Gynecologic Oncology 152 (2019) 629–637

Contents lists available at ScienceDirect

Gynecologic Oncology

journal homepage: www.elsevier.com/locate/ygyno

Genomic characterization of cervical cancer based on human papillomavirus status

Liya Zhang [a,1], Yi Jiang [b,1], Xiaofan Lu [a], Huiling Zhao [a], Chen Chen [a], Yaoyan Wang [a], Wenjun Hu [a], Yue Zhu [a], Hangyu Yan [a], Fangrong Yan [a,*]

[a] Research Center of Biostatistics and Computational Pharmacy, China Pharmaceutical University, Nanjing, PR China
[b] Department of Gynecology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, PR China

HIGHLIGHTS

- HPV− tumors were likely to occur at an older age and were often adenocarcinomas or adenosquamous carcinomas.
- 5 genes were significantly differentially mutated between HPV+ and HPV− tumors.
- *CDO1*, *PCDHB2* and *MYOD1* demonstrated different response to radiotherapy between HPV+ and HPV− tumors.
- 4 genes were different drug-response to cisplatin between HPV+ and HPV− cervical cancer tumors.

ARTICLE INFO

Article history:
Received 5 September 2018
Received in revised form 10 December 2018
Accepted 16 December 2018
Available online 20 December 2018

Keywords:
Cervical cancer
HPV+
HPV−
Biomarkers
Radiotherapy
Drug susceptibility

ABSTRACT

Objective. It is uncommon for cervical cancer patients to be diagnosed without a human papillomavirus (HPV) infection. As prophylactic vaccines against high-risk HPV types are an ineffective preventive measure for these patients it is essential to identify differential biomarkers that may be associated with detection, prognosis and novel targeted therapies. The objective of this study was to compare the two entities, HPV+ and HPV− cervical cancers, based on TCGA public data.

Methods. We collected and analyzed clinical information of 299 cervical cancer patients as the first step, then identified differential expressed genes and conducted downstream analyses to characterize this tumor based on HPV status, including functional annotation, pathway mapping, survival analysis and comparative somatic mutation landscapes. We further inferred the likelihood of responding to traditional treatment including radiotherapy and chemotherapy.

Results. It was found that HPV− tumors were likely to occur at an older age and were often adenocarcinomas or adenosquamous carcinomas, and there was no significant overall survival difference between HPV+ vs. HPV− tumors. Gene expression profiles of HPV+ and HPV− tumors differed especially in *ANKRD7*, *SERPINB3*, *EMX2*, *MEI1*, *RNF212*, *RP11-13 K12.5*, *RP11-325F22.2* and *ZFR2* which were significantly relevant to cervical cancer prognosis. *TP53*, *ARID5B*, *ARID1A*, *CTNNB1* and *PTEN* were significantly differentially mutated between HPV+ and HPV− tumors. Results of radiotherapy analyses demonstrated that *CDO1*, *PCDHB2* and *MYOD1* were different between the two subsets. In addition, *RP11-299 L17.3*, *SLC14A2*, *FGF18* and *OASL* represented different drug-sensitivity to cisplatin between both.

Conclusions. These potential biomarkers may offer insights to further personalize therapeutic decision-making to improve survival in HPV− cervical cancer patients.

Abbreviations: CESC, Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma; TCGA, The Cancer Genome Atlas; DEGs, differential expressed genes; HPV, human papillomavirus; HPV+, HPV-positive; HPV−, HPV-negative; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; DAVID, Database for Annotation, Visualization and Integrated Discovery; GDSC, Genomics of Drug Sensitivity in Cancer.

* Corresponding author at: Research Center of Biostatistics and Computational Pharmacy, China Pharmaceutical University, Nanjing 210009, PR China.
E-mail address: f.r.yan@163.com (F. Yan).

[1] The authors wish it to be known that, in their opinion, the first 2 authors should be regarded as joint First Authors.

https://doi.org/10.1016/j.ygyno.2018.12.017
0090-8258/

1. Introduction

Cervical cancer, arising from the cervix, is one of the most commonly diagnosed tumors and the major cause of cancer death among females, especially in developing countries. Globally, approximately 528,000 cases of cervical cancer emerge each year, causing a total of 266,000 deaths, which is more than any other gynecologic tumors [1]. There are about 13,240 new cervical cancer cases and 4170 deaths in 2018 in the United States [2]. As early as 1983, the research group of Harald zur Hausen extracted human papillomavirus (HPV) type16 DNA which appeared in about half of cervical cancer biopsies samples and cell lines [3]. While the majority of HPV infections are cleared spontaneously by the human immune system, a few persist for years and eventually cause cancer, so many researchers had a consensus that persistent HPV infection, which is HPV+, was the main etiological cause of cervical cancer [4–6]. With the rapid development of genomic sequencing technology, numerous studies have focused on the molecular features and mechanism of HPV+ cervical cancers. For instance, it has been manifested that *PIK3CA*, *PTEN*, *TP53*, *STK11*, *KRAS*, *SHKBP1*, *ERBB3*, *HLA-A*, *CASP8* and *TGFBR2* are significantly mutated genes in cervical cancer [7–11].

Recently, researchers from The Cancer Genome Atlas Research Network (CGARN) identified novel genomic and molecular features of cervical cancer in order to support the molecular classification technique and realize targeted therapy. They also found that about 5% of primary cervical cancers were not caused by persistent HPV infection and might be triggered by genetic alteration or other factors [11,12]. Although lots of studies on cervical cancer have yielded excellent results, few studies have deconvoluted the molecular differences across cervical cancer patients based on different HPV status, that is, HPV+ and HPV−. Additionally, the potential genomic biomarkers of HPV− cervical cancer remain unclear.

Similar to recent reports in head and neck cancers [13,14], this study sought to identify biomarkers that distinguish HPV+ and HPV− patients, and made an integrative comparison from clinical level, genomic level and treatment level across cervical cancer patients with different HPV status. Above all, the study analyzed clinical information and identified differential expressed genes (DEGs) between the two groups, then GO enrichment analysis, KEGG pathway enrichment analysis and survival analysis were performed to reveal the behind biological functions among DECs. Subsequently, the mutation rates of genes were obtained and the per gene somatic mutation frequencies were compared statistically between the two classes. Finally, the sensitivity of response to radiotherapy and chemotherapy between these two entities was also compared. The findings may help reveal molecular differences between HPV+ and HPV− samples and offer potential biomarkers for cervical cancer diagnosis and treatment. With the rapid pace of discovery and developments in biologic and computational research we are poised for a better understanding of HPV− cervical cancer and anticipate these discoveries will improve patient treatment and provide novel therapies.

2. Materials and methods

2.1. Overall workflow

The overall research workflow shown in Fig. 1 is broken down into two parts, representing distinct levels of analysis, including clinical level, genomic level and treatment level. In Fig. 1A, clinical information was analyzed to identify human mRNAs and miRNAs that were differentially expressed in cervical cancer HPV+ versus HPV− samples. Gene function annotation and GO enrichment analysis, KEGG pathway enrichment analysis and survival analysis were performed on these differential expressed genes. Besides, somatic mutation landscapes were compared between multiple tumor cohorts by analyzing MAF files.

In addition, treatment analyses of cervical cancer, including radiotherapy analysis and drug susceptibility analysis, were conducted in the second-level research (Fig. 1B). Differential expressed genes were identified between cervical cancer patients who responded completely to radiotherapy and those who failed to completely respond, and on these genes downstream analyses were performed, such as gene function annotation, pathway analysis, and so on. Then cervical cancer cell lines data downloaded from Genomics of Drug Sensitivity in Cancer (GDSC) database (http://www.cancerrxgene.org/) was used to analyze drug sensitivity, which aims to answer how HPV status-based differentially expressed genes affect the response to cisplatin, a commonly dosed chemo in clinical treatment for cervical cancer.

2.2. Data sets

Five data sets were used in our study: human mRNA expression data, human miRNA expression data, somatic mutation data, clinical data and 14 cervical cancer cell lines data with drugs response information (abbreviated as GDSC set). The mRNA-Seq genes expression data (Level 3, 304 samples) and miRNA-Seq expression data/(Level 3, 312 samples) of Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (CESC) were downloaded from The Cancer Genome Atlas (TCGA [15]) GDC Data Portal (https://portal.gdc.cancer.gov/). Somatic mutations

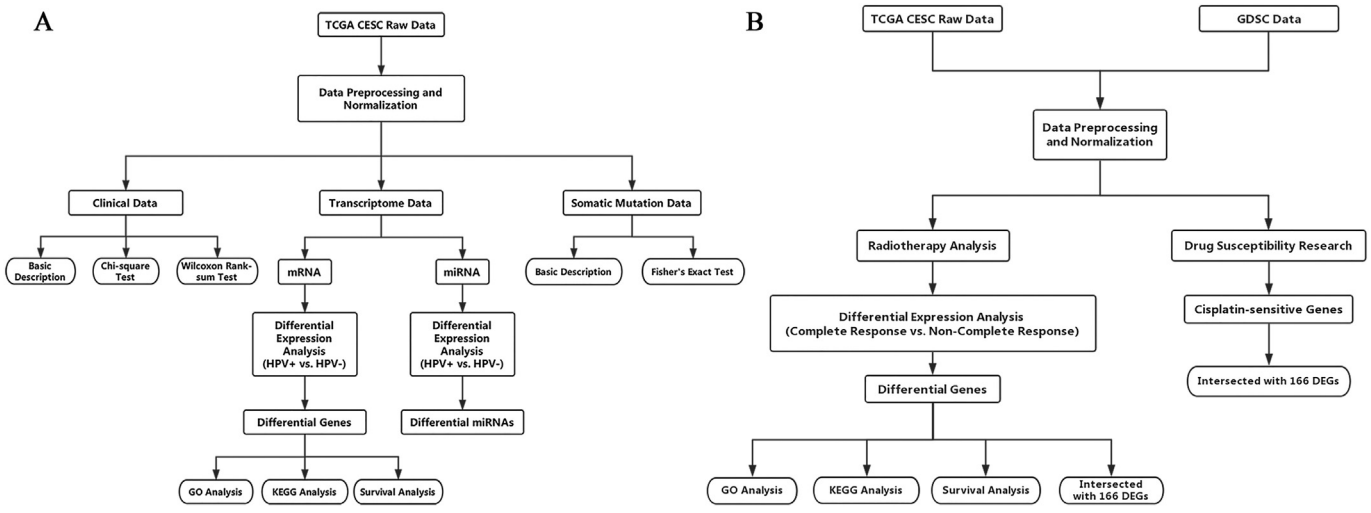

Fig. 1. Data analysis overall workflow. (A)The workflow of CESC clinical, transcriptome and somatic mutation data analyses between HPV+ and HPV−. (B) The workflow of treatment analyses, including radiotherapy analysis and drug susceptibility analysis.

for CESC were extracted from the gsc_CESC_pairs.aggregated.capture.tcga.uuid.automated.somatic.maf file available for download on the TCGA website, with somatic mutations available for primary tumors of 297 samples. The CESC clinical data set that comprises a total of 307 patients was obtained from Firehose (http://firebrowse.org). Moreover, 14 cervical cancer cell lines data and corresponding drugs response information was collected from GDSC database [16,17].

2.3. Data preprocessing and normalization

Various data sets were processed first as follows: First, we collated and checked the clinical data of cervical cancer patients, such as HPV test results, HPV type, histological type, radiotherapy information, etc. Second, the normal samples (3 samples in CESC TCGA data set) and tumor samples (301 samples in mRNA set and 309 samples in miRNA set) were identified. For tumor samples, 301 samples were selected from three data sets (except somatic mutation data set and GDSC set) for further analysis. 279 HPV+ samples (the specific HPV type distribution in Table S1) and 22 HPV− samples were identified from these three data sets. Furthermore, tumor samples and HPV test results were matched among mRNA, miRNA, clinical and somatic mutation data sets. By mapping them to TCGA barcode, 276 HPV+ samples and 21 HPV− samples in somatic mutation data set were detected.

The study used raw read counts of mRNA and miRNA expression data to detect differentially expressed mRNA and miRNA. As raw read counts follow a negative binomial distribution, the voom methodology [18] was applied for data normalization by using limma R package. The voom method estimates the mean-variance of the log-counts and produces a precision weight for each observation. The final normalized counts of mRNA and miRNA expression data were generated for all further bioinformatics analysis.

2.4. Statistical analysis of clinical data

R software was utilized to conduct basic descriptive analysis of the clinical information and Wilcoxon rank-sum test was performed for real-valued data, such as age and survival time. The Chi-square or Fisher's exact tests for categorical data were executed by SPSS 22.0 with calculation of a two-sided p-value, and samples with invalid record were not included in the test for significance. Statistically significance was determined with a threshold of p-value <0.05.

2.5. Differential expression analysis

Differentially expressed features based on different HPV status for CESC were found using DESeq [19] and edgeR [20] R packages with threshold of log-foldchange >2 and FDR < 0.05. The intersections of results calculated by these two methods were regarded as the final results for accuracy. The sensitivity and specificity of the top 5 most significant genes with reference to HPV status were calculated by pROC R package.

2.6. Gene functional enrichment analysis and KEGG pathway analysis

Gene functional annotation, i.e., Gene Ontology (GO) enrichment analysis was performed through DAVID 6.8 (https://david.ncifcrf.gov/) to reveal the biological meaning for a list of genes which were discovered through differential expression analysis. In order to synthetically understand the biological functions of multiple genes, KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway database information was applied and visualized via clusterProfiler [21] R package and Cytoscape software [22].

2.7. Survival analysis

Survival analysis of cervical cancer patients was conducted based on the differential genes expression between HPV+ and HPV− samples. Kaplan-Meier survival curves were used to display the overall survival differences between patients with higher and lower gene expression compared to the median values of selected gene expression. If gene expression level in a patient was lower than median, the patient should be classified as the lower-expression group; otherwise as the higher-expression group. The log-rank test for Kaplan-Meier curve was calculated, and significant survival time differences between two groups of patients was defined as p-value <0.05.

2.8. Somatic mutation data analysis

GenVisR [23] R package was applied to calculate the mutation rate of genes and to draw the somatic mutation oncoprint between HPV+ and HPV− samples. The somatic mutation frequencies per gene between classes were compared using Fisher's exact test, and the corresponding conditional maximum likelihood estimate of odds ratio and 95% CI were plotted in Fig. 4. A logistic regression model was constructed by pooling the top 10 significantly differentially mutated genes between HPV+ and HPV− tumor. We accessed the prediction performance of this model by calculating its AUROC (the Area Under the curve of the Receiver Operating Characteristic) by pROC R package.

2.9. Radiotherapy analysis

The treatment information used in the study came from the TCGA CESC clinical data set. It contained the data of 49 patients who only received radiotherapy treatment, including 40 cases that completely responded to radiation therapy and 9 cases of partial response or disease progression (referred to as Non-complete Response) of patients. Differentially expressed genes between complete response and non-complete response to radiotherapy were identified, and downstream analyses were performed to find the genes that are sensitive to radiotherapy. Then the intersections of differential expression genes between complete response vs. non-complete response group and HPV+ vs. HPV− group represented the difference in radiotherapy response between HPV+ and HPV− samples.

2.10. Drug susceptibility research

Since treatment data for cervical cancer patients in TCGA only provided information of radiotherapy and combination of chemotherapy and radiotherapy, it is necessary to find specialized drug response data to conduct a comprehensive drug susceptibility analysis between HPV + and HPV−. Data of 14 cervical cancer cell lines including medication information was selected and downloaded from GDSC database to analyze drug sensitivity. In order to understand whether the expression of HPV+ and HPV− differentially expressed genes affect the efficacy of platinum drugs, focus was mainly placed on the analysis of cisplatin treatment information.

3. Results

3.1. Clinical correlation characteristics

No significant difference was found in the distributions of patient ethnicity, survival status, BMI, tobacco use, tobacco smoking history, total number of pregnancies, number of successful pregnancies, tumor grade, clinical stage and lymph node metastasis between HPV+ and HPV− cervical cancer patients (Table 1). However, Table 1 shows that there is high relevance between histological type and HPV infection (p-value = 0.001), which suggests that adenocarcinomas and adenosquamous carcinomas might be more common in HPV− than in HPV+ cervical cancer. Additionally, patients with HPV− were older at initial diagnosis than patients with HPV+ (Wilcoxon rank-sum test, p-value = 0.012), and Fig. S1 more intuitively depicts the age difference between HPV− and HPV+ patients. As shown in Fig. S2, no significant

Table 1
Overview of clinical characteristics of patients.

	Total	HPV+ (HPV+)	HPV− (HPV−)	p-Value
Patients	299	278	21	
Age at Initial Diagnosis				
Mean ± Sd	48.20 ± 13.87	47.65 ± 13.77	55.48 ± 13.48	
Ethnicity				0.139
American Indian or Alaska Native	7 (2.6%)	5 (2.0%)	2 (11.1%)	
Asian and Pacific Islander	22 (8.3%)	21 (8.5%)	1 (5.6%)	
Black or African American	30 (11.3%)	28 (11.4%)	2 (11.1%)	
White	206 (77.8%)	193 (78.1%)	13 (72.2%)	
Unknown	34	31	3	
Survival Status				0.562
Alive	229 (76.6%)	214 (77.0%)	15 (71.4%)	
Dead	70 (23.4%)	64 (23.0%)	6 (28.6%)	
BMI (WHO Standards)				0.834
Underweight (<18.50)	12 (4.7%)	11 (4.7%)	1 (5.3%)	
Normal range (18.50–24.99)	86 (33.7%)	78 (33.1%)	8 (42.1%)	
Overweight (25.00–29.99)	73 (28.7%)	69 (29.1%)	4 (21.0%)	
Obesity (≥30.00)	84 (32.9%)	78 (33.1%)	6 (31.6%)	
Unknown	44	42	2	
Tobacco Use				0.813
Never/light smoker (<10 pack years)	30 (32.6%)	28 (32.9%)	2 (28.6%)	
Heavy smoker (≥10 pack years)	62 (67.4%)	57 (67.1%)	5 (71.4%)	
Pack years unknown	207	193	14	
Tobacco Smoking History				0.586
1	142 (55.3%)	132 (55.5%)	10 (52.6%)	
2–3	72 (28.0%)	65 (27.3%)	7 (36.9%)	
4–5	43 (16.7%)	41 (17.2%)	2 (10.5%)	
Unknown	42	40	2	
Total Number of Pregnancies				0.064
0–3	149 (55.8%)	133 (53.8%)	16 (80.0%)	
4–7	100 (37.5%)	96 (38.9%)	4 (20.0%)	
≥8	18 (6.7%)	18 (7.3%)	0 (0.0%)	
Unknown	32	31	1	
Number of Successful Pregnancies				0.103
0–2	143 (53.6%)	128 (51.8%)	15 (75.0%)	
3–5	104 (38.9%)	99 (40.1%)	5 (25.0%)	
≥6	20 (7.5%)	20 (8.1%)	0 (0.0%)	
Unknown	32	31	1	
Histological Type				0.001
Adenocarcinoma	46 (15.4%)	39 (14.0%)	7 (33.3%)	
Adenosquamous	5 (1.7%)	3 (1.1%)	2 (9.5%)	
Squamous Cell Carcinoma	248 (82.9%)	236 (84.9%)	12 (57.2%)	
Grade				0.352
G1	18 (6.7%)	18 (7.2%)	0 (0.0%)	
G2	133 (49.4%)	125 (50.2%)	8 (40.0%)	
G3	117 (43.5%)	105 (42.2%)	12 (60.0%)	
G4	1 (0.4%)	1 (0.4%)	0 (0.0%)	
Unknown	30	29	1	
Clinical Stage				0.418
I	161 (55.1%)	151 (55.7%)	10 (47.6%)	
II	67 (22.9%)	61 (22.5%)	6 (28.6%)	
III	44 (15.1%)	42 (15.5%)	2 (9.5%)	
IV	20 (6.9%)	17 (6.3%)	3 (14.3%)	
Unknown	7	7	0	
Lymph node metastasis				0.936
Metastasis	52 (32.9%)	49 (33.3%)	3 (27.3%)	
No Metastasis	106 (67.1%)	98 (66.7%)	8 (72.7%)	
Unknown	141	131	10	

overall survival difference between the HPV+ vs. HPV− patients could be observed (p-value = 0.254).

3.2. Differential expression analysis in HPV+ vs. HPV− samples

Differential expression analysis was first performed to find molecular differences between HPV+ and HPV− cervical cancer samples. After comparing the human mRNA expression levels of 27,959 genes, 167 genes were identified as significantly differentially expressed via DESeq R-package (Fig. 2A). 903 genes were found by edgeR R-package. Ultimately, 166 differential expression genes (DEGs) were obtained from the intersection of results, which did not contain *ACSM3*, calculated by these two methods. In these 166 DEGs, 151 genes (90.96%) were overexpressed in the HPV− CESC samples. The AUROC (area under ROC), sensitivity and specificity of the top 5 most significant DEGs with reference to HPV status are shown in Table S2, and Fig. S3 demonstrates the receiver operating characteristic (ROC) curves of these five genes. Almost all the AUROC are >0.75 except for CADM1 (AUROC = 0.696).

Based on gene expression data for 166 DEGs, the cervical cancer samples were divided into high-age (age ≥ 48.20, n = 129) and low-age (age < 48.20, n = 172) groups for differential expression analysis and no age-oriented significant difference could be found. Similarly, we divided samples into adenocarcinoma or adenosquamous carcinomas (n = 52) and squamous cell carcinomas (n = 249) groups and found that 12 of 166 DEGs were significantly related to histological

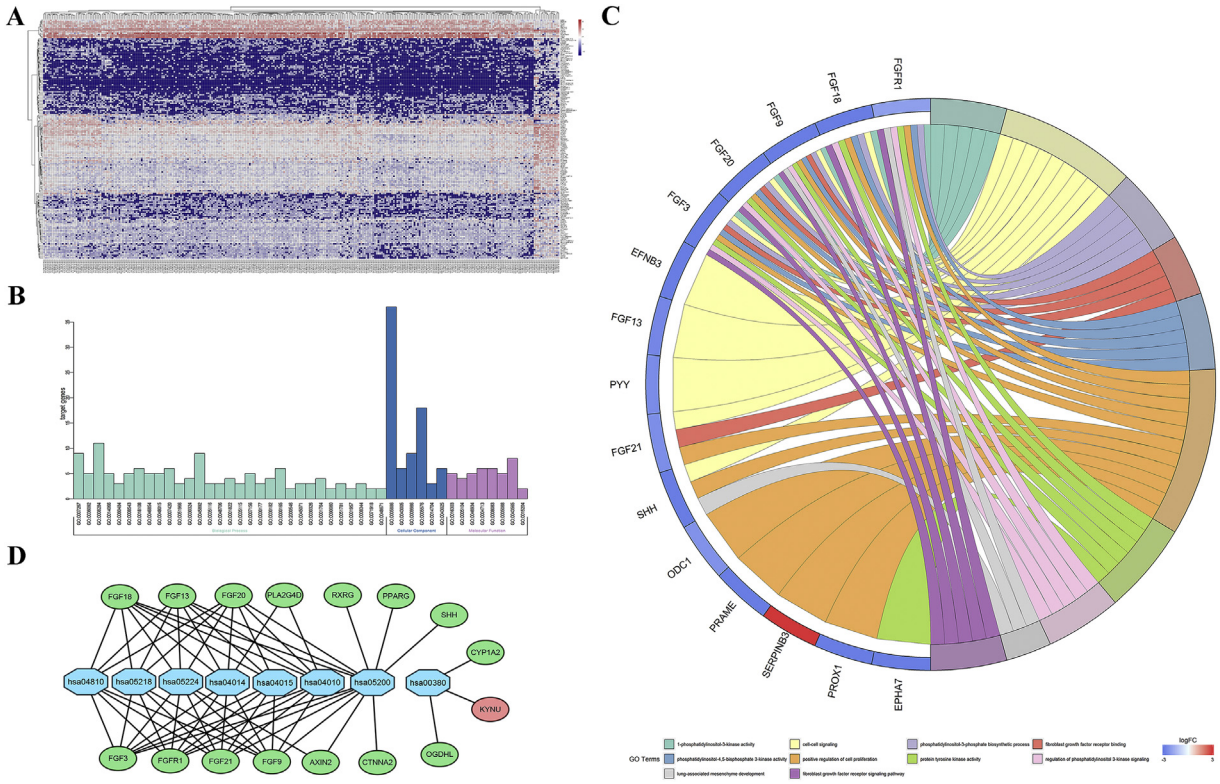

Fig. 2. (A) Heatmap of 167 genes that were differentially expressed between HPV+ and HPV− cervical tumors via DESeq R-package. (B) The histogram of 45 significantly enriched GO terms. 31 GO terms (green) represented biological process, 6 GO terms (blue) displayed cellular component and 8 GO terms (purple) denoted molecular function. (C) GO annotation and enrichment plot of the top ten most significant GO terms and 15 genes from 166 DEGs. Among them, only *SERPINB3* was lower expressed in HPV− samples (red in the left semicircle) and the remaining 14 genes were overexpressed in HPV− cervical cancer samples (blue in the left semicircle). The different colors on the right half circle represent diverse GO terms. (D) The graph of KEGG pathway analysis by using Cytoscape software. Eight statistically significant pathways (blue) and 16 corresponding genes were obtained. *KYNU* was significantly lower in HPV− samples than in HPV+ samples (red) and others (green) were reversed.

type. Among them, *KRT14*, *AIM2*, *LY6D*, *SERPINB3*, *AQP3*, *MMP28* and *PLA2G4D* were up-regulated in squamous cell carcinoma, while *STXBP6*, *KLHL14*, *DACH1*, *SLC26A7* and *UCMA* were down-regulated.

In addition, human miRNA-level differential expression analysis between HPV+ and HPV− tumors revealed that *has-mir-182*, *has-mir-183* and *has-mir-5002* which were significantly overexpressed were only found in the HPV− samples (p-value <0.05).

3.3. Gene functional enrichment analysis and KEGG pathway analysis for 166 DEGs

In order to explore the behind biological function of genes, Gene Ontology (GO) enrichment analysis and KEGG pathway analysis were performed for 166 DEGs discovered above. First of all, 45 significantly enriched GO terms was identified (Fig. 2B) and a maximum of 38 genes were enriched in GO: 000586~plasma membrane (p-value = 0.0073) which is related with the cellular component. It is also found that the most prominent GO term was GO:0016303—1-phosphatidylinositol-3-kinase activity (p-value = 0.0001), and interestingly, there were six GO terms associated with phosphatidylinositol including GO:0016303, GO:0036092—phosphatidylinositol-3-phosphate biosynthetic process (p-value = 0.0002), GO:0046934—phosphatidylinositol-4,5-bisphosphate 3-kinase activity (p-value = 0.0005), GO:0014066—regulation of phosphatidylinositol 3-kinase signaling (p-value = 0.0014), GO:0046854—phosphatidylinositol phosphorylation (p-value = 0.0028) and GO:0048015—phosphatidylinositol-mediated signaling (p-value = 0.0043). Phosphatidylinositol 3-kinase (PI3K) is a lipid kinase and a key signal transduction enzyme involving in cell survival and proliferation, cell motility and adhesion, cytoskeletal rearrangement and vesicle trafficking [24]. Fig. 2C demonstrates the top ten most significant GO terms and 15 genes related to these terms, wherein *FGFR1*, *FGF18*, *FGF9*, *FGF20* and *FGF3* were associated with the aforementioned phosphatidylinositol, whereas only *SERPINB3* correlated with GO: 0008284~positive regulation of cell proliferation (p-value = 0.0006) was lower expressed in HPV− samples, and the remaining 14 genes, such as *SHH*, *PRAME* and *EPHA7* etc., were overexpressed in HPV− cervical cancer patients.

As shown in Table S3 and Fig. 2D, through the KEGG pathway analysis, eight statistically significant pathways and 16 corresponding genes were obtained. About 75% (12/16) of the genes were assigned to hsa05200: pathways in cancer (p-value = 0.00029), and it is found that *KYNU* involved in hsa00380: tryptophan metabolism was significantly lower in HPV− samples than in HPV+ samples.

3.4. Survival analysis for 166 DEGs

To reveal correlation between 166 differential genes expression levels and CESC prognosis, survival analysis was conducted and 8 genes were found significantly associated with CESC prognosis (log-rank test, p-value < 0.05). Kaplan-Meier curves (Fig. 3) show that cervical cancer patients with lower expression levels of *ANKRD7* and *SERPINB3* have better overall survival prognoses than those with higher expression levels of these 2 genes. Similarly, patients with higher expression levels of 6 genes including *EMX2*, *MEI1*, *RNF212*, *RP11-13 K12.5*, *RP11-325F22.2* and *ZFR2* have better overall survival prognoses than those with lower expression levels of these 6 genes, which were indicated as having significantly different expression in HPV+ vs. HPV− tumors. It is worth noting that *ANKRD7*, *EMX2*, *RP11-13 K12.5* and *RP11-325F22.2* were higher expressed in HPV− samples, while *MEI1*, *RNF212*, *SERPINB3* and *ZFR2* were higher expressed in HPV+ CESC samples.

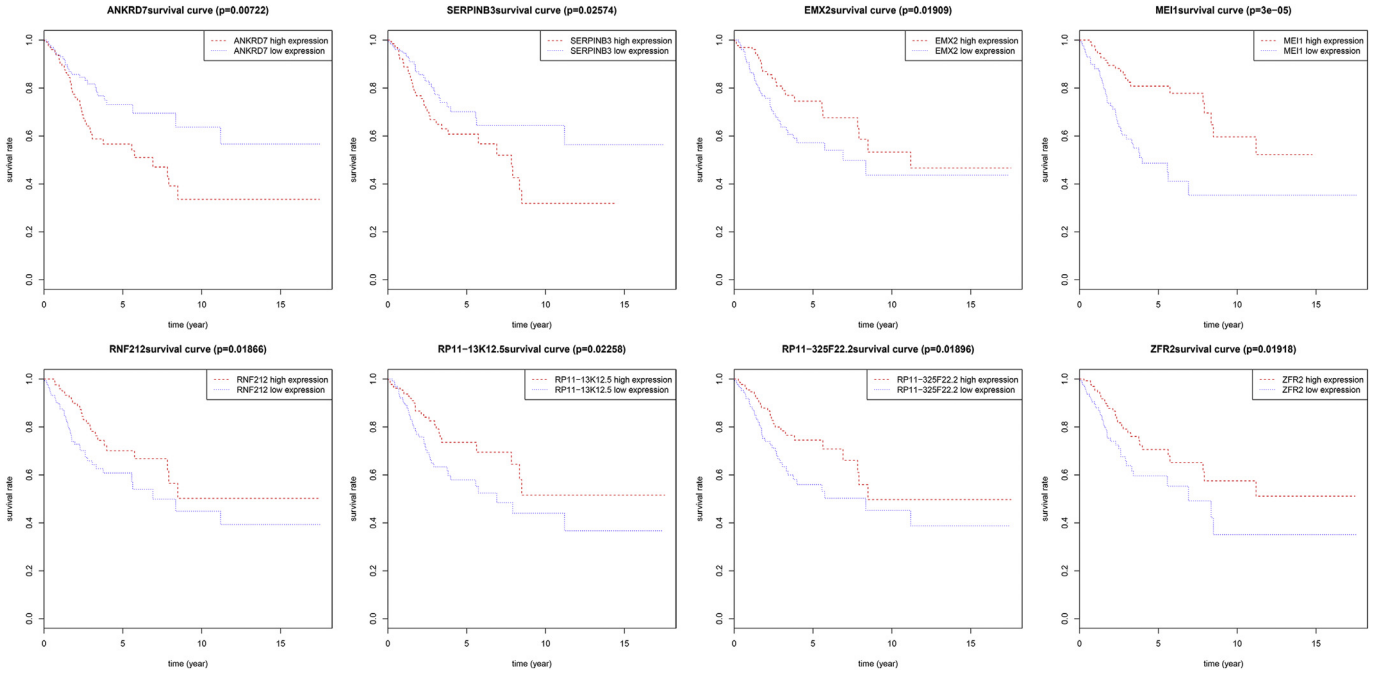

Fig. 3. Kaplan-Meier (KM) survival curves for 8 genes. KM survival curves show significant overall survival differences between higher-expression levels and lower-expression levels of cervical cancer patients.

3.5. Somatic mutation landscape differences

The mutation rate of 18,749 genes between HPV+ and HPV− samples was calculated and tested via Fisher's exact test. From Fig. S4, it can be seen that the top ten mutated genes ranked by mutation rate in HPV + samples are *TTN* (42.75%), *MUC4* (37.32%), *PIK3CA* (31.16%), *MUC16* (24.28%), *KMT2C* (22.83%), *BAGE2* (21.74%), *KMT2D* (16.67%), *SYNE1* (15.22%), *CROCCP2* (15.22%), *SNHG14* (14.86%), and the top ten mutated genes in HPV− samples are *PIK3CA* (52.38%), *TP53* (47.62%), *TTN* (38.10%), *SYNE1* (38.10%), *MUC4* (38.10%), *ARID1A* (38.10%), *SNHG14* (33.33%), *PTEN* (33.33%), *PCLO* (33.33%), *KMT2D* (33.33%). In general, the frequencies with which cervical cancer genes are mutated are similar between the two classes. However, among the18749 genes, 2133 genes had significantly different distributions in the mutation frequencies between HPV+ and HPV− samples (a two-tailed Fisher's exact p-value <0.05). 46 genes had different somatic mutation rates and were all more frequently mutated in HPV− tumors (Fig. 4). For example, *TP53*, a well-known tumor suppressor, which prevents cancer initiation by maintaining genomic stability [25–27], was more likely to be mutated in HPV− tumors (47.62%) than in HPV+ tumors (3.99%) (a two-tailed Fisher's exact p-value = 6.63E−08). This is consistent with the absence of E6 HPV oncogene expression in HPV− tumors. The enrichment of *TP53* mutations in the HPV− tumors provides convincing evidence that these tumors have escaped HPV dependence. The prediction performance of the logistic regression model represented by AUROC was up to 0.895 (Fig. S5).

In addition, the results indicate that *ARID5B*, *ARID1A*, *CTNNB1* and *PTEN* are significantly differentially mutated between HPV+ and HPV − tumors.

3.6. Radiotherapy

98 differentially expressed genes between complete response samples and non-complete response samples to radiotherapy were identified, where *ALOX15* and *IGLV3–19* were up-regulated and 96 were down-regulated. According to downstream analyses of these 98 genes, 23 significantly enriched GO terms were obtained, but no statistical significant KEGG pathway was found. Then, the intersection of 166 DEGs mentioned above and the 98 genes was analyzed, and only three genes, *CDO1*, *PCDHB2* and *MYOD1*, demonstrated different response to radiotherapy between HPV+ and HPV− samples. As shown in Fig. S6, *PCDHB2* correlated with GO: 0005887—integral component of plasma membrane (p-value = 0.0018), GO: 0016339—calcium-dependent cell-cell adhesion via plasma membrane cell adhesion molecules (p-value = 0.0029), GO: 0007416—synapse assembly (p-value = 0.0132) and GO: 0005886—plasma membrane (p-value = 0.0159) was higher expressed in HPV− samples. Furthermore, it is noticed that patients with lower expression levels of *MYOD1* have better overall survival prognoses than those with higher expression levels of this genes when treated with radiotherapy (Fig. S7, p-value = 0.0189).

3.7. Drug sensitivity

Based on the analysis of 14 cervical cancer cell lines, 685 genes were associated with the drug efficacy of cisplatin, among which four genes, *RP11-299 L17.3*, *SLC14A2*, *FGF18* and *OASL*, also belonged to the 166 DEGs dataset (Table 2). In other words, these four genes demonstrated different drug-response to cisplatin between HPV+ and HPV− cervical cancer samples. Patients with higher expression levels of *RP11-299 L17.3* and *OASL* had significantly stronger resistance to cisplatin than those with lower expression levels of these genes. In contrast, patients with higher expression levels of *SLC14A2* and *FGF18* were significantly less resistant to cisplatin than those with lower levels of these genes. It also noted that only *RP11-299L17.3* was resistant to cisplatin in HPV − cervical cancer samples.

4. Discussion

According to the latest research [11], HPV+ and HPV− cervical cancers are distinct clinical entities. Although early detection and targeted therapy have dramatically improved identification and survival probability of persons at risk, it remains necessary to develop novel or improved treatments for all cervical carcinomas. The aim of the present study was to discover the extent of biologic and clinical differences

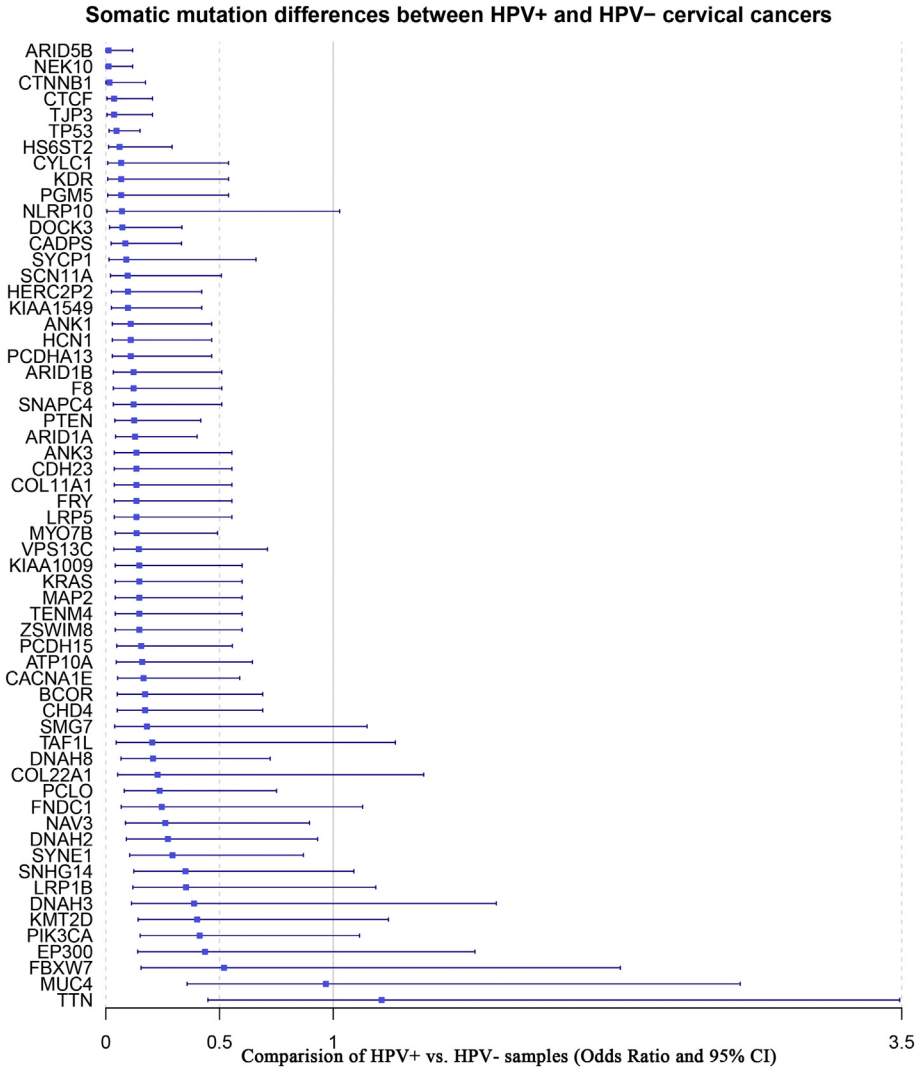

Fig. 4. Somatic mutation differences between HPV+ and HPV− cervical cancers. Among of these genes, *COL22A1*, *SNHG14*, *DNAH3*, *KMT2D*, *PIK3CA*, *EP300*, *FBXW7*, *MUC4* and *TTN* were not significantly different between two classes (a two-tailed Fisher's exact p-value >0.05).

between HPV+ and HPV− tumors and the potential to refine patient management based on this distinction. This study, for the first time, summarized and compared the clinical, biological molecular, radiotherapy and drug sensitivity differences between these two entities by using various types of data including TCGA public data and GDSC database information.

In the study, clinical analysis shows that HPV− tumors occur in older women and are often adenocarcinomas or adenosquamous carcinomas, which is reasonable and consistent with actual clinical research and previous studies [28], indicating that it is necessary to further study the histological subtypes of cervical cancer. 166 distinct genes were identified between HPV -positive and -negative tumors, and 90.96% of them were overexpressed in the HPV− tumors. Specifically, six GO terms were associated with phosphatidylinositol, including *FGFR1*, *FGF18*, *FGF9*, *FGF20* and *FGF3*. Prior studies classified them as members of the basic fibroblast growth factor (FGF) gene family which possess broad mitogenic and cell survival activities and are involved in diverse biological processes including embryonic development, cell growth, morphogenesis, tissue repair, tumor growth and invasion [29,30]. It is found that FGF family genes are usually higher expressed in HPV− tumors, and FGF and PI3K pathway aberrations may be potential therapeutic targets [31]. Moreover, *PRAME*, higher-expressed in the HPV− tumors, is also expressed in a variety of tumors, including melanoma, non-small cell lung cancer, breast cancer, head and neck cancer, and

Table 2
Correlation between gene expression and cisplatin's IC50.

Genes	Cor*	p-Value	Expression in HPV−	Results
RP11-299L17.3	0.689	0.011349	High Expression	The higher gene expression, the more drug resistance.
SLC14A2	−0.643	0.028347	High Expression	The lower gene expression, the more drug resistance.
FGF18	−0.611	0.037622	High Expression	The lower gene expression, the more drug resistance.
OASL	0.556	0.039395	Low Expression	The higher gene expression, the more drug resistance.

Cor* represents the correlation coefficient of gene and cisplatin in drug sensitivity analysis.

renal cell carcinoma [32–36], and it may be one of the biomarkers that distinguish HPV− patients from HPV+ patients. Additionally, *ANKRD7*, *SERPINB3*, *EMX2*, *MEI1*, *RNF212*, *RP11-13 K12.5*, *RP11-325F22.2* and *ZFR2* were found significantly relevant to cervical cancer prognosis. Among these genes, *EMX2*, empty spiracles homeobox 2, is expressed in epithelial tissues and is negatively regulated by *HOXA10* [37] which may function in fertility, embryo viability, and regulation of hematopoietic lineage commitment. Several studies have shown that *HOXA10* is relevant to ovarian cancer [38] and endometrial cancer, and is able to promote the proliferation and invasion of ovarian clear cell carcinoma. It can also affect the invasion and migration of breast cancer cells by regulating the expression of *P53* gene in breast cancer cells [39,40]. Similarly, *HOXA10* may be associated with cervical cancer, that is, high expression of *EMX2* in HPV− tumors may be one of the prognostic indicators of HPV− cervical cancer. However, this conclusion needs to be examined by further experimental research. Apart from this, *has-mir-182*, *has-mir-183* and *has-mir-5002* which were overexpressed in the HPV− samples were also identified, whereas the results have not been confirmed.

Furthermore, there were profound differences in global somatic mutation levels, and this study provided strong evidence that a fraction of cervical cancers can slowly evolve to HPV by accumulating somatic mutations in cancer driver genes. For instance, cancer driver genes *TP53*, *PTEN*, *CTNNB1*, *ARID5BA* and *ARID1A* were far more likely to be mutated in HPV− tumors. Finally, *CDO1*, *PCDHB2* and *MYOD1* demonstrated that the response to radiotherapy between HPV+ and HPV− samples were different, and *RP11-299 L17.3*, *SLC14A2*, *FGF18* and *OASL* showed different drug-response to cisplatin between the two entities, but these results need to be further validated.

Briefly, we sought here to provide a comprehensive understanding of molecular and clinicopathological characteristics of cervical cancers that are independent of HPV oncogene activity. The molecular differences between HPV+ and HPV− tumors may favor the opportunity to be targeted under specific therapeutic approaches separately. However, this study has a few limitations. Firstly, the small sample size of HPV− cervical cancer may make it difficult to detect the overall survival difference between the two entities. Additionally, the TCGA data enrolled for analysis was mostly collected from patients with cervical cancer in developed countries, but lacked data from developing countries. To put it in a nutshell, this study mainly reveals the distinct genomic differences between HPV+ and HPV− cervical cancers, which could encourage further studies to better understand the molecular mechanism of cervical cancer without HPV infection and to develop novel drugs or targeted therapies for clinical use.

Acknowledgements

F. Yan was supported by the National Social Science Foundation of China (No. 16BTJ021). Y. Jiang was supported by the National Natural Science Foundation of China (No. 81702566) and National Key Projects of Precision Medicine (No. 2016YFC0902901). H. Yan was supported by the Fundamental Research Funds for the Central Universities grants of China (No. 2632018FY04).

Conflicts of interest

There is no conflict of financial interest in this study.

Author contribution

L.Z. and Y.J. conceived of the presented idea. H.Z, C.C., Y.W., W.H. and Y.Z. contributed to data collection L.Z. and X.L. carried out the analysis. H.Y. assisted with all the statistical analysis. L.Z. and Y.J. wrote the manuscript in consultation with F.Y. All authors discussed the results and contributed to the final manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygyno.2018.12.017.

References

[1] J. Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D.M. Parkin, D. Forman, F. Bray, Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012, Int. J. Cancer 136 (5) (2015) 359–386.
[2] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, 2018, CA Cancer J. Clin. 68 (1) (2018) 7–30.
[3] M. Rusan, Y.Y. Li, P.S. Hammerman, Genomic landscape of human papillomavirus–associated cancers, Clin. Cancer Res. 21 (9) (2015) 2009–2019.
[4] H. Zur Hausen, Papillomaviruses and cancer: from basic studies to clinical application, Nat. Rev. Cancer 2 (5) (2002) 342–350.
[5] M. Schiffman, N. Wentzensen, S. Wacholder, W. Kinney, J.C. Gage, P.E. Castle, Human papillomavirus testing in the prevention of cervical cancer, J. Natl. Cancer Inst. 103 (5) (2011) 368–383.
[6] E.J. Crosbie, M.H. Einstein, S. Franceschi, H.C. Kitchener, Human papillomavirus and cervical cancer, Lancet 382 (9895) (2013) 889–899.
[7] K.W. Tang, B. Alaei-Mahabadi, T. Samuelsson, M. Lindh, E. Larsson, The landscape of viral expression and host gene fusion and adaptation in human cancer, Nat. Commun. 4 (2013) 2513.
[8] K. Akagi, J. Li, T.R. Broutian, H. Padillanash, W. Xiao, B. Jiang, J.W. Rocco, T.N. Teknos, B. Kumar, D. Wangsa, D. He, T. Ried, D.E. Symer, M.L. Gillison, Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability, Genome Res. 24 (2) (2014) 185–199.
[9] A.I. Ojesina, L. Lichtenstein, S.S. Freeman, et al., Landscape of genomic alterations in cervical carcinomas, Nature 506 (7488) (2014) 371–375.
[10] Z. Hu, D. Zhu, W. Wang, et al., Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism, Nat. Genet. 47 (2) (2015) 158–163.
[11] C.G.A. Network, Integrated genomic and molecular characterization of cervical cancer, Nature 543 (7645) (2017) 378–384.
[12] K.U. Petry, C. Liebrich, A. Luyten, M. Zander, T. Iftner, Surgical staging identified false HPV-negative cases in a large series of invasive cervical cancers, Papillomavirus Res. 4 (2017) 85–89.
[13] S. Tomar, C.A. Graves, D. Altomare, S. Kowli, S. Kassler, N. Sutkowski, M.B. Gillespie, K.E. Creek, L. Pirisi, Human papillomavirus status and gene expression profiles of oropharyngeal and oral cancers from European American and African American patients, Head Neck 38 (S1) (2016) 694–704.
[14] T.Y. Seiwert, Z. Zuo, M.K. Keck, A. Khattri, C.S. Pedamallu, T. Stricker, C. Brown, T.J. Pugh, P. Stojanov, J. Cho, M.S. Lawrence, G. Getz, J. Brägelmann, R. DeBoer, R.R. Weichselbaum, A. Langerman, L. Portugal, E. Blair, K. Stenson, M.W. Lingen, E.E. Cohen, E.E. Vokes, K.P. White, P.S. Hammerman, Integrative and comparative genomic analysis of HPV-positve and HPV-negative head and neck squamous cell carcinomas, Clin. Cancer Res. 21 (3) (2015) 632–641.
[15] K. Tomczak, P. Czerwińska, M. Wiznerowicz, The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge, Contemp. Oncol. 19 (1A) (2015) 68–77.
[16] M.J. Garnett, E.J. Edelman, S.J. Heidorn, et al., Systematic identification of genomic markers of drug sensitivity in cancer cells, Nature 483 (7391) (2012) 570–575.
[17] W. Yang, J. Soares, P. Greninger, E.J. Edelman, H. Lightfoot, S. Forbes, N. Bindal, D. Beare, J.A. Smith, I.R. Thompson, S. Ramaswamy, P.A. Futreal, D.A. Haber, M.R. Stratton, C. Benes, U. McDermott, M.J. Garnett, Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells, Nucleic Acids Res. 41 (D1) (2013) 955–961.
[18] C.W. Law, Y. Chen, W. Shi, G.K. Smyth, voom: precision weights unlock linear model analysis tools for RNA-seq read counts, Genome Biol. 15 (2) (2014) R29.
[19] S. Anders, W. Huber, Differential expression analysis for sequence count data, Genome Biol. 11 (10) (2010) R106.
[20] M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, Bioinformatics 26 (1) (2010) 139–140.
[21] G. Yu, L.G. Wang, Y. Han, Q.Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters, OMICS 16 (5) (2012) 284–287.
[22] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome Res. 13 (11) (2003) 2498–2504.
[23] Z.L. Skidmore, A.H. Wagner, R. Lesurf, K.M. Campbell, J. Kunisaki, O.L. Griffith, M. Griffith, GenVisR: genomic visualizations in R, Bioinformatics 32 (19) (2016) 3012–3014.
[24] L.C. Cantley, The phosphoinositide 3-kinase pathway, Science 31 (296) (2002) 1655–1657.
[25] K.H. Khoo, C.S. Verma, D.P. Lane, Drugging the p53 pathway: understanding the route to clinical efficacy, Nat. Rev. Drug Discov. 13 (3) (2014) 217–236.
[26] A. Rufini, P. Tucci, I. Celardo, G. Melino, Senescence and aging: the critical roles of p53, Oncogene 32 (43) (2013) 5129–5143.
[27] J. Chen, Signaling pathways in HPV-associated cancers and therapeutic implications, Rev. Med. Virol. 25 (S1) (2015) 24–53.
[28] C.E. Banister, C. Liu, L. Pirisi, K.E. Creek, P.J. Buckhaults, Identification and characterization of HPV-independent cervical cancers, Oncotarget 8 (8) (2017) 13375–13386.

[29] M. Mohammadi, S.K. Olsen, O.A. Ibrahimi, Structural basis for fibroblast growth factor receptor activation, Cytokine Growth F. R. 16 (2) (2005) 107–137.
[30] N. Turner, R. Grose, Fibroblast growth factor signalling: from development to cancer, Nat. Rev. Cancer 10 (2) (2010) 116–129.
[31] A. Akinleye, P. Avvaru, M. Furqan, Y. Song, D. Liu, Phosphatidylinositol 3-kinase (PI3K) inhibitors as cancer therapeutics, J. Hematol. Oncol. 6 (1) (2013) 88.
[32] L.J. Van't Veer, H. Dai, M.J. Van De Vijver, Y.D. He, A.A. Hart, M. Mao, H.L. Peterse, K. van der Kooy, M.J. Marton, A.T. Witteveen, G.J. Schreiber, R.M. Kerkhoven, C. Roberts, P.S. Linsley, R. Bernards, S.H. Friend, Gene expression profiling predicts clinical outcome of breast cancer, Nature 415 (6871) (2002) 530–536.
[33] C.M. Li, M. Guo, A. Borczuk, C.A. Powell, M. Wei, H.M. Thaker, R. Friedman, U. Klein, B. Tycko, Gene expression in Wilms' tumor mimics the earliest committed stage in the metanephric mesenchymal-epithelial transition, Am. J. Pathol. 160 (6) (2002) 2181–2190.
[34] M.J. Szczepanski, T.L. Whiteside, Elevated PRAME expression: what does this mean for treatment of head and neck squamous cell carcinoma? Biomark. Med 7 (4) (2013) 575–578.
[35] Z. Sun, Z. Wu, F. Zhang, Q. Guo, L. Li, K. Li, H. Chen, J. Zhao, D. Song, Q. Huang, L. Li, J. Xiao, PRAME is critical for breast cancer growth and metastasis, Gene 594 (1) (2016) 160–164.
[36] J. Roszik, W.L. Wang, J.A. Livingston, C.L. Roland, V. Ravi, C. Yee, P. Hwu, A. Futreal, A.J. Lazar, S.R. Patel, A.P. Conley, Overexpressed PRAME is a potential immunotherapy target in sarcoma subtypes, Clin. Sarcoma Res. 7 (1) (2017) 11.
[37] P.J. Troy, G.S. Daftary, C.N. Bagot, H.S. Taylor, Transcriptional repression of peri-implantation EMX2 expression in mammalian reproduction by HOXA10, Mol. Cell. Biol. 23 (1) (2003) 1–13.
[38] W. Cheng, J. Liu, H. Yoshida, D. Rosen, H. Naora, Lineage infidelity of epithelial ovarian cancers is controlled by HOX genes that specify regional identity in the reproductive tract, Nat. Med. 11 (5) (2005) 531–537.
[39] H. Yoshida, R. Broaddus, W. Cheng, S. Xie, H. Naora, Deregulation of the HOXA10 homeobox gene in endometrial carcinoma: role in epithelial-mesenchymal transition, Cancer Res. 66 (2) (2006) 889–897.
[40] M.C. Chu, F.B. Selam, H.S. Taylor, HOXA10 regulates p53 expression and matrigel invasion in human breast cancer cells, Cancer Biol. Ther. 3 (6) (2004) 568–572.