Gynecologic Oncology xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

Gynecologic Oncology

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

Integrated analysis of HPV-mediated immune alterations in cervical cancer

Long Chen [a,*], Shaohong Luan [a], Baoguo Xia [a], Yansheng Liu [a], Yuan Gao [a], Hongyan Yu [a], Qingling Mu [a], Ping Zhang [a], Weina Zhang [a], Shengmiao Zhang [a], Guopeng Wei [b], Min Yang [b], Ke Li [b]

[a] Department of Gynecology, Qingdao Municipal Hospital, Qingdao 266011, PR China
[b] Gezhi Research Lab, Building T1, No.722 Yizhou Avenue, Chengdu 610000, PR China

HIGHLIGHTS

- Immune pathway is activated in HPV+ cervical patients compared with the HPV−.
- 9 genes were identified changed in HPV+ cervical cancer.
- The 9 genes are associated with early relapse of cervical cancer.

ARTICLE INFO

Article history:
Received 18 November 2017
Received in revised form 25 January 2018
Accepted 28 January 2018
Available online xxxx

Keywords:
Cervical cancer
HPV infection
Immune process
Relapse

ABSTRACT

Objective. Human papillomavirus (HPV) infection is the primary cause of cervical cancer. HPV-mediated immune alterations are known to play crucial roles in determining viral persistence and host cell transformation. We sought to thoroughly understand HPV-directed immune alterations in cervical cancer by exploring publically available datasets.

Methods. 130 HPV positive and 7 HPV negative cervical cancer cases from The Cancer Genome Atlas were compared for differences in gene expression levels and functional enrichment. Analyses for copy number variation (CNV) and genetic mutation were conducted for differentially expressed immune genes. Kaplan-Meier analysis was performed to assess survival and relapse differences across cases with or without alterations of the identified immune signature genes.

Results. Genes up-regulated in HPV positive cervical cancer were enriched for various gene ontology terms of immune processes ($P = 1.05E\text{-}14 \sim 1.00E\text{-}05$). Integrated analysis of the differentially expressed immune genes identified 9 genes that displayed either CNV, genetic mutation and/or gene expression changes in at least 10% of the cases of HPV positive cervical cancer. Genomic amplification may cause elevated levels of these genes in some HPV positive cases. Finally, patients with alterations in at least one of the nine signature genes overall had earlier relapse compared to those without any alterations. The altered expression of either TFRC or MMP13 may indicate poor survival for a subset of cervical cancer patients ($P = 1.07E\text{-}07$).

Conclusions. We identified a novel immune gene signature for HPV positive cervical cancer that is potentially associated with early relapse of cervical cancer.

1. Introduction

Cervical cancer remains to be one of the most common causes of cancer-related deaths in women globally. A hallmark of cervical cancer etiology is human papillomavirus (HPV) infection, especially the high-risk types HPV16 and HPV18 [1,2]. Keratinocytes (KCs) of the basal layer of the epidermis and mucosal epithelia are the exclusive target of HPV infection; despite the fact that KCs are well equipped to detect and initiate immune responses to invading pathogens, high-risk HPVs have evolved to actively interfere with KCs innate and adaptive immune mechanisms to achieve viral persistence [3] and may integrate into the host cell genome thus allowing for persistent expression of the viral oncogene [1,4–6]. It has been shown that HPV E6 and E7 oncoproteins dysregulate gene expression, signaling transduction and cellular trafficking of critical host immune modulators [3,7]. They also cause inactivation of p53 and Rb, which leads to increased instability of host cell genome and accumulation of somatic mutations [8]. In addition, the E6 and E7 proteins of high-risk HPV effectively bind to cell cycle regulatory proteins and interfere with multiple cell cycle checkpoints, which

* Corresponding author.
E-mail address: chenlong6517@vip.sina.com (L. Chen).

https://doi.org/10.1016/j.ygyno.2018.01.031
0090-8258/

promotes ultimate transformation [8]. Yet while expression of the E6 and E7 proteins quickly upregulates cell cycle- and proliferation-related genes and at the same time inactivates several tumor suppressors, the development of invasive cancer requires many years of viral persistence and disease progression in immunocompetent individuals [7,9]. By contrast, genes altered gradually by HPV persistence are enriched for immune response regulators, and studies suggest that those immune alterations are critical for the virus to prevent the elimination of HPV-infected cells during cancer progression [7]. Further examination of virus-mediated immune dysregulation would be essential to develop preventive and therapeutic tools for treating virus-associated cancer as well as eliminating virus-infected cells [7].

The recent introduction of next generation sequencing (NGS) and other 'omics' approaches have rapidly and comprehensively improved our understanding of the pathogenesis of HPV-driven tumors including cervical cancer [10]. Hu et al. performed whole-genome sequencing and high-throughput viral integration detection in cervical cancer, which identified 3667 HPV integration breakpoints that include many new hot spots [5]. Other studies discovered novel recurrent genetic and epigenetic alterations, and demonstrated associations of HPV16 and HPV18 integration with structural aberrations and increased target-gene expression [4,6]. Such studies generated unprecedentedly large datasets most of which are now publicly available; further exploration is warranted for molecular characterization of the devastating disease. Here, in order to thoroughly understand HPV-directed alterations of immune responses in cervical cancer, we carried out detailed analyses of relevant datasets deposited in The Cancer Genome Atlas (TCGA). Our integrated analysis of the genetic and genomic changes of immune response genes revealed novel signatures for cervical cancer.

2. Methods

2.1. Data resources

Publicly available data for the TCGA-CESC and TCGA-HNSC cohort were downloaded on September 10, 2017 from http://www.cbioportal.org and https://cancergenome.nih.gov. We restricted our analysis on 137 CESC cases with known information on HPV infection history, composed of 130 HPV infected and 7 HPV negative cases. The clinical and other case information were from Supplementary Table 3 of the paper published by TCGA (reference [4]).

2.2. Gene expression analysis

The gene expression data was obtained as raw count values from TCGA public level 2 transcription profiles. We applied R packages (edgeR version 3.18.1) to the transcription profiles and tested for differential expression between eight HPV-positive/-negative group samples. *P*-values were corrected for multiple testing by computing q-values (false discovery rates). Then the significant differential expressed genes (DEGs, $P < 0.01$ and Fold change value larger than two) involved in immunological pathway were selected out for the next step analysis.

2.3. Gene function enrichment analysis

The Gene Ontology (GO, version 0.9) functional annotation of DEGs was accomplished by Biomart Database (http://plants.ensembl.org/biomart) and KEGG (Kyoto Encyclopedia of Genes and Genomes, version 84.0). Pathway annotation of DEGs was accomplished by using BLASTP (version 2.7.1) to align to KEGG database (www.kegg.jp) with a cutoff e-value of 10^{-5}. GO enrichment analysis provides all GO terms that were significantly enriched by DEGs comparing to the genome background, and filter the DEGs that correspond to biological functions. This method firstly mapped all DEGs to GO terms in the database (http://www.geneontology.org/), calculating gene numbers for every term, then using hyper geometric test to find significantly enriched GO terms in DEGs comparing to the genome background. The calculating formula is:

$$P = 1 - \sum_{i=0}^{m-1} \frac{\binom{M}{i}\binom{N-M}{i-i}}{\binom{N}{n}}$$

Where N is the number of all genes with GO annotation; n is the number of DEGs in N; M is the number of all genes that are annotated to the certain GO terms; m is the number of DEGs in M. The calculated *P*-value goes through Bonferroni Correction, taking corrected *P*-value ≤ 0.05 as a threshold. GO terms fulfilling this condition are defined as significantly enriched GO terms in DEGs. KEGG pathway enrichment analysis identifies significantly enriched metabolic pathways or signal transduction pathways in DEGs comparing with the whole genome back ground. The calculating formula was the same as that in GO enrichment analysis.

2.4. Integrative analysis

Integrative analysis of RNA-seq, copy number, and mutation was performed using OncoPrinter (version 1.0.1) at cBioPortal. The core set of samples was used since all samples in this set had data available across these three platforms. For analysis involving the RNA-seq datasets, a log2-transformation was used in order to deal with skewness in the data. mRNA z-Scores were calculated with RNA Seq V2 RSEM for each gene, and z-score of ±2 was set as threshold to identify cases with altered expression of a given gene. The CNV data included the regions determined from GISTIC 2.0 (values: −2 = homozygous deletion; −1 = hemizygous deletion; 0 = neutral/no change; 1 = gain; 2 = high level amplification), with CNVs treated as a continuous measurement based on the segmentation mean value for the region.

2.5. Kaplan-Meier analysis

Kaplan-Meier analysis (R package 'survival' version 2.41–3) was performed to assess survival and relapse difference across cases with or without alterations of the nine immune genes. Cases with changes in either mRNA expression, mutation, or CNVs in at least one of the 9 genes were considered with alteration.

3. Results

3.1. Differentially expressed genes in HPV positive cervical cancer were enriched for various immune processes

In an effort to systemically understand HPV-mediated immune alterations in cervical cancer, we first examined gene expression level changes in patients with cervical cancer. We obtained RNA-sequencing data of 130 samples of cervical cancer excision from TCGA, of which 103 were HPV16 positive and 27 HPV18 positive (Table S1). Sequencing data for a set of 7 HPV-negative cervical cancer samples were also obtained from TCGA as a control group (Fig. 1), as differences between the groups would presumably be closely related to HPV persistence. EdgeR analysis of the RNA-seq data identified 2368 differentially expressed genes (DEGs) at the cut-off *P* value <0.01, with 823 being up-regulated and 1545 down-regulated by at least 2-fold (Figure S1 and Table S2). Intriguingly, when Gene Ontology (GO) analysis was applied for these DEGs, "immune response" (GO:0006955) was found to be one of the most significantly enriched GO term (corrected *P* value (P_c) = 1.11E-05). Similarly, various immune signaling pathways such as "cytokine-cytokine receptor interaction" (P_c = 3.19E-04) and "Jak-STAT signaling pathway" (P_c = 1.02E-02) were enriched as revealed by KEGG pathway analysis (Table S3). Further, separate GO analysis showed that 78 GO terms of biological processes were significantly enriched by the up-regulated genes and 4 out of the top 5 are directly implicated in immune functions including "immune response", "innate immune

Please cite this article as: L. Chen, et al., Integrated analysis of HPV-mediated immune alterations in cervical cancer, Gynecol Oncol (2018), https://doi.org/10.1016/j.ygyno.2018.01.031