HPVDB- Human Papilloma Virus Database a Sequence Annotation Database

Kumari vibha¹ Praveen Kumar²
¹,²M.Tech Student
¹, ²Department of Computer Science & Engineering
1, 2School of Science and Engineering, Galgotias University, Greater Noida, Uttar Pradesh, India

Abstract— Human papilloma virus (HPV) is a DNA virus from the papillomavirus family that is capable of infecting humans. Like all papillomaviruses, HPVs establish productive infections only in keratinocytes of the skin or mucous membranes. While the majority of the known types of HPV cause no symptoms in most people, some types can cause warts (verrucae), while others can be in a minority of cases lead to cancers of the cervix, vulva, vagina, penis, oropharynx and anus. Genital human papillomavirus (also called HPV) is the most common sexually transmitted infection (STI). These HPV types can also infect the mouth and throat. Most people who become infected with HPV do not know they have it. HPV is not the same as herpes or HIV (virus results AIDS). There are more than 40 HPV types that can infect males and females. These are all viruses that can be passed on during sex, but they cause different symptoms and health problems.

I. INTRODUCTION
A study that covered 100% of the U.S. population during 2004–2008 estimated that about 33,300 HPV associated cancers occur each year. Cervical cancer is the most common HPV-associated cancer among women, and oropharyngeal cancers [Figure 1] [14] are the most common among men. About 21,300 HPV associated cancers occur each year among females, and about 12,100 occur each year among males. This study used cancer registry data to estimate the amount of potentially HPV-associated cancer in the United States by examining cancer in parts of the body and cancer cell types that are more likely to be caused by HPV. Cancer registries do not collect data on the presence or absence of HPV in cancer tissue at the time of diagnosis. In general, HPV is thought to be responsible for more than 90% of anal and cervical cancers and more than 50% of vaginal, vulvar, and penile cancers. Cancers of the head and neck are mostly caused by tobacco and alcohol, but recent studies show that about 60%–70% of cancers of the oropharynx may be linked to HPV. Many of these may be caused by a combination of tobacco, alcohol, and HPV.

II. HPV (HUMAN PAPILLOMA VIRUS)
Human papilloma virus (HPV) is a DNA virus from the papillomavirus family that is capable of infecting humans. Like all papillomaviruses, HPVs establish productive infections only in keratinocytes of the skin or mucous membranes. While the majority of the known types of HPV cause no symptoms in most people, some types can cause warts (verrucae), while others can be in a minority of cases lead to cancers of the cervix, vulva, vagina, penis, oropharynx and anus. [1] Recently, HPV has been linked with an increased risk of cardiovascular disease. [2] In addition, HPV 16 and 18 infections are strongly associated with an increased odds ratio of developing oropharyngeal (throat) cancer. [3]

More than 30 to 40 types of HPV are typically transmitted through sexual contact and infect the anogenital region. Some sexually transmitted HPV types may cause genital warts. Persistent infection with “high-risk” HPV types—different from the ones that cause skin warts—may progress to precancerous lesions and invasive cancer.[4] HPV infection is a cause of nearly all cases of cervical cancer.[5] However, most infections do not cause disease.

Most HPV infections in young women are temporary and have little long-term significance. Seventy percent of infections are gone in 1 year and ninety percent in 2 years.[6] However, when the infection persists—in 5% to 10% of infected women—there is high risk of developing precancerous lesions of the cervix, which can progress to invasive cervical cancer. This process usually takes 10–15 years, providing many opportunities for detection and treatment of the pre-cancerous lesion. Progression to invasive cancer can be almost always prevented when standard prevention strategies are applied, but the lesions
still cause considerable burden necessitating preventive surgeries, which do in many cases involve loss of fertility.

In more developed countries, cervical screening using a Papanicolaou (Pap) test or liquid-based cytology is used to detect abnormal cells that may develop into cancer. If abnormal cells are found, women are invited to have a colposcopy. During a colposcopic inspection, biopsies can be taken and abnormal areas can be removed with a simple procedure, typically with a cautering loop or, more commonly in the developing world by freezing (cryotherapy). Treating abnormal cells in this way can prevent them from developing into cervical cancer.

III. FACT SHEETS

Pap smears have reduced the incidence and fatalities of cervical cancer in the developed world, but even so there were 11,000 cases and 3,900 deaths in the U.S. in 2008. Cervical cancer has substantial mortality in resource-poor areas; worldwide, there are an estimated 490,000 cases and 270,000 deaths each year. [7][8]


execute a quick and efficient search on HPVDB data. The database can be queried comprehensively through argument functions such as NCBI Locus number, different protein name, different predicted functional family and stability data. HPVDB is an extremely useful resource for computational and experimental biologist working in related areas.

IV. METHODOLOGY

System architecture and design a relational database was constructed in MySQL which facilitate storage, query and visualization of annotation information. It includes three key entities: ‘functional analysis’, ‘molecular analysis’ and ‘cleavage sites’, for proteins. This information is managed at a protein level to provide a general view of the data. The HPVDB data and related information are stored in MySQL relational database tables. Meta-information for different types of biological data is stored in layers of tables. The application layer between the web interface and the backend relational tables has been implemented using PHP. Database features Data access HPVDB can be queried to obtain the information about the protein sequences in many ways.

Data stored in HPVDB can be accessed in the following ways:

1. Search by protein name: The user can enter the desired protein name to access the Meta information about the protein sequences;

2. Search by protein functional family: The user can select the different protein functional family to find out the protein functional group of different structural and non-structural proteins;

3. Search by NCBI locus ID: The user can enter the NCBI locus ID to obtain Human Papiloma virus protein sequence information;

4. Search by Instability Index: To find out the stable and unstable protein, user can search by instability index;

5. Compare two proteins: HPVDB can be queried to obtain the information about protein-protein comparison. The user can enter the corresponding NCBI locus ID or select the protein name to compare two proteins.

Database visualization helps the user to process, interpret and act upon large stored data sets. HPVDB provides a number of web-based forms for querying the dataset and selecting either a more detailed view of molecular annotation, cleavage sites or functional family or for viewing the comparison between two selected proteins. In an effort to improve access to diverse HPV data, the HPVDB has been modified to include an abundance of linkage to other database including PUBMED [13] for related paper abstracts and NCBI for corresponding sequences. Data analysis the protein function family predicted by SVMProt is different for each structural and non-structural protein of Human Papiloma virus strain, some of which may be responsible for virulence or pathogenicity of the virus and others for replication of the virus in the host. Prediction of the functional roles of lipid binding proteins is important for facilitating the study of various biological processes and the search for new therapeutic targets. Comparison of two amino acid sequences of any Human Papiloma virus protein will reveal the user, the distinguished functional properties of the corresponding protein, if there is any amino acid change at any position as SVM works on the basis of physico-chemical properties of the amino acids of the protein. In an example, when comparing functional assignment of two different proteins (RNA-dependent RNA polymerase (RDRP) and matrix protein M), where functions assigned to each protein is different (functions like phosphotransferases, glycosyltransferases and mRNA capping are specific to RDRP and whereas, functions like zinc-binding, metal-binding, lyases (carbon-oxygen lysases), calcium-binding, DNA repair, copper-binding and magnesium-binding are specific for matrix protein M). It is indicated from this analysis that each protein performs a specific function assigned and evolved by the viral genome. Comparison of functions of other sequences e.g. NS protein and matrix protein reveals that zinc-binding and DNA repair functions are common to both the proteins, whereas the hydrolases (acting on ester bonds) function is specific to NS protein. However, metal-binding, magnesium-binding & carbon-oxygen lysases are specific to matrix protein. Patterns of restriction sites for all types of restriction enzymes in Human Papiloma virus are visualized using the web server.

V. CONCLUSION

HPVDB has been designed to manage and explore the vast amount of viral protein data analysis. The current version of HPVDB has provides the information on the molecular and functional analysis of data in Human Papiloma virus. HPVDB has been developed with the availability Human Papiloma virus proteins in public domains. We plan to include the modeled structures of different Human Papiloma virus proteins and analyze quantitative structure-activity relationship of novel ligands targeting different proteins in the future. The database will be updated monthly on the basis of additional data availability from analysis of the Human Papiloma virus sequences from other reliable resources.

REFERENCES

[14] http://www.cdc.gov/std/stats12/figures/45.htm Figure [1][2].