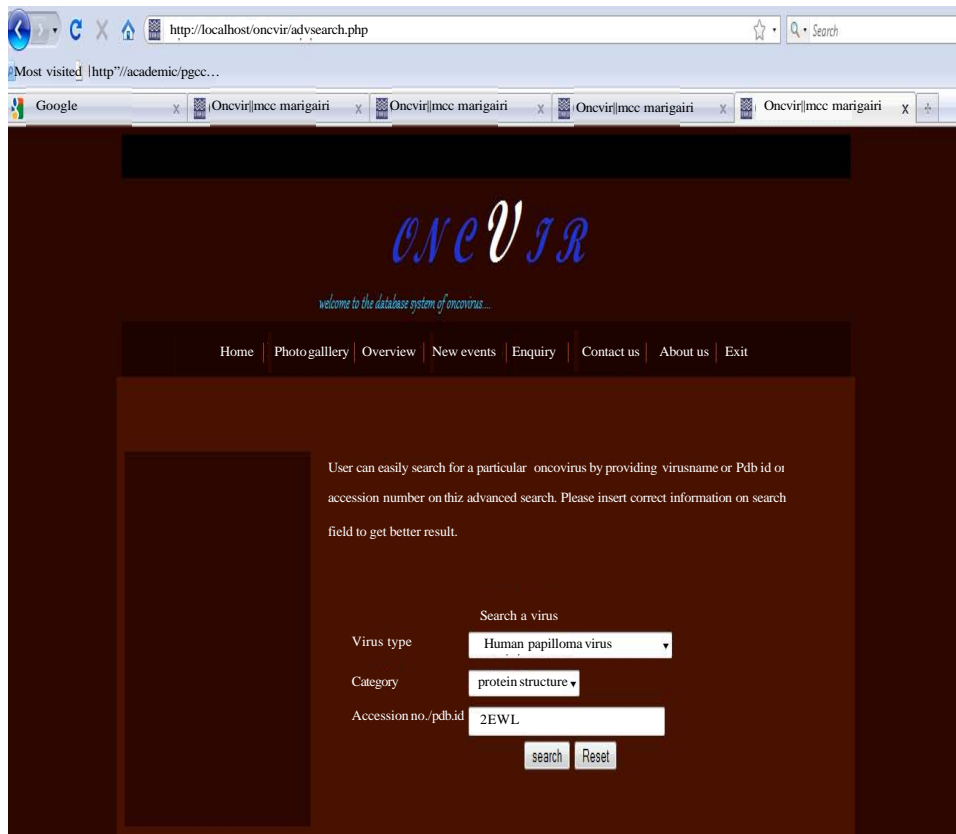


Fig. 5: Advanced keyword search

Entry number, PDB ID, Primary citation (Title), Classification, Structure weight, molecule, length, organism, Experimental method, Resolution and structure.

The database houses two submission tools such as OncSeq (Fig. 6) and OncTer. OncSeq stands for OncVir sequence submitter is web based sequence submission

**ONCVIR**  
welcome to the database system of oncovirus...

Home | Photogallery | Overview | New events | Enquiry | Contact us | About us | Exit

Wellcome aruna

Insert sequence information

\*First name

\*Last name

Department

Institution

\*Qualification

\*Category

\*Type of submission

\*Organism

Other, specify organism

\*Public domain database

Other, specify database

Public Domain id

Fig. 6: Sequence submission tool “OneSeq”

Table 2: Features of protein sequence

Feature	Description
Entry No.	It is a unique number given to each protein sequence entry in to OncVir. It takes the form OVPSE0000.
Identification	It describes the identification number of the public domain databases.
Accession No.	It is a unique identifier given to a protein sequence record to allow for tracking of different versions of that sequence record and the associated sequence over time in a single data repository.
Gene Name	This line gives information on the name of the gene.
Description	The description line(s) is the general descriptive information about the sequence stored, starting with the proposed official name of the protein.
Organism species	The Organism Species line specifies the organism(s) which were the source of the stored sequence.
Organism classification	The Organism Classification line(s) contains the taxonomic classification of the source organism.
Sequence length	The line contains the total length of the amino acid sequence. The length is followed by the letter code ‘AA’ (Amino Acids).
Function	It describes the function of the protein sequence.
Sequence	It shows the protein sequence.

tools which conveniently and quickly submit the sequence information. OncTer stands for OncVir submitter is web based submission tools which conveniently and quickly submit the structure information. This is the first database which comprehends the information about human oncogenic viruses. The database also facilitates the following options:

- **Overview:** It gives overall information about OncVir database
- **Description:** It describes the various human oncoviruses
- **Gallery:** It lists the pictorial representation of human oncoviruses
- **References:** It gives cross references with literature and online resources for further information
- **Help:** It helps the user in analyzing the protein sequence, nucleotide sequence and protein structure information data file
- **New events:** It shows the recent information on oncovirus and provide web link to access that information
- **Enquiry:** It facilitates user to query any general information on OncVir

Few databases relating to viruses such as RNA Virus Database, HCV sequence database, Viral Genome DataBase and VirGen were publicly available. The genome organisation of about 938 known species of RNA virus and certain analysis tools were provided by RNA Virus Database (Belshaw *et al.*, 2009). This database mainly facilitates the comparative analysis of RNA viruses. HCV sequence database (Kuiken *et al.*, 2005) provides

annotated sequences and analysis tools for significant public threat virus such as Human Hepatitis C Virus. Genes and predicted protein sequences of about 15 viruses were dealt in the database named Viral Genome DataBase (Hiscock and Upton, 2000). The genome of these viruses were completely sequenced and of larger in size of about 100 kb. The database includes information such as nucleotide frequency, codon use, isoelectric point (pI), amino acid content, protein sequence, DNA sequence and GenBank and user-entered notes about the viruses of large genomic size. VirGen (Kulkarni-Kale *et al.*, 2004) is an annotated and curated database which provides whole genomic sequences of viruses and various tools for data mining. OncVir differs from all the existing databases by integrating the information concerning only about human oncoviruses.

### CONCLUSION

The challenges posed by viral induced cancer have a global attention. Researchers are trying to solve the mystery and stigma associated with this disease and the discipline Bioinformatics can help them in a greater extend. The development of high-throughput technologies has generated large amount of sequence and structural data. These huge amounts of data are available in public domain databases but mining the specific data is quite a Herculean task. These challenges can be overcome from the database "OncVir".

"OncVir" is a database which was developed using PHP, HTML and JavaScript as front-end and MySQL as back-end. It contains information on human oncoviruses. The information includes the nucleotide sequence, protein sequence, protein structural data as well as some additional features such as description on human oncoviruses, a gallery of oncoviruses, references and new events showing the updated information. The database can be searched by using simple keyword or through advanced search. It also facilitates user to upload sequence and structural data of oncoviruses through OncSeq and OncTer, respectively. It also provides cross-reference to other public domain databases. Thus "OncVir" helps researchers by providing a specific and integrated data. Moreover it remains as a primary accession point for retrieving information related to oncoviruses.

### ACKNOWLEDGMENTS

The authors thank Anjumol Thomas, Aruna S I, Aswathy SJ and Raisen for their technical support. The authors are also grateful to Malankara Catholic College.

### REFERENCES

- Bairoch, A. and R. Apweiler, 1999. The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. *Nucleic Acids Res.*, 27: 49-54.
- Barbacid, M., D. Bolognesi, S.A. Aaronson, 1980. Humans have antibodies capable of recognizing oncoviral glycoproteins: Demonstration that these antibodies are formed in response to cellular modification of glycoproteins rather than as consequence of exposure to virus. *Proc. Natl. Acad. Sci.*, 77: 1617-1621.
- Belshaw, R., T. de Oliveira, S. Markowitz and A. Rambaut, 2009. The RNA virus database. *Nucleic Acids Res.*, 37: D431-D435.
- Benson, D.A., M.S. Boguski, D.J. Lipman, J. Ostell, B.F. Ouellette, B.A. Rapp and D.L. Wheeler, 1999. Gen bank. *Nucleic Acids Res.*, 27: 12-17.
- Bernstein, F.C., T.F. Koetzle, G.J. Williams, E.E. Jr. Meyer and M.D. Brice *et al.*, 1977. The protein data bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.*, 112: 535-542.
- Bryan, B., 2009. Bioinformatics Computing. PHI Learning Pvt. Ltd., New Delhi, India, pp: 42-102.
- Damania, B. and J.M. Pipas, 2009. DNA Tumor Viruses. Springer, New York, USA., pp: 1-79.
- Didier, J., 1983. Human oncogenic viruses. *Sem. Hop.*, 59: 2021-2023.
- Gurtsevitch, V.E., 2008. Human oncogenic viruses: Hepatitis band hepatitis C viruses and their role in hepatocarcinogenesis. *Biochemistry*, 73: 627-639.
- Halimi, M. and S.M. Asl, 2011. Human papillomavirus infection in lung vs. Oral squamous cell carcinomas: A polymerase chain reaction study. *Pak. J. Biol. Sci.*, 14: 641-646.
- Hames, B.D and N.M. Hooper, 2000. Instant Notes in Biochemistry. 2nd Edn., BIOS Scientific Publishers Limited, Oxford, UK.
- Hiscock, D. and C. Upton, 2000. Viral genome dataBase: Storing and analyzing genes and proteins from complete viral genomes. *Bioinformatics*, 16: 484-485.
- Joseph, B. and V.M. Nair, 2012. Onemi Rs: Small noncoding RNA with multifaceted role in cancer. *Res. J. Recent Sci.*, 1: 70-76.
- Kasper, M., G. Regl, A.M. Frischauf and F. Aberger, 2006. GLI transcription factors: Mediators of oncogenic Hedgehog signalling. *Eur. J. Cancer*, 42: 437-445.
- Kuiken, C., K. Yusim, L. Boykin and R. Richardson, 2005. The los alamos HCV sequence database. *Bioinformatics*, 21: 379-384.
- Kulkarni-Kale, U., S. Bhosle, G.S. Manjari and A.S. Kolaskar, 2004. VirGen: A comprehensive viral genome resource. *Nucleic Acids Res.*, 32: D289-D292.



- Leinonen, R., F.G. Diez, D. Birns, W. Fleischmann, R. Lopez and R. Apweiler, 2004. UniProt archive. *Bioinformatics*, 20: 3236-3237.
- Parkin, D.M., 2006. The global health burden of infection-associated cancers in the year 2002. *Int. J. Cancer*, 118: 3030-3044.
- Prajapat, R., A. Marwal, Z. Shaikh and R.K. Gaur, 2012. *Geminivirus* Database (GVDB): First database of family geminiviridae and its genera *Begomovirus*. *Pak. J. Biol. Sci.*, 15: 702-706.
- Zheng, Z.M., 2010. Viral oncogenes, noncoding RNAs and RNA splicing in human tumor viruses. *Int. J. Biol. Sci.*, 6: 730-755.