

SCIENTIFIC REPORTS

OPEN

HPVbase – a knowledgebase of viral integrations, methylation patterns and microRNAs aberrant expression: As potential biomarkers for Human papillomaviruses mediated carcinomas

Received: 04 November 2014

Accepted: 26 June 2015

Published: 24 July 2015

Amit Kumar Gupta & Manoj Kumar

Human papillomaviruses (HPVs) are extremely associated with different carcinomas. Despite consequential accomplishments, there is still need to establish more promising biomarkers to discriminate cancerous progressions. Therefore, we have developed HPVbase (<http://crdd.osdd.net/servers/hpvbase/>), a comprehensive resource for three major efficacious cancer biomarkers i.e. integration and breakpoint events, HPVs methylation patterns and HPV mediated aberrant expression of distinct host microRNAs (miRNAs). It includes clinically important 1257 integrants and integration sites from different HPV types i.e. 16, 18, 31, 33 and 45 associated with distinct histological conditions. An inclusive HPV integrant and breakpoints browser was designed to provide easy browsing and straightforward analysis. Our study also provides 719 major quantitative HPV DNA methylation observations distributed in 5 distinct HPV genotypes from higher to lower in numbers namely HPV 16 (495), HPV 18 (113), HPV45 (66), HPV 31 (34) and HPV 33 (11). Additionally, we have curated and compiled clinically significant aberrant expression profile of 341 miRNAs including their target genes in distinct carcinomas, which can be utilized for miRNA therapeutics. A user-friendly web interface has been developed for easy data retrieval and analysis. We foresee that HPVbase an integrated and multi-comparative platform would facilitate reliable cancer diagnostics and prognosis.

The human papillomaviruses (HPVs) are circular, double-stranded DNA genome of approximately 8.0 kb in length. It belongs to the papillomaviridae family, which is further taxonomically classified into distinct genera namely alpha, beta, gamma, mu and nu¹. HPVs encode eight well-defined open reading frames (ORFs) along with one non-coding long control region (LCR) or regulatory region. HPV proteins are mainly divided into two coding regions classified as early (E) and late (L) region. E region includes six ORFs encoding 3 functional regulatory genes (E1, E2, E4), 3 oncogenes (E5, E6, E7) and the L region encodes the two viral capsid genes (L1 and L2)¹⁻³.

Bioinformatics Centre, Institute of Microbial Technology, Council of Scientific and Industrial Research (CSIR), Sector 39-A, Chandigarh; 160036, India. Correspondence and requests for materials should be addressed to M.K. (email: manojk@imtech.res.in)

HPVs are well adapted in host differentiating epithelial tissues and prevalent in cervical carcinogenesis. They employ cellular machinery to survive in host. Expression and functions of viral proteins primarily HPV oncoproteins E6 and E7 is very essential for the tumor progression. Upregulation of these oncoproteins in the infected cells have been shown to expedite proliferation, immortalization and malignant transformation. They interact with numerous cellular targets mainly inhibits two essential tumor suppressor proteins, p53 and retinoblastoma (Rb). Interaction of E6 with p53 prohibits proteolytic degradation and abrogates apoptosis⁴. In addition, binding of E7 to pRb inhibits cell cycle arrest^{5,6}.

Cervical carcinoma (CC) is the fourth most common cancer and cause of death in women worldwide (GLOBOCAN 2012). Series of steps involves in cervical oncogenesis are HPVs infection and transmission, viral persistence, progression of infected epithelium to precancer or high-grade precursor lesion and invasion. Persistence infection with distinct carcinogenic HPVs is a critical factor in the etiology of cervical precursor lesions and cancer^{2,7,8}. Prodigious involvement and influence of HPVs in distinct carcinomas has accelerated the research in screening and prevention strategies for cancers. Moreover, cytology and HPV-based screening methods have greatly reduced the cancer morbidity and mortality rate among women⁹. However, despite of consequential accomplishments, sensitivity and specificity of these methods remains a challenge to make decision for colposcopic referral of women. Furthermore, these are not sufficient to discriminate transient infection with high-grade carcinomas¹⁰. Additionally, there is no effective treatment and therapeutics for HPV infection is available to eradicate the CC up to now and it still remains prevailing cause of cancer death^{8,10}.

HPV infection and associated consequences are valuable determinants of multistage cervical neoplastic progression. In this context, there are alternative biomarkers, which could discriminate latent infection with high-grade precursor lesion and eventually cancer⁸. Potential biomarkers pertinent to HPV mediated diseases are mainly viral DNA integration, viral methylation and cellular miRNAs expression. These could improve the reliability of cancer screening and prevention^{8,11}.

The infection and integration of HPVs is a well-known factor and event in the development of CC, human genital and head and neck squamous cell carcinoma (HNSCC)^{12–18}. They are known to be integrating into human chromosomes during the progression of cancers, which provide enhancement and stabilization of HPV oncogenes (E6 and E7) transcription^{19,20}. These integration sites found to be distributed all over the human genome with a preference in or near to fragile sites (FS) or oncogenes and translocation breakpoints²¹. It usually rises in high-grade lesions and carcinomas with uneven integration frequency in distinct cases^{12,22}. However, in case of low-grade lesions and benign, HPV genomes persist mainly in episomal state²³.

In addition, epigenetic mechanisms are also known to influence chromatin conformation and alter gene expression. CpG methylation is a fundamental and well-recognized molecular regulatory mechanism in epigenetics^{24,25}. Methylations play a crucial role in the HPV transcriptional modulation. Although the underline molecular basis and determinant of methylation at individual CpG sites during the transformation is still poorly understood²⁶. The major observations from previous studies, guide towards two distinct types of molecular mechanisms. One, targeted methylation may contribute in neoplastic progression by blocking binding sites (E2BSs) of HPV E2 repressor, which stimulate oncogene (E6 and E7) expression, thus enhance the carcinogenesis. Second, it may be a consequence of host defense mechanism against integrated HPV genome for the inhibition of viral replication and transcription through de novo methylation^{27–29}.

Apart from the general concepts of genomic instability govern by carcinogenic HPVs the mechanism of transformation is more complex. The viral oncoproteins (E6 & E7) are known to interact with several cellular components and may affect different cellular pathways⁶. The viral oncogenes E6 and E7 deregulate tumor suppressive and oncogenic miRNAs (Oncomirs)^{30,31}. In addition, host miRNAs also influence the papillomavirus genes expression by targeting viral RNA transcripts thus provide an interplay mechanism between cellular miRNAs and HPV oncogenes³². Mature miRNAs are approximately 18–25 nucleotides in length and are found in the wide range of organisms including viruses³³. MiRNAs alters the gene expression either by repressing mRNA translation or by catalyzing the mRNA cleavage^{34,35}. They usually affect various molecular processes such as apoptosis, cell proliferation, morphogenesis, human tumorigenesis, chromatin modifications etc³⁶.

Despite of numerous experimental studies in the field, there is a paucity of computational resources with a unique focus on cancer progression and therapeutically potential biomarkers. Moreover, it is still unclear and less known about the viral mediated processes, such as integration events, methylation pattern in different epidemiological conditions and above all importance of various cellular miRNAs expressions in the etiology of oncogenesis. Hence, it is most important and necessary to have a proper multi-comparative platform of these specific biomarkers. Here, the goal of “HPVbase” is to provide a comprehensive web-based resource of the extensively cross-analyzed different biomarkers.

Results

HPV integration and breakpoints. The importance and crucial role of HPV integration and human genome integration sites in the progression and etiology of CC were identified and analyzed in several studies. However, complete knowledge and understanding of integration mechanism is still unclear. An inclusive HPV integrant and breakpoints browser were designed to provide analysis and interpretation. In total, 1257 integrants and integration sites were collected which are associated with distinct histological

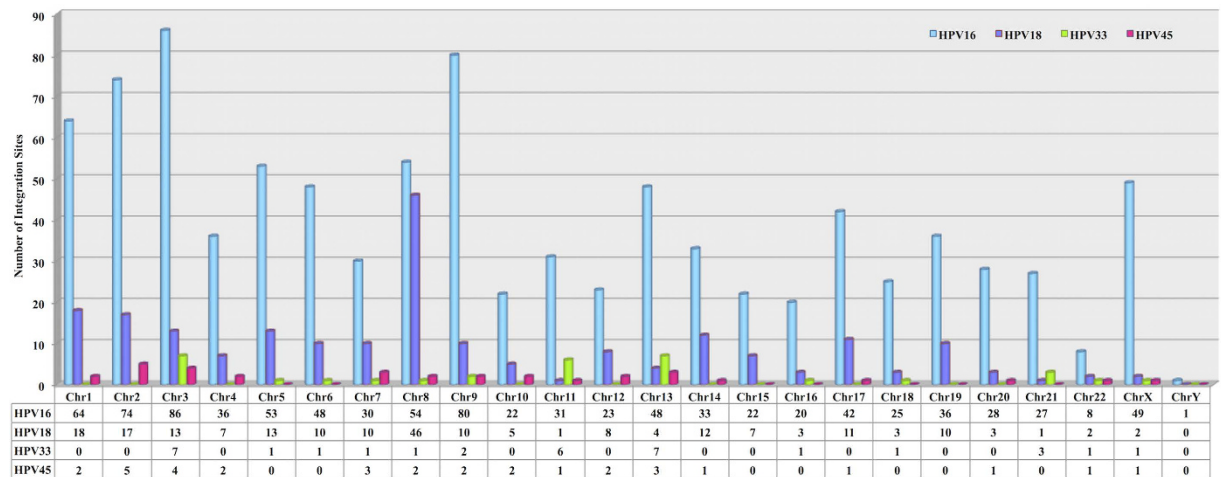


Figure 1. Bar chart representing distribution of HPV16, HPV18, HPV33 and HPV45 integration sites on human genome.

conditions. These integration sites are distributed among several HPV types mainly on HPV16, 18, 45, 33 and 31. It comprises of various clinical information such as viral integrant, host integration sites, gene target region, fragile sites, cancer specimen type, sample size, carcinoma types, detection techniques etc.

Using the highly interactive browser user can find and browse distinct HPV integrants. Different color-coded blocks provide corresponding information and existing analysis of HPV integration (see Supplementary Figure S1). All clinical histological information was comprehended using the set of tracks. Users can easily find respective details regarding integration events and utilize it to map and compare with new experimental data. Supplementary Figure S2 shows the distribution of integration sites among distinct HPV types. Most of these sites belong to HPV16 (954), HPV18 (216), HPV45 (33) and HPV33 (33). Additionally, we have also reported and depicted frequency distribution of HPV integration events based on specific disease (cancer) types and chromosomal loci (see Supplementary Figure S3-S4). As shown in Supplementary Figure S5, integration data in HPVbase is also represented in user-friendly tabular format with supporting additional annotations with reference links. Users can find exact chromosomal location, cytoband and target host genes using the HPV integration and breakpoints table. This knowledge can be utilized to optimize and accelerate clinical biomarker research.

The most important and critical part of integration events is their frequency to integrate on different chromosomes and genomic region. To facilitate this, we have also analyzed HPV type wise distribution and frequency of integration and breakpoints on different human chromosomes. Figure 1 depicts chromosomal distribution frequency of HPV16, 18, 33 and 45 integration sites respectively. Using this we can find prominent genomic target regions in human genome liable for higher instability. As delineated in Fig. 1, HPVs integration sites and breakpoints are distributed all over the human genome but some genomic regions show higher liability towards disruption and can be used as the hot spot for cancer clinical research. We found that HPV16 is used to prefer and highly integrate at chromosome 3, 9, 2, 1, 8 and 5. Similarly, HPV18 mainly favors chromosome 8, 1, 2, 5 and 3. Likewise, HPV33 and HPV45 integration sites are almost evenly distributed among different locations.

Moreover, we have also analyzed the complete gene compendium altered/disrupted due to integration events presented in HPVbase. First of all, we have extracted missing gene name information based on human genomic coordinates using UCSC genome browser. Subsequently, from entire set after removing duplicates and other non-genes (pseudo genes, LINC, non-coding RNAs) we have listed 481 unique genes. Furthermore, disruption frequencies of each unique 481 genes from whole dataset were calculated (see additional file 2 Table S3). Figure 2 portrays frequencies of disrupted genes, which are at least represented twice in complete gene set. These genes are predominantly involved in metabolic, biological regulations and apoptotic processes.

HPV methylation patterns. Extensive research in the quantitation of methylation in the carcinogenesis allowed searching methylation patterns in different diseased stages. Our study provides major quantitative HPV DNA methylation observations together and the probable combination of possible CpG methylation sites that could be used to discriminate normal and cancerous progressions. Distinct studies provide different notations for methylation status. To simplify the terminologies, based on methylation levels we have used three categories: hypo or low methylation (HypoM), hyper or high methylation (HyperM) and significantly hyper methylation (HyperM##). Here, we have cataloged 719 methylation entries along with respective important and clinical information. It comprises of CpG methylation sites, gene region, differential methylation patterns, detection techniques, sample size and disease sample type

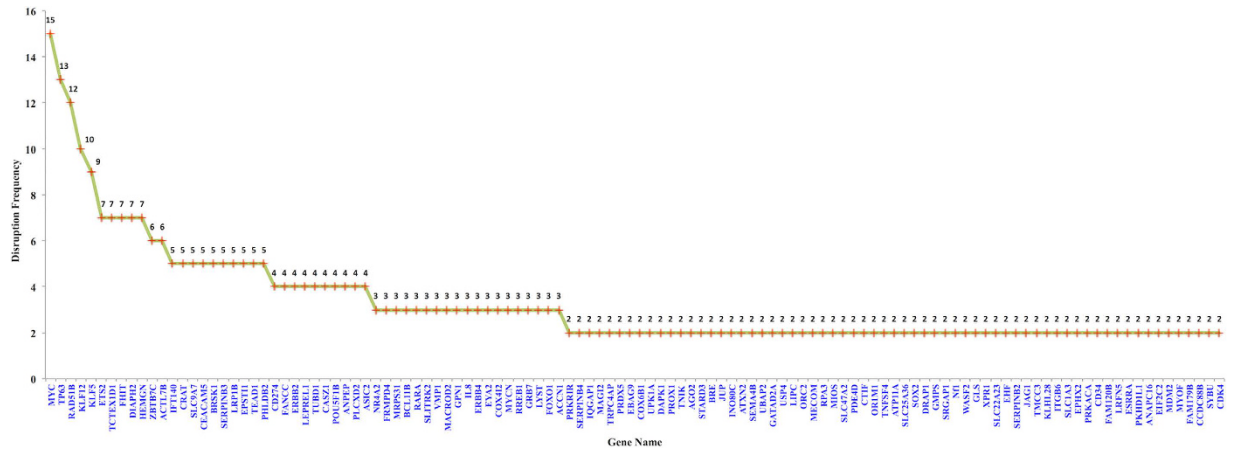


Figure 2. Depicting frequency of genes disrupted due to viral integration events.

use to identify methylation status (see Supplementary Figure S6). These are distributed in 5 distinct HPV genotypes from higher to lower in numbers namely HPV16 (495), HPV18 (113), HPV45 (66), HPV31 (34) and HPV33 (11) (see Supplementary Figure S7).

Assorted cancer studies have advocated an association between CpG methylation patterns and the carcinogenic development. Integration of episomal HPV genomes in carcinoma usually correlates with elevated DNA methylation. Our resource also provide diverse pattern in methylation outcome in corresponding specimen type, histological conditions and detection techniques used. In HPV positive women, the methylation level at specific CpGs increase with consistent viral infection and even enhance more remarkably in high-grade lesions. Maximum sites from late HPV regions L1 and L2 exhibit conclusive and significant hyper methylation patterns. However, methylation pattern in adjacent long control region (LCR) shows relative inconsistency. It could provide a comprehensive base for comparative examination between distinctive features and cancers, which in turn facilitate advancement in clinical screening tests. Furthermore, it could be used as a cost-effective molecular tool for the triage of HPV⁺ women.

Cellular miRNAs aberrant expression. In various studies, miRNAs are suggested being involved in pathogenesis of cancers usually in CC and HNSCC including other cancer types i.e. breast cancer, thyroid papillary malignancy, ovarian cancer, prostate cancer etc^{37,38}. The aberrant (increased or decreased) expression of miRNAs is evidentiary in diverse HPV mediated cancers and viral diseases^{37,38}. Thus, the screening for differential expression profile of miRNAs in cancerous and normal tissues may help to examine the role of miRNAs in cancer pathogenesis.

To facilitate this, we developed an integrated source for the cellular miRNAs with description, which could be explored for greater insight between miRNAs and HPV related tumorigenesis. Regulation of cellular miRNAs is key to prognosis. We have curated and compiled clinically significant aberrant expression profile of 341 miRNAs from different carcinomas. It includes 142 miRNAs, which comprises of 80 upregulated and 62 downregulated in CC, 176 miRNAs among which 85 and 91 are up and downregulated respectively from HNSCCs and 22 miRNAs includes 9 upregulated and 13 downregulated from vulvar carcinoma. Beside this, 1 downregulated miRNA is related to penile carcinoma. It includes miRNAs and their mirbase id, host chromosomal location, genome coordinates, their potential target genes as shown in Supplementary Figure S8. These miRNAs affects various target genes, which mainly influence apoptosis, metastasis, cell proliferation, senescence, host defense mechanism, immune recognition and genomic stability with other host mechanisms. An expression profile of these miRNAs and their target genes can be utilized as the prognostic marker for aggressiveness of cancer.

We have cross-analyzed chromosomal distribution of these aberrantly (up and down) regulated miRNAs in different carcinomas i.e. CC, HNSCC and vulvar carcinoma. Figure 3 shows category wise distribution of these miRNAs on genomic region. Using this we can find chromosomal distribution relationship between integration hot spots and various miRNAs expression profile in distinct carcinoma. Along with this, we have also explored intra-relationship between these miRNAs expression profile and found set of commonly up regulated (see Fig. 4a) and down regulated (see Fig. 4b) miRNAs in diverse cancers. It is also interesting to notice some miRNAs shows both down and up regulation with in specific cancer type. We have identified 23 (see Fig. 4c) and 18 (see Fig. 4d) miRNAs, which are both over and under expressed in HNSCC and CC respectively. Dual expression can be a further area of research to find their specific role in varied conditions.

Utility as potential biomarker. The vital role of HPV integrations and human genome chromosomal sites were analyzed and illustrated in various studies. HR-HPVs integration is a key incident in HPV-mediated carcinogenesis, which leads to increase expression and stability of viral oncogenes (E6/E7)

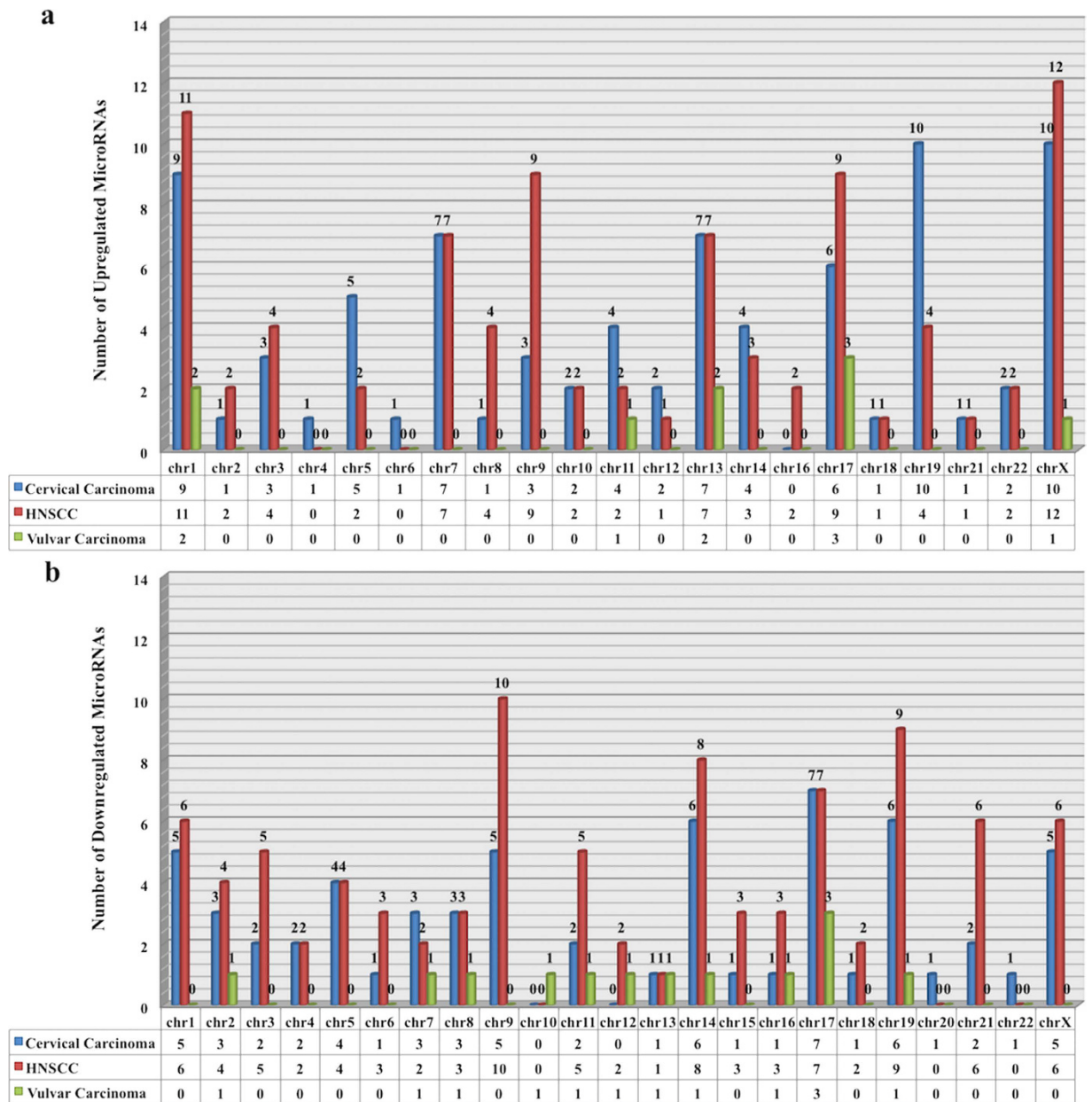


Figure 3. Plot showing chromosomal distribution of aberrantly expressed miRNAs in HPV associated carcinomas. (a) Upregulated miRNAs, (b) Downregulated miRNAs.

as well as disruption of various key cellular proteins. It is usually considered as one of the main events takes place during the transition from low grade to high-grade lesions. Additionally, it was also observed that these are frequently allied with FS normally play a role in the structural alterations, translocations and large deletions. Thus, it is one of the important mechanisms that enhances and contributes in aberrant proliferation, genomic instability and cellular immortalization, which enhance malignant progression. It plays an acute function not only to alter viral genome as well as host genomic regions mainly associated with cancer related genes.

From our analysis, we have identify and reported certain specific chromosomal loci regions such as 8q24.21, 3q28, 13q22.1, 9q22.33, 14q24.1 and many more which shows higher level of viral integration indicating preferential target in diverse carcinomas (see Supplementary Figure S4). In this study, we have also catalogued and represented a comprehensive set of disrupted genes due to viral integration events (see Fig. 2). Among these genes most of them are interrelated with tumor development and various regulatory processes that could jointly become involved in viral induced oncogenesis. Such as RAD51B, TP63, MYC, ETS2, FHIT etc.

For example, TP63 is a tumor protein 63 belongs to family of transcription factors. It mainly involves in complex events and regulates cellular proliferation and neoplastic progression. MYC, it is a v-myc

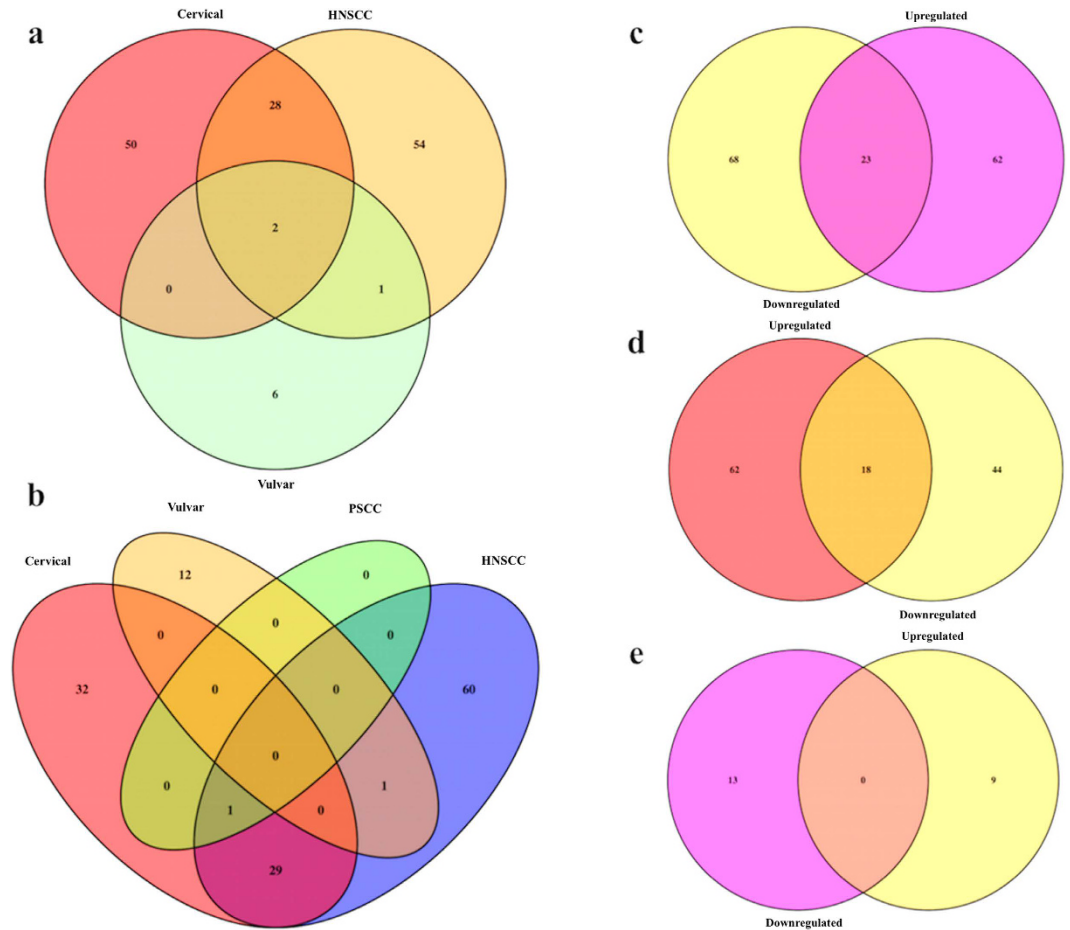


Figure 4. Figure illustrating commonly regulated miRNAs in diverse carcinomas. (a) Upregulated miRNAs, (b) Downregulated miRNAs, (c) Up and down regulated miRNAs in HNSCC, (d) Up and down regulated miRNAs in CC, and (e) Up and down regulated miRNAs in vulvar carcinoma.

avian myelocytomatosis viral oncogene homolog usually plays a role in cellular transformation, apoptosis and cell cycle. ETS2, v-ets avian erythroblastosis virus E26 oncogene homolog 2 is a tumor suppressor gene mainly regulates genes involved in development and apoptosis. It also found to be involved in regulation of telomerase. FHIT, belongs to histidine triad protein family is a tumor suppressor gene. It encloses the common fragile site FRA3B and highly susceptible to the translocations and aberrant expression which leads to the various carcinomas. RAD51B (RAD51 paralog B) is a member of the RAD51 family. It is essential element of the DNA repair pathway and found to be associated with cell cycle delay and cell apoptosis. Hence, integration at unstable regions and at certain genes concomitant with DNA repair mechanism also signifies its relation with various chromosomal machineries.

These interpretations from existing knowledge suggest that prevalence of certain viral integration sites on specific genomic locations could play an essential role to determine eventual cancer phenotype. Therefore, revealing these imperative events in the transforming lesions from low grade to invasive carcinomas might be useful and precious to be used as potential cancer biomarker.

Apart, from genetic changes due to integration, HPVs are also known to cause epigenetic alterations mainly modification in DNA methylation pattern and miRNAs aberrant expression. These mechanisms contribute to the development of cancer either reducing expression of tumor suppressor genes or the enhancing oncogene expression supporting cancer progression. With the advent of high throughput technologies, researchers are now focusing on the application of these alterations to explore their prognostic and diagnostic value to determine disease progression^{39,40}. The careful quest suggests that methylation patterns and rate differ according to pathological conditions and severities^{41,42}. The profound research in quantification of HPV DNA methylation can be utilized as a predictive prominent biomarker to distinguish transient infection from those that evolve to cancerous state. Data from our resource could provide broad spectrum of HPV methylome from diverse cancers. This may help research community to identify certain combination of HPV methylation pattern to discriminate malignant from normal tissues.

Additionally, differential expression of oncogenic and tumor suppressive miRNAs has confounding affect on cancer paradigm. They generally post transcriptionally modulate or regulate expression of target

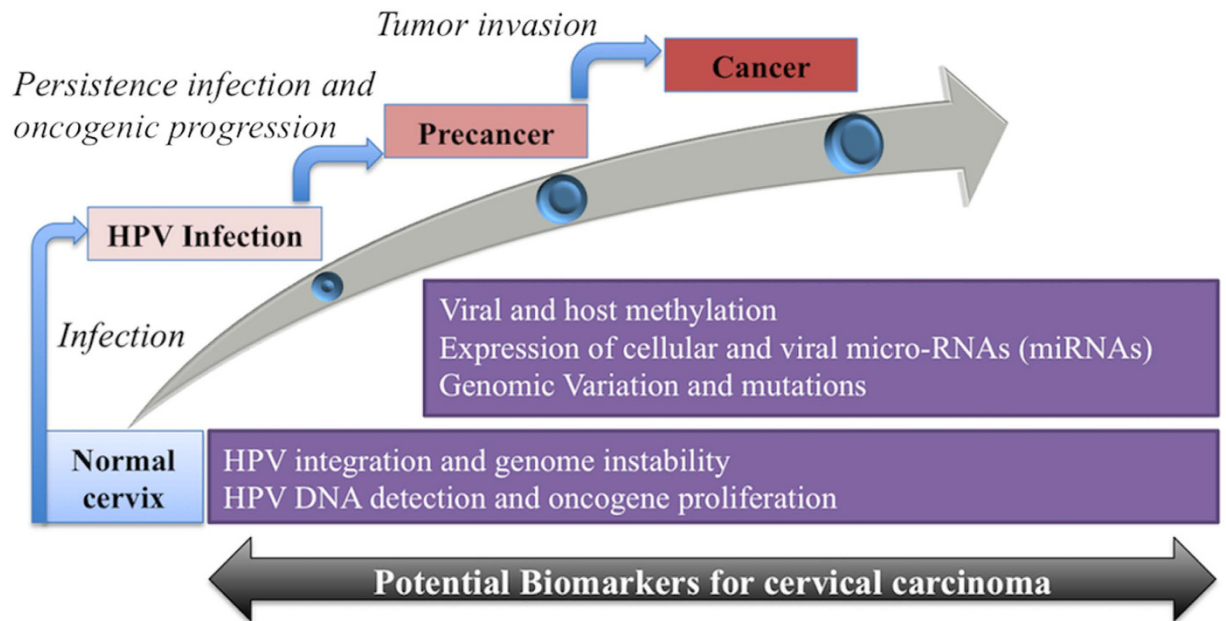


Figure 5. Multistage neoplastic progression and biomarkers overview.

genes and play significant role in cell cycle progression, apoptosis, angiogenesis etc. the observation of different circulatory miRNA signatures might be useful as diagnostic tool as well as could serve as therapeutic target in HPV induced oncogenesis. Furthermore, distinct characteristics of miRNAs such as resistant to temperature, degradation and non-immunogenic nature to utilize host regulatory machinery make them more appropriate to serve as prognostic indicator of cancer aggressiveness.

Some approaches like restoring miRNA level or blocking its function can be used in drugs development. Concluding same, certain miRNA based cancer-targeted drugs such as miR-34-mimic (MRX34); a liposome-based drug is in clinical trial for the treatment of hepatocellular carcinoma. Likewise, “miravirsen”, antisense oligonucleotides based an antiviral compound known as locked nucleic acids (LNA anti-miRs) is in clinical evaluation for the treatment of hepatitis c virus (HCV)^{43,44}. Here, we are providing set of HPV regulated miRNAs from various carcinomas and their expression profile, which could be useful in further research to develop novel cancer screening strategies and therapeutic measures.

Discussion

HPVs play a cardinal role in the etiology of distinct carcinomas including CC, HNSCC etc. CC is still account as fourth most common cancer globally and remained as leading cause of cancer mortalities. However, regardless of substantial efforts in the field to develop screening and prevention methods, there is no effectual solution is available to combat against these lethal diseases. The general progression of HPV mediated cancer is associated with the various events and changes with other risk factors. These events can be used as biomarkers. Here, the main aim and feature of HPVbase is to deliver integrative and multi comparative platform for HPV integrations, methylation patterns and cellular miRNAs expression to facilitate identification and analysis of valuable cancer biomarkers. These are very essential factors especially in cervical progression (see Fig. 5).

The HPVbase is the first web knowledgebase to provide highly interactive and manually curated resource for the clinically significant viral and cellular biomarkers. Nevertheless, we have also searched and compared existing sources related to these events and papillomaviruses. For viral integration sites, DR. VIS^{45,46} is the only database reported till now, which also host HPV associated entries. However, in HPVbase, we have compiled 1257 integration and breakpoint events associated with different clinical and histological conditions. Whereas, we found 8 methylation related databases (see additional file 1, Table S1). All these databases only address host specific methylation patterns. However, we are first time providing integrated resource for the HPV methylome profile in different carcinomas with relevant clinical information. Simultaneously, there are mainly 19 data sources is available related to microRNAs (see additional file 1, Table S2), from which no one provides information pertinent to HPV mediated aberrant expression of cellular microRNAs in distinct carcinomas. Here, we are presenting a comprehensive catalogue of 341 deregulated miRNAs with additional knowledge such as histological condition, potential target genes; their chromosomal locations etc. This can be valuable for the development of miRNA-based therapeutics.

Some sources also developed which provide HPV genomic and proteomic information. The Papillomavirus Episteme (PaVE) contains 303 genomes, 3126 gene region, 2823 protein sequences and

48 protein structures (<http://pave.niaid.nih.gov>)⁴⁷. The Human papillomavirus T cell Antigen Database (HPVdb) hosts 2781 antigen entries of antigenic proteins, 191 T cell epitopes and 45 human leukocytes antigen (HLA) ligands with alignment and search tools (<http://cvc.dfc.harvard.edu/hpv/>)⁴⁸. However, no resource is available related to HPV mediated cancer biomarkers. To support cancer biomarker discovery and utilize existing knowledge and information, we have developed a web-based interactive resource “**HPVbase**” for the extensive comparative analysis of different biomarkers and genomic information. It will provide systematic compilation of knowledge and user-friendly visualization of clinical content. It can be exploited for assessment of clinically significant biomarker data and will provide a comprehensive base for examination between distinct features and cancers, which in turn facilitate advancement in clinical screening tests. Finally, combined impact of multiple events can be utilized for the future prevention and therapeutics. Thus, in turn this exhaustive platform promises to be precious to facilitate and accelerate reliable cancer diagnostics and prognosis.

Materials and methods

Biomarker data collection and curation. A systematic and extensive search was performed using keywords from PubMed. All exhaustive clinical studies reporting (1) the integration and breakpoint events in HPV associated carcinomas (2) HPVs methylation events relevant to the cancer clinical outcomes and (3) the HPV mediated aberrant expression of distinct host miRNAs that are crucial in cancer etiology were included and analyzed.

HPV integration and breakpoints. Collection of integration and breakpoints were derived from scientific literatures searched using set of keywords “((((HPVs) OR human papillomaviruses) OR human papillomavirus) OR HPV*) AND (((cancer) OR carcinoma)) AND ((integration*) OR breakpoint*)” as query. Overall, 755 articles were found, out of which 117 review articles were excluded from study and 638 research papers were screened for extensive data mining. Furthermore, maximum studies only state HPV status or the information related to presence or absence of HPV genes such studies does not provide integration sites, coordinates and further clinical details were filtered and removed. After vigilant reading, comprehensive knowledge associated with integration events were extracted such as HPV type (e.g., HPV16, HPV18), HPV region (e.g., E6), viral integrant/breakpoints (e.g., 84–107), human chromosome number (e.g., 3) and coordinates (e.g., 47984368–47984397), cytoband (e.g., q21.3), target genes (e.g., SYN3, PROX1), fragile sites (e.g., FRA2K), detection technique or methods (e.g., RNA-seq, APOT assay), histological conditions (e.g., CC, HNSCC), specimen or sample type (e.g., tumor sample, cell line) etc. The collected data were manually checked and rectified for errors and inconsistencies.

HPV DNA methylation patterns. The HPV DNA methylation studies were extensively searched using combination of keywords “((((HPVs) OR human papillomaviruses) OR human papillomavirus) OR HPV*) AND (((cancer) OR carcinoma)) AND methylation”. In total, 320 articles excluding 31 review articles were screened. Most of the literature covers host genome or gene methylations, which are not in the scope of this study and hence excluded. All-inclusive information related to HPV methylation i.e. HPV type (e.g., HPV16), its genomic region (e.g., E2), methylation status (e.g., HyperM), technique used for detecting methylation (e.g., bisulphite sequencing), sample type (e.g., clinical biopsy samples), associated cancer type (e.g., CC) or stage (e.g., cervical intraepithelial neoplasia (CIN)) etc. were mined and compiled.

Cellular miRNAs aberrant expression. All literature information were collected using keywords query “((((HPVs) OR human papillomaviruses) OR human papillomavirus) OR HPV*) AND ((cancer) OR carcinoma) AND (((microRNA) OR miRNA) OR microRNAs) OR miRNAs)” from PubMed. Overall, 99 articles excluding 24 review articles were searched. We have extracted information comprises of miRNAs, their expression profile, mirbase id, host chromosomal location and genome coordinates from distinct human carcinomas. Apart from manually curated expression profiles, these are further integrated and linked with the external resources. Along with this, we also explored and obtained known target genes of different miRNAs using mirtarbase⁴⁹.

HPVbase interface and construction. The web interface (front and back end) is developed using the open source LAMP (Linux-Apache-MySQL-PHP) solution stack reinforced by combination of programming and scripting languages i.e. HTML, javascript and PERL. We have also utilized the JBrowse⁵⁰, a java-script based genome browser to develop descriptive sections related to specific biomarkers using highly lightweight interchangeable JavaScript Object Notation (JSON) data format. The all data files were converted using custom Perl script in gene feature files (GFF3) format with some manual editing and configured using JBrowse to visualize interactively. The complete system hosted on IBM machine under Red Hat Enterprise Linux 5 environment using MySQL client version 5.0.51 b, PHP 5.2.14 and Apache 2.2.17 server.

HPVbase organization and search tools. HPVbase is a first web-based, HPV knowledgebase for the effective analysis and assessment of viral and cellular biomarkers involves in pathogenicity of different HPV mediated carcinomas. It has well-designed user interface for the interactive visualization and

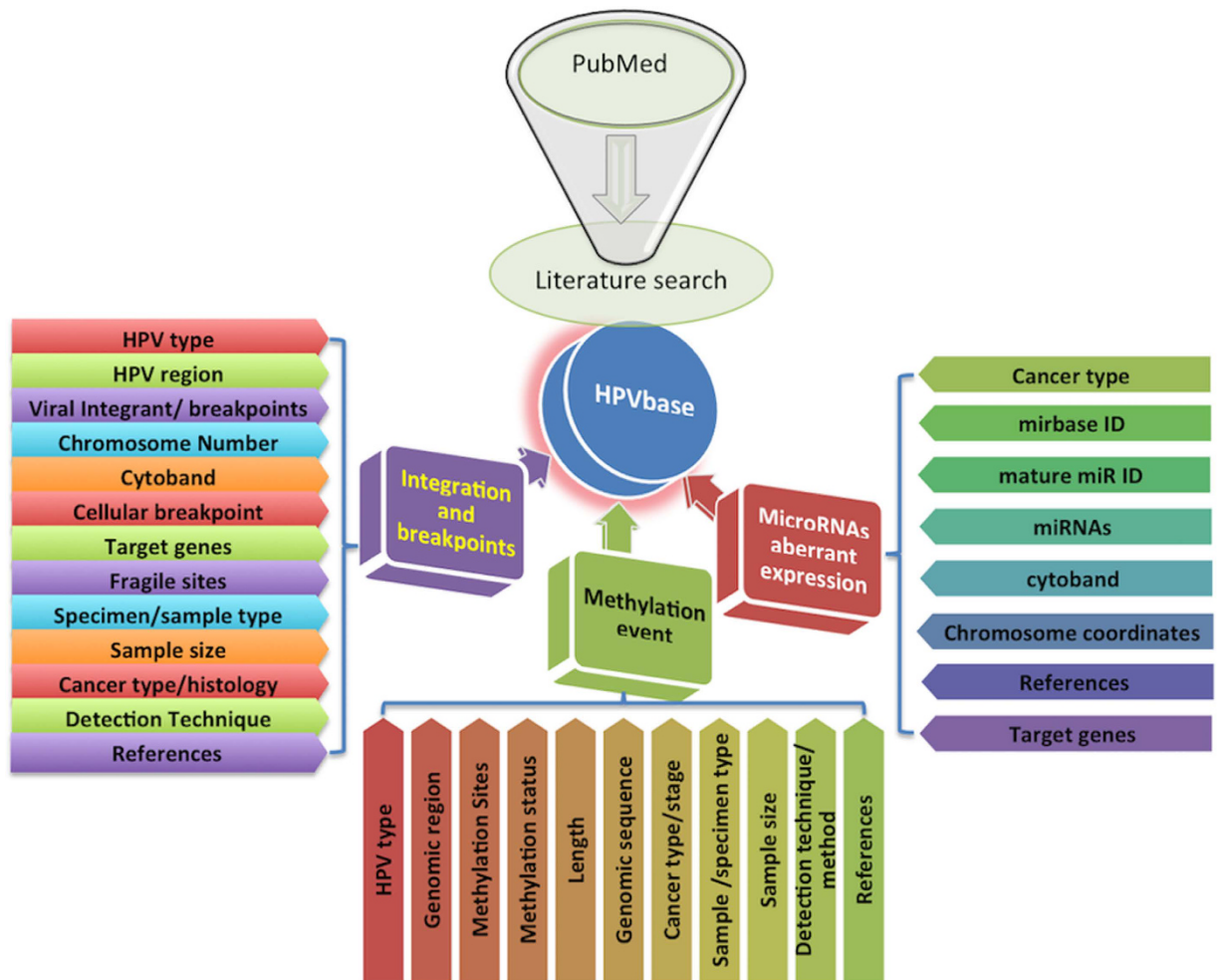


Figure 6. Layout depicting overall organization of HPVbase.

evaluation. In HPVbase, data is classified and organized in three different sections- integration sites and breakpoints, methylation patterns and miRNAs expression profiles (see Fig. 6). Subsequently, these biomarkers are further divided into various subsections based on the viral types and diverse carcinomas. Each biomarker sections are represented in highly interactive browsers and user intensive tabular interfaces with other useful external links. It provides user friendly searching and browsing facility for data retrieval. Each section encompasses search box and sorting ability. User can search and filter data using any relevant keywords like specific HPV type, HPV genomic region, cytoband, target genes, detection technique used etc. (e.g., HPV18, E7, q32.3, RNA-seq). It also facilitates browsing of clinically important data based on the HPV genotypes and individual carcinoma.

Alongside, HPVbase also facilitate user affable data searching and retrieval options. The resulting two search tools (a) integration search and (b) advance search have been implemented at website.

(a) Integration search. User can extensively search integration sites based on various keywords with exact or containing sub-options from number of provided fields including integration site, HPV type, chromosomal locations, cancer type etc.

(b) Advance search. This tool permits user to perform systematic search using combination of logical operators (OR/AND). It allows user to build a query using different combination of keywords to narrow down search.

Updating of HPVbase. We have incorporated the latest and highly curated data in HPVbase. To cope with this increasing amount of valuable clinical data, we would like to further update our resource in every six-month.

References

- de Villiers, E. M., Fauquet, C., Broker, T. R., Bernard, H. U. & zur Hausen, H. Classification of papillomaviruses. *Virology* **324**, 17–27, doi: 10.1016/j.virol.2004.03.033 (2004).
- Doorbar, J. Molecular biology of human papillomavirus infection and cervical cancer. *Clin Sci (Lond)* **110**, 525–541, doi: 10.1042/CS20050369 (2006).
- Munoz, N. *et al.* Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N Engl J Med* **348**, 518–527, doi: 10.1056/NEJMoa021641 (2003).
- Mantovani, F. & Banks, L. The human papillomavirus E6 protein and its contribution to malignant progression. *Oncogene* **20**, 7874–7887, doi: 10.1038/sj.onc.1204869 (2001).
- Munger, K. *et al.* Biological activities and molecular targets of the human papillomavirus E7 oncoprotein. *Oncogene* **20**, 7888–7898, doi: 10.1038/sj.onc.1204860 (2001).
- Moody, C. A. & Laimins, L. A. Human papillomavirus oncoproteins: pathways to transformation. *Nat Rev Cancer* **10**, 550–560, doi: 10.1038/nrc2886 (2010).
- Doorbar, J. *et al.* The biology and life-cycle of human papillomaviruses. *Vaccine* **30 Suppl 5**, F55–70, doi: 10.1016/j.vaccine.2012.06.083 (2012).
- Schiffman, M. & Wentzensen, N. Human papillomavirus infection and the multistage carcinogenesis of cervical cancer. *Cancer Epidemiol Biomarkers Prev* **22**, 553–560, doi: 10.1158/1055-9965.EPI-12-1406 (2013).
- Boulet, G. A., Horvath, C. A., Berghmans, S. & Bogers, J. Human papillomavirus in cervical cancer screening: important role as biomarker. *Cancer Epidemiol Biomarkers Prev* **17**, 810–817, doi: 10.1158/1055-9965.EPI-07-2865 (2008).
- Woodman, C. B., Collins, S. I. & Young, L. S. The natural history of cervical HPV infection: unresolved issues. *Nat Rev Cancer* **7**, 11–22, doi: 10.1038/nrc2050 (2007).
- Sahasrabudde, V. V., Luhn, P. & Wentzensen, N. Human papillomavirus and cervical cancer: biomarkers for improved prevention efforts. *Future Microbiol* **6**, 1083–1098, doi: 10.2217/fmb.11.87 (2011).
- Schwarz, E. *et al.* Structure and transcription of human papillomavirus sequences in cervical carcinoma cells. *Nature* **314**, 111–114 (1985).
- zur Hausen, H. Papillomaviruses and cancer: from basic studies to clinical application. *Nat Rev Cancer* **2**, 342–350, doi: 10.1038/nrc798 (2002).
- Schmitz, M., Driesch, C., Jansen, L., Runnebaum, I. B. & Durst, M. Non-random integration of the HPV genome in cervical cancer. *PLoS One* **7**, e39632, doi: 10.1371/journal.pone.0039632 (2012).
- Akagi, K. *et al.* Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. *Genome Res* **24**, 185–199, doi: 10.1101/gr.164806.113 (2014).
- Parfenov, M. *et al.* Characterization of HPV and host genome interactions in primary head and neck cancers. *Proc Natl Acad Sci U S A* **111**, 15544–15549, doi: 10.1073/pnas.1416074111 (2014).
- Ojesina, A. I. *et al.* Landscape of genomic alterations in cervical carcinomas. *Nature* **506**, 371–375, doi: 10.1038/nature12881 (2014).
- Xu, B. *et al.* Multiplex Identification of Human Papillomavirus 16 DNA Integration Sites in Cervical Carcinomas. *PLoS One* **8**, e66693, doi: 10.1371/journal.pone.0066693 (2013).
- Jeon, S. & Lambert, P. F. Integration of human papillomavirus type 16 DNA into the human genome leads to increased stability of E6 and E7 mRNAs: implications for cervical carcinogenesis. *Proc Natl Acad Sci U S A* **92**, 1654–1658 (1995).
- Wentzensen, N., Vinokurova, S. & von Knebel Doeberitz, M. Systematic review of genomic integration sites of human papillomavirus genomes in epithelial dysplasia and invasive cancer of the female lower genital tract. *Cancer Res* **64**, 3878–3884, doi: 10.1158/0008-5472.CAN-04-0009 (2004).
- Thorland, E. C., Myers, S. L., Gostout, B. S. & Smith, D. I. Common fragile sites are preferential targets for HPV16 integrations in cervical tumors. *Oncogene* **22**, 1225–1237, doi: 10.1038/sj.onc.1206170 (2003).
- Matovina, M., Sabol, I., Grubisic, G., Gasperov, N. M. & Grce, M. Identification of human papillomavirus type 16 integration sites in high-grade precancerous cervical lesions. *Gynecol Oncol* **113**, 120–127, doi: 10.1016/j.ygyno.2008.12.004 (2009).
- Klaes, R. *et al.* Detection of high-risk cervical intraepithelial neoplasia and cervical cancer by amplification of transcripts derived from integrated papillomavirus oncogenes. *Cancer Res* **59**, 6132–6136 (1999).
- Bird, A. The essentials of DNA methylation. *Cell* **70**, 5–8 (1992).
- Robertson, K. D. DNA methylation and human disease. *Nat Rev Genet* **6**, 597–610, doi: 10.1038/nrg1655 (2005).
- Clarke, M. A. *et al.* Human papillomavirus DNA methylation as a potential biomarker for cervical cancer. *Cancer Epidemiol Biomarkers Prev* **21**, 2125–2137, doi: 10.1158/1055-9965.EPI-12-0905 (2012).
- Turan, T. *et al.* High-throughput detection of human papillomavirus-18 L1 gene methylation, a candidate biomarker for the progression of cervical neoplasia. *Virology* **361**, 185–193, doi: 10.1016/j.virol.2006.11.010 (2007).
- Brandsma, J. L. *et al.* Distinct human papillomavirus type 16 methylomes in cervical cells at different stages of premalignancy. *Virology* **389**, 100–107, doi: 10.1016/j.virol.2009.03.029 (2009).
- Crosbie, E. J., Einstein, M. H., Franceschi, S. & Kitchener, H. C. Human papillomavirus and cervical cancer. *Lancet* **382**, 889–899, doi: 10.1016/S0140-6736(13)60022-7 (2013).
- Reshmi, G. & Pillai, M. R. Beyond HPV: oncomirs as new players in cervical cancer. *FEBS Lett* **582**, 4113–4116, doi: 10.1016/j.febslet.2008.11.011 (2008).
- Gomez-Gomez, Y., Organista-Nava, J. & Gariglio, P. Deregulation of the miRNAs expression in cervical cancer: human papillomavirus implications. *Biomed Res Int* **2013**, 407052, doi: 10.1155/2013/407052 (2013).
- Zheng, Z. M. & Wang, X. Regulation of cellular miRNA expression by human papillomaviruses. *Biochim Biophys Acta* **1809**, 668–677, doi: 10.1016/j.bbagr.2011.05.005 (2011).
- Kim, V. N. Small RNAs: classification, biogenesis, and function. *Mol Cells* **19**, 1–15 (2005).
- Engels, B. M. & Hutvagner, G. Principles and effects of microRNA-mediated post-transcriptional gene regulation. *Oncogene* **25**, 6163–6169, doi: 10.1038/sj.onc.1209909 (2006).
- Lee, J. W. *et al.* Altered MicroRNA expression in cervical carcinomas. *Clin Cancer Res* **14**, 2535–2542, doi: 10.1158/1078-0432.CCR-07-1231 (2008).
- Winter, J. & Diederichs, S. MicroRNA biogenesis and cancer. *Methods Mol Biol* **676**, 3–22, doi: 10.1007/978-1-60761-863-8_1 (2011).
- Zhang, L. *et al.* microRNAs exhibit high frequency genomic alterations in human cancer. *Proc Natl Acad Sci U S A* **103**, 9136–9141, doi: 10.1073/pnas.0508889103 (2006).
- Wang, X. *et al.* microRNAs are biomarkers of oncogenic human papillomavirus infections. *Proc Natl Acad Sci U S A* **111**, 4262–4267, doi: 10.1073/pnas.1401430111 (2014).
- Wentzensen, N. *et al.* Methylation of HPV18, HPV31, and HPV45 genomes and cervical intraepithelial neoplasia grade 3. *J Natl Cancer Inst* **104**, 1738–1749, doi: 10.1093/jnci/djs425 (2012).
- Patel, D. A. *et al.* Patterns of cellular and HPV 16 methylation as biomarkers for cervical neoplasia. *J Virol Methods* **184**, 84–92, doi: 10.1016/j.jviromet.2012.05.022 (2012).

41. Mirabello, L. *et al.* Elevated methylation of HPV16 DNA is associated with the development of high grade cervical intraepithelial neoplasia. *Int J Cancer* **132**, 1412–1422, doi: 10.1002/ijc.27750 (2013).
42. Brandsma, J. L. *et al.* Methylation of Twelve CpGs in Human Papillomavirus Type 16 (HPV16) as an Informative Biomarker for the Triage of Women Positive for HPV16 Infection. *Cancer Prev Res (Phila)*, doi: 10.1158/1940-6207.CAPR-13-0354 (2014).
43. Ling, H., Fabbri, M. & Calin, G. A. MicroRNAs and other non-coding RNAs as targets for anticancer drug development. *Nature reviews. Drug discovery* **12**, 847–865, doi: 10.1038/nrd4140 (2013).
44. Qureshi, A., Thakur, N., Monga, I., Thakur, A. & Kumar, M. VIRmiRNA: a comprehensive resource for experimentally validated viral miRNAs and their targets. *Database (Oxford)* **2014**, doi: 10.1093/database/bau103 (2014).
45. Zhao, X. *et al.* Dr.VIS: a database of human disease-related viral integration sites. *Nucleic Acids Res* **40**, D1041–1046, doi: 10.1093/nar/gkr1142 (2012).
46. Yang, X. *et al.* Dr.VIS v2.0: an updated database of human disease-related viral integration sites in the era of high-throughput deep sequencing. *Nucleic Acids Res* **43**, D887–892, doi: 10.1093/nar/gku1074 (2015).
47. Van Doorslaer, K. *et al.* The Papillomavirus Episteme: a central resource for papillomavirus sequence data and analysis. *Nucleic Acids Res* **41**, D571–578, doi: 10.1093/nar/gks984 (2013).
48. Zhang, G. L. *et al.* HPVdb: a data mining system for knowledge discovery in human papillomavirus with applications in T cell immunology and vaccinology. *Database (Oxford)* **2014**, bau031, doi: 10.1093/database/bau031 (2014).
49. Hsu, S. D. *et al.* miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. *Nucleic Acids Res* **42**, D78–85, doi: 10.1093/nar/gkt1266 (2014).
50. Skinner, M. E., Uzilov, A. V., Stein, L. D., Mungall, C. J. & Holmes, I. H. JBrowse: a next-generation genome browser. *Genome Res* **19**, 1630–1638, doi: 10.1101/gr.094607.109 (2009).

Acknowledgement

This work was supported by the Council of Scientific and Industrial Research, India (GENESIS- BSC0121) and Department of Biotechnology, Government of India (GAP001).

Author Contributions

A.G. acquisition of data and development of web resource. A.G. and M.K. analysis and interpretation of data. A.G. and M.K. drafting manuscript and revising it critically. M.K. conceived the study. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Kumar Gupta, A. and Kumar, M. HPVbase- a knowledgebase of viral integrations, methylation patterns and microRNAs aberrant expression: As potential biomarkers for Human papillomaviruses mediated carcinomas. *Sci. Rep.* **5**, 12522; doi: 10.1038/srep12522 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>