



OPEN

# Inferring Virus-Host relationship between HPV and its host *Homo sapiens* using protein interaction network

Qurat ul Ain Farooq<sup>1</sup>, Zeeshan Shaukat<sup>2</sup>, Tong Zhou<sup>1</sup>, Sara Aiman<sup>1</sup>, Weikang Gong<sup>1</sup> & Chunhua Li<sup>1</sup>✉

Human papilloma virus (HPV) is a serious threat to human life globally with over 100 genotypes including cancer causing high risk HPVs. Study on protein interaction maps of pathogens with their host is a recent trend in 'omics' era and has been practiced by researchers to find novel drug targets. In current study, we construct an integrated protein interaction map of HPV with its host human in Cytoscape and analyze it further by using various bioinformatics tools. We found out 2988 interactions between 12 HPV and 2061 human proteins among which we identified MYLK, CDK7, CDK1, CDK2, JAK1 and 6 other human proteins associated with multiple viral oncoproteins. The functional enrichment analysis of these top-notch key genes is performed using KEGG pathway and Gene Ontology analysis, which reveals that the gene set is enriched in cell cycle a crucial cellular process, and the second most important pathway in which the gene set is involved is viral carcinogenesis. Among the viral proteins, E7 has the highest number of associations in the network followed by E6, E2 and E5. We found out a group of genes which is not targeted by the existing drugs available for HPV infections. It can be concluded that the molecules found in this study could be potential targets and could be used by scientists in their drug design studies.

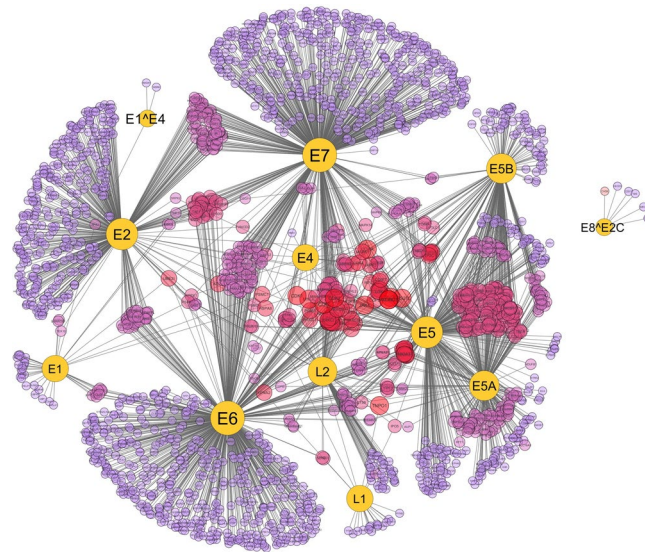
Human papilloma virus (HPV) is associated with approximately 5% of all human cancers affecting 0.6 million people worldwide with cervical, anal, oropharyngeal, penile and vulvovaginal cancers<sup>1-3</sup>. Among these cancers, cervical cancer ranks 4th in affecting women worldwide<sup>4</sup> while in developing countries it ranks second<sup>5</sup>. According to World Health Organization (WHO) current factsheets, there are more than 100 genotypes of HPV, out of which 14 strains are high-risk. The most talked about high-risk HPV strains are HPV 6, 11, 16, 18, 31, 33, 35, 45, 52 and 58 with type 16 and 18 responsible for 70% of cervical cancer cases<sup>6-8</sup>. HPV is a serious threat to human life and it is causing 250,000 deaths annually, among which 85% of cases are occurring in low and middle-income countries<sup>9</sup>.

HPV is a small ~8 kb in size, non-enveloped circular dsDNA virus<sup>5,10</sup>. The HPV genome encodes 8 proteins among which 2 are structural viral capsid proteins (L1 and L2) while 6 are non-structural viral proteins (E1, E2, E4, E5, E6, E7)<sup>10,11</sup>. Besides these 8 proteins, there are a few other macromolecules found in literature which are actually the transcripts made by the fusion of two existing HPV proteins. E8^E2, a transcript, is created by the fusion of E8 with carboxy terminal of E2<sup>12</sup>, and E1^E4 is generated by the fusion of E1 to the Open Reading Frame (ORF) of E4<sup>13</sup>.

Protein interaction network provides a plethora of information when it comes to virus-host relationship because viruses entirely depend upon the host factors for their survival<sup>14,15</sup>. Viruses tend to regulate host biological processes by manipulating its cell proteome. Researchers have been using network biology for designing novel antiviral drug therapies<sup>16</sup>. We can dig deep down into the molecular mechanisms of the infections by analyzing the protein-protein interactions (PPIs) between the pathogen and its host. Researchers use various experimental and computational methods to identify PPIs and to study the 'interactome' of organisms.

A plethora of biological data has been generated because of the advent of "omics" data by high throughput technologies including genomics, proteomics and transcriptomics. In current postgenomic biomedical research,

<sup>1</sup>College of Life Science and Bioengineering, Beijing University of Technology, Beijing, 100124, China. <sup>2</sup>Faculty of Information Technology, Beijing University of Technology, Beijing, 100124, China. ✉e-mail: [chunhuali@bjut.edu.cn](mailto:chunhuali@bjut.edu.cn)



**Figure 1.** HPV-Human protein interaction network constructed in Cytoscape. The network comprises of 2073 nodes (proteins) with 2988 edges (interactions) between them. HPV proteins form highly connected hubs. The network is mapped according to node size. The lesser the number of interactions, the smaller the size of the node. HPV proteins E7 and E6 are of largest node size, indicating their substantial numbers of interactions. Viral proteins are represented in a uniform yellow color while human proteins in different colors. Host proteins with high numbers of interactions have larger node size (colored in bright red) compared to the proteins with lower numbers of associations (light purple).

systems biology and network biology approaches are at the core and have been used by scientists to model disease networks for better understanding of the pathogen-host relationship and disease pathways<sup>17–19</sup>. By constructing pathogen-host interactome maps, we can find functionally important proteins and further study them by doing their functional enrichment analysis.

In current study, we integrate all the previous works on interactome of Human Papilloma virus to construct a unified and comprehensive protein interaction network of HPV with its host *Homo sapiens*. From small-scale experimental investigations to large-scale computational studies, we try to fetch every single interaction from the scientific literature. The study includes protein-protein interactions between HPV and Human, HPV and HPV identified by co-immunoprecipitation, mass spectrometry, Yeast 2 Hybrid, GST pull-down assays, tandem affinity purification, chromatin immunoprecipitation, indirect immunofluorescence assay and various other approaches. The data is also gathered from several well-established PPI databases including IMEx databases (DIP, HPIDB, BioGrid and IntAct) and databases which are not a part of IMEx consortium including VirHostNet and VirusMint. By constructing such a comprehensive interactome map of HPV with human, we will be able to study and understand the relationship between the pathogen with the host proteins and can find potential molecules involved in causing life-threatening infections and diseases. In a human body, every disease is triggered not just because of a single element but because of mutual interactions between various molecules.

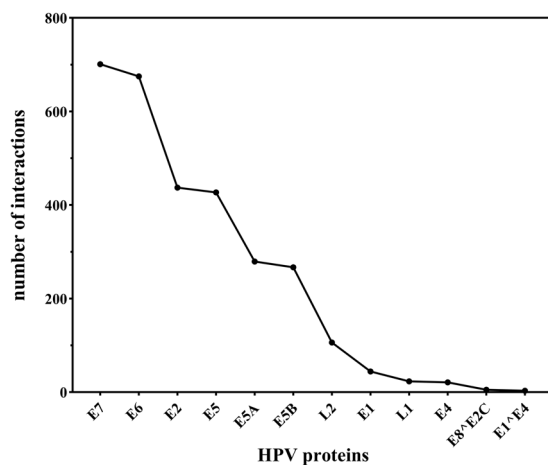
The aim of this research is to find the molecules which are highly involved in provoking the HPV infections and to find out human proteins that are highly associated with viral proteins. We come up with certain host genes which are found to have distinct interactions with viral proteins and could be potential drug targets for scientists in their molecular docking studies.

## Results

**The HPV-Human protein interaction map and its statistical significance.** The integrated protein interaction network map of HPV with its host *Homo sapiens* constructed in Cytoscape is shown in Fig. 1. The network comprises of 2988 interactions between 12 HPV and 2061 human proteins (Supplementary table S1) in which highly associated proteins can be seen forming hubs in the overall network. The figure can be zoomed to see individual protein and its associations with other viral/host proteins in the network. It is noted that the interaction network of HPV with human is not much dense i.e. the average number of neighbors per node is 2.883 which means that the average number of interactions a node can have with the other nodes is not greater than 3, so the network density is less and it can be analyzed and explored easily. We can explore the network and dig deep down into the molecular associations of each protein in the network. The network was analyzed by using network analyzer tool of Cytoscape and we found out several important statistical features of the HPV-Human network including clustering coefficient, shortest path, average number of neighbors etc. Table 1 shows the important statistical measures of the network calculated by the network analyzer tool. The high scoring viral proteins are also mentioned in the table. CytoHubba is a built-in plugin of Cytoscape by which we can analyze hubs and even single nodes. By using this tool in our network, we found out that non-structural viral protein E7 is interacting with the most number of host proteins i.e. E7 has 701 associations in the overall network followed

Attributes	Values
Num. of Nodes	2073
Num. of Edges	2988
Avg. Degree	2.883
High score interacting viral proteins	E7, E6, E2, E5
Clustering Coefficient	0.001
Network Density	0.001
Shortest Paths	4262192 (99%)

**Table 1.** Statistics of the PPI network constructed between HPV-Human.



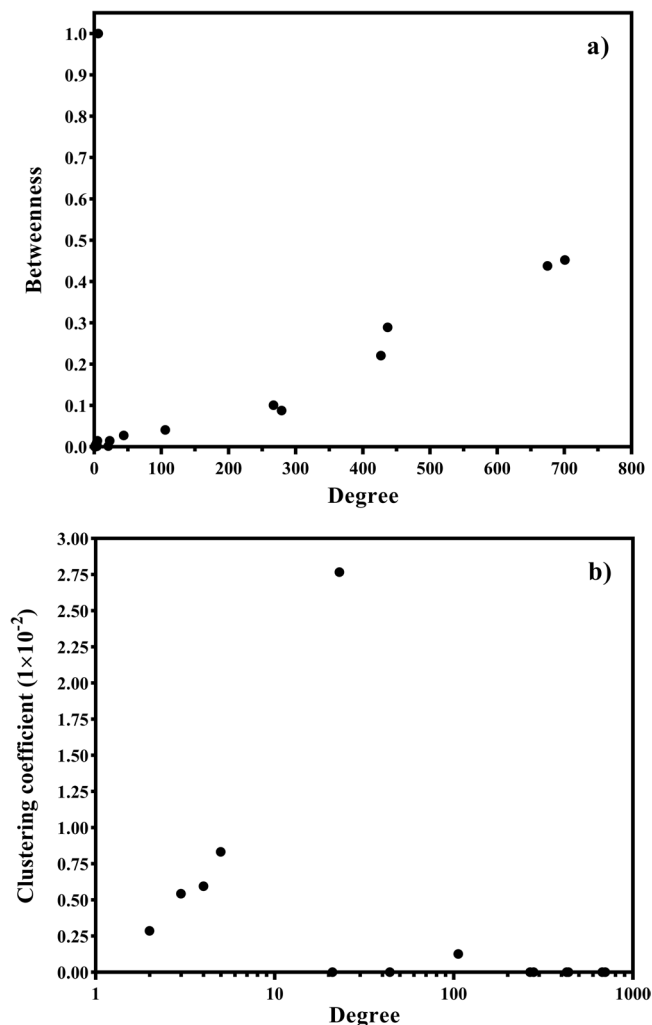
**Figure 2.** Numbers of interactions HPV proteins have in the overall network.

by non-structural proteins E6, E2 and E5 with a degree of 675, 437 and 427 respectively. Figure 2 represents a line graph showing the distribution of each HPV protein with their respective number of associations in the network. Figure 3 shows their betweenness centrality and average clustering coefficient graphs of the HPV-Human protein interaction network along with their average number of neighbors (degree). HPV is a virus and it is small in size, so the scatter plots are not very dense and is easily understandable.

With its versatile and flexible layouts, Cytoscape helps user visualize and explore the macromolecular network in their best possible ways. Figure 1 also represents the analysis of the network and its visualization with respect to degree parameter. The node size is mapped according to its number of associations with other proteins in the network. The lesser the number of interactions, the smaller the size of the node while large size nodes represent proteins with high number of interactions. HPV oncoproteins E7 and E6 can be clearly seen having the largest node size, indicating their high number of associations with other proteins. Compared to viral proteins, the highest number of interactions a human protein has in the network is 5. MYLK, CDK7, CDK1, CDK2, JAK1, SQOR, RPS27L, MT-CO2, VKORC1, TNPO1 and COPA are the human proteins which are interacting with 5 viral proteins including oncoproteins E7 and E6 in the overall network. The enrichment analysis of these highly HPV-associated human genes was performed and is explained in the next section.

**Functional enrichment analysis.** We performed pathway enrichment analysis of the top 25 host proteins scored and ranked by CytoHubba on the basis of Maximal Clique Centrality (MCC) which perceives on the concept that essential proteins tend to make clusters in a PPI network. Table 2 shows the pathway enrichment analysis results of the set of 25 human genes. The list of 10 enriched pathways is shown in the table together with their respective gene set size, p value and the value of corresponding False Discovery Rate (FDR). FDR<sup>20</sup>, also known as Benjamini-Hochberg (BH) method is a multiple-testing correction statistical method which uses the un-corrected p-value threshold and the number of tests to evaluate the fraction of falsely enriched pathways over the enriched pathways. The set of 25 genes ranked by CytoHubba on the basis of MCC was found to be enriched in vascular smooth muscle contraction, platelet activation, viral carcinogenesis, gap junction and several others mentioned in Table 2.

Additionally, we selected a group of 11 proteins with highest number of interactions in the HPV-Human protein interaction network and performed their pathway enrichment analysis and Gene Ontology. Among the group of proteins, each protein is interacting with 5 viral proteins in the network. KEGG knowledgebase lets us identify the pathways in which our input genes are over-represented and playing a pivotal role. KEGG pathway analysis of top 11 human proteins (Fig. 4) which have the highest number of interactions with HPV viral proteins reveals the fact that the gene set is actively involved in cell cycle which is one of the most important pathways for survival of an organism. The second most important pathway in which the gene set is actively playing a part



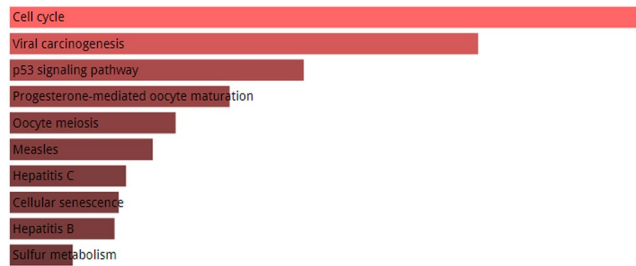
**Figure 3.** (a) Betweenness centrality and (b) clustering coefficient values of HPV proteins in the interaction network with the x-axis representing their degrees.

Gene set	Description	Size	P value	FDR value
hsa04270	Vascular smooth muscle contraction	121	$5.4 \times 10^{-6}$	$1.8 \times 10^{-3}$
hsa04611	Platelet activation	123	$1.4 \times 10^{-4}$	$2.3 \times 10^{-2}$
hsa05203	Viral carcinogenesis	201	$9.2 \times 10^{-4}$	$5.9 \times 10^{-2}$
hsa04540	Gap junction	88	$9.5 \times 10^{-4}$	$5.9 \times 10^{-2}$
hsa04666	Fc gamma R-mediated phagocytosis	91	$1.0 \times 10^{-3}$	$5.9 \times 10^{-2}$
sa04658	Th1 and Th2 cell differentiation	92	$1.1 \times 10^{-3}$	$5.9 \times 10^{-2}$
hsa04914	Progesterone-mediated oocyte maturation	99	$1.3 \times 10^{-3}$	$6.2 \times 10^{-2}$
hsa04659	Th17 cell differentiation	107	$1.7 \times 10^{-3}$	$6.8 \times 10^{-2}$
hsa04110	Cell cycle	124	$2.6 \times 10^{-3}$	$8.3 \times 10^{-2}$
hsa04114	Oocyte meiosis	124	$2.5 \times 10^{-3}$	$8.3 \times 10^{-2}$

**Table 2.** Pathway enrichment analysis of the top 25 host genes ranked by CytoHubba on the basis of MCC.

is viral carcinogenesis, followed by p53 signaling pathway, progesterone-mediated oocyte maturation and so on. Gene Ontology enrichment analysis of a gene set reveals useful information regarding the biological processes the gene set is involved in, the molecular function of the genes and the cellular component of which the gene set is a part. Figure 5 shows GO biological process, cellular component and molecular function of the gene set having high connections with viral proteins.

We searched for the current drugs available to treat HPV infections on drug repurposing hub (<https://clue.io/repurposing>) and found out that the existing drugs Imiquimod, Podofilox and Trichloroacetic Acid target several



**Figure 4.** KEGG pathway analysis of top 11 highly HPV-associated human genes. The genes are a part of one of the chief cellular pathways i.e. cell cycle, viral carcinogenesis and p53 signalling pathway. The genes are also involved in several other pathways including progesterone-mediated oocyte maturation, oocyte meiosis, measles and HCV but the intensity of their involvement in these pathways is lesser. The length and color of the bar represent the intensity of the genes enriched in the pathway/process. The longer the bar and lighter the color, the more enriched the gene set in the specific pathway.

proteins detected out in our study including TLR7, TUBA4A and TUBB respectively but there is no drug targeting the human proteins MYLK, CDK7, CDK1 and few more highly connected host proteins which are found to be having multiple associations with viral oncoproteins. We found out some drugs CaMKII-IN-1, BS 181 dihydrochloride, PF-573228, which will target the proteins MYLK, CDK7, and CDK1 respectively detected in our study but they are in their preclinical phases yet.

**Clustering analysis of the network.** To find out protein complexes and functional modules in a protein-protein interaction network, we performed clustering analysis. Cytoscape has various clustering apps including CytoCluster<sup>21</sup>, ClusterMaker<sup>22</sup>, ClusterViz<sup>23</sup>, MCODE<sup>24</sup> and ClusterOne<sup>25</sup>. All of these apps integrates different clustering algorithms based on different methods to analyze biological networks and identify crucial protein complexes. We used CytoCluster, which is a combination of six clustering algorithms with a mutual goal of identification of functional modules and protein complexes in the network. The clusters in the whole network were found to be making smaller sub-networks after clustering analysis performed by CytoCluster. One of the cluster (Fig. 6a) contains 7 nodes among which E8<sup>^</sup>E2C is a viral protein while ZZEF1, IDE, BIRC6, HECD3, CHD6, CNTN6 belong to human. Another cluster (Fig. 6b) consists of 9 nodes in which L1 and L2 are viral proteins while human proteins in the cluster are HSPA8, PPIB, KPNA2, TNPO1, KPNA1, KPNB1 and IPO5. The pathways in which these clusters are enriched are represented in Table S2.

## Discussion

The two most important pathways in which the set of human genes are involved are cell cycle and viral carcinogenesis which depicts the oncogenic nature of the Human papilloma virus proteins. The interactions of cell cycle-associated human genes with the oncoproteins of HPV give us an idea of how they might be involved in interfering the normal cell cycle and might alter the process which in turn can cause cancer. The third notable pathway in which the gene set is involved is p53 signaling pathway which is induced by many factors, some of which include activated oncogenes. Other disease pathways in which the gene set is found to be enriched include measles, hepatitis C and hepatitis B. The human genes which are highly associated with HPV and have maximum numbers of interactions with the viral proteins are: MYLK, CDK7, CDK1, CDK2, JAK1, SQOR, RPS27L, MT-CO2, VKORC1, TNPO1 and COPA. We integrated protein-protein interaction data from most pathogenic genotypes of HPV including HPV1, 3, 5, 6, 8, 9, 11, 16, 18, 31, 33 and HPV 39 from small-scale studies to large-scale extensive investigations carried out either experimentally or computationally.

Various researchers have been using network biology to study the disease and pathogen-host relationship on molecular level which ultimately helps in identifying key viral proteins and their human targets and helps scientists in further biological investigations. Kumar *et al.*<sup>17</sup> studied the pathogenesis of *Leptospira interrogans* with Homo sapiens by analyzing their pathogen-host interactions after constructing a protein interaction network. Similarly, in our previous study, we constructed a comprehensive protein interaction map of HCV with its host human<sup>26</sup> and then investigated it and found multiple potential drug targets. Besides infectious diseases, protein interaction networks have been used to inspect complex diseases including multiple sclerosis<sup>27</sup>, breast cancer metastasis<sup>28</sup>, Alzheimer's disease<sup>29</sup>, HBV and hepatocellular carcinoma (HCC)<sup>30</sup>.

Previously, HPV-Human protein interaction network has been studied by other group of scientists too. Eckhardt *et al.*<sup>31</sup> performed mass spectrometry analysis to find out interactions between HPV and human proteins. They identified 137 interactions between 9 viral proteins belonging to HPV genotype 31. On the other hand, our study presents an integrated network between multiple HPV genotypes and host proteins. We constructed a comprehensive protein interaction network and identified a total of 2988 interactions between 12 HPV and 2061 human proteins. We also incorporated the interactions identified by Eckhardt *et al.* in this study. Thus, our study presents a new approach of constructing a comprehensive protein interaction map of HPV with its host Homo sapiens by integrating all the previous small-scale and large-scale investigations performed on HPV PPI. We further have analyzed the PPI network by using various bioinformatics tools and have grouped the genes according to their most to least number of associations with human papilloma viral proteins and then performed KEGG pathway analysis to find out the biological pathways targeted by HPV. We also performed pathway enrichment of



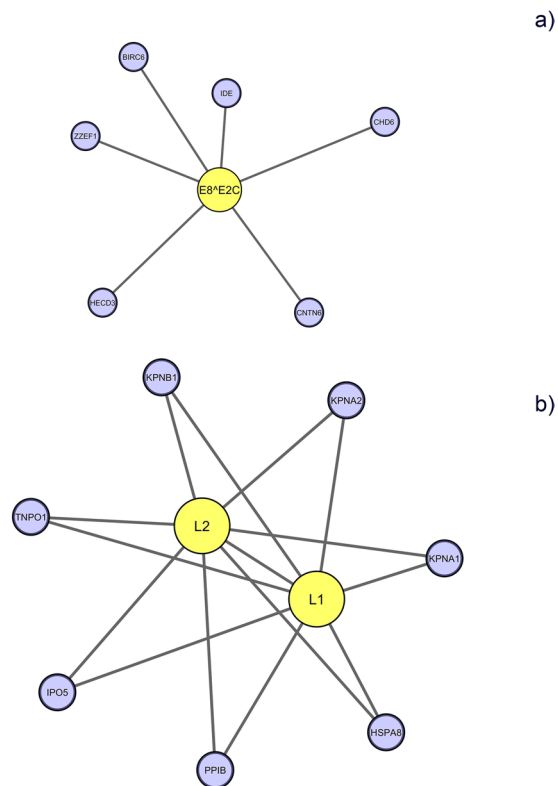


**Figure 5.** Gene Ontology of top 11 highly HPV-associated human genes. **(a)** The gene set is an important part of the DNA damage response which involves a cascade of processes activated by the p53 protein (GO:0030330), and also a part of cell cycle checkpoints including G1/S transition checkpoint (GO:0044819), G1 DNA damage checkpoint (GO:0044783), and a part of various other transition phases of cell cycle. **(b)** GO cellular component analysis reveals that the gene set is a major component of cyclin-dependent kinase activating kinase complexes responsible for cell cycle progression regulation (GO:0019907). The gene set is also a part of holo TFIID complex (GO:0005675), a transcription factor essential for initiation of promoters. **(c)** GO molecular function analysis of the gene set shows that it is responsible for catalysis of a group of reactions and requires a CDK activity. The gene set is clearly involved in the regulation of cell cycle and the molecular functions they exhibit are cyclin-dependent protein serine/threonine kinase activity (GO:0004693), cyclin-dependent protein kinase activity (GO:0097472), histone kinase activity (GO:0035173) and several others as mentioned in the figure.

top 25 human genes ranked by CytoHubba on the basis of MCC to find out pathways they are enriched in along with their important statistical significance i.e. p-value and FDR. Additionally, we performed Gene Ontology analysis of the host genes which have highest number of associations with HPV proteins in the network to find out the biological processes targeted by HPV. We also found out that there are several human proteins which are currently not targeted by the available drugs available for HPV treatment. Some of the compounds targeting the high-connected proteins detected in this study are in their preclinical phases yet but there is no drug available currently to target COPA, TNPO1, RPS27L and SQOR, although they interact with multiple oncogenic HPV proteins.

## Conclusion

The infections caused by HPV are serious threats to human life. With emerging high throughput 'omics' technologies, we have huge amount of biological data which need to be investigated and analyzed by scientists from all over the world. Network biology is quite a new direction in the world of life sciences and it has been immensely used by researchers from past decade to unveil important information about pathogens, diseases and their impact on organisms. Our research concentrated on the protein-protein interactions between human papilloma virus and *Homo sapiens*. We constructed an integrated network of HPV with human and then analyzed it further using various techniques, which manifests potential target molecules. We are able to identify novel infection-related genes and further KEGG pathway and Gene Ontology enrichment analyses reveal the important biological pathways and processes targeted by the pathogenic infection at molecular level. In our research, we focus not only on protein-protein interactions from large database sources but also on individual interactions detected by scientists in their investigations. We are able to find out the involvement of target genes in several other disease pathways including measles, HCV and HBV. Though the intensity of their enrichment in these disease pathways is low but there is still a chance of knocking down these diseases by targeting these potential molecules.



**Figure 6.** Clusters formed by CytoCluster using a hierarchical clustering algorithm. Viral proteins are represented as yellow nodes while human proteins are denoted in blue.

## Methods

**Data collection.** The first step of our study is to obtain a cutting edge picture of what is out there and how much has been done on protein interaction data and interactome of HPV with its host *Homo sapiens*. For this purpose, a PubMed advanced search is performed using multiple keywords. The search successfully yields 1222 results. The next step is to refine our search and to carefully analyze the studies specifically relating to protein-protein interaction (PPI) data of Human papilloma virus. Every study is carefully explored to get the maximum possible interactions between HPV and host proteins. From small-scale researches to extensive large-scale studies on protein-protein interactions involved in HPV and human, every possible interaction is gathered from the literature. Finally, 32 studies are chosen which contain exclusively protein-protein interaction data of HPV with its host and with its own viral proteins. The scientists used various potential protein-protein interaction detection methods and we gathered the information regarding every single interaction detected in their laboratories. Table 3 shows the list of selected studies performed on HPV-host protein-protein interactions (PINs). The list includes all the small-scale and large-scale investigations carried out either experimentally or computationally for detecting the PPIs.

The data gathered from these studies are merged and integrated, and after removing duplicates, 2988 unique interactions are found. Another challenge in this study is to bring all the data gathered from distinct studies into one common format, because for the network construction in Cytoscape, the data must be in one single format. Different scientists used different identifiers; some of them used protein/gene name while others used database identifiers e.g. UniProt, EMBL/GenBank or any other biological database identifier. For example, Dong *et al.*<sup>32</sup> identified 877 protein-protein interactions between HPV and human proteins and he used UniProt identifiers in his large-scale study. To bring the data into a uniform format, an online UniProt ID Mapping tool<sup>33</sup> is used to convert every single identifier into one common format which in our case is gene name.

**Construction and analysis of the virus-host PPI network in cytoscape.** After having a uniform set of protein-protein interaction data, we use the latest version of Cytoscape (v3.7.1) to construct the integrated network between HPV and its host. Cytoscape<sup>24</sup> is a most popular, user-friendly and freely available platform for the construction and exploration of biomolecular networks. Within the new version, it improves its performance in the context of versatility and interactivity for better analyses of the networks. Cytoscape has various tools and plugins by which the biomolecular networks can be analyzed and extensively explored on the basis of the attributes including degree distribution, betweenness centrality, clustering coefficient and so on. In current study, the network comprises of 2073 proteins (nodes) with 2988 Interactions (edges) between them. The number of human proteins in the network is 2061 while 12 proteins belong to HPV. The network is analyzed based on various topological measures including degree, diameter, betweenness centrality and clustering coefficient. Functionally important hubs are also explored which constitute the backbone of the HPV-Human protein interaction network.

No.	Paper	Num. of Interactions	Method of PPI detection	Ref
1	Gulati <i>et al.</i> (2019)	1	Mass spectrometry	47
2	Drews <i>et al.</i> (2019)	1	Co-transfection	48
3	Yang <i>et al.</i> (2019)	1906	HPIDB	16
4	Eckhardt <i>et al.</i> (2018)	137	Mass Spectrometry	49
5	DeSmet <i>et al.</i> (2018)	4	Co-immunoprecipitation, Mass Spectrometry	50
6	Sankovski <i>et al.</i> (2018)	1	Co-immunoprecipitation, Mass Spectrometry	51
7	Poirson <i>et al.</i> (2017)	47	Co-immunoprecipitation, GPCA	52
8	Spriggs <i>et al.</i> (2017)	3	Chromatin Immunoprecipitation (ChIP)	53
9	Dong <i>et al.</i> (2015)	877	IntAct, APID, VirHostNet	32
10	Tang <i>et al.</i> (2014)	1	Yeast 2 Hybrid, Co-immunoprecipitation	54
11	Jang <i>et al.</i> (2015)	253	Tandem affinity purification, Mass spectrometry	55
12	Kanginakudru <i>et al.</i> (2015)	1	Chromatin immunoprecipitation (ChIP)	56
13	Muller <i>et al.</i> (2012)	57	VirHostNet, VirusMint, PubMed	57
14	Woodham <i>et al.</i> (2012)	1	Co-immunoprecipitation, Electron Paramagnetic Resonance	58
15	Muller <i>et al.</i> (2012)	53	Yeast 2 Hybrid	59
16	Yaginuma <i>et al.</i> (2010)	1	Yeast 2 Hybrid, Co-immunoprecipitation	60
17	Xu <i>et al.</i> (2010)	1	Chromatin immunoprecipitation (ChIP)	61
18	Fertey <i>et al.</i> (2010)	1	Yeast 2 Hybrid, Co-immunoprecipitation	62
19	Côté-Martin <i>et al.</i> (2008)	1	Tandem affinity purification, Mass spectrometry	63
20	Wu <i>et al.</i> (2007)	2	GST pull down assay, Co-transfection	64
21	Zhang <i>et al.</i> (2005)	2	Immunoprecipitation, Transient transfection	65
22	Bernat <i>et al.</i> (2003)	1	Mammalian two-hybrid assay, Chloramphenicol acetyltransferase (CAT) reporter assay, GST pull-down assay and coimmunoprecipitation	66
23	Finnen <i>et al.</i> (2003)	1	Co-immunoprecipitation	67
24	Yang <i>et al.</i> (2003)	1	Immunoprecipitation, Indirect Immunofluorescence	68
25	Mantovani <i>et al.</i> (1999)	1	GST assay	69
26	Massimi <i>et al.</i> (1999)	1	Binding assays, Transient DNA replication assay, co-immunoprecipitation.	70
27	Thomas <i>et al.</i> (1999)	1	Co-transfection	71
28	Patel <i>et al.</i> (1999)	2	Co-immunoprecipitation, GST pull-down assay	72
29	Daniels <i>et al.</i> (1998)	2	Immunoprecipitation, Indirect immunofluorescence analysis	73
30	Swindle <i>et al.</i> (1998)	1	Co-immunoprecipitation	74
31	Jones <i>et al.</i> (1996)	1	Co-immunoprecipitation	75
32	Antinore <i>et al.</i> (1996)	4	Immunoprecipitation, Yeast two-hybrid system	76

**Table 3.** List of studies with accessible protein-protein interaction data of HPV-host.

- Degree of a node refers to the number of associations it has in the overall network, and it is one of the most basic properties of a network.
- Diameter of the network is the characteristic path length i.e. the shortest distance between two nodes in the network.
- Betweenness centrality is a notable parameter to identify essential and critical nodes in a network. It calculates the involvement of a node in the shortest paths in a network.
- Clustering coefficient determines the degree of the connectedness of the node's neighbors.

**Network exploration and enrichment analysis.** The next step in this study is to identify the pivotal human genes in the network and to perform their functional enrichment analysis. The set of genes which are interacting with multiple viral proteins should be playing a crucial part in the pathogenesis of the infections caused by HPV. Through functional enrichment analysis of the top-notch proteins, we can understand their molecular functions and can get to know the important biological processes they are involved in. Protein-protein interactions (PPI) are considered as unbiased data to explore the biological significance of gene sets, and have been used by many researchers for the investigation of gene sets<sup>19,34,35</sup>. Biological networks and pathways have been combined by various scientists in their analyses and they are proved to be successful in practice<sup>36,37</sup>. Enrichment analysis helps us to identify the pathways and biological processes in which a particular gene or gene set might be over-represented. The pathways are ordered on the basis of p value computed by a Fisher's exact test ranging from 0 to 1 which accounts from highly significant to not significant, respectively. We first perform pathway enrichment analysis of top 25 host proteins selected on the basis of the scores computed by CytoHubba. CytoHubba<sup>38</sup> is a novel Cytoscape app which ranks nodes in a network on the basis of network



features. The score of every single node in the network is calculated by eleven network parameters which include degree, Maximum Neighborhood Component, Edge Percolated Component, Maximal Clique Centrality and so on. Through CytoHubba we can find the individual node scores including its degree, clustering coefficient, betweenness, and can make subgraphs of top-ranked nodes.

We also perform KEGG pathway enrichment analysis and Gene Ontology of the highly connected human proteins, filtered from the list of top 25 on the basis of their numbers of interactions with viral proteins in the pathogen-host protein interaction network. The online tool Enrichr<sup>39</sup> (<http://amp.pharm.mssm.edu/Enrichr>), a comprehensive open-source web server, is used for gene set enrichment analysis. Enrichr is user-friendly and currently contains more than 100 gene set libraries with more than 180,000 gene sets. We can perform various analyses using this platform including KEGG (Kyoto Encyclopedia of Genes and Genomics), GO (Gene Ontology), Gene Expression Omnibus (GEO), Online Mendelian Inheritance In Man (OMIM) and several others. We will perform KEGG pathway analysis and GO analysis of the gene set which is associated with most number of viral proteins and has high interactions compared to the other host proteins from the network. The results obtained from Enrichr are based on the combined score of p-value and Z-score where p-value is computed using fisher exact test while z-score is computed as the deviation from the expected rank.

- KEGG<sup>40,41</sup>: KEGG, initially developed in 1995 is now an integrated knowledgebase for analyzing molecular datasets including genomics, metagenomics, proteomics, transcriptomics, metabolomics and other high-throughput biological data<sup>42</sup>. Through pathway enrichment analysis, we can identify the infectious pathways in which the gene set might be playing a crucial part.
- GO<sup>43</sup>: Gene Ontology is the most widely used and comprehensive knowledgebase containing over 7 million gene or gene product annotations of more than 3200 species. GO lets us identify the molecular function and subcellular localization of a gene product and also the biological process in which the gene product might be enriched or utilizing its molecular function in<sup>44,45</sup>.

Pathway enrichment analysis and Gene Ontology are at the core of bioinformatics for analyzing omics data generated by various genome-scale experiments, and with on-going research happening in this field, it is necessary to analyze macromolecular data with the same pace at which they are being produced by high-throughput methodologies. There are a lot of bioinformatics tools available to deal with the omics raw data and yield useful results.

To search for the current drugs available for the treatment of infections caused by HPV and to find out whether or not the existing drugs target the pivotal genes found out in current research, we use a next-generation drug library, the Drug Repurposing Hub<sup>46</sup> which is a collection of more than 3000 clinical drugs targeting 2,247 proteins comprehensively annotated. It has comprehensive annotations of compounds and the information about clinical development status of drugs is available including the drugs that have been launched, or are in their pre-clinical stages or have been withdrawn from use along with the knowledge of their disease area.

Received: 17 December 2019; Accepted: 11 May 2020;

Published online: 26 May 2020

## References

1. Groves, I. J. & Coleman, N. Human papillomavirus genome integration in squamous carcinogenesis: what have next-generation sequencing studies taught us? *J. Pathol.* **245**, 9–18, <https://doi.org/10.1002/path.5058> (2018).
2. Gao, G. *et al.* Whole genome sequencing reveals complexity in both HPV sequences present and HPV integrations in HPV-positive oropharyngeal squamous cell carcinomas. *BMC Cancer* **19**, 352, <https://doi.org/10.1186/s12885-019-5536-1> (2019).
3. Bansal, A., Singh, M. P. & Rai, B. Human papillomavirus-associated cancers: A growing global problem. *Int. J. Appl. Basic. Med. Res.* **6**, 84–89, <https://doi.org/10.4103/2229-516X.179027> (2016).
4. Manikandan, S. *et al.* Knowledge and Awareness Toward Cervical Cancer Screening and Prevention Among the Professional College Female Students. *J. Pharm. Bioallied Sci.* **11**, S314–S320, [https://doi.org/10.4103/JPBS.JPBS\\_21\\_19](https://doi.org/10.4103/JPBS.JPBS_21_19) (2019).
5. Harari, A., Chen, Z. & Burk, R. D. Human papillomavirus genomics: past, present and future. *Curr. Probl. Dermatol.* **45**, 1–18, <https://doi.org/10.1159/000355952> (2014).
6. Paz-Zulueta, M. *et al.* Prevalence of high-risk HPV genotypes, categorised by their quadrivalent and nine-valent HPV vaccination coverage, and the genotype association with high-grade lesions. *BMC Cancer* **18**, 112, <https://doi.org/10.1186/s12885-018-4033-2> (2018).
7. Li, N., Franceschi, S., Howell-Jones, R., Snijders, P. J. F. & Clifford, G. M. Human papillomavirus type distribution in 30,848 invasive cervical cancers worldwide: Variation by geographical region, histological type and year of publication. *Int. J. Cancer* **128**, 927–935, <https://doi.org/10.1002/ijc.25396> (2011).
8. SALINA ZHANG, B., Batur, P. & NCMP, C. J. C. j. o. m. Human papillomavirus in 2019: An update on cervical cancer prevention and screening guidelines. **86**, 173 (2019).
9. Araldi, R. P. *et al.* The human papillomavirus (HPV)-related cancer biology: An overview. *Biomedicine Pharmacotherapy* **106**, 1537–1556, <https://doi.org/10.1016/j.biopha.2018.06.149> (2018).
10. Zheng, Z.-M. & Baker, C. C. Papillomavirus genome structure, expression, and post-transcriptional regulation. *Front. Biosci.* **11**, 2286–2302, <https://doi.org/10.2741/1971> (2006).
11. Graham, S. V. Human papillomavirus: gene expression, regulation and prospects for novel diagnostic methods and antiviral therapies. *Future Microbiol.* **5**, 1493–1506, <https://doi.org/10.2217/fmb.10.107> (2010).
12. Zobel, T., Iftner, T. & Stubenrauch, F. The papillomavirus E8-E2C protein represses DNA replication from extrachromosomal origins. *Mol. Cell Biol.* **23**, 8352–8362, <https://doi.org/10.1128/mcb.23.22.8352-8362.2003> (2003).
13. Wilson, R., Fehrman, F. & Laimins, L. A. Role of the E1–E4 protein in the differentiation-dependent life cycle of human papillomavirus type 31. *J. virology* **79**, 6732–6740, <https://doi.org/10.1128/JVI.79.11.6732-6740.2005> (2005).
14. Ackerman, E. E., Alcorn, J. F., Hase, T. & Shoemaker, J. E. A dual controllability analysis of influenza virus-host protein-protein interaction networks for antiviral drug target discovery. *BMC Bioinforma.* **20**, 297–297, <https://doi.org/10.1186/s12859-019-2917-z> (2019).
15. Goodacre, N., Devkota, P., Bae, E., Wuchty, S. & Uetz, P. Protein-protein interactions of human viruses. *Seminars in Cell & Developmental Biology*, <https://doi.org/10.1016/j.semcdb.2018.07.018> (2018).
16. Yang, S., Fu, C., Lian, X., Dong, X. & Zhang, Z. Understanding Human-Virus Protein-Protein Interactions Using a Human Protein Complex-Based Analysis Framework. *mSystems* **4**, e00303–00318, <https://doi.org/10.1128/mSystems.00303-18> (2019).

17. Kumar, S. *et al.* Inferring pathogen-host interactions between *Leptospira interrogans* and *Homo sapiens* using network theory. *Sci. Rep.* **9**, 1434, <https://doi.org/10.1038/s41598-018-38329-1> (2019).
18. Barabási, A.-L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101–113, <https://doi.org/10.1038/nrg1272> (2004).
19. Wu, X., Jiang, R., Zhang, M. Q. & Li, S. Network-based global inference of human disease genes. *Mol. Syst. Biol.* **4**, 189–189, <https://doi.org/10.1038/msb.2008.27> (2008).
20. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Society. Ser. B* **57**, 289–300 (1995).
21. Li, M., Li, D., Tang, Y., Wu, F. & Wang, J. CytoCluster: A Cytoscape Plugin for Cluster Analysis and Visualization of Biological Networks. *Int. J. Mol. Sci.* **18**, 1880, <https://doi.org/10.3390/ijms18091880> (2017).
22. Morris, J. H. *et al.* clusterMaker: a multi-algorithm clustering plugin for Cytoscape. *BMC Bioinforma.* **12**, 436, <https://doi.org/10.1186/1471-2105-12-436> (2011).
23. Wang, J. *et al.* ClusterViz: A Cytoscape APP for Cluster Analysis of Biological. *Network. IEEE/ACM Trans. computational Biol. bioinformatics/IEEE, ACM* **12**, 815–822, <https://doi.org/10.1109/TCBB.2014.2361348> (2015).
24. Su, G., Morris, J. H., Demchak, B. & Bader, G. D. Biological network exploration with Cytoscape 3. *Curr. Protoc. Bioinforma.* **47**(8), 13.11–18.13.24, <https://doi.org/10.1002/0471250953.bi0813s47> (2014).
25. Nepusz, T., Yu, H. & Paccanaro, A. Detecting overlapping protein complexes in protein-protein interaction networks. *Nat. methods* **9**, 471–472, <https://doi.org/10.1038/nmeth.1938> (2012).
26. Farooq, Qu. A. & Khan, F. F. Construction and analysis of a comprehensive protein interaction network of HCV with its host *Homo sapiens*. *BMC Infect. Dis.* **19**, 367, <https://doi.org/10.1186/s12879-019-4000-9> (2019).
27. Satoh, J. I., Tabunoki, H. & Yamamura, T. Molecular network of the comprehensive multiple sclerosis brain-lesion proteome. *Multiple Scler. J.* **15**, 531–541, <https://doi.org/10.1177/1352458508101943> (2009).
28. Yao, C. *et al.* Multi-level reproducibility of signature hubs in human interactome for breast cancer metastasis. *BMC Syst. Biol.* **4**, 151–151, <https://doi.org/10.1186/1752-0509-4-151> (2010).
29. Soler-López, M., Zanzoni, A., Lluís, R., Stelzl, U. & Aloy, P. Interactome mapping suggests new mechanistic details underlying Alzheimer's disease. *Genome Res.* **21**, 364–376, <https://doi.org/10.1101/gr.114280.110> (2011).
30. Wu, Z.-J., Zhu, Y., Huang, D.-R. & Wang, Z.-Q. Constructing the HBV-human protein interaction network to understand the relationship between HBV and hepatocellular carcinoma. *J. Exp. Clin. Cancer Res.* **29**, 146–146, <https://doi.org/10.1186/1756-9966-29-146> (2010).
31. Eckhardt, M. *et al.* Multiple Routes to Oncogenesis Are Promoted by the Human Papillomavirus-Host Protein Network. *Cancer discovery* **8**, 1474–1489, <https://doi.org/10.1158/2159-8290.CD-17-1018> (2018).
32. Dong, Y. *et al.* Improving the Understanding of Pathogenesis of Human Papillomavirus 16 via Mapping Protein-Protein Interaction Network. *Biomed. Res. Int.* **2015**, 890381–890381, <https://doi.org/10.1155/2015/890381> (2015).
33. Pundir, S., Martin, M. J., O'Donovan, C. & UniProt, C. UniProt Tools. *Curr. Protoc. Bioinforma.* **53**, 1.29.21–21.29.15, <https://doi.org/10.1002/0471250953.bi0129s53> (2016).
34. Chuang, H.-Y., Lee, E., Liu, Y.-T., Lee, D. & Ideker, T. Network-based classification of breast cancer metastasis. *Mol. Syst. Biol.* **3**, 140–140, <https://doi.org/10.1038/msb4100180> (2007).
35. Ruan, J. *et al.* In *Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine* 418–425 (Association for Computing Machinery, Orlando, Florida, 2012).
36. Alexeyenko, A. *et al.* Network enrichment analysis: extension of gene-set enrichment analysis to gene networks. *BMC Bioinforma.* **13**, 226–226, <https://doi.org/10.1186/1471-2105-13-226> (2012).
37. Glaab, E., Baudot, A., Krasnogor, N., Schneider, R. & Valencia, A. EnrichNet: network-based gene set enrichment analysis. *Bioinformatics* **28**, i451–i457, <https://doi.org/10.1093/bioinformatics/bts389> (2012).
38. Chin, C.-H. *et al.* cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* **8**(Suppl 4), S11–S11, <https://doi.org/10.1186/1752-0509-8-S4-S11> (2014).
39. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids Res.* **44**, W90–W97, <https://doi.org/10.1093/nar/gkw377> (2016).
40. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids Res.* **28**, 27–30, <https://doi.org/10.1093/nar/28.1.27> (2000).
41. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic acids Res.* **44**, D457–D462, <https://doi.org/10.1093/nar/gkv1070> (2016).
42. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361, <https://doi.org/10.1093/nar/gkw1092> (2016).
43. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29, <https://doi.org/10.1038/75556> (2000).
44. Hill, D. P., Smith, B., McAndrews-Hill, M. S. & Blake, J. A. Gene Ontology annotations: what they mean and where they come from. *BMC Bioinforma.* **9**, S2, <https://doi.org/10.1186/1471-2105-9-S5-S2> (2008).
45. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330–D338, <https://doi.org/10.1093/nar/gky1055> (2018).
46. Corsello, S. M. *et al.* The Drug Repurposing Hub: a next-generation drug library and information resource. *Nat. Med.* **23**, 405–408, <https://doi.org/10.1038/nm.4306> (2017).
47. Gulati, N. M., Miyagi, M., Wiens, M. E., Smith, J. G. & Stewart, P. L.  $\alpha$ -Defensin HD5 Stabilizes Human Papillomavirus 16 Capsid/Core Interactions. *Pathog. Immun.* **4**, 196–234, <https://doi.org/10.20411/pai.v4i2.314> (2019).
48. Drews, C. M., Case, S. & Vande Pol, S. B. E6 proteins from high-risk HPV, low-risk HPV, and animal papillomaviruses activate the Wnt/ $\beta$ -catenin pathway through E6AP-dependent degradation of NHERF1. *PLoS Pathog.* **15**, e1007575–e1007575, <https://doi.org/10.1371/journal.ppat.1007575> (2019).
49. Eckhardt, M. *et al.* Multiple Routes to Oncogenesis Are Promoted by the Human Papillomavirus-Host Protein Network. *Cancer Discovery* **8**, 1474, <https://doi.org/10.1158/2159-8290.CD-17-1018> (2018).
50. DeSmet, M. *et al.* Papillomavirus E2 protein is regulated by specific fibroblast growth factor receptors. *Virology* **521**, 62–68, <https://doi.org/10.1016/j.virol.2018.05.013> (2018).
51. Sankovski, E., Abroi, A., Ustav, M. & Ustav, M. Nuclear myosin 1 associates with papillomavirus E2 regulatory protein and influences viral replication. *Virology* **514**, 142–155, <https://doi.org/10.1016/j.virol.2017.11.013> (2018).
52. Poirson, J. *et al.* Mapping the interactome of HPV E6 and E7 oncoproteins with the ubiquitin-proteasome system. *FEBS J.* **284**, 3171–3201, <https://doi.org/10.1111/febs.14193> (2017).
53. Spriggs, C. C. & Laimins, L. A. FANCD2 Binds Human Papillomavirus Genomes and Associates with a Distinct Set of DNA Repair Proteins to Regulate Viral Replication. *MBio* **8**, e02340–02316, <https://doi.org/10.1128/mBio.02340-16> (2017).
54. Tang, S. Y. *et al.* Interaction of Daxx and human papillomavirus type 16 E2 protein. *Mol. Biol.* **48**, 594–598, <https://doi.org/10.1134/S0026893314040165> (2014).
55. Jang, M. K., Anderson, D. E., van Doorslaer, K. & McBride, A. A. A proteomic approach to discover and compare interacting partners of papillomavirus E2 proteins from diverse phylogenetic groups. *Proteomics* **15**, 2038–2050, <https://doi.org/10.1002/pmic.201400613> (2015).

56. Kanginakudru, S., DeSmet, M., Thomas, Y., Morgan, I. M. & Androphy, E. J. Levels of the E2 interacting protein TopBP1 modulate papillomavirus maintenance stage replication. *Virology* **478**, 129–135, <https://doi.org/10.1016/j.virol.2015.01.011> (2015).
57. Muller, M. & Demeret, C. The HPV E2-Host Protein-Protein Interactions: A Complex Hijacking of the Cellular Network. *Open. Virol. J.* **6**, 173–189, <https://doi.org/10.2174/1874357901206010173> (2012).
58. Woodham, A. W. *et al.* The S100A10 subunit of the annexin A2 heterotetramer facilitates L2-mediated human papillomavirus infection. *PLoS one* **7**, e43519–e43519, <https://doi.org/10.1371/journal.pone.0043519> (2012).
59. Muller, M. *et al.* Large scale genotype comparison of human papillomavirus E2-host interaction networks provides new insights for e2 molecular functions. *PLoS Pathog.* **8**, e1002761–e1002761, <https://doi.org/10.1371/journal.ppat.1002761> (2012).
60. Yaginuma, Y., Yoda, K. & Ogawa, K. Characterization of Physical Binding between Human Papillomavirus 18 Protein E7 and Centromere Protein C. *Oncology* **79**, 219–228, <https://doi.org/10.1159/000322188> (2010).
61. Xu, M., Katzenellenbogen, R. A., Grandori, C. & Galloway, D. A. NFX1 plays a role in human papillomavirus type 16 E6 activation of NFkappaB activity. *J. virology* **84**, 11461–11469, <https://doi.org/10.1128/JVI.00538-10> (2010).
62. Fertey, J. *et al.* Interaction of the papillomavirus E8–E2C protein with the cellular CHD6 protein contributes to transcriptional repression. *J. virology* **84**, 9505–9515, <https://doi.org/10.1128/JVI.00678-10> (2010).
63. Côté-Martin, A. *et al.* Human papillomavirus E1 helicase interacts with the WD repeat protein p80 to promote maintenance of the viral genome in keratinocytes. *J. virology* **82**, 1271–1283, <https://doi.org/10.1128/JVI.01405-07> (2008).
64. Wu, M.-H. *et al.* Physical and functional interactions of human papillomavirus E2 protein with nuclear receptor coactivators. *Biochemical Biophysical Res. Commun.* **356**, 523–528, <https://doi.org/10.1016/j.bbrc.2007.02.162> (2007).
65. Zhang, Y. *et al.* BRC1 Interaction with Human Papillomavirus Oncoproteins. **280**, 33165–33177, <https://doi.org/10.1074/jbc.M505124200> (2005).
66. Bernat, A., Avvakumov, N., Mymryk, J. S. & Banks, L. Interaction between the HPV E7 oncoprotein and the transcriptional coactivator p300. *Oncogene* **22**, 7871–7881, <https://doi.org/10.1038/sj.onc.1206896> (2003).
67. Finnen, R. L., Erickson, K. D., Chen, X. S. & Garcea, R. L. Interactions between Papillomavirus L1 and L2 Capsid Proteins. *J. Virology* **77**, 4818, <https://doi.org/10.1128/JVI.77.8.4818-4826.2003> (2003).
68. Yang, R., Yutzy, W. H., Viscidi, R. P. & Roden, R. B. S. Interaction of L2 with  $\beta$ -Actin Directs Intracellular Transport of Papillomavirus and Infection. **278**, 12546–12553, <https://doi.org/10.1074/jbc.M208691200> (2003).
69. Mantovani, F. & Banks, L. The interaction between p53 and papillomaviruses. *Semin. Cancer Biol.* **9**, 387–395, <https://doi.org/10.1006/scbi.1999.0142> (1999).
70. Massimi, P., Pim, D., Bertoli, C., Bouvard, V. & Banks, L. Interaction between the HPV-16 E2 transcriptional activator and p53. *Oncogene* **18**, 7748–7754, <https://doi.org/10.1038/sj.onc.1203208> (1999).
71. Thomas, M., Pim, D. & Banks, L. The role of the E6-p53 interaction in the molecular pathogenesis of HPV. *Oncogene* **18**, 7690–7700, <https://doi.org/10.1038/sj.onc.1202953> (1999).
72. Patel, D., Huang, S. M., Baglia, L. A. & McCance, D. J. The E6 protein of human papillomavirus type 16 binds to and inhibits co-activation by CBP and p300. *EMBO J.* **18**, 5061–5072, <https://doi.org/10.1093/emboj/18.18.5061> (1999).
73. Daniels, P. R., Sanders, C. M. & Maitland, N. J. Characterization of the interactions of human papillomavirus type 16 E6 with p53 and E6-associated protein in insect and human cells. **79**, 489–499, <https://doi.org/10.1099/0022-1317-79-3-489> (1998).
74. Swindle, C. S. & Engler, J. A. Association of the human papillomavirus type 11 E1 protein with histone H1. *J. virology* **72**, 1994–2001 (1998).
75. Jones, D. L. & Münger, K. Interactions of the human papillomavirus E7 protein with cell cycle regulators. *Semin. Cancer Biol.* **7**, 327–337, <https://doi.org/10.1006/scbi.1996.0042> (1996).
76. Antinore, M. J., Birrer, M. J., Patel, D., Nader, L. & McCance, D. J. The human papillomavirus type 16 E7 gene product interacts with and trans-activates the AP1 family of transcription factors. *EMBO J.* **15**, 1950–1960 (1996).

## Acknowledgements

I would like to thank my parents for always believing in me and being with me in every situation, my supervisor Professor Chunhua Li for guiding me throughout this research. I would also like to thank all my friends for their immense support. This work was supported by the National Natural Science Foundation of China [31971180].

## Author contributions

Q.F. conceived of the study. C.L. provided assistance with experimental design and sample section. Q.F., Z.S., S.A. performed the experiment and analyzed the results. T.Z. and W.G. refined the figures. All authors read and reviewed the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-65837-w>.

**Correspondence** and requests for materials should be addressed to C.L.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020