Bioinformatics Master's Degree

1

Identification of Hub Genes and Key Pathways Associated with Human Papillomavirus Status in Cervical Squamous Cell Carcinoma Based on Gene Expression Profiling via Integrated Bioinformatics

Graduation project

Student name: Marah Massoud

Student ID: marah_154259

Supervised by:

Dr.Lama Yousef

Dr.Majd Al-Jamali

Contents

Summary

Using integrated bioinformatics to screen differentially expressed genes (DEGs) associated with two HPV status (HPV positive and HPV negative) in Cervical Squamous Cell Carcinoma could reveal valuable information about the pathogenic mechanism underlying the tumor progression. Moreover, the identification of significant differentially expressed genes, enrichment of their biological functions and key pathways, and visualization of the network of DEGs and hub genes will provide more accurate and reliable biomarkers and therapeutic targets for early diagnosis, individualized prevention measures, and improvement of therapeutic efficacy.

In this study, a series of analyses was conducted using R2 software of HPV status in squamous cervical carcinoma-related data in TCGA database to screen and identify prognostic biomarkers related to differentially expressed genes. Then, the up- and downregulated DEGs were classified into three groups (biological processes, molecular functions, and cellular components) according to Gene Ontology(GO) terms, and KEGG pathway enrichment analysis was conducted using DAVID website.

To further study the potential relationships between the genes, the Protein-Protein Interaction(PPI) network was created using STRING database and reconstructed via Cytoscape software, with the degree of connectivity of each node in the network calculated. The Cytoscape plug-in CytoHubba was used to identify the hub genes by finding the intersections of the first 30 genes from 12 topological analysis methods. To evaluate the biomarkers, Kaplan-Meier Plotter could use gene expression data to assess the survival rate of cancer patients and the prognostic value of the extracted key genes. Afterwards, the expression of these key genes in squamous cell cervical cancer was verified using UALCAN online tool. Subsequently, eleven hub genes were defined (STAT1, CTNNB1 IRF9, EGFR, RSAD2, IRF7, MX1, IFIH1, IRF5, IRF1, DDX58) and the expression level of each protein in different tissues was obtained using the Cancer Atlas part from The Human Protein Atlas portal tool. The results showed that four proteins were differently stained in normal and tumor tissues (RSAD2, MX1, STAT1, DDX58). Next, functional enrichment analysis was conducted using WebGestalt server through a suggested functional database (DrugBank) to predict gene-drug interactions. Finally, Pathway-based Rational Drug Repositioning analysis was done by Gene2drug online tool for four genes that were obtained using immunohistochemistry analysis, and the results demonstrated a number of drugs that may be involved in the future investigation of this disease.

Keywords: bioinformatics analysis, biomarkers, Cervical Squamous Cell Carcinoma, prognosis, differentially expressed genes

Background:

Cervical cancer remains the second-most common cause of cancer-related deaths in women worldwide, with about 450,000 new cases diagnosed each year (1). Defining the type of cervical cancer helps determine the prognosis and treatment. However, most cervical cancers are squamous cell carcinomas(2). Squamous Cell Carcinoma is a type of cervical cancer that begins in the thin, flat cells (squamous cells) lining the outer part of the cervix, which projects into the vagina. In developing countries, cervical cancer is the most common cancer in women and may constitute up to 25% of all female cancers. Cervical cancer is preceded only by breast cancer as the most common cause of death from cancer in women worldwide and fourth most common type of cancer (15.1 per 100,000) and cause of cancer mortality (8.2 per 100,000) among women worldwide in 2018. In 1996, the World Health Association, along with the European Research Organization on Genital Infection and Neoplasia and the National Institutes of Health Consensus Conference on Cervical Cancer, recognized HPV as an important cause of cervical cancer. Additionally, HPV has been implicated in 99.7% of cervical squamous cell cancer cases worldwide (3) (4).

Morphologic variants:

- Keratinizing
    - Keratin pearls, abundant keratohyaline granules and intercellular bridges
    - Large, hyperchromatic nuclei with coarse chromatin and inconspicuous nucleoli
- Non-keratinizing
    - Polygonal cells forming sheets or nests
    - Intercellular bridges but not keratin pearls
    - Large nuclei with unevenly distributed, coarsely granular chromatin and one or multiple nucleoli
    - Numerous mitoses. (7)

Various strains of Human papillomavirus (HPV), a sexually transmitted infection, represents the most important risk factor for the development of cervical cancer (1), the magnitude of the association between HPV and cervical squamous cell carcinoma is higher than that for the association between smoking and lung cancer. When exposed to HPV, the body's immune system typically prevents the virus from doing harm. In a small percentage of people, the virus survives for years, contributing to the process that causes some cervical cells to become cancer cells. Prevalence surveys, large case-control studies, and case series have unequivocally shown that HPV DNA can be detected in cervical cancer specimens in 90–100% of cases, compared with a prevalence of 5–20% in cervical specimens from women identified as suitable epidemiological controls (5). Although more than 100 distinct HPV genotypes have been described, and at least 20 are associated with cervical cancer, HPV types 16 and 18 are the most frequently detected in cervical cancer regardless of the geographical origin of the patients (5). Although early stage cervical cancer can be cured by radical surgery or radiotherapy with equal effectiveness (6), pelvic radiation represents the standard therapy for the treatment of locally advanced disease. Despite technological advances, however, up to 35% of patients overall will develop advanced, metastatic disease, for which treatment results are poor. A deeper understanding of the molecular basis of cervical cancer has the potential to refine significantly the diagnosis and management of these tumors and may eventually lead to the development of novel, more specific and more effective treatments for prevention of disease progression following first-line therapy.

Human papillomavirus (HPV) infection of the genital tract is common in young sexually active individuals, the majority of whom clear the infection without overt clinical disease. Most of those who do develop benign lesions eventually mount an effective cell-mediated immune (CMI) response, and the lesions regress. Regression of anogenital warts is accompanied histologically by a $CD4^+$ T cell-dominated Th1 response; animal models support this and provide evidence that the response is modulated by antigen-specific $CD4^+$ T cell-dependent mechanisms. Failure to develop an effective CMI response to clear or control infection results in persistent infection and, in the case of the oncogenic HPVs, an increased probability of progression to high-grade intraepithelial neoplasia and invasive carcinoma. Effective evasion of innate immune recognition seems to be the hallmark of HPV infections. HPV globally downregulates the innate immune signaling pathways in the infected keratinocyte. Proinflammatory cytokines, particularly the type I interferons, are not released, and the signals for Langerhans cell (LC) activation and migration, together with recruitment of stromal dendritic cells and macrophages, are either not present or inadequate. This immune ignorance results in chronic infections that persist over weeks and months. Progression to high-grade intraepithelial neoplasia with concomitant

upregulation of the E6 and E7 oncoproteins is associated with further deregulation of immunologically relevant molecules, particularly chemotactic chemokines and their receptors, on keratinocytes and endothelial cells of the underlying microvasculature, limiting or preventing the ingress of cytotoxic effectors into the lesions (8).

In the largest study reporting on the prognostic importance of keratinization in Squamous Cell Carcinoma(SCC). keratinization in Squamous Cell Carcinoma (KSCC) may be less radiosensitive and associated with shorter overall survival. Furthermore, in the natural history of the SCC, keratinization signifies striking reduction in survival (9). Non-keratinizing Squamous Cell Carcinoma (NKSCC) was strongly associated with high-risk HPV genotypes and $p16^{INK4A}$ positivity (69% and 100%, respectively) compared to KSCC (8% and 36%, respectively)(10). HPV-related tumors have alternately been described as 'basaloid,' 'basal-like,' 'poorly-differentiated,' as well as 'non-keratinizing' (11).

Demographic and exposure differences between HPV-positive (HPV^+) and negative (HPV^-) Cervical Squamous Cell Carcinoma suggest that HPV^+ tumors may constitute a subclass with different biology, whereas clinical differences have also been observed. Further investigation of differentially expressed genes may reveal the unique pathways in HPV+ tumors that may explain the different natural history and biological properties of these tumors. These properties may be exploited as a target of novel therapeutic agents in Cervical Squamous Cell Carcinoma treatment. The obtained hub genes and key pathways could be the therapeutic targets for the precise treatment of these two HPV status with different prognoses. using integrated bioinformatics to screen differentially expressed genes (DEGs) in cervical squamous cell cancer could benefit us for understanding the pathogenic mechanism underlying the tumor progression. Also to identify significantly upregulated genes in differentially expressed genes (DEGs) from both HPV positive and HPV negative status, which might be used as the biomarkers for early diagnosis and prevention of the disease.

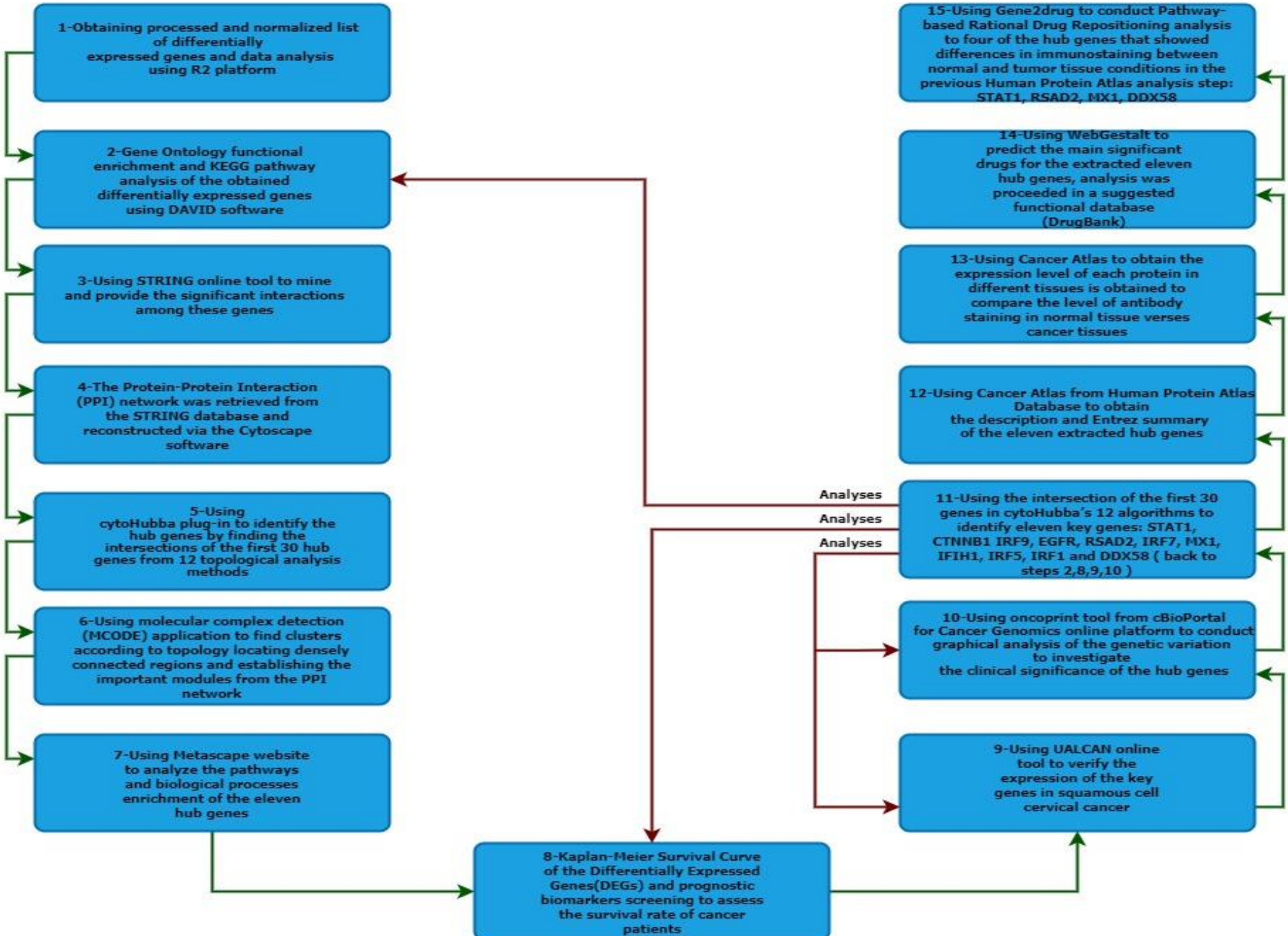
The aim of the study

Gene regulatory networks reveal how genes work together to carry out their biological functions. Consequently, reconstructions of gene networks from gene expression data greatly facilitate the understanding of underlying biological mechanisms and provide new opportunities for biomarker and drug discoveries. Eventually, identifying genes that have many interactions with other genes (hub gene), which usually plays an essential role in gene regulation and biological processes.
The obtained hub genes and key pathways could be the therapeutic targets of the precise treatment of HPV positive and HPV negative status in Cervical Squamous Cell Carcinoma with different prognoses, and could improve our understanding of the process of tumorigenesis and the underlying molecular events.

Materials and Methods

Flow chart clarifies the integrated bioinformatics tools that were applied to screen hub genes and key pathways associated with HPV status (HPV-positive and HPV-negative) in cervical squamous cell carcinoma(CSCC) samples derived from TCGA database:

Flow chart of the integrated bioinformatics tools that were applied to screen hub genes and key pathways associated with HPV status (HPV-positive and HPV-negative) in cervical squamous cell carcinoma(CSCC) samples derived from TCGA database



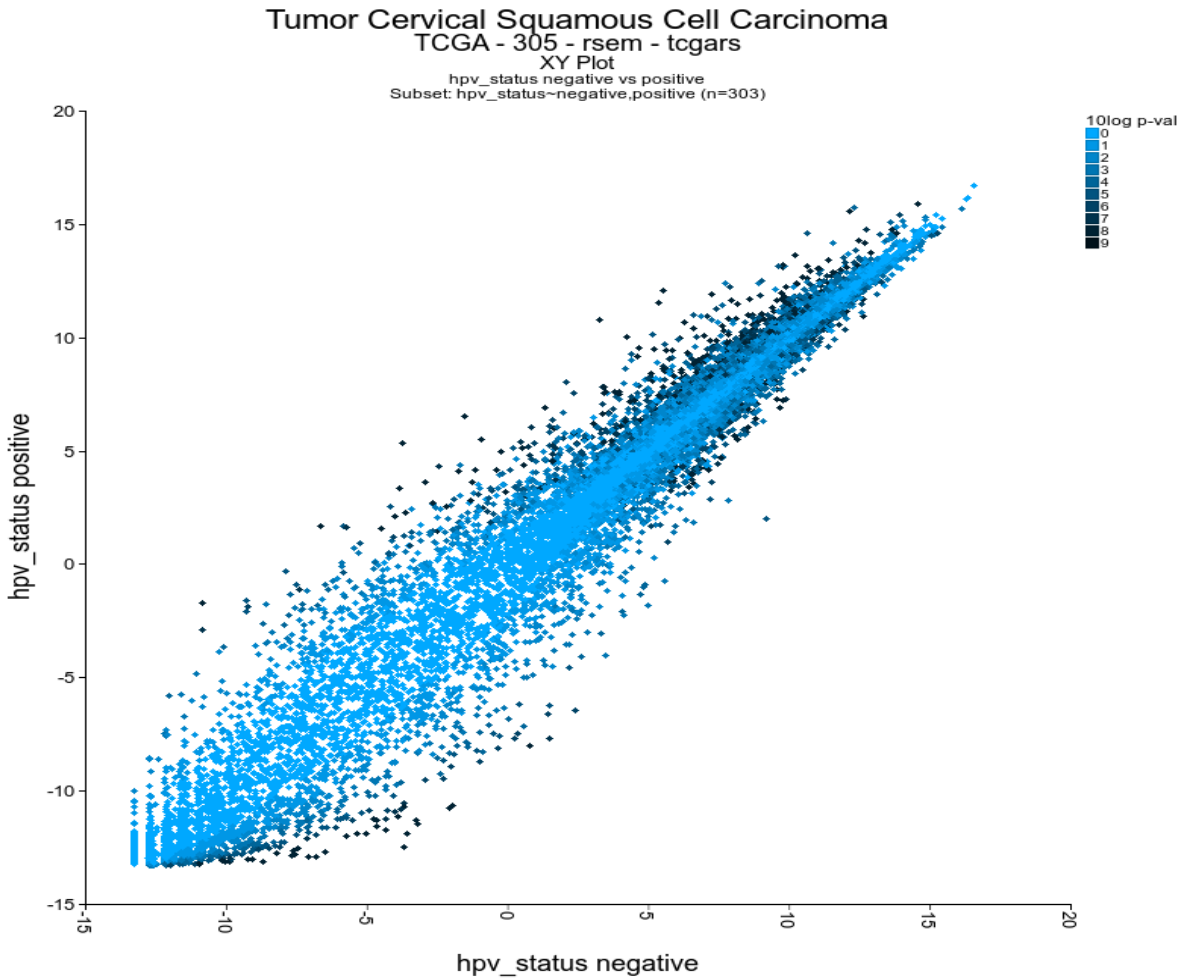
1. Obtaining processed and normalized list of differentially expressed genes and data analysis using R2 platform
2. Gene Ontology functional enrichment and KEGG pathway analysis of the obtained differentially expressed genes using DAVID software
3. Using STRING online tool to mine and provide the significant interactions among these genes
4. The Protein-Protein Interaction (PPI) network was retrieved from the STRING database and reconstructed via the Cytoscape software
5. Using cytoHubba plug-in to identify the hub genes by finding the intersections of the first 30 hub genes from 12 topological analysis methods
6. Using molecular complex detection (MCODE) application to find clusters according to topology locating densely connected regions and establishing the important modules from the PPI network
7. Using Metascape website to analyze the pathways and biological processes enrichment of the eleven hub genes
8. Kaplan-Meier Survival Curve of the Differentially Expressed Genes(DEGs) and prognostic biomarkers screening to assess the survival rate of cancer patients
9. Using UALCAN online tool to verify the expression of the key genes in squamous cell cervical cancer
10. Using oncoprint tool from cBioPortal for Cancer Genomics online platform to conduct graphical analysis of the genetic variation to investigate the clinical significance of the hub genes
11. Using the intersection of the first 30 genes in cytoHubba's 12 algorithms to identify eleven key genes: STAT1, CTNNB1 IRF9, EGFR, RSAD2, IRF7, MX1, IFIH1, IRF5, IRF1 and DDX58.
12. Using Cancer Atlas from Human Protein Atlas Database to obtain the description and Entrez summary of the eleven extracted hub genes
13. Using Cancer Atlas to obtain the expression level of each protein in different tissues is obtained to compare the level of antibody staining in normal tissue verses cancer tissues
14. Using WebGestalt to predict the main significant drugs for ten hub genes that were generated by CytoHubba with their scores, analysis was proceeded in a suggested functional database (DrugBank)
15. Using Gene2drug to conduct Pathway-based Rational Drug Repositioning analysis to four of the hub genes that showed differences in immunostaining between normal and tumor tissue conditions in the previous Human Protein Atlas analysis step: STAT1, RSAD2, MX1, DDX58

Obtaining processed differentially expressed genes and data analysis

R2 platform (https://hgserver1.amc.nl/cgi-bin/r2) was used to obtain the normalized RNA-seq data samples of patients with Cervical Squamous Cell Carcinoma derived from TCGA database (https://gdc.cancer.gov/), data were collected from 305 samples including both HPV status (including HPV negative: 22 patient, and HPV positive: 281 patient). Analysis was performed by R2 to obtain differentially expressed genes (DEGs) and their expression levels in HPV positive and HPV negative samples (Figure 1), multiple testing correction applied is False Discovery Rate(FDR), p-value cutoff =< 0.01 and |log2 fold change (FC)| >=1. The results were considered statistically significant.

- HPV status data stands for:

1. HPV positive for high risk HPV subtypes as they are associated with cervical cancer development: 16, 18, 35, 33, 31, 45, 52, and 58
2. HPV negative: not associated with any HPV subtype infection.



(Figure 1) scatter plot of differentially expressed genes (DEGs) and their expression levels associated with HPV positive and HPV negative in Cervical Squamous Cell Carcinoma samples colored by p-value via R2 software. Each dot represents a different gene, the darkness of the color is proportional to the level of expression

Gene set analysis was applied via R2 software to demonstrate the upregulated genes associated with each HPV status separately, using KEGG as gene set collection database and p value=0.05 (Figure 2 A-B). This step may give an important indication in advanced stages of analysis.

From your input (n=**1333**), **424** genes were also present in the current geneset collection (kegg).
Within the dataset ps_avgpres_tcgacesc305_tcgars, 6650 genes were detected in the current geneset selection (hugoonce=yes and minimal present call=1)
The table below lists genesets where the number of genes from your list are present more than expected (**p<1** from 2X2 contingency table analysis with continuity correction)
hpv_status: negative >= positive,

set	R#	#	p_value_	Genelist
over-representation KAH:Basal_cell_carcinoma	55	14	3.1e-08	, AXIN2, BMP4, CTNNB1, FZD3, FZD5, FZD7, HHIP, LEF1, PTCH1, PTCH2, SHH, SMO, STK36, WNT11
over-representation KAH:Wnt_signaling_pathway	140	22	1.1e-05	, AXIN2, BAMBI, CCND1, CTNNB1, CUL1, DAAM2, FZD3, FZD5, FZD7, GPC4, LEF1, LRP6, NFATC4, NKD1, NKD2, NLK, PLCB4, SFRP5, SKP1, SOX17, WIF1, WNT11
over-representation KAH:Synaptic_vesicle_cycle	63	11	7.8e-04	, AP2A2, ATP6V0A2, ATP6V0E2, ATP6V1C1, ATP6V1H, CPLX2, SLC17A6, SLC18A3, STX3, STXBP1, SYT1
over-representation KAH:Melanoma	71	11	3.5e-03	, BRAF, CCND1, E2F3, FGF13, FGF18, FGF19, FGF20, FGF3, FGF4, FGF9, FGFR1
over-representation KAH:Hedgehog_signaling_pathway	50	8	0.01	, BMP4, HHIP, PTCH1, PTCH2, SHH, SMO, STK36, WNT11
over-representation KAH:Vibrio_cholerae_infection	51	8	0.01	, ATP6V0A2, ATP6V0E2, ATP6V1C1, ATP6V1H, GNAS, PDIA4, PLCG1, SEC61A2
over-representation KAH:Vitamin_digestion_and_absorption	24	5	0.01	, CUBN, FOLH1, RBP2, SLC19A1, SLC5A6
over-representation KAH:Glycerophospholipid_metabolism	94	12	0.02	, AGPAT1, AGPAT6, CHKA, CRLS1, ETNK1, GPAM, LPIN2, PCYT1B, PEMT, PLA2G12A, PTDSS1, PTDSS2
over-representation KAH:Pathways_in_cancer	397	36	0.03	, AXIN2, BMP4, BRAF, CCND1, CTNNA2, CTNNB1, E2F3, FGF13, FGF18, FGF19, FGF20, FGF3, FGF4, FGF9, FGFR1, FZD3, FZD5, FZD7, GNAS, HDAC2, HHIP, HSP90AB1, LAMA1, LAMB2, LAMC1, LEF1, PLCB4, PLCG1, PTCH1, PTCH2, RXRG, SHH, SMO, STK36, VHL, WNT11
over-representation KAH:Thyroid_cancer	29	5	0.04	, BRAF, CCND1, CTNNB1, LEF1, RXRG
over-representation KAH:Citrate_cycle__TCA_cycle_	30	5	0.05	, ACLY, DLD, IDH1, MDH2, OGDHL
over-representation KAH:Vitamin_B6_metabolism	6	2	0.06	, PDXP, PHOSPHO2
over-representation KAH:Cysteine_and_methionine_metabolism	43	6	0.08	, AHCY, BCAT1, CTH, DNMT3A, LDHB, MDH2
over-representation KAH:Aminoacyl_tRNA_biosynthesis	44	6	0.10	, EARS2, GARS, LARS2, NARS2, VARS, YARS2
over-representation KAH:Pathogenic_Escherichia_coli_infection	55	7	0.10	, ARHGEF2, CTNNB1, FYN, NCL, TUBA1A, TUBB2B, TUBB3

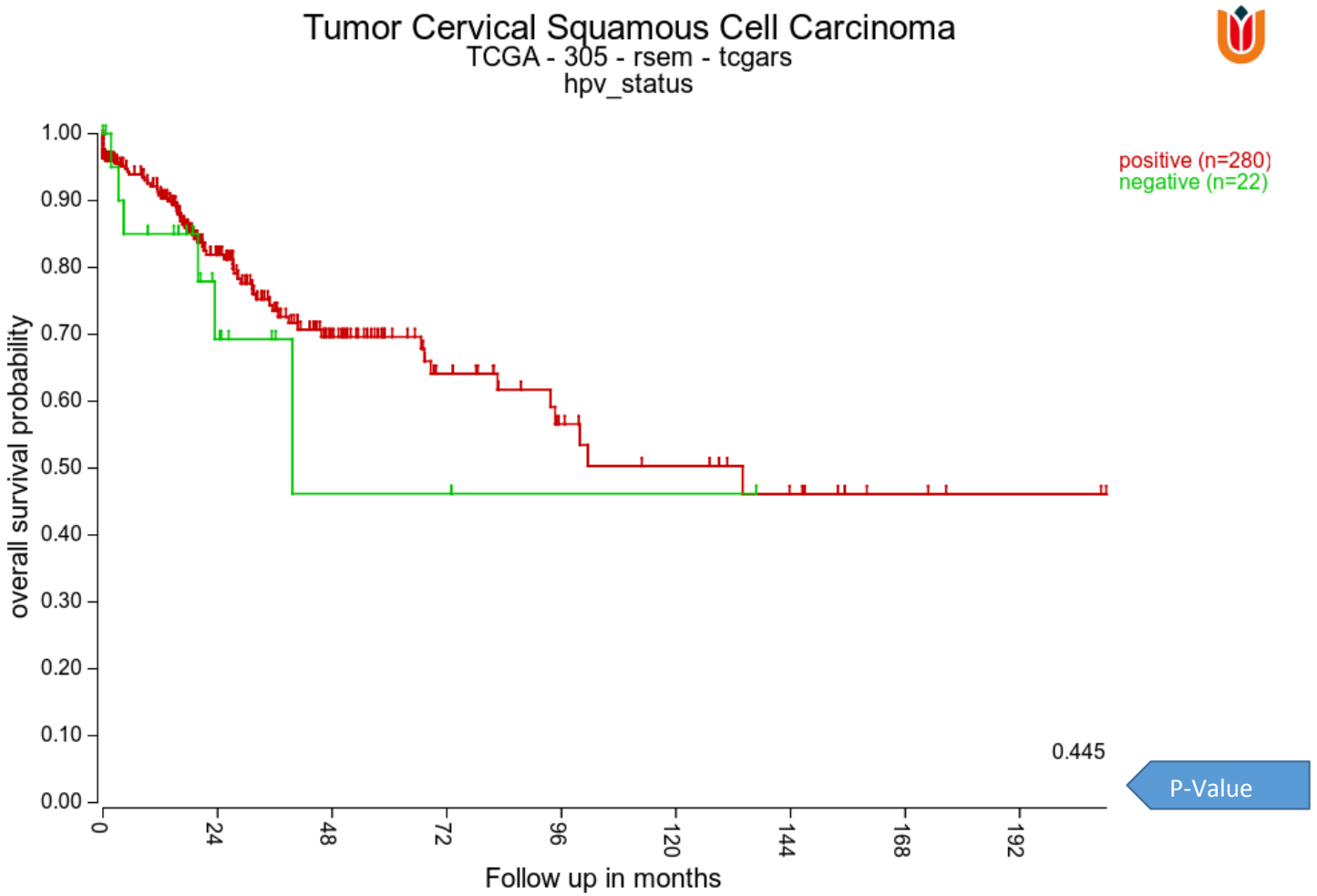
(Figure 2-A)KEGG pathways of upregulated genes associated with HPV negative status in Cervical Squamous Cell Carcinoma via R2 software

From your input (n=**1521**), **558** genes were also present in the current geneset collection (kegg).
Within the dataset ps_avgpres_tcgacesc305_tcgars, 6650 genes were detected in the current geneset selection (hugoonce=yes and minimal present call=1)
The table below lists genesets where the number of genes from your list are present more than expected (**p<1** from 2X2 contingency table analysis with continuity correction)
hpv_status: negative < positive,

set	R#	#	p_value	Genelist
over-representation KEGG:Herpes_simplex_infection	181	52	5.6e-23	, ARNTL, C3, C5, CCL5, CD74, CDK1, CDK2, DDX58, FAS, FASLG, FOS, HCFC2, HLA-B, HLA-C, HLA-DMA, HLA-DOB, HLA-DPA1, HLA-DPB1, HLA-DQA2, HLA-DRA, HLA-DRB1, HLA-DRB5, HLA-E, HLA-F, IFIH1, IFIT1, IFNA13, IFNGR1, IL12B, IL15, IRF7, IRF9, JAK2, NFKB1, NFKBIA, OAS1, OAS2, OAS3, PER1, PER2, PER3, PML, RNASEL, SP100, STAT1, STAT2, TAP1, TAP2, TBK1, TLR2, TLR3, TNFSF14
over-representation KEGG:Graft_versus_host_disease	37	18	1.1e-17	, FAS, FASLG, HLA-B, HLA-C, HLA-DMA, HLA-DOB, HLA-DPA1, HLA-DPB1, HLA-DQA2, HLA-DRA, HLA-DRB1, HLA-DRB5, HLA-E, HLA-F, KIR2DL1, KIR3DL2, KLRC1, KLRD1
over-representation KEGG:Inflammatory_bowel_disease__IBD_	63	24	9.1e-17	, FOXP3, GATA3, HLA-DMA, HLA-DOB, HLA-DPA1, HLA-DPB1, HLA-DQA2, HLA-DRA, HLA-DRB1, HLA-DRB5, IFNGR1, IL12B, IL12RB2, IL18RAP, IL4R, NFATC1, NFKB1, RORC, SMAD3, STAT1, STAT3, STAT4, TLR2, TLR5
over-representation KEGG:Antigen_processing_and_presentation	69	25	3.2e-16	, B2M, CD74, CIITA, HLA-B, HLA-C, HLA-DMA, HLA-DOB, HLA-DPA1, HLA-DPB1, HLA-DQA2, HLA-DRA, HLA-DRB1, HLA-DRB5, HLA-E, HLA-F, IFI30, KIR2DL1, KIR3DL2, KLRC1, KLRC2, KLRC3, KLRD1, NFYC, TAP1, TAP2
over-representation KEGG:Allograft_rejection	35	16	1.6e-14	, CD40LG, FAS, FASLG, HLA-B, HLA-C, HLA-DMA, HLA-DOB, HLA-DPA1, HLA-DPB1, HLA-DQA2, HLA-DRA, HLA-DRB1, HLA-DRB5, HLA-E, HLA-F, IL12B
over-representation KEGG:Influenza_A	170	40	1.5e-12	, CASP1, CCL5, CIITA, CXCL10, DDX58, FAS, FASLG, HLA-DMA, HLA-DOB, HLA-DPA1, HLA-DPB1, HLA-DQA2, HLA-DRA, HLA-DRB1, HLA-DRB5, ICAM1, IFIH1, IFNA13, IFNGR1, IL12B, IRF7, IRF9, JAK2, MX1, NFKB1, NFKBIA, OAS1, OAS2, OAS3, PIK3CD, PML, PRKCB, PYCARD, RNASEL, RSAD2, STAT1, STAT2, TBK1, TLR3, TNFSF10
over-representation KEGG:Type_I_diabetes_mellitus	41	16	9.5e-12	, FAS, FASLG, GAD2, HLA-B, HLA-C, HLA-DMA, HLA-DOB, HLA-DPA1, HLA-DPB1, HLA-DQA2, HLA-DRA, HLA-DRB1, HLA-DRB5, HLA-E, HLA-F, IL12B
over-representation KEGG:Natural_killer_cell_mediated_cytotoxicity	126	32	1.1e-11	, CD244, CD247, CD48, FAS, FASLG, HCST, HLA-B, HLA-C, HLA-E, ICAM1, IFNA13, IFNGR1, KIR2DL1, KIR3DL2, KLRC1, KLRC2, KLRC3, KLRD1, KLRK1, NCR1, NCR3, NFATC1, PIK3CD, PRKCB, PTK2B, RAC2, RAET1E, RAET1G, RAET1L, SH2D1B, TNFSF10, ULBP2
over-representation KEGG:DNA_replication	36	14	2.7e-10	, LIG1, MCM2, MCM5, MCM6, PCNA, POLA1, POLD1, POLD3, POLE2, PRIM1, RFC4, RFC5, RPA2, RPA3
over-representation KEGG:Fanconi_anemia_pathway	52	17	1.1e-09	, BRCA1, BRIP1, FANCA, FANCB, FANCE, FANCG, FANCI, FANCL, FANCM, HES1, MLH1, PALB2, RAD51C, REV1, RMI2, RPA2, RPA3
over-representation KEGG:HTLV_I_infection	256	48	2.2e-09	, ADCY3, ADCY9, ATF3, CCNB2, CD3D, CD3E, CDKN1A, CDKN2A, CDKN2B, CDKN2C, E2F2, ETS2, FOS, FZD6, HLA-B, HLA-C, HLA-DMA, HLA-DOB, HLA-DPA1, HLA-DPB1, HLA-DQA2, HLA-DRA, HLA-DRB1, HLA-DRB5, HLA-E, HLA-F, ICAM1, IL15, IL15RA, IL2RB, MAP3K14, NFATC1, NFKB1, NFKB2, NFKBIA, PCNA, PIK3CD, POLD1, POLD3, POLE2, RRAS, SMAD3, STAT5A, TSPO, WNT3A, WNT4, WNT7B, ZFP36
over-representation KEGG:Leishmaniasis	70	20	3.5e-09	, C3, FOS, HLA-DMA, HLA-DOB, HLA-DPA1, HLA-DPB1, HLA-DQA2, HLA-DRA, HLA-DRB1, HLA-DRB5, IFNGR1, IL12B, JAK2, NCF2, NCF4, NFKB1, NFKBIA, PRKCB, STAT1, TLR2
over-representation KEGG:Autoimmune_thyroid_disease	49	16	3.9e-09	, CD40LG, FAS, FASLG, HLA-B, HLA-C, HLA-DMA, HLA-DOB, HLA-DPA1, HLA-DPB1, HLA-DQA2, HLA-DRA, HLA-DRB1, HLA-DRB5, HLA-E, HLA-F, IFNA13
over-representation KEGG:Measles	133	30	6.9e-09	, CCNE2, CD3D, CD3E, CDK2, DDX58, FAS, FASLG, IFIH1, IFNA13, IFNGR1, IL12B, IL2RB, IRF7, IRF9, JAK2, MX1, NFKB1, NFKBIA, OAS1, OAS2, OAS3, PIK3CD, STAT1, STAT2, STAT3, STAT5A, TBK1, TLR2, TNFSF10, TP73

(Figure 2-B) KEGG pathways of upregulated genes associated with HPV positive status in Cervical Squamous Cell Carcinoma via R2 software

Kaplan-Meier Survival Curve of the DEGs and Screening of Prognostic Biomarkers Kaplan-Meier Plotter can use gene expression data to assess the survival rate of different types of cancer patients. Overall survival plot (OS) for the DEGs in HPV status was drawn (Figure 3). Using the Kaplan-Meier chart, I could evaluate to some extent the effects of the genes' expression for each HPV status on the prognosis of cervical cancer patients. As it's shown in the plot, the overall survival of HPV positive cervical cancer patients is not significantly better than the overall survival of HPV negative patients (p value is 0.445 calculated from chi-square test, chi=0.58 at freedom degree df=1). However, when hub genes are identified, I will be able to figure out those genes that can be used as prognostic biomarkers for this disease.



(Figure3) Kaplan-Meier overall survival plot of the DEGs associated with HPV positive and HPV negative of cervical cancer patients via R2 software, the overall survival of HPV positive cervical cancer patients is not significantly better than the overall survival of HPV negative patients (p value is 0.445 calculated from chi-square test, chi=0.58 at freedom degree df=1

Functional Enrichment and Pathway Analysis of Differentially Expressed Genes (DEGs)

DAVID (20) (21) was utilized to conduct Gene ontology (GO) and Kyoto Gene and Genome Encyclopedia (KEGG) enrichment analysis (http://www.DAVID.org).

The DAVID online analysis was used to conducted biological annotation of the identified common differentially expressed genes from integrated analysis of RNA-seq data in squamous cervical cancer. I obtained Gene ontology (GO) functional enrichments of up- and downregulated genes with a P-value<0.05 as a cutoff value. Three functional groups, including molecular function (MFs), biological processes (BPs), and cell composition (CCs) and KEGG pathways of the enriched genes, were divided in GO analysis of the common DEGs (Table1). In the biological processes group, the identified DEGs were mainly enriched in: defense response to virus, regulation of transcription from RNA polymerase II promoter, negative regulation of transcription from RNA polymerase II promoter, negative regulation of viral genome replication, and in the cellular composition group, genes were mainly categorized in cytoplasm, nucleus, nucleoplasm, plasma membrane. In the molecular function group genes were mainly concentrated in protein binding, RNA polymerase II core promoter proximal region sequence-specific DNA binding, RNA polymerase II transcription factor activity, sequence-specific DNA binding, transcription factor activity, sequence-specific DNA binding, metal ion binding.

Pathway enrichment analysis revealed the enrichment of the differentially expressed genes in many pathways: Herpes simplex virus 1 infection, Pathways in cancer, Epstein-Barr virus infection, Inflammatory bowel disease.

(Table 1)Gene ontology and pathway enrichment analysis of the differentially expressed genes associated with HPV positive and HPV negative in Cervical Squamous Cell Carcinoma via DAVID Software

CATEGORY	TERM	COUNT	%	P-VALUE
GOTERM_CC_DIRECT	cytoplasm	840	30.80308	1.76E-13
GOTERM_CC_DIRECT	nucleus	878	32.19655	2.06E-11
GOTERM_CC_DIRECT	cytosol	806	29.55629	5.63E-10
GOTERM_CC_DIRECT	nucleoplasm	599	21.96553	1.95E-09
GOTERM_CC_DIRECT	plasma membrane	718	26.3293	1.91E-06
GOTERM_MF_DIRECT	protein binding	1841	67.51008	1.14E-19
GOTERM_MF_DIRECT	RNA polymerase II core promoter proximal region sequence-specific DNA binding	233	8.544188	4.86E-10
GOTERM_MF_DIRECT	RNA polymerase II transcription factor activity, sequence-specific DNA binding	233	8.544188	9.19E-08
GOTERM_MF_DIRECT	transcription factor activity, sequence-specific DNA binding	118	4.327099	1.04E-07
GOTERM_MF_DIRECT	MHC class II protein complex binding	14	0.513385	1.11E-05
GOTERM_MF_DIRECT	sequence-specific double-stranded DNA binding	108	3.960396	2.52E-05
GOTERM_MF_DIRECT	metal ion binding	380	13.93473	0.028443
GOTERM_BP_DIRECT	defense response to virus	77	2.823616	3.9E-16
GOTERM_BP_DIRECT	regulation of transcription from RNA polymerase II promoter	299	10.96443	3.1E-10
GOTERM_BP_DIRECT	negative regulation of transcription from RNA polymerase II promoter	187	6.857352	3.22E-10
GOTERM_BP_DIRECT	negative regulation of viral genome replication	24	0.880088	5.75E-10

GOTERM_BP_DIRECT	positive regulation of transcription from RNA polymerase II promoter	221	8.104144	7.51E-10
GOTERM_BP_DIRECT	response to virus	39	1.430143	1.91E-09
GOTERM_BP_DIRECT	apoptotic process	121	4.43711	9.59E-08
GOTERM_BP_DIRECT	signal transduction	207	7.590759	1.69E-05
KEGG_PATHWAY	Herpes simplex virus 1 infection	112	4.107077	1.95E-07
KEGG_PATHWAY	Antigen processing and presentation	29	1.06344	1.18E-06
KEGG_PATHWAY	Epstein-Barr virus infection	55	2.016868	1.48E-06
KEGG_PATHWAY	Pathways in cancer	112	4.107077	7.16E-06
KEGG_PATHWAY	Inflammatory bowel disease	24	0.880088	1.39E-05
KEGG_PATHWAY	Graft-versus-host disease	18	0.660066	2.49E-05
KEGG_PATHWAY	Basal cell carcinoma	20	0.733407	0.000796

Construction of the Protein-Protein Interaction (PPI) Network for the DEGs:

To further study the potential relationships between the genes, STRING online database (https://string-db.org/) was used to mine and provide the significant interactions among these genes (24) (25).

Gene network was constructed with minimum required interaction score as high confidence (0.7), the network contains 307 nodes, 1313 edges, average node degree: 8.55, avg. local clustering coefficient: 0.454PPI, and enrichment p-value< 1.0e-16. This means that the proteins have more interactions among themselves than what would be expected for a random set of proteins of the same size and degree distribution drawn from the genome. Such an enrichment, indicates that the proteins are at least partially biologically connected, as a group.

Cytoscape was used for visual exploration of interactive networks with a confidence score> 0.4 as the cut-off criterion. The PPI network was retrieved from the STRING database and reconstructed via the Cytoscape software, and the degree of connectivity of each node of the network was calculated. The Cytoscape plug-in cytoHubba was used to identify the hub genes by finding the intersections of the first 30 genes from 12 topological analysis methods (Figures 4 and 5). Afterward, I used molecular complex detection (MCODE) to find clusters according to topology locating densely connected regions. MCODE was used to establish the important modules from the PPI network with a degree cutoff = 2, node score cutoff = 0.2, k- core = 2, and max depth = 100. Cluster 1 includes 27 nodes and 351 edges (Figure 6)

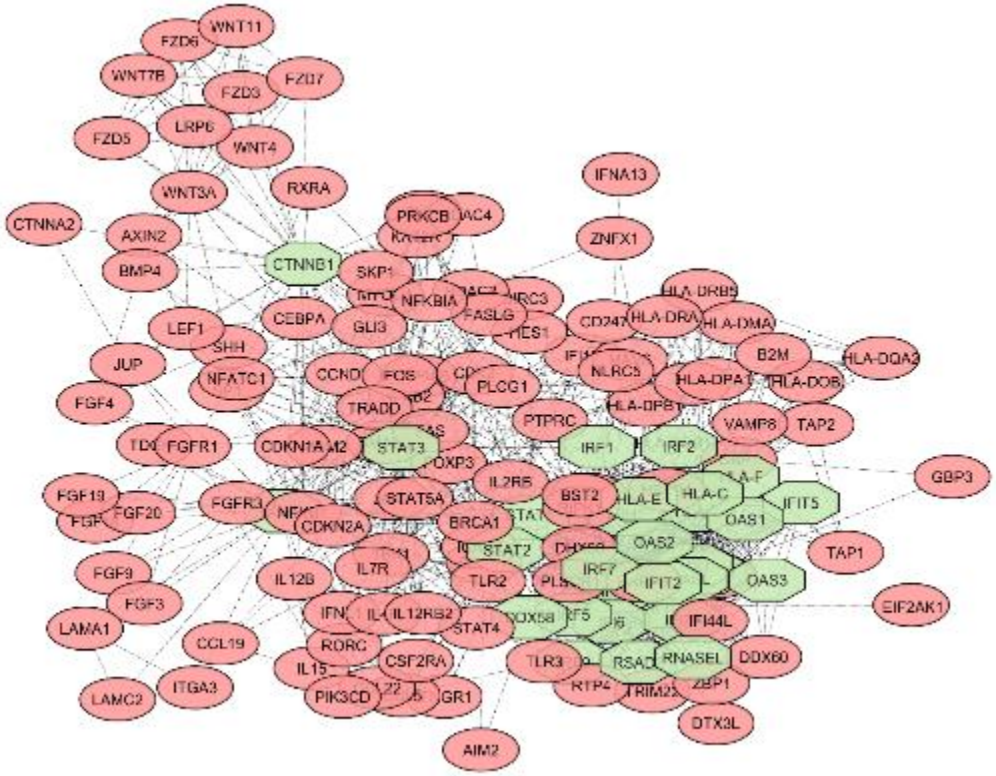
(Table 2) 30 differentially expressed Hub genes associated with HPV positive and HPV negative status in Cervical Squamous Cell Carcinoma resulted from CytoHubba plug-in

Rank	*Name*	*Score*
1	STAT1	64
2	STAT3	42
3	IRF7	40
4	HLA-C	39
4	HLA-E	39
4	HLA-B	39
7	IFIT3	38
7	DDX58	38
9	IFIT1	37
9	RSAD2	37
9	MX1	37
9	HLA-F	37
13	STAT2	36
14	OASL	35

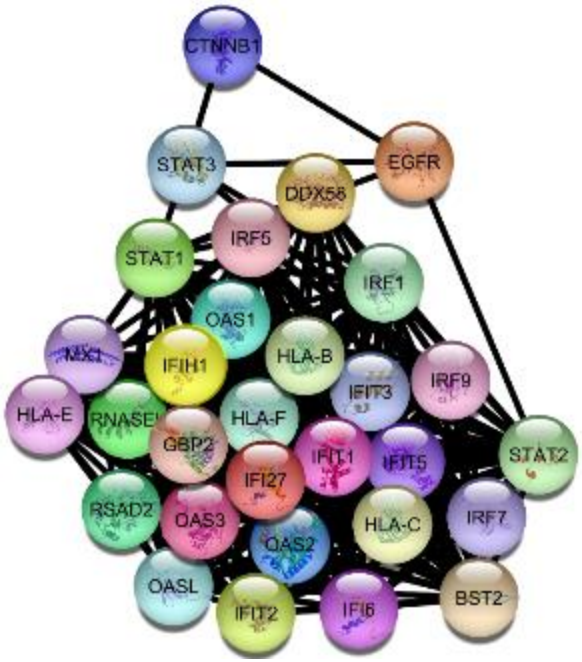
14	OAS2	35
14	CTNNB1	35
14	IFIH1	35
18	IFIT2	34
18	IRF9	34
18	OAS1	34
18	IRF1	34
18	OAS3	34
23	IFI6	32
24	EGFR	31
25	IFIT5	30
26	RNASEL	29

(Table 3) MCODE App results to find best clusters (genes network) to establish the important modules from the protein-protein interaction network of the differentially expressed genes associated with HPV positive and HPV negative in Cervical Squamous Cell Carcinoma.

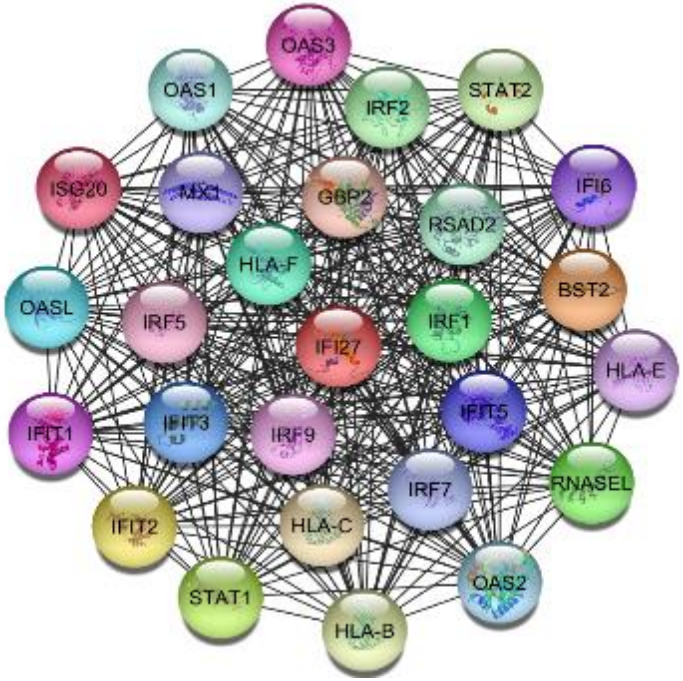
Cluster	*Score*	*Nodes*	*Edges*	*Node IDs*
1	27	27	351	IRF2, IRF9, HLA-B, IFI27, OAS3, OASL, HLA-F, STAT2, BST2, IFIT2, HLA-E, MX1, RNASEL, HLA-C, RSAD2, IRF1, ISG20, OAS1, IFIT5, IRF7, STAT1, OAS2, GBP2, IFIT1, IFIT3, IRF5, IFI6



(Figure 4) The network of Differentially expressed genes associated with HPV positive and HPV negative status in Cervical Squamous Cell Carcinoma is shown in pink via Cytoscape software, hub genes network is highlighted in green generated via CytoHubba plug-in

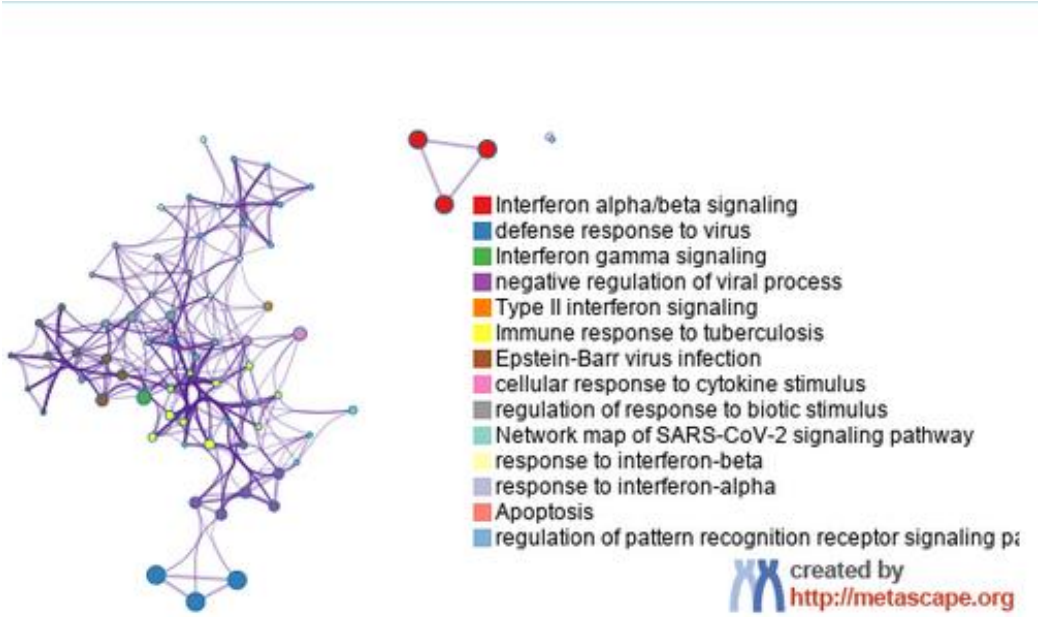


(Figure 5) hub genes network of Differentially expressed genes associated with HPV positive and HPV negative status in Cervical Squamous Cell Carcinoma generated via CytoHubba plug-in

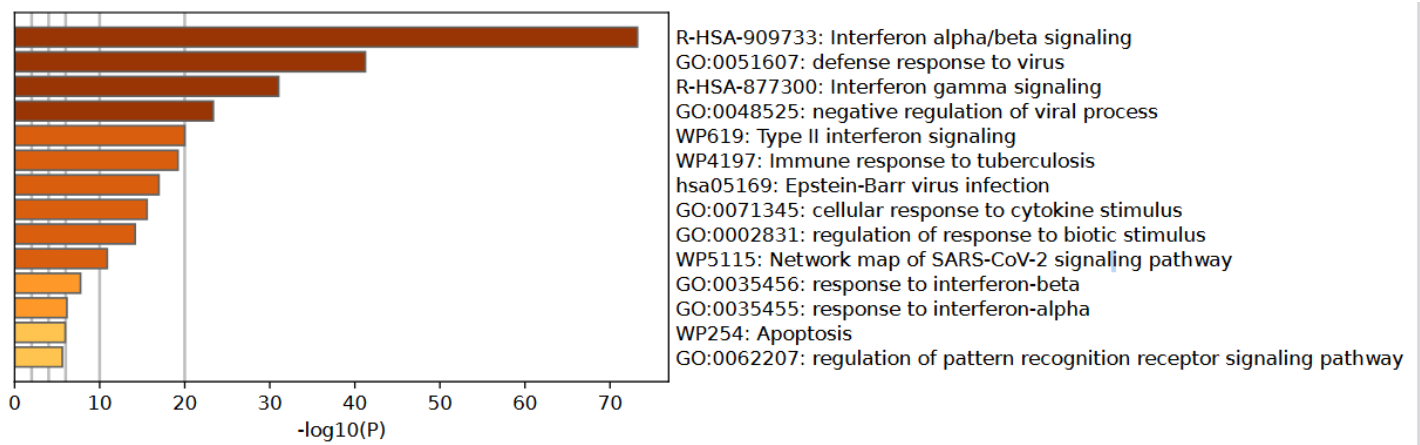


(Figure 6) Best ranked cluster of the differentially expressed genes associated with HPV positive and HPV negative status in Cervical Squamous Cell Carcinoma according to MCODE app results

Metascape tools (https://metascape.org/) were used to analyze the pathways and biological processes enrichment of hub genes (30). It was observed that key genes are enriched in defense response to virus, Interferon alpha/beta signaling, cellular response to cytokine stimulus, etc (Figure 7 and Figure 8).



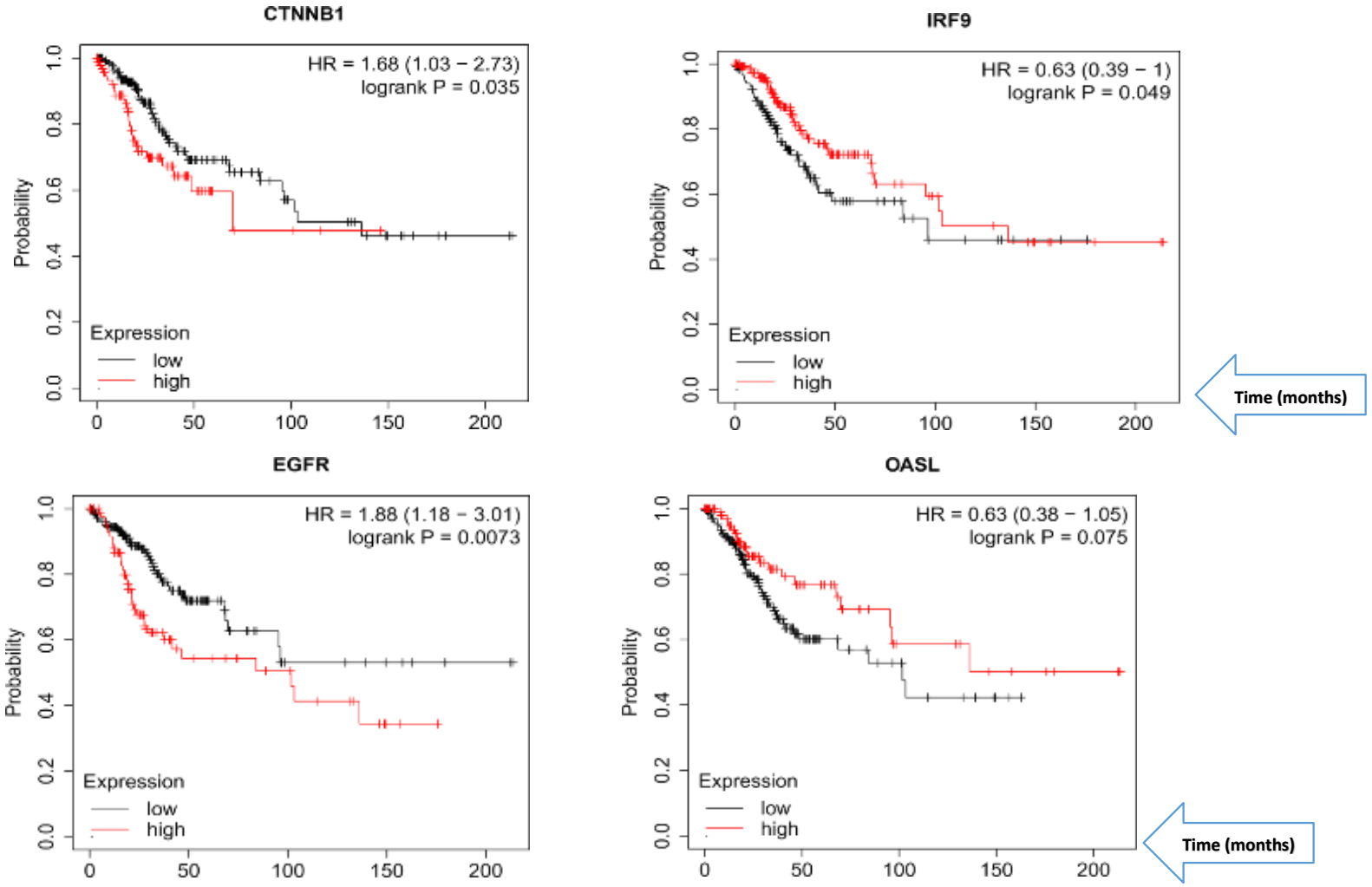
(Figure 7) Network of enriched pathways and biological processes of differentially expressed hub genes (key genes that were extracted from CytoHubba in the previous step) associated with HPV positive and HPV negative status in Cervical Squamous Cell Carcinoma samples created by Metascape website: colored by cluster ID, where nodes that share the same cluster ID are typically close to each other

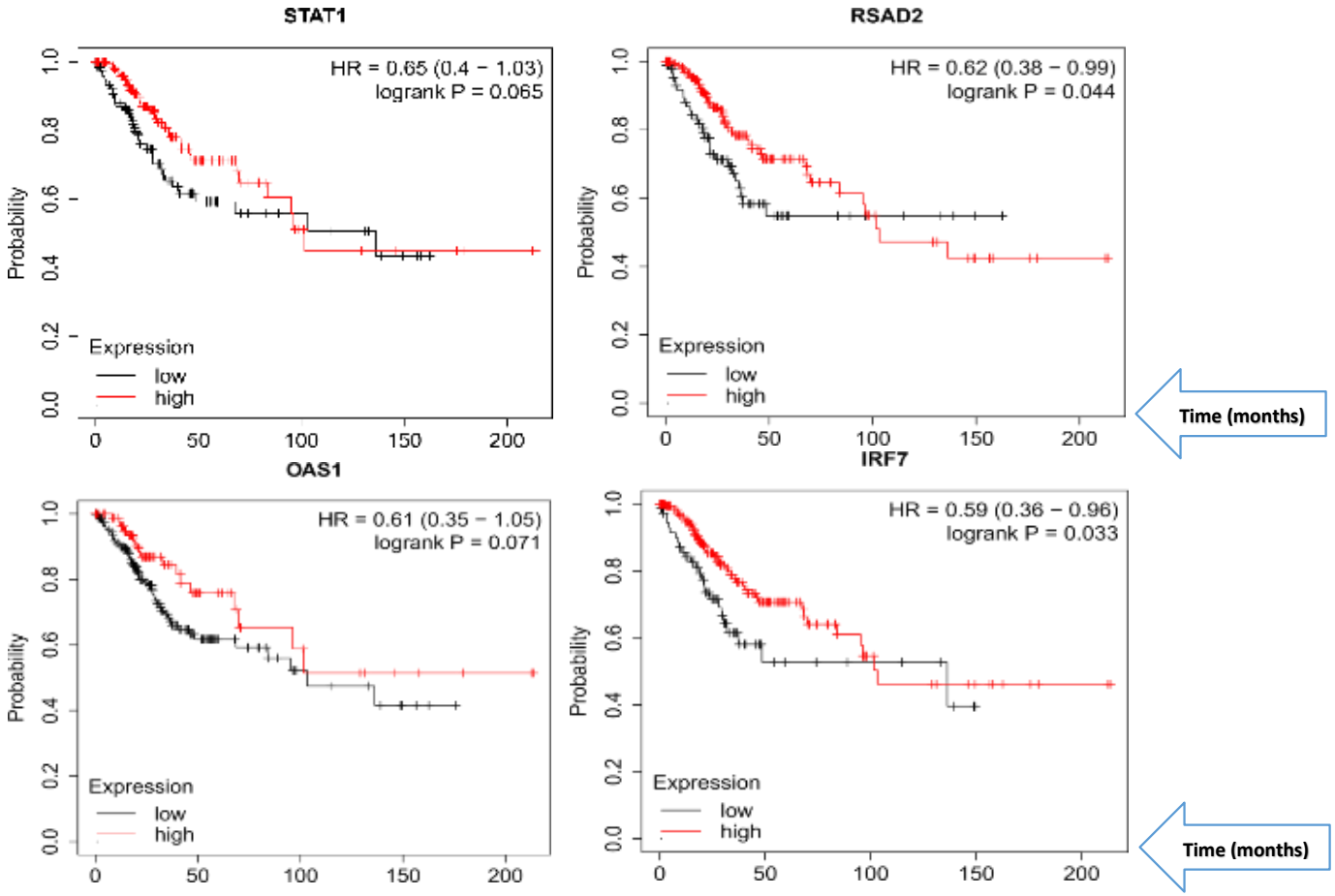


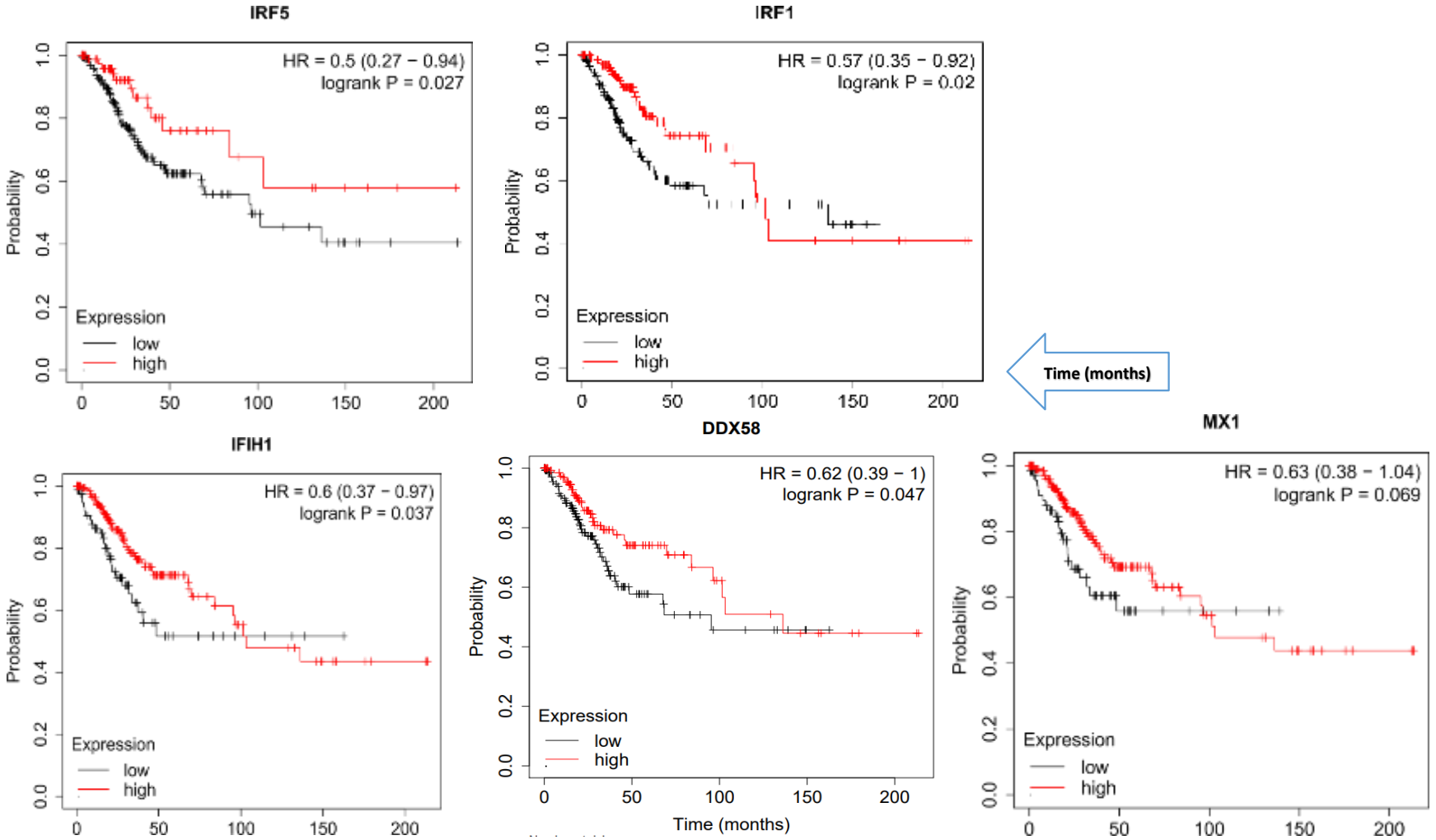
(Figure 8) Bar graph of enriched pathways and biological processes of the differentially expressed hub genes (key genes that were extracted from CytoHubba in the previous step) associated with HPV positive and HPV negative status in Cervical Squamous Cell Carcinoma samples, colored by p-values created via Metascape website

Kaplan-Meier Survival Curve of the Differentially Expressed Genes and Screening of Prognostic Biomarkers

Kaplan-Meier Plotter (https://kmplot.com/analysis/) can use gene expression data to assess the survival rate of cancer patients. The main purpose of this tool is biomarkers evaluation based on meta-analysis (29). Using the Kaplan-Meier chart, the effects of the hub genes is evaluated on the prognosis of squamous cervical cancer patients, in order to identify genes that can be used as prognostic biomarkers for this disease.


CTNNB1
HR = 1.68 (1.03 − 2.73)
logrank P = 0.035
Probability
Expression
low
high
IRF9
HR = 0.63 (0.39 − 1)
logrank P = 0.049
Time (months)
EGFR
HR = 1.88 (1.18 − 3.01)
logrank P = 0.0073
OASL
HR = 0.63 (0.38 − 1.05)
logrank P = 0.075
Time (months)

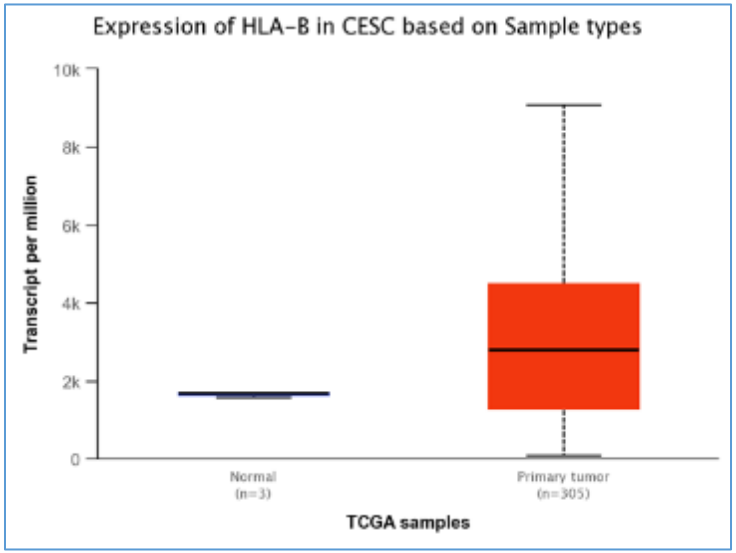

STAT1
HR = 0.65 (0.4 − 1.03)
logrank P = 0.065
Probability
Expression
low
high
RSAD2
HR = 0.62 (0.38 − 0.99)
logrank P = 0.044
Time (months)
OAS1
HR = 0.61 (0.35 − 1.05)
logrank P = 0.071
IRF7
HR = 0.59 (0.36 − 0.96)
logrank P = 0.033
Time (months)



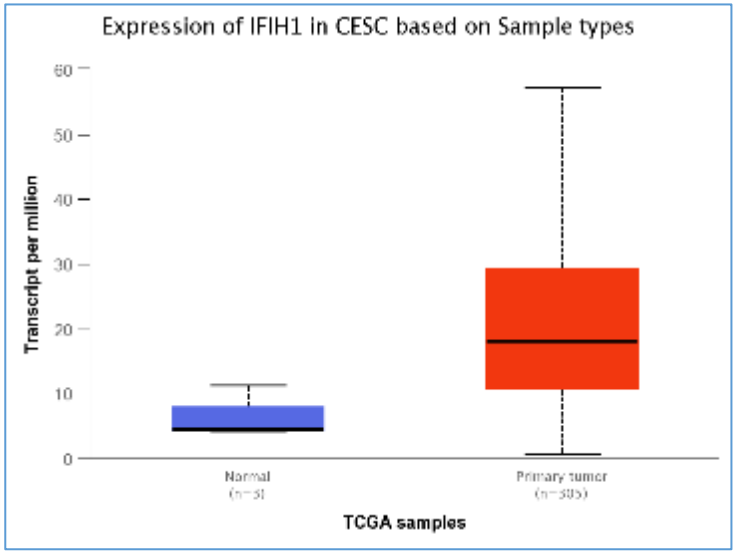
Kaplan-Meier plots of differentially expressed hub genes associated with HPV positive and HPV negative status in Cervical Squamous Cell Carcinoma samples with significant p-value. Kaplan-Meier Plotter use gene expression data to assess the survival rate of cancer patients

Hub Gene Verification Through UALCAN

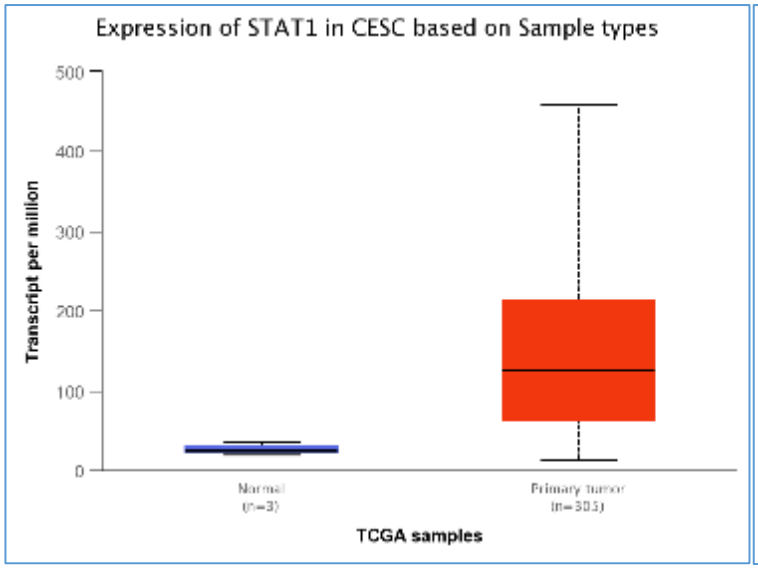
UALCAN is an online tool (http://ualcan.path.uab.edu/) with data from TCGA and GTEx(22) (23), was used to verify the expression of these key genes in squamous cell cervical cancer. In this study, according to the RNA sequence data from TCGA database, the mRNA expression levels of 30 hub genes were compared between the squamous cervical cancer samples and the adjacent normal tissues. Twelve genes were found to be highly expressed at the transcriptional level in cancer tissues compared with normal tissues with statistical significance (Figure 9 **A-L**).



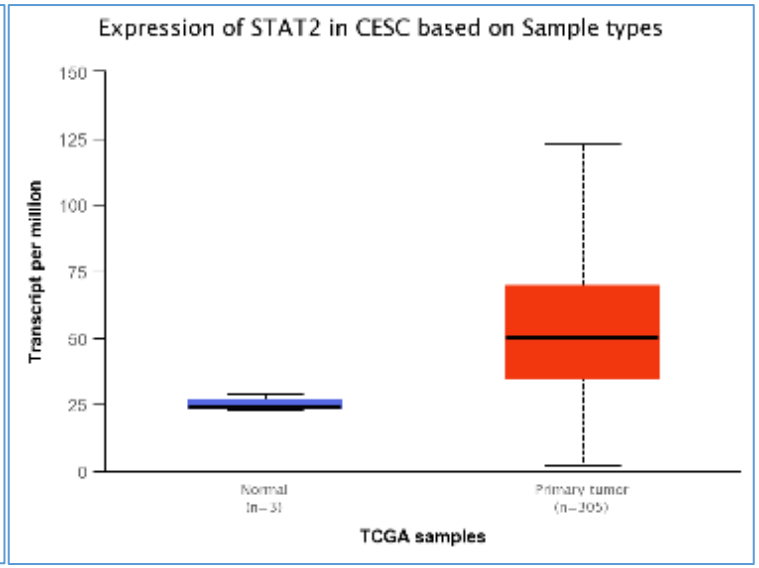
A



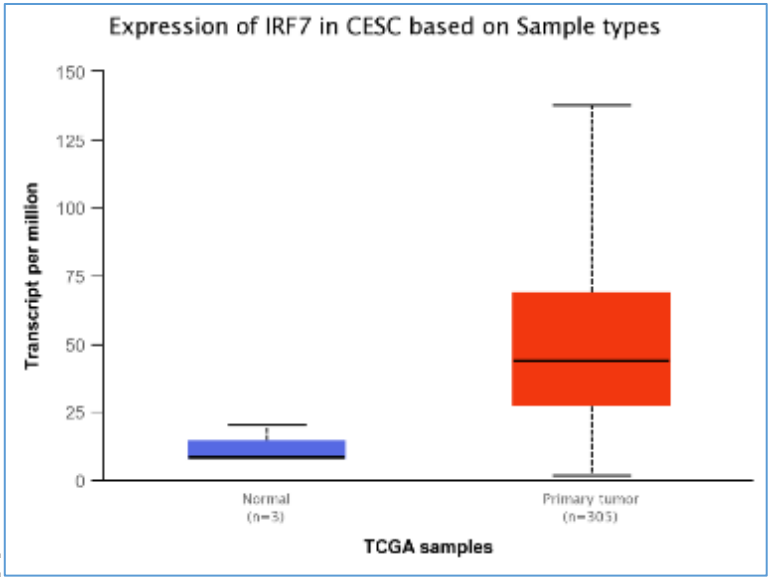
B


Expression of STAT1 in CESC based on Sample types
Transcript per million
500
400
300
200
100
0
Normal
(n=3)
Primary tumor
(n=305)
TCGA samples

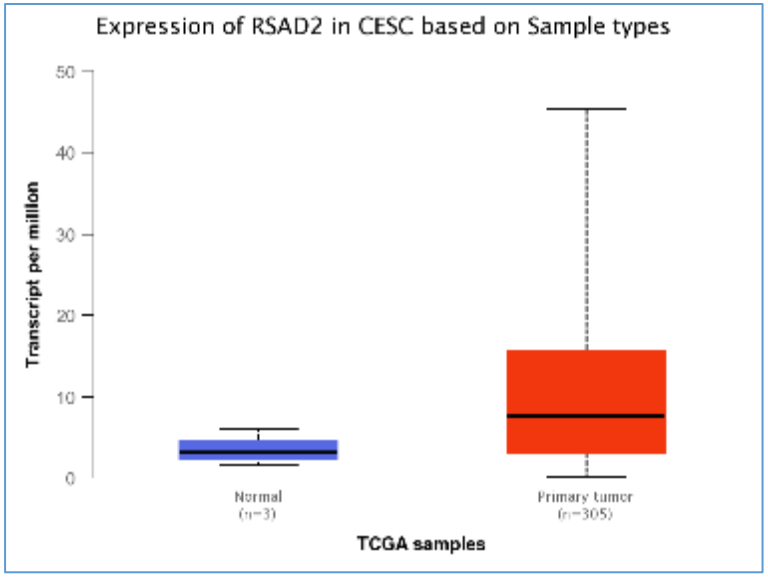
C


Expression of STAT2 in CESC based on Sample types
Transcript per million
150
125
100
75
50
25
0
Normal
(n=3)
Primary tumor
(n=305)
TCGA samples

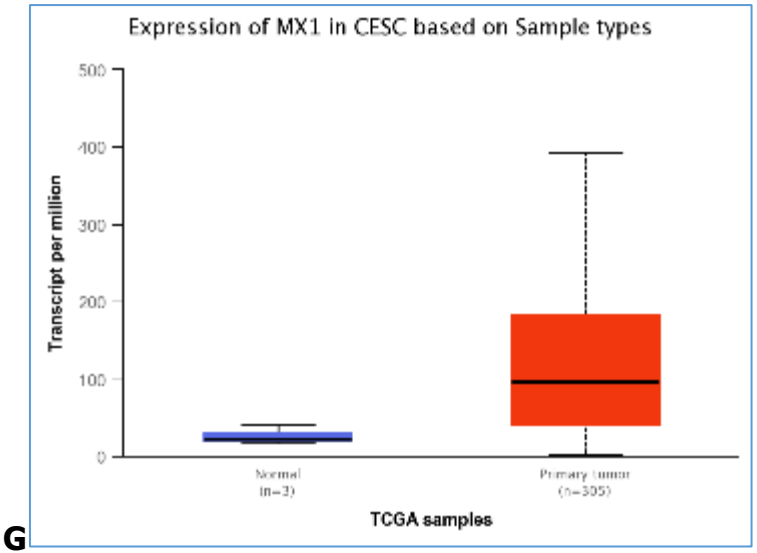
D


Expression of IRF7 in CESC based on Sample types
Transcript per million
150
125
100
75
50
25
0
Normal
(n=3)
Primary tumor
(n=305)
TCGA samples

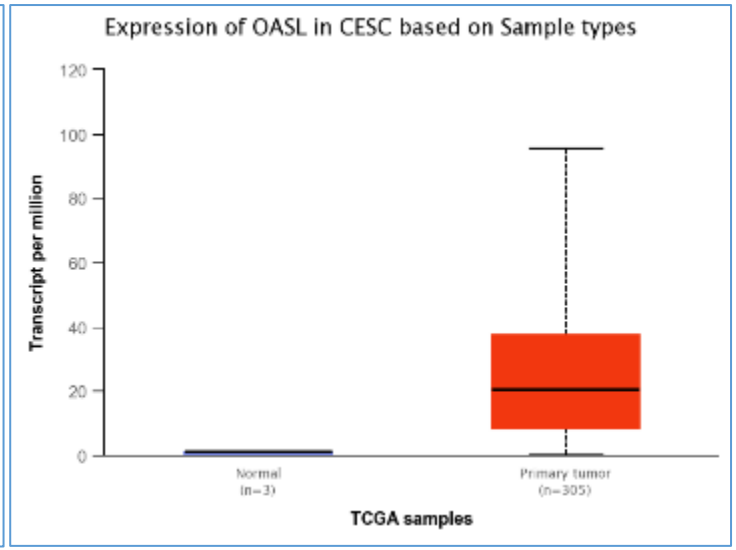
E


Expression of RSAD2 in CESC based on Sample types
Transcript per million
50
40
30
20
10
0
Normal
(n=3)
Primary tumor
(n=305)
TCGA samples

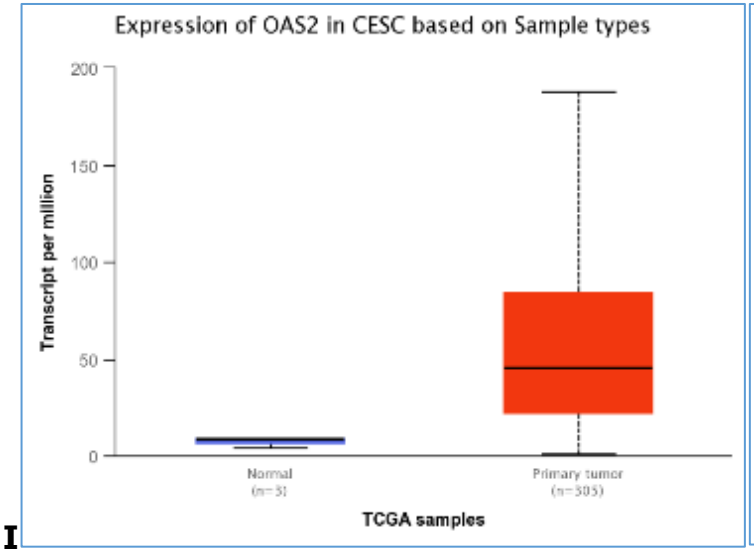
F


Expression of MX1 in CESC based on Sample types
Transcript per million
500
400
300
200
100
0
Normal
(n=3)
Primary tumor
(n=305)
TCGA samples

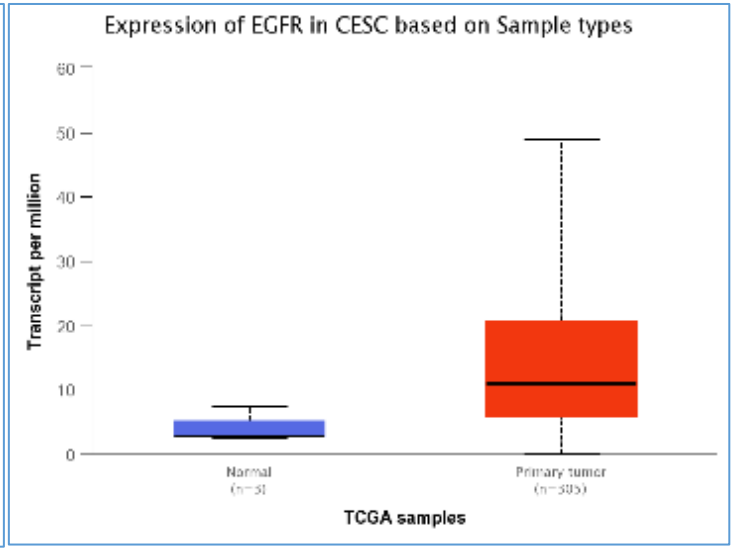
G


Expression of OASL in CESC based on Sample types
Transcript per million
120
100
80
60
40
20
0
Normal
(n=3)
Primary tumor
(n=305)
TCGA samples

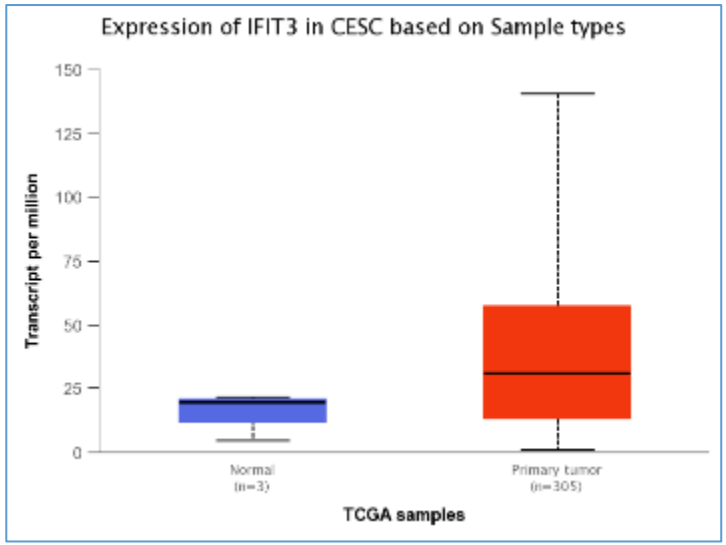
H


Expression of OAS2 in CESC based on Sample types
Transcript per million
200
150
100
50
0
Normal
(n=3)
Primary tumor
(n=305)
TCGA samples

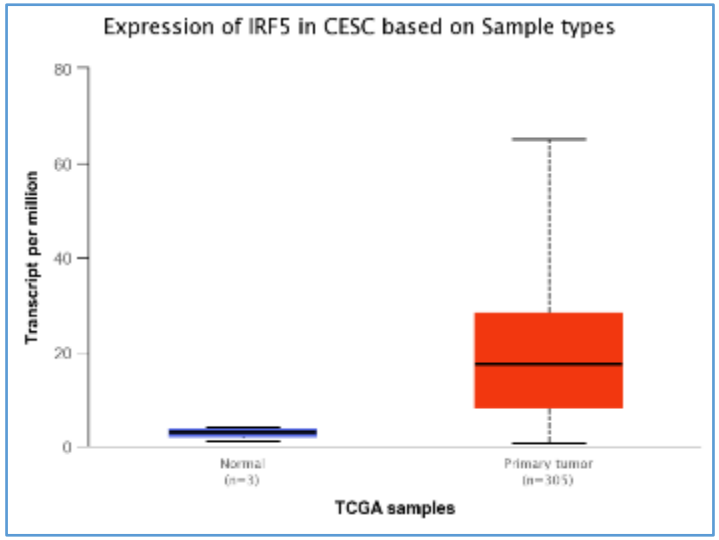
I


Expression of EGFR in CESC based on Sample types
Transcript per million
60
50
40
30
20
10
0
Normal
(n=3)
Primary tumor
(n=305)
TCGA samples

J



K



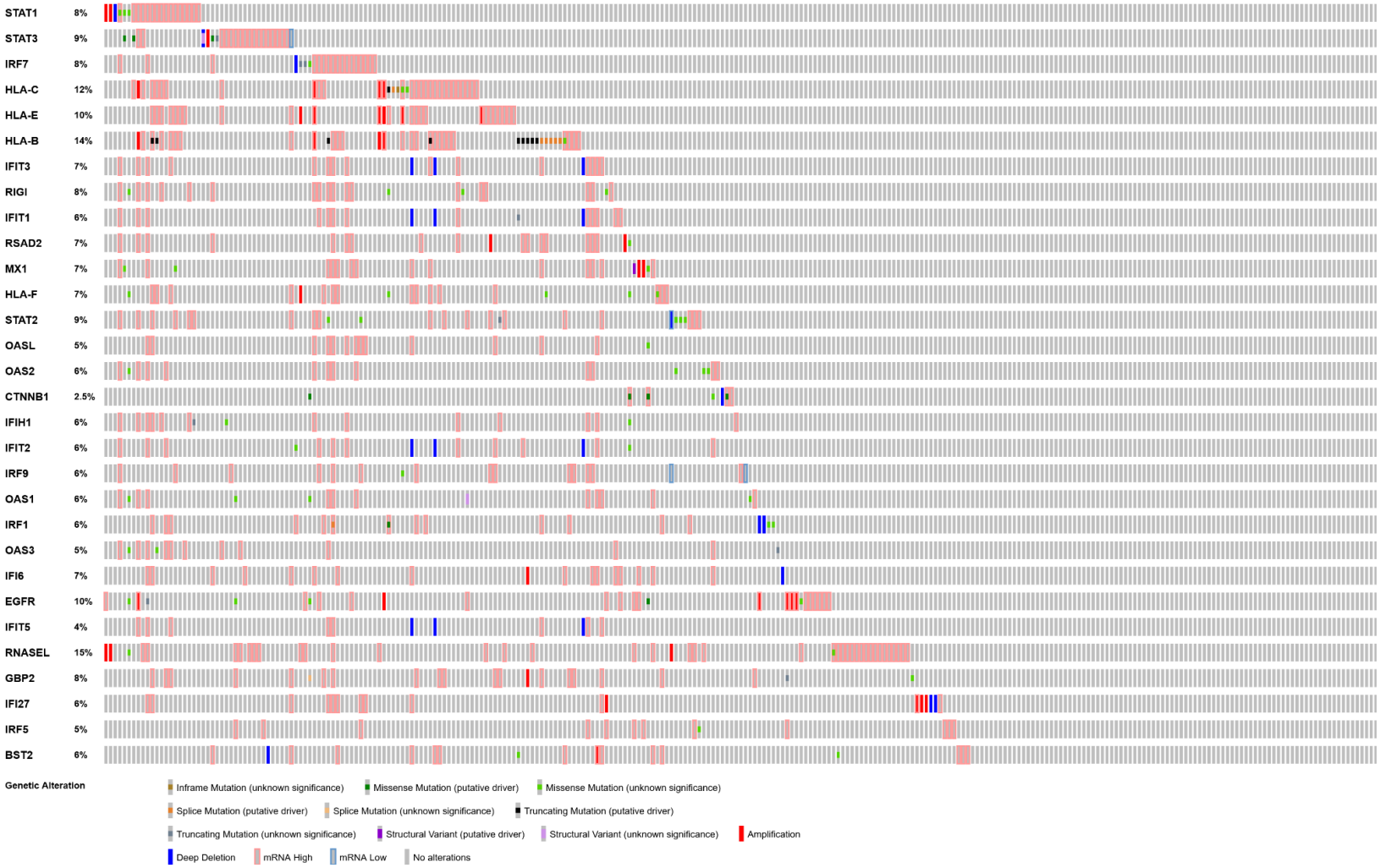
L

Figure 9 | Analysis of the expression of key mutated genes of HPV positive and HPV negative status in Cervical Squamous Cell Carcinoma samples with significant p-value. Twelve key genes are highly expressed in Cervical Squamous Cell Carcinoma tissues compared with normal tissues. **(A)***HLA-B,* **(B)** *IFIH1,* **(C)** *STAT1,* **(D)** *STAT2,* **(E)** *IRF7,* **(F)** *RSAD2,* **(K)** *IFIT3,* **(L)** *IRF5,* **(I)** *OAS2,* **(J)** *EGFR,* **(G)** *MX1,* **(H)** *OASL*

Analysis of Cancer Genomics Data Through cBioPortal

The cBioPortal (19) for Cancer Genomics online platform (http://www.cbioportal.org/) provides resources for visualizing and analyzing multidimensional cancer genomics data. A graphical analysis of the genetic variation was done via oncoprint tool to investigate the clinical significance of the hub genes. As shown in the figure, 30 key genes all showed a high mutation rate in squamous cell cervical cancer, with a rate of genome change ranging from 2.5% to 15% (Figure 10). OncoPrint from cBioPortal revealed that 68% of cases (187 out of 275) exhibited genetic alterations, including amplification, deep deletion, missense mutation, severe depletion, truncating mutation, and various mutations.

One limitation of the cBioportal analysis was that none of the cases were divided into HPV+ and HPV- groups. However, genetic alterations and up-or downregulation of these hub genes could be demonstrated.



(Figure 10) Graphic analysis of the genetic alteration of key genes in squamous cell cervical cancer via cBioPortal platform. OncoPrint from cBioPortal revealed that 68% of cases (187 out of 275) exhibited genetic alterations, including amplification, deep deletion, missense mutation, severe depletion, truncating mutation, and various mutations

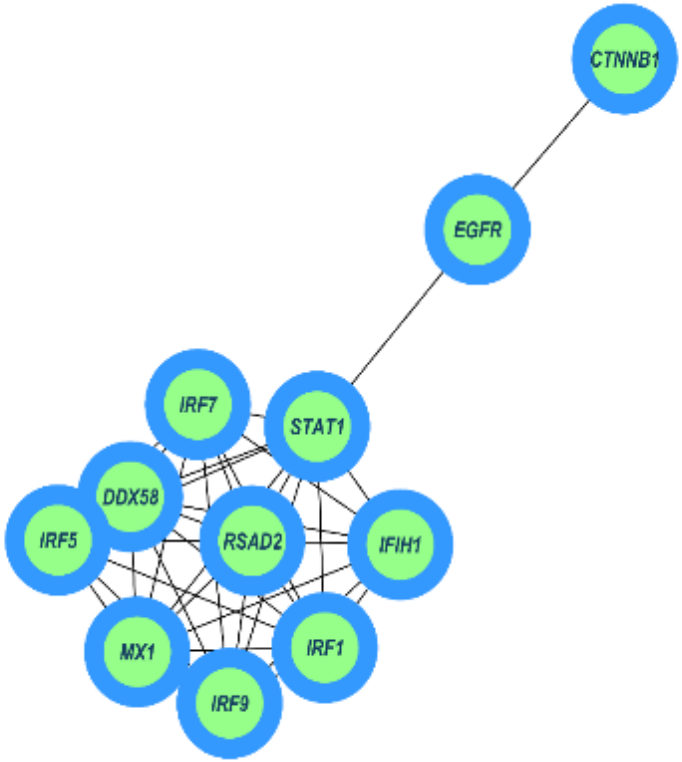
Screening and Survival Analysis of Pivotal Genes

Using the intersection of the first 30 genes in **cytoHubba's** 12 algorithms, I identified 11 key genes: STAT1, CTNNB1 IRF9, EGFR, RSAD2, IRF7,MX1, IFIH1, IRF5, IRF1 and DDX58 (Figure 11). **DAVID** software was used to analyze the pathway and biological process enrichment of these eleven hub genes. It was observed that key genes are enriched in the defense response to virus, positive regulation of transcription from RNA polymerase II promoter, transcription factor activity, sequence-specific DNA binding, Hepatitis C pathway. The Functional Annotation Clustering option in DAVID was also used to cluster the hub genes into different functional groups which makes the data more interpretable (Figure 12).

The **cBioPortal** online platform provided a graphic analysis of the genetic variation of the hub genes. As shown in Figure10, eleven key DEGs all showed a high mutation rate in SCC, with a rate of genome change ranging from 4% to 7%. To determine whether the selected hub genes have clinical correlations, I used **Kaplan-Meier curves** to analyze the univariate survival of these genes and found that the expression of those genes was correlated with prognosis (*as shown in the plots above*). Thus, these genes can be used as prognostic indicators of Squamous Cervical Cancer.

UALCAN, an online tool with data from TCGA and GTEx, was used to verify the expression of these key genes in Squamous Cervical Cancer. In this study, according to the RNA sequence data from TCGA database, the mRNA expression levels of 11 genes were compared between the tumor samples and the adjacent normal tissues (Figure11). These genes were found to be highly expressed at the transcriptional level, except one gene" CTNNB1" was found to be highly expressed in normal tissues than tumor tissues with statistical significance=7.31E-01. I chose P Value=0.05 as a threshold.

Human Protein Atlas could be used to analyze the hub genes at the histological level, since it is a discovery tool for potential oncology biomarkers of clinical relevance.

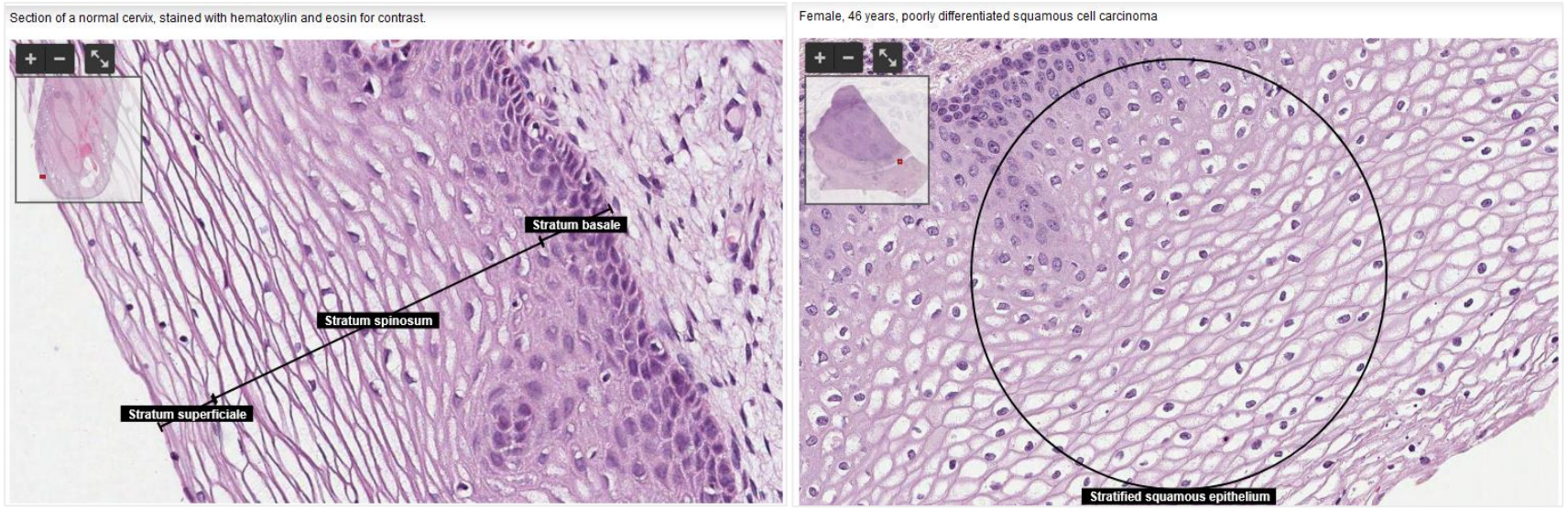


(Figure 11) Eleven key genes network associated with HPV positive and HPV negative status in Cervical Squamous Cell Carcinoma samples after using the intersection of the first 30 genes in cytoHubba's 12 algorithms

(Figure 12) 2D View of Functional Annotation Clustering via DAVID created to reveal the hub genes associated with HPV positive and HPV negative status in Cervical Squamous Cell Carcinoma samples that were corresponding to the functional enrichment terms. The green squares indicate to the positive corresponding, while the dark ones refer to the negative corresponding, ***Enrichment Score of the cluster: 1.96***

Human Protein Atlas as a tool to facilitate clinical biomarker discovery

Tissue-based diagnostics and research is incessantly evolving with the development of new molecular tools. It has long been realized that immunohistochemistry can add an important new level of information on top of morphology and that protein expression patterns in a cancer may yield crucial diagnostic and prognostic information (26). An immunohistochemistry-based map was generated of protein expression profiles was in normal tissues, cancer and cell lines. For each antibody, altogether 708 spots of tissues and cells are analyzed and the resulting protein expression data, including underlying high-resolution images, are published on the free and publically available Human Protein Atlas portal (www.proteinatlas.org/). The gene-centric database includes a putative classification of proteins in various protein classes, both functional classes, such as kinases or transcription factors and project-related classes, such as candidate genes for cancer. Protein profiling of cancer-associated signatures thus requires the possibility of analyzing several proteins in tumor tissue samples, and it is essential to analyze a multitude of tumor types in order to widen the screening for cancer biomarkers. The major forms of human cancer analyzed in the Human Protein Atlas are represented by 12 different tumors for each tumor type, this enables the perception of a potential protein signature for each given type of cancer and a starting point for further analyses of cancer type–specific proteins. Extended analyses of clinically well-defined tumor tissue samples from clinical biobanks containing large retrospectively collected patient cohorts with long-term follow-up are required to test and validate the diagnostic, prognostic and predictive value with regard to the treatment of a candidate cancer marker. Thus, the presented Human Protein Atlas provides a resource for pathology-based biomedical research, including protein science and biomarker discovery with in silico-based methods. Moreover, the global analysis of how our genome is expressed at the protein level has provided basic knowledge on the ubiquitous expression of a large proportion of our proteins and revealed the paucity of cell- and tissue-type-specific proteins. Taking the aforementioned information into consideration, I queried the eleven hub genes that were obtained from the previous step separately to reveal the description and Entrez summary of each one using the Cancer Atlas. Afterward, the expression level of each protein in different tissues is obtained, and the comparison is done by detecting the level of antibody staining in normal tissue verses cancer tissues (Figures 13 from A to K).



These figures from Cancer Atlas show Normal Squamous epithelium elements to the left, and poorly differentiated Squamous Cell Carcinoma to the right

Results

Gene name	IFIH1 (HGNC Symbol)
Synonyms	Hlcd, IDDM19, MDA-5, MDA5
Description	Interferon induced with helicase C domain 1 (HGNC Symbol)
Entrez gene summary	DEAD box proteins, characterized by the conserved motif Asp-Glu-Ala-Asp (DEAD), are putative RNA helicases. They are implicated in a number of cellular processes involving alteration of RNA secondary structure such as translation initiation, nuclear and mitochondrial splicing, and ribosome and spliceosome assembly. Based on their distribution patterns, some members of this family are believed to be involved in embryogenesis, spermatogenesis, and cellular growth and division. This gene encodes a DEAD box protein that is upregulated in response to treatment with beta-interferon and a protein kinase C-

	activating compound, mezerein. Irreversible reprogramming of melanomas can be achieved by treatment with both these agents; treatment with either agent alone only achieves reversible differentiation. Genetic variation in this gene is associated with diabetes mellitus insulin-dependent type 19

Gene name	IRF5
Description	Interferon regulatory factor 5 (HGNC Symbol)
Entrez gene summary	This gene encodes a member of the interferon regulatory factor (IRF) family, a group of transcription factors with diverse roles, including virus-mediated activation of interferon, and modulation of cell growth, differentiation, apoptosis, and immune system activity. Members of the IRF family are characterized by a conserved N-terminal DNA-binding domain containing tryptophan (W) repeats. Multiple transcript variants encoding different isoforms have been found for this gene, and a 30-nt indel polymorphism (SNP rs60344245) can result in loss of a 10-aa segment.

Gene name	IRF1 (HGNC Symbol)
Synonyms	MAR
Description	Interferon regulatory factor 1 (HGNC Symbol)
Entrez gene summary	IRF1 encodes interferon regulatory factor 1, a member of the interferon regulatory transcription factor (IRF) family. IRF1 serves as an activator of interferons alpha and beta transcription, and in mouse it has been shown to be required for double-stranded RNA induction of these genes. IRF1 also functions as a transcription activator of genes induced by interferons alpha, beta, and gamma. Further, IRF1 has been shown to play roles in regulating apoptosis and tumor-suppression

Gene name	DDX58 (HGNC Symbol)
Synonyms	DKFZp434J1111, FLJ13599, RIG-I

Description	DEAD (Asp-Glu-Ala-Asp) box polypeptide 58 (HGNC Symbol)
Entrez gene summary	DEAD box proteins, characterized by the conserved motif Asp-Glu-Ala-Asp (DEAD), are putative RNA helicases which are implicated in a number of cellular processes involving RNA binding and alteration of RNA secondary structure. This gene encodes a protein containing RNA helicase-DEAD box protein motifs and a caspase recruitment domain (CARD). It is involved in viral double-stranded (ds) RNA recognition and the regulation of immune response

Gene name	STAT1 (HGNC Symbol)
Synonyms	ISGF-3, STAT91
Description	Signal transducer and activator of transcription 1, 91kDa (HGNC Symbol)
Entrez gene summary	The protein encoded by this gene is a member of the STAT protein family. In response to cytokines and growth factors, STAT family members are phosphorylated by the receptor associated kinases, and then form homo- or heterodimers that translocate to the cell nucleus where they act as transcription activators. This protein can be activated by various ligands including interferon-alpha, interferon-gamma, EGF, PDGF and IL6. This protein mediates the expression of a variety of genes, which is thought to be important for cell viability in response to different cell stimuli and pathogens. Two alternatively spliced transcript variants encoding distinct isoforms have been described.

Gene name	CTNNB1 (HGNC Symbol)
Synonyms	armadillo, beta-catenin, CTNNB
Description	Catenin (cadherin-associated protein), beta 1, 88kDa (HGNC Symbol)
Entrez gene summary	The protein encoded by this gene is part of a complex of proteins that constitute adherens junctions (AJs). AJs are necessary for the creation and maintenance of epithelial cell layers by regulating cell growth and adhesion between cells. The encoded protein also anchors the actin cytoskeleton and may be responsible for transmitting the contact inhibition signal that causes cells to stop dividing once the

	epithelial sheet is complete. Finally, this protein binds to the product of the APC gene, which is mutated in adenomatous polyposis of the colon. Mutations in this gene are a cause of colorectal cancer (CRC), pilomatrixoma (PTR), medulloblastoma (MDB), and ovarian cancer. Three transcript variants encoding the same protein have been found for this gene

Gene name	IRF9 (HGNC Symbol)
Synonyms	ISGF3G
Description	Interferon regulatory factor 9

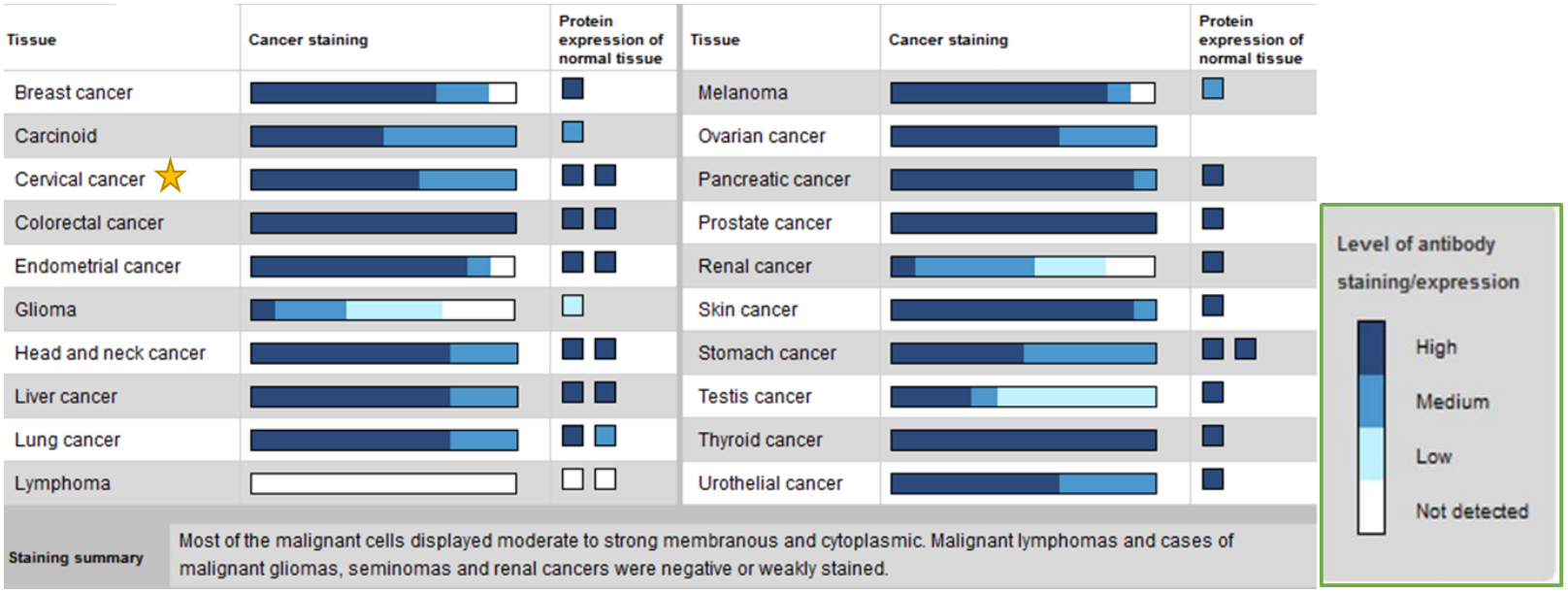
Gene name	EGFR (HGNC Symbol)
Synonyms	ERBB, ERBB1
Description	Epidermal growth factor receptor (HGNC Symbol)
Entrez gene summary	The protein encoded by this gene is a transmembrane glycoprotein that is a member of the protein kinase superfamily. This protein is a receptor for members of the epidermal growth factor family. EGFR is a cell surface protein that binds to epidermal growth factor. Binding of the protein to a ligand induces receptor dimerization and tyrosine autophosphorylation and leads to cell proliferation. Mutations in this gene are associated with lung cancer. Multiple alternatively spliced transcript variants that encode different protein isoforms have been found for this gene.

Gene name	RSAD2 (HGNC Symbol)
Synonyms	cig5, vig1, viperin
Description	Radical S-adenosyl methionine domain containing 2 (HGNC Symbol)

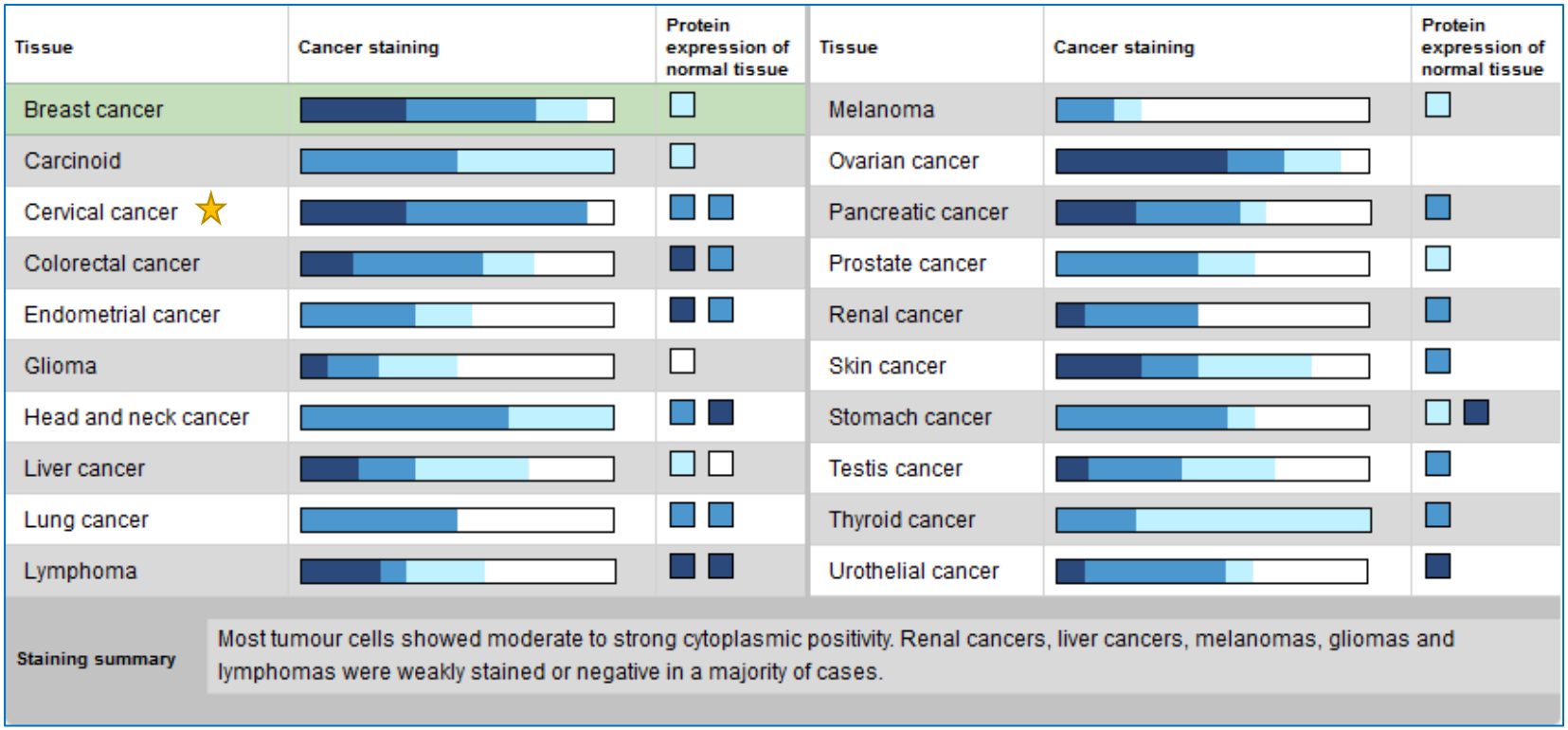
Gene name	IRF7

Description	Interferon regulatory factor 7 (HGNC Symbol)
Entrez gene summary	IRF7 encodes interferon regulatory factor 7, a member of the interferon regulatory transcription factor (IRF) family. IRF7 has been shown to play a role in the transcriptional activation of virus-inducible cellular genes, including interferon beta chain genes. Inducible expression of IRF7 is largely restricted to lymphoid tissue. Multiple IRF7 transcript variants have been identified, although the functional consequences of these have not yet been established.

Gene name	MX1 (HGNC Symbol)
Synonyms	IFI-78K, MxA
Description	MX dynamin-like GTPase 1 (HGNC Symbol)
Entrez gene summary	This gene encodes a guanosine triphosphate (GTP)-metabolizing protein that participates in the cellular antiviral response. The encoded protein is induced by type I and type II interferons and antagonizes the replication process of several different RNA and DNA viruses. There is a related gene located adjacent to this gene on chromosome 21, and there are multiple pseudogenes located in a cluster on chromosome 4. Alternative splicing results in multiple transcript variants



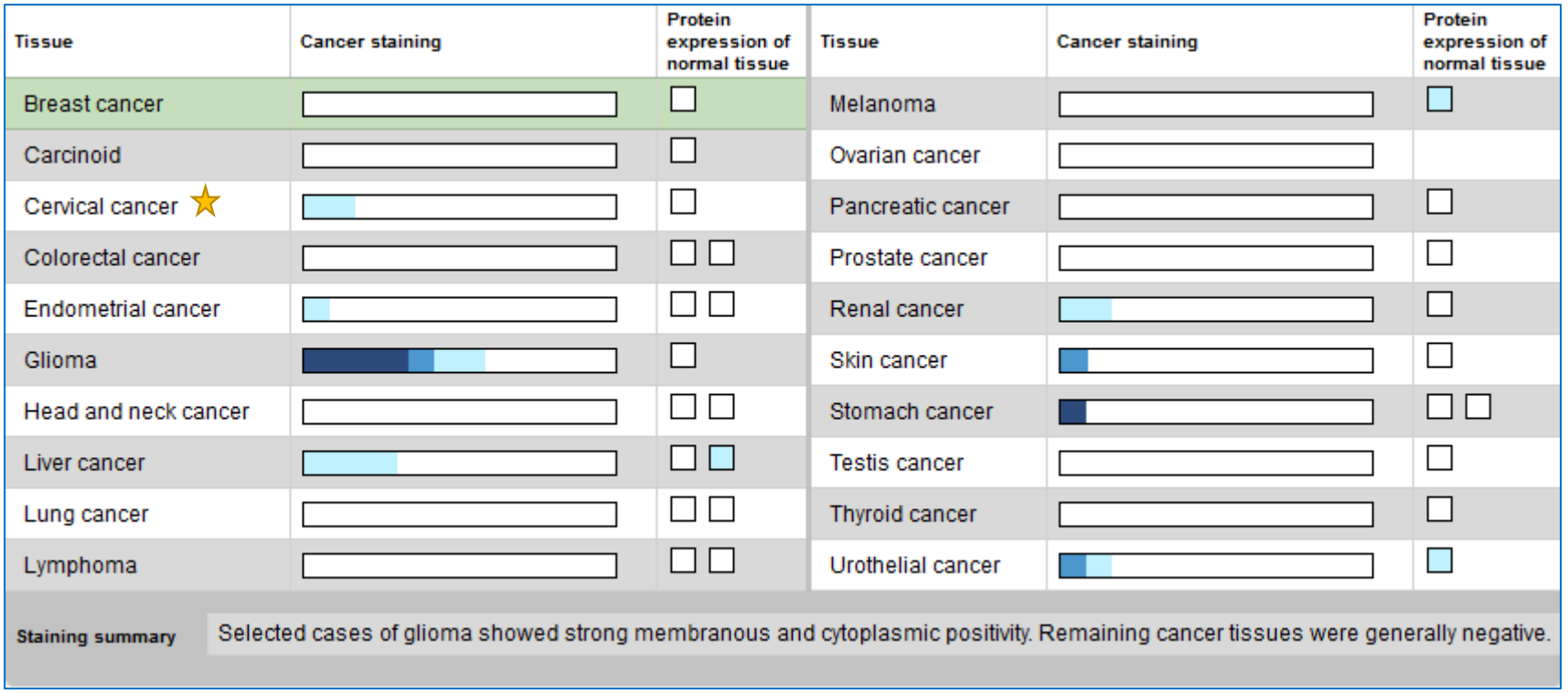
(Figure 13-A) CTNNB1 protein levels of expression compared between normal and cancer tissues in different cancer types, the level of antibody staining is grading from high (dark blue) to not detected (white)



Tissue	Cancer staining	Protein expression of normal tissue	Tissue	Cancer staining	Protein expression of normal tissue
Breast cancer			Melanoma		
Carcinoid			Ovarian cancer		
Cervical cancer ★			Pancreatic cancer		
Colorectal cancer			Prostate cancer		
Endometrial cancer			Renal cancer		
Glioma			Skin cancer		
Head and neck cancer			Stomach cancer		
Liver cancer			Testis cancer		
Lung cancer			Thyroid cancer		
Lymphoma			Urothelial cancer		

Staining summary Most tumour cells showed moderate to strong cytoplasmic positivity. Renal cancers, liver cancers, melanomas, gliomas and lymphomas were weakly stained or negative in a majority of cases.

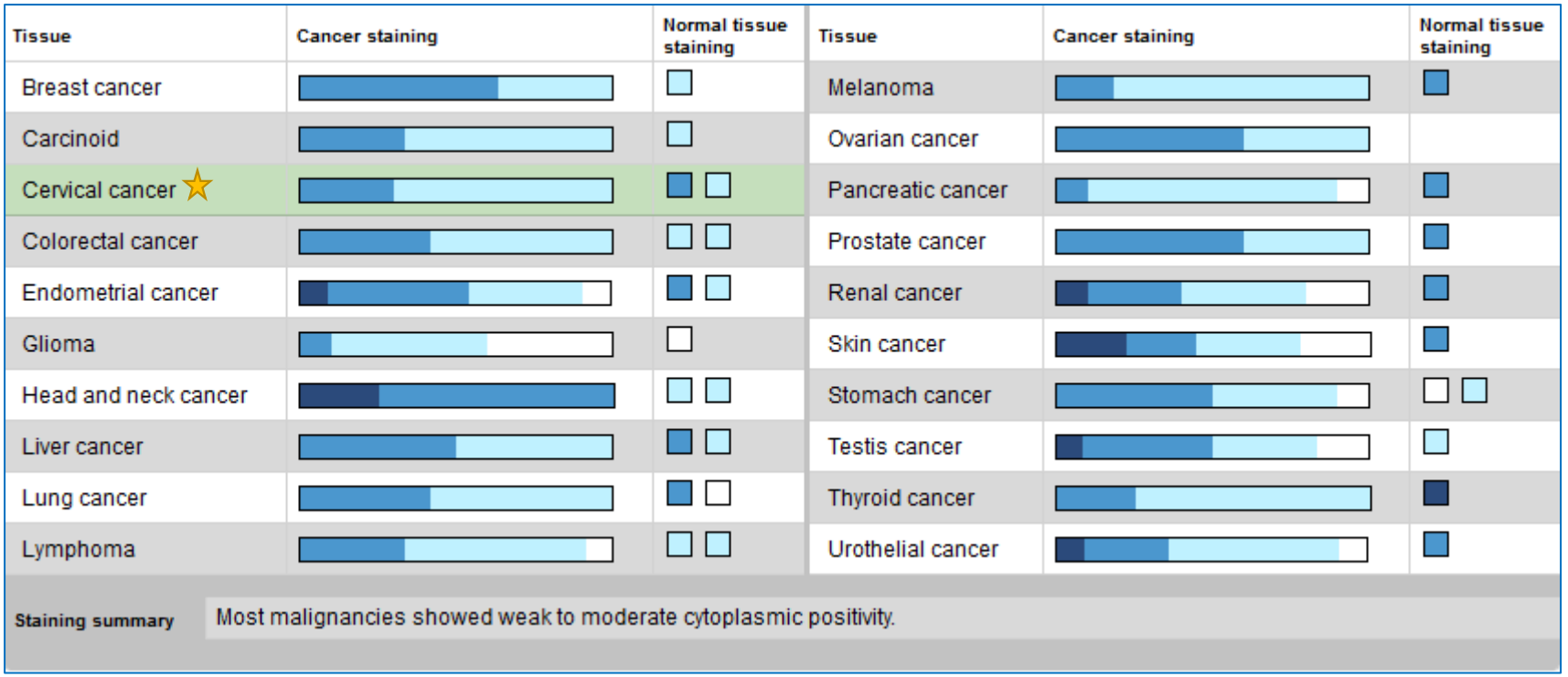
(Figure 13-B) DDX58 protein levels of expression compared between normal and cancer tissues in different cancer types, the level of antibody staining is grading from high (dark blue) to not detected (white)



Tissue	Cancer staining	Protein expression of normal tissue	Tissue	Cancer staining	Protein expression of normal tissue
Breast cancer			Melanoma		
Carcinoid			Ovarian cancer		
Cervical cancer ★			Pancreatic cancer		
Colorectal cancer			Prostate cancer		
Endometrial cancer			Renal cancer		
Glioma			Skin cancer		
Head and neck cancer			Stomach cancer		
Liver cancer			Testis cancer		
Lung cancer			Thyroid cancer		
Lymphoma			Urothelial cancer		

Staining summary Selected cases of glioma showed strong membranous and cytoplasmic positivity. Remaining cancer tissues were generally negative.

(Figure 13-C) EGFR protein levels of expression compared between normal and cancer tissues in different cancer types, the level of antibody staining is grading from high (dark blue) to not detected (white)



Tissue	Cancer staining	Normal tissue staining	Tissue	Cancer staining	Normal tissue staining
Breast cancer			Melanoma		
Carcinoid			Ovarian cancer		
Cervical cancer ★			Pancreatic cancer		
Colorectal cancer			Prostate cancer		
Endometrial cancer			Renal cancer		
Glioma			Skin cancer		
Head and neck cancer			Stomach cancer		
Liver cancer			Testis cancer		
Lung cancer			Thyroid cancer		
Lymphoma			Urothelial cancer		

Staining summary Most malignancies showed weak to moderate cytoplasmic positivity.

(Figure 13-D) IFIH1 protein levels of expression compared between normal and cancer tissues in different cancer types, the level of antibody staining is grading from high (dark blue) to not detected (white)

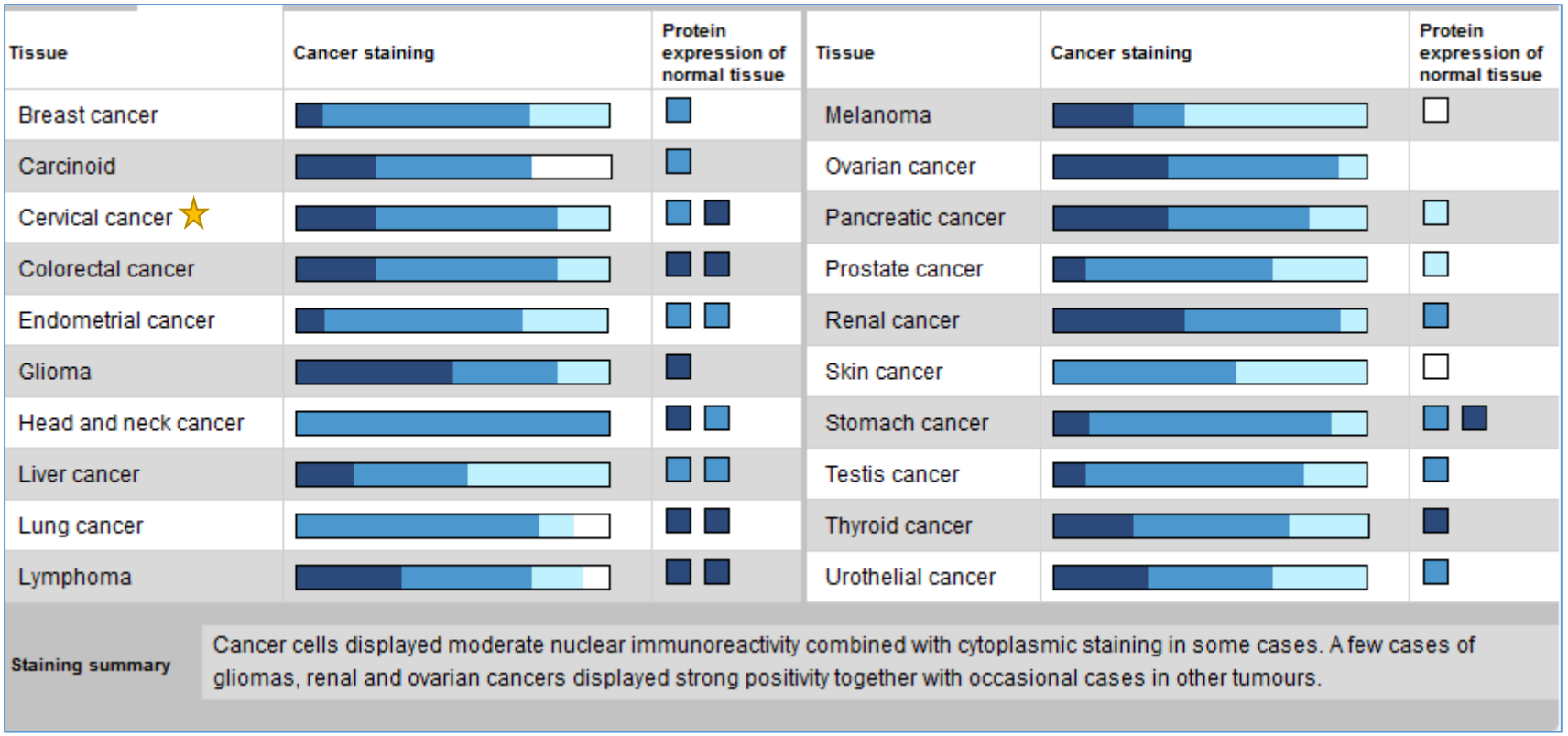
Tissue	Cancer staining	Protein expression of normal tissue	Tissue	Cancer staining	Protein expression of normal tissue
Breast cancer			Melanoma		
Carcinoid			Ovarian cancer		
Cervical cancer ★			Pancreatic cancer		
Colorectal cancer			Prostate cancer		
Endometrial cancer			Renal cancer		
Glioma			Skin cancer		
Head and neck cancer			Stomach cancer		
Liver cancer			Testis cancer		
Lung cancer			Thyroid cancer		
Lymphoma			Urothelial cancer		
Staining summary	A majority of cancer cells showed moderate nuclear or nucleolar positivity.				

(Figure 13-E) IRF1 protein levels of expression compared between normal and cancer tissues in different cancer types, the level of antibody staining is grading from high (dark blue) to not detected (white)

Tissue	Cancer staining	Normal tissue staining	Tissue	Cancer staining	Normal tissue staining
Breast cancer			Melanoma		
Carcinoid			Ovarian cancer		
Cervical cancer ★			Pancreatic cancer		
Colorectal cancer			Prostate cancer		
Endometrial cancer			Renal cancer		
Glioma			Skin cancer		
Head and neck cancer			Stomach cancer		
Liver cancer			Testis cancer		
Lung cancer			Thyroid cancer		
Lymphoma			Urothelial cancer		

Staining summary Non-Hodgkin's lymphomas showed moderate to strong cytoplasmic staining. Inflammatory cells exhibited strong cytoplasmic positivity. Remaining cancer tissues were negative.

(Figure 13-F) IRF5 protein levels of expression compared between normal and cancer tissues in different cancer types, the level of antibody staining is grading from high (dark blue) to not detected (white)



Tissue	Cancer staining	Protein expression of normal tissue	Tissue	Cancer staining	Protein expression of normal tissue
Breast cancer			Melanoma		
Carcinoid			Ovarian cancer		
Cervical cancer ★			Pancreatic cancer		
Colorectal cancer			Prostate cancer		
Endometrial cancer			Renal cancer		
Glioma			Skin cancer		
Head and neck cancer			Stomach cancer		
Liver cancer			Testis cancer		
Lung cancer			Thyroid cancer		
Lymphoma			Urothelial cancer		

Staining summary Cancer cells displayed moderate nuclear immunoreactivity combined with cytoplasmic staining in some cases. A few cases of gliomas, renal and ovarian cancers displayed strong positivity together with occasional cases in other tumours.

(Figure 13-G) IRF7 protein levels of expression compared between normal and cancer tissues in different cancer types, the level of antibody staining is grading from high (dark blue) to not detected (white)

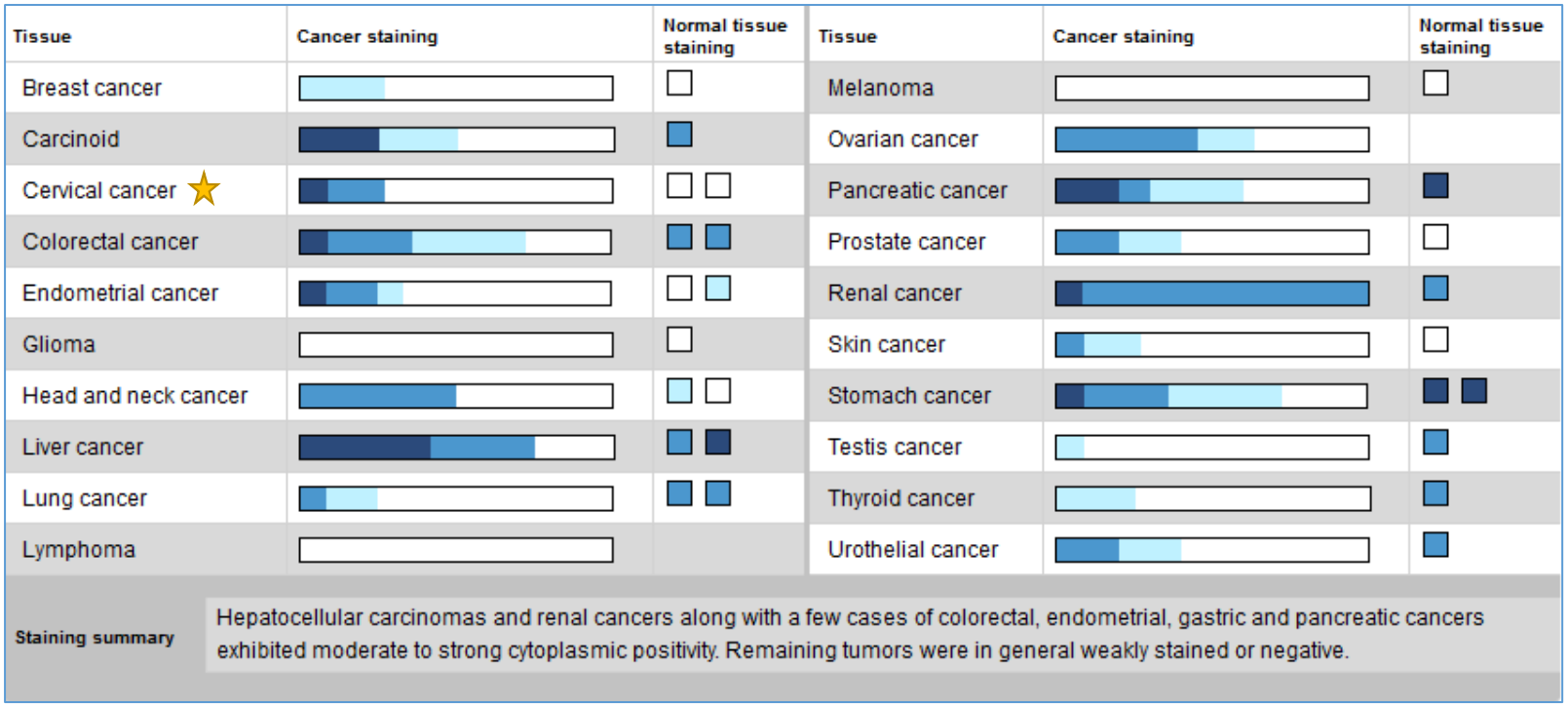
Tissue	Cancer staining	Normal tissue staining	Tissue	Cancer staining	Normal tissue staining
Breast cancer			Melanoma		
Carcinoid			Ovarian cancer		
Cervical cancer ★			Pancreatic cancer		
Colorectal cancer			Prostate cancer		
Endometrial cancer			Renal cancer		
Glioma			Skin cancer		
Head and neck cancer			Stomach cancer		
Liver cancer			Testis cancer		
Lung cancer			Thyroid cancer		
Lymphoma			Urothelial cancer		

Staining summary A majority of malignant tissues displayed differential staining pattern, with negative cases and strongly stained cases within the same tumor type. The positivity was mainly observed in nuclei, but often accompanied with cytoplasmic staining of weaker intensity.

(Figure 13-H) IRF9 protein levels of expression compared between normal and cancer tissues in different cancer types, the level of antibody staining is grading from high (dark blue) to not detected (white)

Tissue	Cancer staining	Protein expression of normal tissue	Tissue	Cancer staining	Protein expression of normal tissue
Breast cancer			Melanoma		
Carcinoid			Ovarian cancer		
Cervical cancer ★			Pancreatic cancer		
Colorectal cancer			Prostate cancer		
Endometrial cancer			Renal cancer		
Glioma			Skin cancer		
Head and neck cancer			Stomach cancer		
Liver cancer			Testis cancer		
Lung cancer			Thyroid cancer		
Lymphoma			Urothelial cancer		
Staining summary	Malignant cells displayed moderate to strong cytoplasmic immunoreactivity. Several cases of malignant lymphomas and renal cancers were weakly stained or negative.				

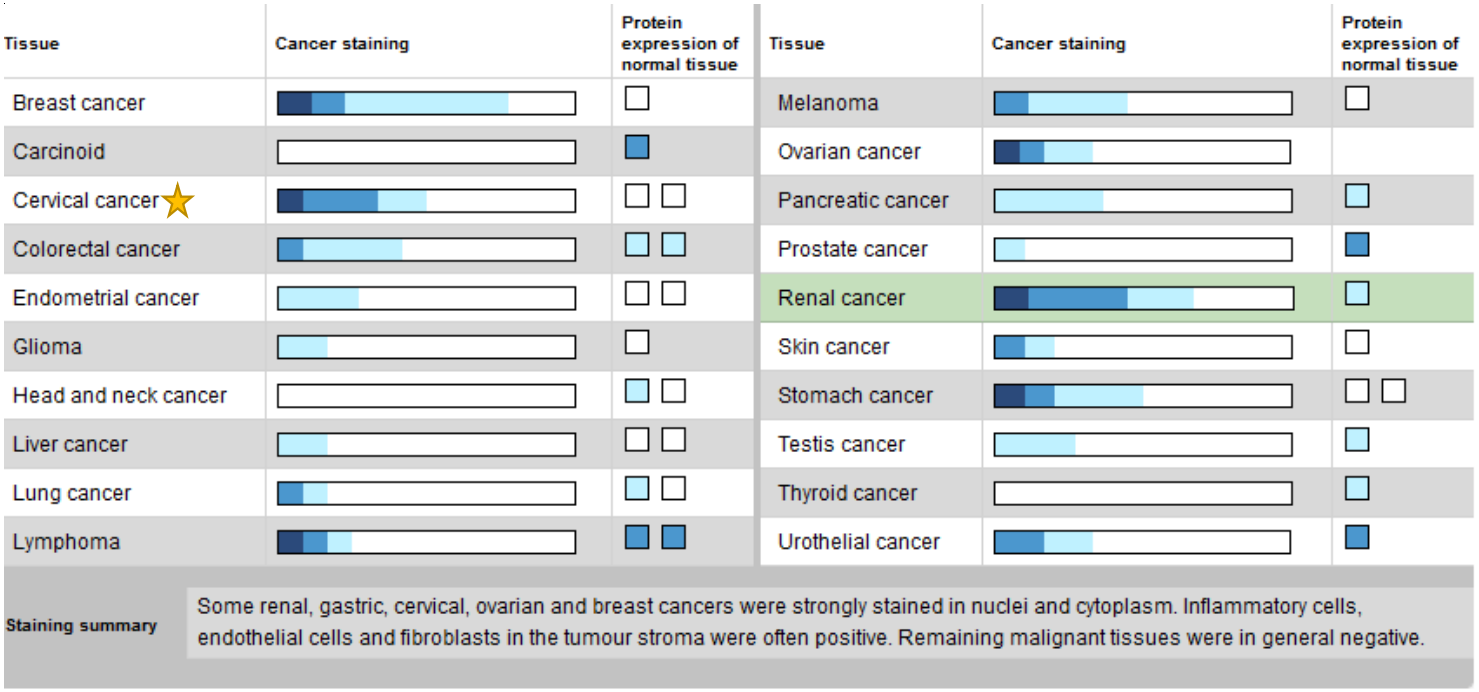
(Figure 13-I) MX1 protein levels of expression compared between normal and cancer tissues in different cancer types, the level of antibody staining is grading from high (dark blue) to not detected (white)



Tissue	Cancer staining	Normal tissue staining	Tissue	Cancer staining	Normal tissue staining
Breast cancer			Melanoma		
Carcinoid			Ovarian cancer		
Cervical cancer			Pancreatic cancer		
Colorectal cancer			Prostate cancer		
Endometrial cancer			Renal cancer		
Glioma			Skin cancer		
Head and neck cancer			Stomach cancer		
Liver cancer			Testis cancer		
Lung cancer			Thyroid cancer		
Lymphoma			Urothelial cancer		

Staining summary Hepatocellular carcinomas and renal cancers along with a few cases of colorectal, endometrial, gastric and pancreatic cancers exhibited moderate to strong cytoplasmic positivity. Remaining tumors were in general weakly stained or negative.

(Figure 13-J) RSAD2 protein levels of expression compared between normal and cancer tissues in different cancer types, the level of antibody staining is grading from high (dark blue) to not detected (white)



Tissue	Cancer staining	Protein expression of normal tissue	Tissue	Cancer staining	Protein expression of normal tissue
Breast cancer			Melanoma		
Carcinoid			Ovarian cancer		
Cervical cancer			Pancreatic cancer		
Colorectal cancer			Prostate cancer		
Endometrial cancer			Renal cancer		
Glioma			Skin cancer		
Head and neck cancer			Stomach cancer		
Liver cancer			Testis cancer		
Lung cancer			Thyroid cancer		
Lymphoma			Urothelial cancer		

Staining summary Some renal, gastric, cervical, ovarian and breast cancers were strongly stained in nuclei and cytoplasm. Inflammatory cells, endothelial cells and fibroblasts in the tumour stroma were often positive. Remaining malignant tissues were in general negative.

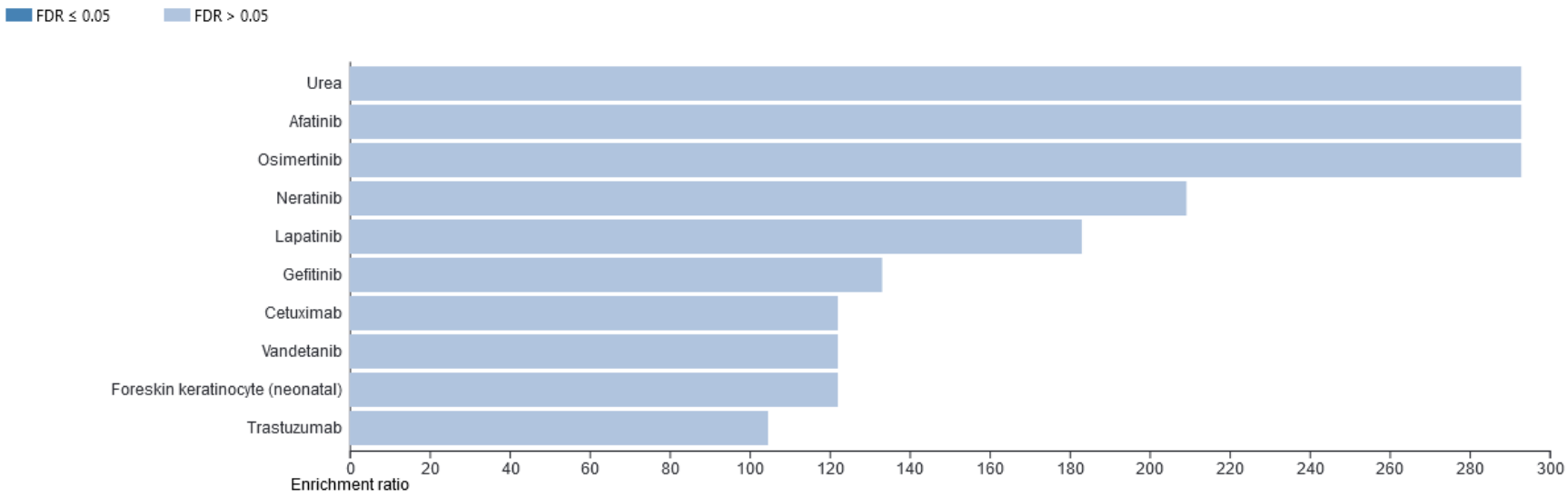
(Figure 13-K) STAT1 protein levels of expression compared between normal and cancer tissues in different cancer types, the level of antibody staining is grading from high (dark blue) to not detected (white)

In the Human Protein Atlas project, antibodies are generated towards all human proteins and used to acquire corresponding protein profiles in both normal human tissues and cancer tissues from different patients, representing the 20 most common forms of human cancer. Immunohistochemically stained TMA sections on glass slides are scanned to create high-resolution images from which pathologists can interpret and annotate the outcome of immunohistochemistry. It is notable that immunohistochemistry is by far the most commonly used application for tissue microarray(TMAs), it can be used in several different settings, e.g. high-throughput screening of protein expression patterns in tissues and cells, biomarker discovery studies based on tissue samples from large patient cohorts and more basic tumor biology studies, including different phenotypes/genotypes, differentiation stages,

progression and metastasis. One advantage of using TMAs is that material from large cohorts can be analyzed at one time, both saving valuable biological material and ensuring more reproducible experiments.
According to the results, it has been realized that STAT1 and RSAD2 are weakly stained or negative in normal tissues, while the expression level is ranged from low, moderate to highly stained in malignant tissues. Those two proteins/genes might abe a novel biomarker for cervical squamous cell carcinoma.
MX1 and DDX58 displayed moderate staining in normal tissues, while malignant tissues showed strong to moderate immunoreactivity.

Computational-Based Drug Repurposing using WebGestalt tool

Functional enrichment analysis plays a critical role in interpreting high-throughput experiment results, which frequently generate a list of interesting genes or proteins. WebGestalt (WEB-based GEne SeT AnaLysis Toolkit) is one of the most widely used gene set enrichment analysis tools that help users extract biological insights from genes of interest (28). WebGestalt supports all three methods available on the WebGestalt server, namely Over Representation Analysis (ORA), Gene Set Enrichment Analysis (GSEA) and Network Topology-based Analysis (NTA). It allows users to locally generate the same results as those from the WebGestalt server and enables batch runs and integration into other pipelines. Pathway and functional enrichment analysis has become the first choice for gaining insight into the underlying biology of differentially expressed genes and proteins, as it reduces complexity and has increased explanatory power.
Analyzing high-throughput molecular measurements at the functional level is very advantageous for two reasons. First, grouping thousands of genes, proteins, and/or other biological molecules by the pathways they are involved in reduces the complexity to just several hundred pathways for the experiment. Second, identifying active pathways that differ between two conditions can have more explanatory power than a simple list of different genes or proteins. Herein, analysis was proceeded in a suggested functional database (ie, Drug Bank) through WebGestalt to predict gene-drug interactions and potential therapeutic options for candidate genes. To predict the main significant drugs for ten hub genes (STAT1, STAT2, IRF1, IRF9, OAS2, IRF7, IFIT5, OAS1, RSAD2, RNASEL), "drug" and "DrugBank" were selected as the functional database and enrichment category, respectively. I considered the analysis as the significance level of P value< 0.05, and over-representation analysis as the method of Interest. WebGestalt gene-drug analysis predicted ten existing drugs (Figure 14)



(Figure 14) Bar graph of the enriched drugs and potential therapeutic options for ten hub genes associated with HPV positive and HPV negative status in Cervical Squamous Cell Carcinoma samples (STAT1, STAT2, IRF1, IRF9, OAS2, IRF7, IFIT5, OAS1, RSAD2, RNASEL) using DrugBank database as a functional database through WebGestalt website

(Table 4) Tabular form the enriched drugs and potential therapeutic options for ten hub genes associated with HPV positive and HPV negative status in Cervical Squamous Cell Carcinoma samples (STAT1, STAT2, IRF1, IRF9, OAS2, IRF7, IFIT5, OAS1, RSAD2, RNASEL) using DrugBank database as a functional database through WebGestalt website

Gene Set	Description	Size	Expect	Ratio	P Value	↑ FDR
DB03904	Urea	5	0.0034141	292.9	0.0034118	0.93237
DB08916	Afatinib	5	0.0034141	292.9	0.0034118	0.93237
DB09330	Osimertinib	5	0.0034141	292.9	0.0034118	0.93237
DB11828	Neratinib	7	0.0047798	209.21	0.0047749	0.93237
DB01259	Lapatinib	8	0.0054626	183.06	0.0054561	0.93237
DB00317	Gefitinib	11	0.0075111	133.14	0.0074983	0.93237
DB00002	Cetuximab	12	0.0081939	122.04	0.0081785	0.93237
DB05294	Vandetanib	12	0.0081939	122.04	0.0081785	0.93237
DB10772	Foreskin keratinocyte (neonatal)	12	0.0081939	122.04	0.0081785	0.93237
DB00072	Trastuzumab	14	0.0095596	104.61	0.0095384	0.93237

- Urea is a keratolytic emollient that works to treat or prevent dry, rough, scaly, itchy skin.
- Afatinib (*Gilotrif*) is an antineoplastic agent used for the treatment of locally advanced or metastatic non-small cell lung cancer (NSCLC) with non-resistant EGFR mutations or resistance to platinum-based chemotherapy.
- Osimertinib (*Tagrisso*) is a tyrosine kinase inhibitor used in the treatment of certain types of non-small cell lung carcinoma.
- Neratinib (*Nerlynx*) is a protein kinase inhibitor used to treat breast cancer that over expresses the HER2 receptor.
- Lapatinib (*Tykerb, Tyverb*) is an antineoplastic agent and tyrosine kinase inhibitor used for the treatment of advanced or metastatic HER-positive breast cancer in patients who received prior chemotherapeutic treatments.

- Gefitinib (*Iressa*) is a tyrosine kinase inhibitor used as first-line therapy to treat non-small cell lung carcinoma (NSCLC) that meets certain genetic mutation criteria. It is used in the treatment of certain types of cancer. gefitinib selectively targets the mutant proteins in malignant cells.
- Cetuximab (*Erbitux*) is an endothelial growth factor receptor binding fragment used to treat colorectal cancer as well as squamous cell carcinoma of the head and neck.
- Vandetanib (*Caprelsa*) is an oral once-daily kinase inhibitor of tumour angiogenesis and tumour cell proliferation with the potential for use in a broad range of tumour types. Vandetanib is an antineoplastic kinase inhibitor used to treat symptomatic or progressive medullary thyroid cancer in patients with unresectable locally advanced or metastatic disease.
- Foreskin fibroblast(neonatal) (*Gintuit*) is a treatment of cells used to treat full-thickness, diabetic foot ulcers that extend through the dermis without tendon, muscle, joint capsule, or bone exposure. Foreskin fibroblast-like stromal cells (FDSCs) are progenitors isolated from human tissue that can differentiate into various cell types (12).
- Trastuzumab (*Herceptin*) is a monoclonal anti-human epidermal growth factor receptor 2 protein antibody used to treat HER2-positive breast, gastroesophageal, and gastric cancers.

Pathway-based Rational Drug Repositioning using Gene2drug tool

The availability of large collections of transcriptional responses to drugs enables computational approaches to drug repositioning directly based on measured molecular effects. A novel computational methodology for rational drug repositioning has been developed as an effective shortcut to drug discovery, which exploits the transcriptional responses following treatment with small molecule. Specifically, given a therapeutic target gene, a prioritization of potential effective drugs is obtained by assessing their impact on the transcription of genes in the pathway(s) including the target. Among the available approaches, rational methods based on a specific molecular target can take advantage of the current availability of genome-wide transcriptional data. Unfortunately, the effects of direct drug targets are usually not detectable at the mRNA level and a more systematic approach is required to link transcriptional data to therapeutic effects. Gene2drug (https://gene2drug.tigem.it) tackles this problem by searching for sets of pathways that are dysregulated by a drug as opposed to single genes. One way to define a set of pathways is to start from a target gene and collect the pathways that it is annotated to, thus assessing the effects of perturbing the target gene indirectly through cellular mechanisms that are expected to be involved. Gene2drug uses a method we called "Pathway-set Enrichment Analysis" (PSEA), analogous to the Gene-set Enrichment Analysis (GSEA) (27). Gene expression profiles from the

Connectivity Map are converted to "pathway expression profiles" and ranked according to the p-value of their Kolmogorov-Smirnov (KS) statistic. Given a set of pathways, the KS statistic is applied again to search for drugs that consistently up-regulate or dysregulate most pathways in the set. Top drugs are those most dysregulating the pathways in the input set. Enrichment score signs report whether the regulation is "up" or "down". The p-value reports how it is unlikely for a random set of ranks to be as extreme (top or bottom) as the ranks of the chosen pathways for each drug. I applied the Pathway-set Enrichment analysis to four genes that showed differences in immunostaining between normal and tumor tissue conditions in the previous Human Protein Atlas analysis step: STAT1, RSAD2, MX1, DDX58.

Results

Gene2drug ranks small molecules according to their ability to dysregulate an input set of pathways. Sets of pathways were defined starting from a gene and exploiting its pathway annotations from a number of publicly available databases.

STAT1

COMPUND NAME	ESCORE	PVALUE
ETOPOSIDE	0.98	1.31e-5
PRAMOCAINE	0.94	4.02e-4

- Etoposide (*Etopophos*) is an antineoplastic agent and an epipodophyllotoxin (a semisynthetic derivative of the podophyllotoxins). It inhibits DNA topoisomerase II, thereby ultimately inhibiting DNA synthesis. Etoposide is cell cycle dependent and phase specific, affecting mainly the S and G2 phases.
- Procaine (*Novocain*) is a topical anesthetic and antipruritic. It is used for many dermatological and anorectal/anogenital conditions including minor cuts/burns, insect bites, hives/rashes due to poison ivy exposure, hemorrhoids and other anorectal/anogenital disorders. Pramocaine is available by itself and in combination with other medications in various topical preparations. Recently, it has been showed that local anesthetics can directly or indirectly affect the progression of tumors and significantly decrease the proliferation of cancers (16).

Pathway set:

- defense response to virus (GO-BP)
- modulation by virus of host morphology or physiology (GO-BP)
- response to virus (GO-BP)

COMPOUND NAME	ESCORE	PVALUE
GABEXATE	1	9.81e-3
PROXYPHYLLINE	0.99	1.47e-2

- Gabexate (*Gabexate*)is a synthetic serine protease inhibitor which has been used as an anticoagulant. It also known to decrease production of inflammatory cytokines. Gabexate has been investigated for use in cancer, ischemia-reperfusion injury, and pancreatitis.

- Proxyphylline is a xanthine derivative that acts as a cardiac stimulant, vasodilator and bronchodilator. Additionally, proxyphylline derivatives displayed moderate antineoplastic activity against human breast adenocarcinoma cell line and high against peripheral blood T lymphoblast and can be employed to prevent cancer-associated biofilm Candida infections (13).

Pathway set:

- Tumorigenesis by ERBB2 CDC25A, downregulation (CGP)

MX1

COMPOUND NAME	ESCORE	PVALUE
OXYTETRACYCLINE	0.76	1.96e-3
LABETALOL	0.75	2.29e-3

- Oxytetracycline (*Terramycin*) is a tetracycline antibiotic used to treat a wide variety of susceptible bacterial infections. Studies suggested that_oxytetracycline could be an effective choice for combining with carboplatin, oxytetracycline improved synergistically with carboplatin in inducing apoptosis (17).

- Labetalol (*Trandate*) is a racemic mixture of 2 diastereoisomers where dilevalol, the R,R' stereoisomer to treat hypertension. Because of its beta-blocking and local anesthetic effect, labetalol via peridural catheter was supposed to reduce pain in patients suffering from gynecologic cancers (14).
  According to a study that analyzed data on outcomes of cancer patients treated with beta-blocker drugs from the Food and Drug Administration Adverse Event Reporting System. They found that mortality in patients treated with a beta-blocker fell by an average of 17% across all major cancer types, with a nearly 15% decrease in mortality among patients with ovarian and cervical cancers (15).

Pathway set:

- Systemic Lupus Erythematosus (CGP)
- Apoptosis Reversed by Il6 (CGP)
- Silenced During Tumor Angiogenesis (CGP)
- Antiviral Response to Ribavirin, upregulation (CGP)

RSAD2

COMPUND NAME	ESCORE	PVALUE
ETOPOSIDE	0.97	4.35e-5
4,5-DIANILINOPHTHALIMIDE	0.96	9.97e-5

- Etoposide A semisynthetic derivative of podophyllotoxin that exhibits antitumor activity. Etoposide inhibits DNA synthesis by forming a complex with topoisomerase II and DNA. This complex induces breaks in double stranded DNA and prevents repair by topoisomerase II binding. Accumulated breaks in DNA prevent entry into the mitotic phase of cell division, and lead to cell death. Etoposide acts primarily in the G2 and S phases of the cell cycle.

- DAPH (4,5-dianilinophthalimide) is an inhibitor of human Epithelial growth factor. It was originally developed by scientists at CIBA-Geigy Limited as a potential anti-tumor compound, although certain early in vivo studies have been retracted. More recently DAPH-1 has found interest as an anti-amyloid compound.

Pathway set:

- modulation by virus of host morphology or physiology (GO-BP)
- positive regulation of type I interferon production (GO-BP)
- response to virus (GO-BP)

COMPOUND NAME	ESCORE	PVALUE
OXYBUTYNIN	0.67	2.68e-4

- Oxybutynin *(Ditropan)* is an antimuscarinic agent that reduces detrusor muscle activity, relaxing the bladder and preventing the urge to void.

Pathway set:

- Response to Tamoxifen (CGP)
- Apoptosis by Reovirus Infection (CGP)
- Tumor Evasion and Tolerogenicity Up (CGP)
- Antiviral Response to Ribavirin Up (CGP)

DDX58

COMPUOND NAME	ESCORE	PVALUE
TOLBUTAMIDE	1	2.45e-3

- Tolbutamide is an oral antihyperglycemic agent used for the treatment of non-insulin-dependent diabetes mellitus (NIDDM). It belongs to the sulfonylurea class of insulin secretagogues, which act by stimulating β cells of the pancreas to release insulin.

Pathway set:

- Response to virus (GO-BP)

(Table 5) This table demonstrates the references of the top ranked therapeutic options, resulted from drug-gene interaction analyses of Squamous Cervical Cancer related hub genes according to Webgestalt and Gene2drug websites

DRUG	URLs
1. Urea	https://go.drugbank.com/drugs/DB03904
2. Afatinib	https://go.drugbank.com/drugs/DB08916
3. Osimertinib	https://go.drugbank.com/drugs/DB09330
4. Lapatinib	https://go.drugbank.com/drugs/DB01259
5. Gefitinib	https://go.drugbank.com/drugs/DB00317
6. Cetuximab	https://go.drugbank.com/drugs/DB00002
7. Vandetanib	https://go.drugbank.com/drugs/DB05294
8. Foreskin fibroblast (neonatal)	https://go.drugbank.com/drugs/DB10770
9. Trastuzumab	https://go.drugbank.com/drugs/DB00072
10. Etoposide	https://go.drugbank.com/drugs/DB00773
11. Pramocaine	https://go.drugbank.com/drugs/DB09345
12. Procaine	https://go.drugbank.com/drugs/DB00721
13. Methylxanthines	https://www.drugs.com/international/proxyphylline.html
14. Oxytetracycline	https://go.drugbank.com/drugs/DB00595
15. Labetalol	https://go.drugbank.com/drugs/DB00598
16. Dianilinophthalimide (DAPH-1)	https://drugs.ncats.io/drug/Z13D008FZ2
17. Tolbutamide	https://go.drugbank.com/drugs/DB01124

Discussion

Squamous cell carcinoma is an invasive epithelial tumor composed of neoplastic cells with varying degrees of squamous differentiation and the most common type of cervical carcinoma. In 2020, an estimated 604,000 women were diagnosed with cervical cancer worldwide and about 342,000 women died from the disease. The main cause of cervical cancer is persistent infection with high-risk types of human papillomavirus (HPV). High incidence rates and high mortality rates of cervical cancer occur mainly (~90% for both) in low-and middle-income countries (18). To improve the quality of life, prognosis of patients and prolong their survival time, researchers must further clarify the molecular mechanism leading to malignant biological behavior of cervical squamous cell carcinoma and identify key genes that affect the development of this disease. Therefore, bioinformatics analysis of dysregulated genes in clinical samples of cervical cancer may pave the way for development of better prognostic markers and therapeutic targets.

In the present study, a series of analyses were conducted using R2 software to explore downregulated and upregulated genes associated with HPV status in squamous cervical carcinoma. The normalized RNA-seq data samples of patients with Cervical Squamous Cell Carcinoma were derived from TCGA database. Some genes were recognized as HPV positive or HPV negative-specific markers. For instance, the expressions of STAT1, IRF7, DDX58, MX1, BRCA1 and FOS were related to HPV positive according to R2 results. In a previous study, it was confirmed that the expression of antiviral genes (IFIT1 and MX1), genes involved in IFN signaling (STAT1), and proapoptotic genes (TRAIL and XAF1), are inhibited to similar extents by HPV16, -18, and -31(31). In other hand, R2 results indicated that the expression levels of BMP4, CTNNB1, NKD1, WNT11 and STK36 were related to HPV negative status. Furthermore, studies showed that Wnt/CTNNB1 pathway is constitutively activated in Squamous Cell Carcinoma (SCC) cell lines compared to normal keratinocytes(32).

The up- and downregulated DEGs were classified into three groups (biological processes, molecular functions, and cellular components) according to Gene Ontology(GO) terms, and KEGG pathway enrichment analysis was conducted using the DAVID website. The differentially expressed genes were mainly enriched in the nucleus and cytoplasm, mainly regulating the defense response to virus and regulation of transcription from RNA polymerase II promoter biological processes, protein binding and RNA polymerase II core promoter proximal region sequence-specific DNA binding. Pathway analysis showed that differentially expressed genes in cervical carcinoma are mainly involved in the Herpes simplex virus 1 and Epstein-Barr virus infection and pathways in cancer. Researches clarified that HPV-induced cellular changes facilitate the establishment of a latent EBV infection, such latent infections would allow for long-term expression of EBV oncogenes and EBV-induced epigenetic reprogramming that contribute to the progression of HPV-positive oropharyngeal squamous cell carcinoma(33). Additionally, increasing evidence indicates that dysregulation of Wnt signaling pathway cascade is contributed to the development and progression of some solid tumors and

hematological malignancies (34) (35) (36). Hence, the genes involved in these pathways might provide a new direction of research on the original basis. Using integrated bioinformatics including CytoHubba plug-in, UALCAN online analysis and Kaplan-Meier curves survival analysis, I found that the expression of eleven genes in Squamous Cervical Cancer tissues was significantly higher than that in normal tissues and the effects of the hub genes is evaluated on the prognosis of squamous cervical cancer patients, among which only one was upregulated in the HPV negative status (CTNNB1), while the others were upregulated in the HPV positive status (STAT1, IRF9, EGFR, RSAD2, IRF7, MX1, IFIH1, IRF5, IRF1 and DDX58). Interestingly, previous studies revealed that the genetic variants in the CTNNB1 gene might contribute to the development of cervical cancer and it might be a therapeutic target in this disease (37) (38). Also cBioPortal tool was used to study the genomic changes of the key genes in patients with SCC cancer from TCGA database. The eleven key DEGs all showed a high mutation rate in SCC, with a rate of genome change ranging from 4% to 7%.

The description of each hub gene and the expression level of each protein in different tissues were obtained using the Cancer Atlas part from Human Protein Atlas portal tool. It has been realized that four proteins were differently stained in normal and tumor tissues: STAT1, RSAD2, MX1 and DDX58. STAT1 and RSAD2 are weakly stained or negative in normal tissues, while the expression level is ranged from low, moderate to highly stained in malignant tissues. One study used bioinformatics analysis to identify latent biomarkers in connection with progression and prognosis in oral cancer(OC), showed that RSAD2 is one of the independent prognostic indicators of OC (39). Another study showed that high GLUT4 RNA expression in combination with low DDX58 RNA expression levels was significantly correlated with the worst head and neck squamous cell carcinoma patient survival(40). STAT1 has been identified to have prognostic value in patients with solid cancer(41), and a major transcriptional target of human papillomavirus type 31(42).Therefore, those four proteins/genes might be a novel biomarkers and therapeutic targets for the precise treatment of cervical squamous cell carcinoma associated with HPV infection.

Functional enrichment analysis was conducted using WebGestalt server through a suggested functional database (DrugBank) to predict gene-drug interactions and potential therapeutic options for ten ranked key genes that were generated by CytoHubba with their scores (STAT1, STAT2, IRF1, IRF9, OAS2, IRF7, IFIT5, OAS1, RSAD2, RNASEL). WebGestalt gene-drug analysis predicted ten existing drugs that were known to treat several types of cancer like lung cancer and breast cancer. It is worth noting that most of them are tyrosine kinase inhibitor like Afatinib [Giotrif®], which is an orally administered irreversible inhibitor of the ErbB family of tyrosine kinases that provides an important first-line treatment option for advanced non-small cell lung cancer (NSCLC) with activating epidermal growth factor receptor (EGFR) mutations, and an additional treatment option for squamous NSCLC that has progressed following first-line platinum-based chemotherapy(43).

Pathway-based Rational Drug Repositioning analysis was done by Gene2drug online tool for STAT1, RSAD2, MX1 and DDX58.

Gene2drug ranks small molecules according to their ability to dysregulate an input set of pathways related to the aforementioned four genes. Sets of pathways were defined starting from a gene and exploiting its pathway annotations from a number of publicly available databases. The results revealed a number of drugs that may be involved in the future investigation of this disease. Etoposide is one drug that showed high enrichment score according to Gene2drug analysis. Consistent with that outcome, previous study used "big data" derived from patients with head and neck cancer and bioinformatics techniques have also suggested that etoposide in combination with cisplatin, can be used in a regimen of definitive concurrent chemoradiotherapy instead of cisplatin alone regimen for HPV-positive oral and oropharyngeal cancer patients(44).

Conclusion

Taken together, through bioinformatics and computational methods, it has been found that eleven differentially expressed genes (STAT1, CTNNB1 IRF9, EGFR, RSAD2, IRF7, MX1, IFIH1, IRF5, IRF1 and DDX58) were significantly enriched in several different biological processes and certain vital pathways. Besides, they may be key factors in the occurrence and prognosis of Squamous Cervical Cancer and participate in many pathways related to tumor development. Therefore, on this basis, further studies should be performed to detect the polymorphic sites of these genes and explore their corresponding expression levels, which can be used to predict the prognosis of patients. Taken together, the results obtained using immunohistochemistry analysis and Pathway-based Rational Drug Repositioning tools, led to the identification of four hub genes (STAT1, DDX58, MX1, RSAD2), with drugs that consistently up-regulate or dysregulate most pathways in the set for each gene.

These findings will need to be verified in experimental and clinical studies to determine their accuracy and sensitivity in tumorigenesis and to guide the individualized treatment of patients. However, the focus of this study is to provide new ideas for clinical diagnosis and prognostic evaluation through bioinformatics analysis. The results provide an important bioinformatics basis and related theoretical basis for guiding follow-up research on cervical cancer.

References

1. Alessandro D. Santin, Fenghuang Zhan, Eliana Bignotti, Eric R. Siegel, Stefania Cané, Stefania Bellone, Michela Palmieri, Simone Anfossi, Maria Thomas, Alexander Burnett, Helen H. Kay, Juan J. Roman, Timothy J. O'Brien, Erming Tian, Martin J. Cannon, John Shaughnessy, Sergio Pecorelli, Gene expression profiles of primary HPV16- and HPV18-infected early stage cervical cancers and normal cervical epithelium: identification of novel candidate molecular markers for cervical cancer diagnosis and therapy, Virology, Volume 331, Issue 2,2005,Pages 269-291,ISSN 0042-6822,https://doi.org/10.1016/j.virol.2004.09.045. [sciencedirect]

2. AskMayoExpert. Cervical cancer (adult). Rochester, Minn.: Mayo Foundation for Medical Education and Research; 2019. [mayoclinic]

3. Okunade KS. Human papillomavirus and cervical cancer. J Obstet Gynaecol. 2020 Jul;40(5):602-608. doi: 10.1080/01443615.2019.1634030. Epub 2019 Sep 10. Erratum in: J Obstet Gynaecol. 2020 May;40(4):590. PMID: 31500479; PMCID: PMC7062568. [ncbi]

4. Walboomers JM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV, Snijders PJ, Peto J, Meijer CJ, Muñoz N. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J Pathol. 1999 Sep;189(1):12-9. doi: 10.1002/(SICI)1096-9896(199909)189:1<12::AID-PATH431>3.0.CO;2-F. PMID: 10451482. [ncbi]

5. Bosch FX, Lorincz A, Muñoz N, Meijer CJ, Shah KV. The causal relation between human papillomavirus and cervical cancer. J Clin Pathol. 2002 Apr;55(4):244-65. doi: 10.1136/jcp.55.4.244. PMID: 11919208; PMCID: PMC1769629. [ncbi]

6. Fabio Landoni, Andrea Maneo, Alessandro Colombo, Franco Placa, Rodolfo Milani, Patrizia Perego, Giorgio Favini, Luigi Ferri, Costantino Mangioni, Randomised study of radical surgery versus radiotherapy for stage Ib-IIa cervical cancer,The Lancet,Volume 350, Issue 9077,1997,Pages 535-540,ISSN 0140-6736,https://doi.org/10.1016/S0140-6736(97)02250-2(https://www.sciencedirect.com/science/article/pii/S0140673697022502) [sciencedirect]

7. Turashvili G. Squamous cell carcinoma and variants. PathologyOutlines.com website. https://www.pathologyoutlines.com/topic/cervixscc.html. Accessed January 4th, 2023.[pathologyoutlines]

8. Stanley MA. Epithelial cell responses to infection with human papillomavirus. Clin Microbiol Rev. 2012 Apr;25(2):215-22. doi: 10.1128/CMR.05028-11. PMID: 22491770; PMCID: PMC3346303.pithelial Cell Responses to Infection with Human Papillomavirus [pmc]

9. Kumar S, Shah JP, Bryant CS, Imudia AN, Ali-Fehmi R, Malone JM Jr, Morris RT. Prognostic significance of keratinization in squamous cell cancer of uterine cervix: a population based study. Arch Gynecol Obstet. 2009 Jul;280(1):25-32. doi: 10.1007/s00404-008-0851-9. Epub 2008 Nov 29. PMID: 19043726.[pubmed]

10. Broccolo, F., Ciccarese, G., Rossi, A. et al. Human papillomavirus (HPV) and Epstein-Barr virus (EBV) in keratinizing versus non- keratinizing squamous cell carcinoma of the oropharynx. Infect Agents Cancer 13, 32 (2018). https://doi.org/10.1186/s13027-018-0205-6[biomedcentral]

11. Chernock RD. Morphologic features of conventional squamous cell carcinoma of the oropharynx: 'keratinizing' and 'nonkeratinizing' histologic types as the basis for a consistent classification system. Head Neck Pathol. 2012 Jul;6 Suppl 1(Suppl 1):S41-7. doi: 10.1007/s12105-012-0373-4. Epub 2012 Jul 3. PMID: 22782222; PMCID: PMC3394167. [pmc]

12. Huang HI, Chen SK, Wang RY, Shen CR, Cheng YC. Human foreskin fibroblast-like stromal cells can differentiate into functional hepatocytic cells. Cell Biol Int. 2013 Dec;37(12):1308-19. doi: 10.1002/cbin.10175. Epub 2013 Sep 16. PMID: 23956153. [pubmed]

13. Borowiecki P, Wińska P, Bretner M, Gizińska M, Koronkiewicz M, Staniszewska M. Synthesis of novel proxyphylline derivatives with dual Anti-Candida albicans and anticancer activity. Eur J Med Chem. 2018 Apr 25;150:307-333. doi: 10.1016/j.ejmech.2018.02.077. Epub 2018 Feb 27. PMID: 29533875. [pubmed]

14. Margaria E, Gagliardi M, Palieri L, Treves S, Fanzago E. Analgesic effect of peridural labetalol in the treatment of cancer pain. Int J Clin Pharmacol Ther Toxicol. 1983 Jan;21(1):47-50. PMID: 6832867. [pubmed]

15. Scott Merville, MD Anderson Cancer Center, Betting on beta-blockers, 2014. [mdanderson.org]

16. AUTHOR=Zhang Yiguo, Jing Yixin, Pan Rui, Ding Ke, Chen Rong, Meng Qingtao TITLE=Mechanisms of Cancer Inhibition by Local Anesthetics JOURNAL=Frontiers in Pharmacology VOLUME=12 YEAR=2021 URL=https://www.frontiersin.org/articles/10.3389/fphar.2021.770694 DOI=10.3389/fphar.2021.770694 [frontiersin]

17. Jabber EJ, Alzayd AM, Jawad MJ, Sulbi IM, Hassan SM, Jawad MJ, Jasim AM. Synergistic effect of oxytetracycline as a combination treatment with Carboplatin on MCF-7 breast cancer cell line. Braz J Vet Res Anim Sci. 2022;59:e191527. https://doi.org/10.11606/issn.1678-4456.bjvras.2022.191527 [revistas]

18.. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021; 71:209- 249. [pubmed]

19. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013 Apr 2;6(269):pl1. doi: 10.1126/scisignal.2004088. PMID: 23550210; PMCID: PMC4160307. [pubmed]

20. Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, Imamichi T, Chang W. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Res. 2022 Mar 23;50(W1):W216–21. doi: 10.1093/nar/gkac194. Epub ahead of print. PMID: 35325185; PMCID: PMC9252805.[pubmed]

21. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44-57. doi: 10.1038/nprot.2008.211. PMID: 19131956 [pubmed]

22. Chandrashekar DS, Karthikeyan SK, Korla PK, Patel H, Shovon AR, Athar M, Netto GJ, Qin ZS, Kumar S, Manne U, Creighton CJ, Varambally S. UALCAN: An update to the integrated cancer data analysis platform. Neoplasia. 2022 Mar;25:18-27. doi: 10.1016/j.neo.2022.01.001. Epub 2022 Jan 22. PMID: 35078134; PMCID: PMC8788199. [pubmed]

23. Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi BVSK, Varambally S. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. Neoplasia. 2017 Aug;19(8):649-658. doi: 10.1016/j.neo.2017.05.002. Epub 2017 Jul 18. PMID: 28732212; PMCID: PMC5516091. [pubmed]

24. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering CV. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019 Jan 8;47(D1):D607-D613. doi: 10.1093/nar/gky1131. PMID: 30476243; PMCID: PMC6323986. [pubmed]

25. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering CV. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019 Jan 8;47(D1):D607-D613. doi: 10.1093/nar/gky1131. PMID: 30476243; PMCID: PMC6323986. [pubmed]

26. Mathias Uhlén, Erik Björling, Charlotta Agaton, Cristina Al-Khalili Szigyarto, Bahram Amini, Elisabet Andersen, Ann-Catrin Andersson, Pia Angelidou, Anna Asplund, Caroline Asplund, Lisa Berglund, A Human Protein Atlas for Normal and Cancer Tissues Based on Antibody Proteomics, Molecular & Cellular Proteomics,Volume 4, Issue 12, 2005,Pages 1920 1932,ISSN 1535-9476,https://doi.org/10.1074/mcp.M500279-MCP200. [mcponline]

27. Francesco Napolitano, Diego Carrella, Barbara Mandriani, Sandra Pisonero-Vaquero, Francesco Sirci, Diego L Medina, Nicola Brunetti-Pierri, Diego di Bernardo, gene2drug: a computational tool for pathway-based rational drug repositioning, Bioinformatics, Volume 34, Issue 9, 01 May 2018, Pages 1498–1505, https://doi.org/10.1093/bioinformatics/btx800 [academic.oup.com/bioinformatics]

28. Yuxing Liao, Jing Wang, Eric J Jaehnig, Zhiao Shi, Bing Zhang, WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs, Nucleic Acids Research, Volume 47, Issue W1, 02 July 2019, Pages W199–W205, https://doi.org/10.1093/nar/gkz401 [academic.oup.com/nar]

29. Lánczky A, Győrffy B. Web-Based Survival Analysis Tool Tailored for Medical Research (KMplot): Development and Implementation. J Med Internet Res. 2021 Jul 26;23(7):e27633. doi: 10.2196/27633. PMID: 34309564; PMCID: PMC8367126.[pubmed]

30. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun. 2019 Apr 3;10(1):1523. doi: 10.1038/s41467-019-09234-6. PMID: 30944313; PMCID: PMC6447622.[pubmed]

31. Reiser J, Hurst J, Voges M, Krauss P, Münch P, Iftner T, Stubenrauch F. High-risk human papillomaviruses repress constitutive kappa interferon transcription via E6 to prevent pathogen recognition receptor and antiviral-gene expression. J Virol. 2011 Nov;85(21):11372-80. doi: 10.1128\JVI.05279-11. Epub 2011 Aug 17. PMID: 21849431; PMCID: PMC3194958.[pubmed]

32. Zanfi, E.D.; Fantini, S.; Lotti, R.; Bertesi, M.; Marconi, A.; Grande, A.; Manfredini, R.; Pincelli, C.; Zanocco-Marani, T. Wnt/CTNNB1 Signal Transduction Pathway Inhibits the Expression of ZFP36 in Squamous Cell Carcinoma, by Inducing Transcriptional Repressors SNAI1, SLUG and TWIST. Int. J. Mol. Sci. 2020, 21, 5692. [pubmed]

33. Guidry JT, Myers JE, Bienkowska-Haba M, Songock WK, Ma X, Shi M, Nathan CO, Bodily JM, Sapp MJ, Scott RS. Inhibition of Epstein-Barr Virus Replication in Human Papillomavirus-Immortalized Keratinocytes. J Virol. 2019 Jan 4;93(2):e01216-18. doi: 10.1128\JVI.01216-18. PMID: 30381489; PMCID: PMC6321917. [pubmed]

34. Zhu H, Mazor M, Kawano Y, Walker MM, Leung HY, Armstrong K, Waxman J, Kypta RM. Analysis of Wnt gene expression in prostate cancer: mutual inhibition by WNT11 and the androgen receptor. Cancer Res. 2004 Nov 1;64(21):7918-26. doi: 10.1158\0008-5472.CAN-04-2704. Erratum in: Cancer Res. 2005 Sep 1;65(17):8057. PMID: 15520198. [pubmed]

35. Zhang M, Weng W, Zhang Q, Wu Y, Ni S, Tan C, Xu M, Sun H, Liu C, Wei P, et al. The lncRNA NEAT1 activates Wnt\β-catenin signaling and promotes colorectal cancer progression via interacting with DDX5. J Hematol Oncol. 2018;11(1):113 [biomedcentral]

36. GE, X., Wang, X. Role of Wnt canonical pathway in hematological malignancies. J Hematol Oncol 3, 33 (2010). https:\doi.org\10.1186\1756-8722-3-33 [biomedcentral]

37. Wang, B., Li, X., Liu, L. et al. β-Catenin: oncogenic role and therapeutic target in cervical cancer. Biol Res 53, 33 (2020). https:\doi.org\10.1186\s40659-020-00301-7 [biomedcentral]

38. Wang, B., Li, X., Liu, L. et al. β-Catenin: oncogenic role and therapeutic target in cervical cancer. Biol Res 53, 33 (2020). https:\doi.org\10.1186\s40659-020-00301-7[biomedcentral]

39. Reyimu, A., Chen, Y., Song, X. et al. Identification of latent biomarkers in connection with progression and prognosis in oral cancer by comprehensive bioinformatics analysis. World J Surg Onc 19, 240 (2021). https:\doi.org\10.1186\s12957-021-02360-w [biomedcentral]

40. Chang, YC., Chi, LH., Chang, WM. et al. Glucose transporter 4 promotes head and neck squamous cell carcinoma metastasis through the TRIM24-DDX58 axis. J Hematol Oncol 10, 11 (2017). https:\doi.org\10.1186\s13045-016-0372-0 [biomedcentral]

41. Zhang J, Wang F, Liu F, Xu G. Predicting STAT1 as a prognostic marker in patients with solid cancer. Ther Adv Med Oncol. 2020 May 8;12:1758835920917558. doi: 10.1177\1758835920917558. PMID: 32426049; PMCID: PMC7222261[journals]

42. Chang YE, Laimins LA. Microarray analysis identifies interferon-inducible genes and Stat-1 as major transcriptional targets of human papillomavirus type 31. J Virol. 2000 May;74(9):4174-82. doi: 10.1128\jvi.74.9.4174-4182.2000. PMID: 10756030; PMCID: PMC111932.[pubmed]

43. Deeks ED, Keating GM. Afatinib in advanced NSCLC: a profile of its use. Drugs Ther Perspect. 2018;34(3):89-98. doi: 10.1007\s40267-018-0482-6. Epub 2018 Feb 1. Erratum in: Drugs Ther Perspect. 2018;34(4):196. PMID: 29540977; PMCID: PMC5840214.[pubmed]

44. Kim YR, Lee B, Byun MR, Lee JK, Choi JW (2018) Evaluation of pemetrexed and etoposide as therapeutic regimens for human papillomavirus-positive oral and oropharyngeal cancer. PLOS ONE 13(7): e0200509. https:\doi.org\10.1371\journal.pone.0200509[journals]