RNASeq analysis of differentiated keratinocytes reveals a massive response to late events during human papillomavirus type 16 infection, including loss of epithelial barrier function.

Klymenko, T.[1,6], Herbert, I.[1], Stevenson, A.[1], Gu, Q.[1], Iliev V.[1], Bhatia R.[2], Cuschieri, K.[23], Herzyk, P.[4], Gatherer, D.[5] and Graham, S.V.[1*]

[1] MRC-University of Glasgow Centre for Virus Research; Institute of Infection, Immunity and Inflammation; College of Medical, Veterinary and Life Sciences, University of Glasgow, Garscube Estate, Glasgow, G61 1QH, Scotland, UK.

[2] HPV Research Group, University of Edinburgh, 49 Little France Crescent, Edinburgh, EH16 4TJ, Scotland, UK.

[3] Specialist Virology Centre, Royal Infirmary of Edinburgh, 51 Little France Crescent, Edinburgh. EH16 4SA, Scotland, UK.

[4] Institute of Molecular Cell and Systems Biology; Glasgow Polyomics; University of Glasgow, Garscube Estate, Glasgow, G61 1QH, Scotland, UK.

[5]Division of Biomedical & Life Sciences, Faculty of Health & Medicine, Lancaster University, Lancaster, LA1 4YW, UK.

[6]Current address: Barts Cancer Institute, Queen Mary, University of London, John Vane Science Centre, Charterhouse Square, London, EC1M 6BQ, UK.

***Corresponding author.**

Rm 254, Jarrett Building, Garscube Estate, University of Glasgow, Glasgow, G61 1QH, Scotland, UK.

Tel: 44 141 330 6256; Fax: 44 141 330 5602;

e-mail: Sheila.Graham@gla.ac.uk

Running title: RNASeq analysis of keratinocyte response to HPV16 infection.

Abstract: 241 words. **Importance**: 99 words.

Abstract

The human papillomavirus (HPV) replication cycle is tightly linked to epithelial cell differentiation. To examine HPV-associated changes in the keratinocyte transcriptome, RNAs isolated from undifferentiated and differentiated cell populations of normal, spontaneously immortalised, keratinocytes (NIKS), and NIKS stably transfected with HPV16 episomal genomes (NIKS16), were compared using RNASeq. HPV16 infection altered expression of 2862 cellular genes. Next, to elucidate the role of keratinocyte gene expression in late events during the viral life cycle, RNASeq was carried out on triplicate differentiated populations of NIKS (uninfected) and NIKS16 (infected). Of the top 966 genes altered (>$\log_2$ = 1.8, 3.5-fold change) 670 genes were downregulated and 296 genes were up-regulated. HPV down-regulated many genes involved in epithelial barrier function that involves structural resistance to the environment and immunity to infectious agents. For example, HPV infection repressed expression of the differentiated keratinocyte-specific pattern recognition receptor TLR7, the Langerhans cell chemoattractant, CCL20, and proinflammatory cytokines, IL1A and IL1B. However, IRF1, IFNκ and viral restriction factors (IFIT1, 2, 3, 5, OASL, CD74, RTP4) were up-regulated. HPV infection abrogated gene expression associated with the physical epithelial barrier, including keratinocyte cytoskeleton, intercellular junctions and cell adhesion. qPCR and western blotting confirmed changes in expression of seven of the most significantly altered mRNAs. Expression of three genes showed statistically significant changes during cervical disease progression in clinical samples. Taken together, the data indicate that HPV infection manipulates the differentiating keratinocyte transcriptome to create an environment conducive to productive viral replication and egress.

Importance

Human papillomavirus (HPV) genome amplification and capsid formation takes place in differentiated keratinocytes. The viral life cycle is intimately associated with host cell differentiation. Deep sequencing (RNASeq) of RNA from undifferentiated and differentiated uninfected and HPV16-positive keratinocytes showed that almost 3000 genes were differentially expressed in keratinocyte due to HPV16 infection. Strikingly, the epithelial barrier function of differentiated keratinocytes, comprising keratinocyte immune function and cellular structure, was found to be disrupted. These data provide new insights into virus-host interaction crucial for production of infectious virus and reveal that HPV infection remodels keratinocytes for completion of the virus replication cycle.

Key words: human papillomavirus type 16, epithelial differentiation, keratinocyte transcriptome, cervical disease.

Introduction

Human papillomaviruses (HPVs) infect keratinocytes, causing mainly benign lesions or warts (1). Infection is usually transient and is cleared by the immune system (2). However, persistent infection with “high risk” HPV genotypes (HR-HPV) can cause tumour progression to cervical (3), other anogenital (anal, penile, vulvar and vaginal) (4) and oropharyngeal cancers (5). In the case of the cervix, cervical intraepithelial neoplasia (CIN) generally precedes cervical cancer progression (6). CIN1 is thought to represent a transient HPV infection, while CIN3 represents clinically significant, persistent HPV infection that may, if left untreated, progress to cervical cancer (7).

The pathway of epithelial cell differentiation, from basal to granular layer, is tightly controlled by complex patterns of keratinocyte gene expression (8). The HPV infectious life cycle is tightly linked to epithelial differentiation. HPV infects basal epithelial cells where it begins to express its genome. The viral replication factor E1 and its auxiliary protein, E2, which is also the viral transcription factor, together with the regulatory proteins E6 and E7 are expressed early in infection. E2, E6 and E7 have each been shown to control cellular gene expression (6). Viral gene expression required for vegetative viral genome amplification takes place in differentiating keratinocytes in the mid to upper epithelial layers (9). At this stage other viral regulatory proteins E4 and E5 that can regulate the host cell are expressed (6). Finally, L1 and L2 capsid protein synthesis and virion formation takes place in granular layer keratinocytes and virions are shed from the surface of the epithelium in dead squames (10). The epithelium presents a barrier to the environment and to infectious agents (11). Differentiated keratinocytes possess a dense filamentous network comprised of keratins and other molecules such as filaggrin. Moreover, keratinocytes have an important role in innate and adaptive immunity, and cytokines, chemokines and other immune signalling molecules released by these cells are essential for epithelial homeostasis (12). HPVs have evolved to modulate the epithelium to allow infection, virion formation and egress (6), and many means by which HPV evades the immune response have been documented (13). Elucidating the

interactions between HPV and the infected keratinocyte is key to understanding the HPV life cycle and how persistent infection may facilitate development of cervical disease.

A number of previous studies have used a microarray approach to further our understanding of the HPV infectious life cycle and cancer progression. The first compared gene expression in normal keratinocytes with that in HPV31-infected keratinocytes (14). Two subsequent studies examined gene expression changes during tumour progression in HPV18-infected (15) or HPV33-infected keratinocytes (16). A recent study investigated undifferentiated keratinocytes containing HPV16 or HPV18 episomal genomes. However, no studies have analysed how cellular gene expression is altered in differentiating keratinocytes supporting the productive phase of the viral life cycle (17). Here we used Next Generation Sequencing (RNASeq) to examine global changes in the keratinocyte transcriptome due to epithelial differentiation and HPV infection. Our study reveals that HPV infection induces massive changes in the transcriptome during keratinocyte differentiation. In particular, changes in many genes encoding the keratinocyte structural barrier and immune function were altered. Key statistically highly significant changes in gene expression were confirmed by RT-qPCR and western blotting and investigated in clinical samples representing the cervical disease spectrum. These data can be used to understand late events in the viral life cycle and the mechanisms behind cervical disease progression.

Results

The HPV E2 transcription factor (18) and the viral oncoproteins E6 (19), E7 (20) and E5 (21) can all play a role in controlling cellular gene expression, and HPV infection is known to have a significant effect on keratinocyte growth and differentiation (6). In order to elucidate how cellular gene expression is altered during HPV infection we examined changes in the keratinocyte transcriptome during differentiation and HPV16 infection using normal, spontaneously immortalised keratinocytes (NIKS) and the same cells stably transfected with HPV16 genomes (NIKS16). NIKS16 clone 2L maintains ~100 episomal HPV16 genomes per cell (if cultured at low passage (<13)) and forms a CIN1-like (low grade cervical disease) stratified epithelium upon raft culture, suggesting that these cells represent a transient HPV16 infection (22). We also examined a second HPV16 infection model, W12 cells, which are HPV16-infected basal cervical epithelial cells isolated from a patient with a low grade cervical lesion (23). W12 clone 20863 (W12E) cells (if cultured at low passage (<17)) also maintain ~100 episomal HPV16 genomes (24). Both cell lines are capable of differentiation. Differentiated NIKS16 and W12E cell populations expressed involucrin, loricrin and keratin 10, key markers of keratinocyte differentiation,(Figure 1A). NIKS16 cells (and W12 cells (25)) expressed viral late proteins E2, E4 and L1 (Figure 1A, B). A time course of NIKS and NIKS16 differentiation over a 13 day period is shown in Figure 1C. As expected, NIKS cells (Figure 1C lanes 1-4) expressed more involucrin over the time course than NIKS16 cells (Figure 1C lanes 5-8) because HPV infection impairs epithelial differentiation (6).

Global changes in the transcriptome of HPV16-infected keratinocytes.

RNASeq was carried out using RNA prepared from undifferentiated and differentiated NIKS and NIKS16 populations. Comparing undifferentiated with differentiated uninfected NIKS, 809 mRNAs were up-regulated while 422 mRNAs were down-regulated (Figure 1D). In contrast, comparing undifferentiated to differentiated HPV16-infected NIKS16 keratinocytes,

2041 genes were up-regulated while 2052 genes were down-regulated (Figure 1E). Because NIKS16 cells are derived directly from NIKS (22) and were differentiated using the same protocol, the 2862 additional changes observed upon differentiation of HPV16-positive keratinocytes are likely attributable to HPV infection. A similar number of gene expression changes to that for NIKS16 cells were observed between undifferentiated and differentiated W12E cells (data not shown).

HPV16 infection abrogates differentiation and epithelial barrier formation

We are interested in elucidating the link between keratinocyte differentiation and late events during HPV replication. Therefore, we compared the transcriptome of differentiated NIKS to NIKS16 cells. Three, replicate, single-end sequencing experiments were carried out to achieve significance. Supplementary Table 3 lists the top 966 changes in gene expression (log_2>1.8, 3.5-fold change). 670 genes were downregulated while 296 were upregulated, with a range of 184-fold downregulation to 87-fold up-regulated. The data in Figure 2 shows the mean of the results of three separate RNASeq experiments. As expected, key epithelial differentiation markers were down-regulated in NIKS16 cells (Figure 2A). Suprabasal layer keratins were also down-regulated. Keratin 12, which is usually only expressed in the corneal epithelium (26), was the only keratin whose levels were increased in NIKS16 cells (Figure 3B). Expression of cell junction proteins that are key to epithelial barrier function was significantly altered. Desmosome cell-cell junction proteins required for cell adhesion (Figure 3C) (27), and gap junction connexin (Cx) proteins 26, 30 and 32, that allow transfer of small molecules between differentiating epithelial cells (28), were down-regulated (Figure 3D). Claudin proteins control tight junctions, and CLDN3, 10 and 22 were up-regulated while CLDN11 and 17 were down-regulated (Figure 3E). Claudin upregulation can still have a negative impact on the function of tight junctions in a phenomenon referred to as “leaky claudins” (29). Several adherens junction-associated cadherins (27) were also down-

regulated (Figure 3F). Small proline-rich repeat protein (SPRR) family members that contribute to barrier formation by forming the cornified layer in differentiated epithelial cells (30) were down-regulated (Figure 3G). The calcium gradient in the epithelium is altered upon loss of barrier formation (31) and levels of RNAs encoding a range of calcium ion-binding proteins (e.g. S100A8/A9 calgranulin complex, DSG1, matrix Gla protein (MGP), calcium/calmodulin kinase 2B (CAMK2B)) were reduced (Supplementary Table 3). Taken together, the data suggest that HPV infection inhibits epithelial barrier formation and epithelial integrity.

The epithelial barrier also involves immune signalling and significant changes in expression of many genes whose products are involved in intrinsic and innate immunity were also observed (Table 1). Previously, a microarray study revealed that HR-HPV repressed activation of the immune response in undifferentiated epithelial cells through IL-1β. Similarly, in HPV-infected differentiated cells we found IL1B gene expression was down-regulated. IL1A was also down-regulated, as were IL32G and IL36B that activate keratinocyte immune functions. The Langerhans cell chemoattractant CCL20 was down-regulated in the presence of HPV16. However, CCL28 that controls T-cell homing in mucosal epithelia, E6/E7-regulated CXCL12 and CX3CL1 were all up-regulated. The type 1 IFN regulator, IRF1 and the epithelial IFNκ were up-regulated, an unexpected finding since HPVE6 and E7 have been shown to inhibit their expression (32-34). We found a 6-fold down-regulation of the viral DNA pattern recognition receptor TLR7, which is expressed specifically in differentiated keratinocytes (35), together with up-regulation of viral restriction factors APOBEC3B, IFIT1, 2, 3 and 5, CD74, OASL and RTP4 (Table 1). These data indicate that the keratinocyte-mediated immune response is under controlled of HPV16 in the upper epithelial layers, and that there are significant differences to HPV-regulation of immune signalling in differentiated, compared to basal, epithelial cells (17)

Cellular networks involved in the immune response and keratinocyte structure and metabolism are altered by HPV16 infection.

Gene ontology network pathway analysis of the most significantly altered genes (Padj <0.05) revealed distinct classes whose expression was altered by HPV16-infection (Figure 3). Response to type 1 interferon was up-regulated but cytokine and chemokine expression was repressed. Cell matrix adhesion was up-regulated while cell-cell adhesion was down-regulated (reported by the Cytoscape programme as negative regulation of up-regulated leukocyte genes) (Figure 4A). Other significantly down-regulated pathways included keratinization, arachidonic acid metabolism, reactive oxygen and nitric oxide biosynthesis, VEGF and temperature homeostasis (Figure 4B). Network analysis indicated that pathways related to the type 1 interferon response were strongly connected (Figure 4C) while down-regulated genes were associated through cytokine/chemokine/VEGF pathways (Figure 4D). A value of $\log_2$ change>2.5 was chosen to construct a wider pathway linkage diagram. IRF1 and KDR were major HPV-up-regulated genes encoding hub proteins that connected a number of cell growth and apoptosis signalling pathways. IL-1B and REL, an NFkB family transcriptional co-activator, linked HPV-down-regulated cytokine and VEGF (Supplementary Figure 1, Supplementary Table 3).

Verification of gene expression changes due to HPV16 infection

Six genes with statistically highly significant changes (Padj<0.025) were selected for further study (negative: DSG1, SERPINB3, KRT10, positive: VTCN1, KDR, AZGP1) (Table 2). IL1B was also included because expression of this important cytokine was found to be a key gene network hub in both undifferentiated (17) and differentiated HPV-infected cells (Supplementary Figure 1A). These genes all encode proteins with known metabolic or immune/inflammatory roles in the normal epithelium. KRT10 is a differentiation-specific keratinocyte filament protein. DSG1 is a calcium-binding desmosome regulator. KDR

(vascular endothelial growth factor receptor 2, VEGFR-2) has an autocrine function in cell proliferation, adhesion and migration (36). IL1B “node” cytokine activates adaptive immunity. VTCN1 is a T-cell activation inhibitor. SERPINB3 controls epithelial inflammatory responses and AZGP1 is induced by IFNγ in keratinocytes (37). mRNA expression in NIKS versus NIK16 cells was validated by qRT-PCR (Table 2).

Protein levels encoded by these mRNAs were examined in undifferentiated and differentiated NIKS and NIKS16 cells (Figure 5). Levels of KRT10, SERPINB3, DSG1 and involucrin increased upon NIKS cell differentiation, but KRT10 and DSG1 levels were much lower in differentiated NIKS16, compare to NIKS cells, as expected. SERPINB3 levels were greatly reduced following differentiation of NIKS16, but not NIKS cells. VTCN1, KDR and AZGP1 levels were higher in NIKS16 compared to NIKS cells. VTCN1 levels did not alter upon NIKS16 differentiation while, AZGP1 and KDR levels increased in differentiated NIKS16 cells. These data confirm that selected keratinocyte transcriptomic changes due to HPV16 infection are reflected in protein levels.

HPV16 infection-regulated mRNA as biomarkers of cervical disease

It could be argued that the NIKS16 model of the HPV16 life cycle may not directly relate to cervical HPV infection because NIKS16 cells are foreskin, not cervical, keratinocytes. However, the organisation of the HPV life cycle at different anatomical sites is quite similar, suggesting that there may be considerable similarities in gene expression patterns due to HPV infection of cervical, versus foreskin, keratinocytes (38). In support of this, NIKS16 cells appeared to represent a low grade cervical lesion when grown in raft culture (22). HPV16-associated gene expression changes in keratinocytes could be related to the productive life cycle but could equally be associated with cervical disease progression. Therefore, to test whether any of the HPV-related changes in keratinocyte gene expression we detected could have potential as HPV-associated cervical disease biomarkers, we

quantified levels of expression of three up- and three-down-regulated genes (two regulators of the inflammatory response (IL1B, SERPINB3), two proteins involved in cell signalling (KDR, VTCN1), and two involved in barrier function (KRT10, DSG)) by RT-qPCR in liquid based cytology (LBC, Pap smear) samples. Due to lack of mRNA we were unable to test AZGP1. A control cDNA from differentiated W12E cervical keratinocytes was included in each qPCR plate as a standard and absolute levels of RNA in the LBC samples (normalised against GAPDH) were calculated using the Pfaffl standard curve method (39, 40). KRT17 was analysed as a known biomarker of cervical disease progression (41). Figure 6 shows the mean and range of values for each mRNA in 7 no detectable disease (NDD), 10 low grade cervical lesion (CIN1) and 10 high grade cervical lesion (CIN3) samples (Supplementary Table 4). Although we analysed 10 samples graded as NDD, once HPV typing status was revealed, 3 of these were HPV-positive. We decided to exclude these from the analysis in order to compare HPV-negative with HPV-positive clinical samples. KRT10 mRNA levels were very low making analysis of significance difficult, and there was high variability in levels of IL1B and VTCN1. However, very high levels of IL1B mRNA were detected in all patient samples, regardless of disease stage. DSG1 was significantly increased between no detectable disease (NDD) and low grade disease but significantly decreased between low grade and high grade disease. KDR and SERPINB3 levels were significantly up-regulated between low grade and high grade disease, similar to the positive control, KRT17. These data suggest that RNASeq analysis has potential to uncover novel biomarkers of cervical disease.

Discussion

The aim of our work is to examine how human papillomavirus replication is linked to keratinocyte differentiation. In particular we are interested in how differentiating keratinocytes respond to HPV infection during the late, productive phase of the viral life cycle. As a model, we used NIKS and NIKS16 cells. NIKS are spontaneously immortalised neonatal foreskin keratinocytes that have no alterations in differentiation or apoptosis (42). NIKS16 cells were derived directly from NIKS cells by stable transfection of the HPV16 genome isolated from W12 cells (22). We infer that the NIKS16 cells adequately supported the infectious viral life cycle (as previously reported (22)) because several key markers of keratinocyte differentiation and viral life cycle completion, especially capsid protein production, were detected. Moreover, because there was repression VEGF pathways, reduced expression of HOX and MMP proteins, and no general up-regulation of EMT markers, these cells are not undergoing tumour progression. Because NIKS cells are foreskin keratinocytes, they will likely have a number of differences in their gene expression profile compared to cervical keratinocytes. However, such changes must be limited in number because was also carried out RNASeq analysis of undifferentiated and W12 cells, cervical keratinocytes, where very similar profiles of gene expression changes were obtained. That said, NIKS16 is potentially a more robust model for HPV16-associated penile lesions than cervical lesions and it will be interesting in future to compare these data sets with similar sets from differentiated cervical keratinocytes. 3D raft culture would undoubtedly provide a superior approach for examining keratinocyte differentiation and HPV infection. However, for analysis of late events in the viral life cycle in differentiated keratinocytes, this is technically challenging, and difficult to reproduce, because RNA isolation from multiple, microdissected, upper epithelial layer sections would be required for triplicate RNASeq experiments. Our current dataset should provide an important basis for subsequent analysis of raft culture models.

Many transcriptomic studies have analysed cellular changes during HPV-associated tumour progression or due to overexpression of viral proteins (14, 15, 18, 21, 43-50). Of the microarray studies investigating changes due to HPV infection, as opposed to tumourigenesis, one compared expression of HPV31-positive and negative cervical keratinocytes (14), a second examined HPV33-negative and positive vaginal keratinocytes (16) while another compared undifferentiated anogenital keratinocytes with or without episomal HPV16 and 18 genomes (17). All of these studies focused on the effect of HPV on basal keratinocytes, the site of viral entry, and initial replication. No studies to date have examined keratinocyte responses to late events in the viral replication cycle. Moreover, the previous studies used microarray analysis which does not provide the unparalleled depth of information available from RNASeq. To our knowledge, this is the first report comparing the transcriptome of uninfected to HPV-infected differentiated keratinocytes using RNASeq. HPV infection induced massive changes (2862 additional expression changes compared to HPV-negative NIKS cells) in the keratinocyte transcriptome. Desmosomes, adherens, tight and gap junction classes were all down-regulated in the presence of HPV16, likely due to HPV16 E6/E7 reactivation of the cell cycle and decreased keratinocyte differentiation (51) as has been reported previously (17). Together with high level down-regulation of SPRRS, altered arachidonic acid metabolism and changes in mucins (Supplementary Table 3), one can conclude that HR-HPV infection results in a broad abrogation of epithelial barrier function and epithelial integrity. Reduced barrier function could result in increased fragility of cells in the upper epithelial layers to facilitate viral egress.

Keratinocytes are key players in the immune response, and they produce a panoply of molecules involved in host defence against pathogens. In differentiated NIKS16 keratinocytes, HPV infection altered gene expression related to innate immunity, including reduced expression of TLR7, IL1A, IL1B, NLRP3, IL36B, and IL32G. TLR7, a pattern recognition receptor for viral nucleic acids, is upregulated upon keratinocyte differentiation (35) and activates proinflammatory cytokines, and other molecules involved in the adaptive

immune response. There was a 6-fold down-regulation of TLR7 in the presence of HPV16 suggesting that the virus represses pattern recognition during vegetative viral genome amplification, but by a different mechanism to that used in undifferentiated keratinocytes where infection suppresses TLR9 (17). There was a corresponding reduction in NFκB-regulated CCL20, known to be regulated by HPV E7 (52), and required to recruit Langerhans cells. Indeed, NFκB signalling was affected and the NFκB family member, REL, was a major HPV-regulated control node in the pathway analysis of negatively regulated genes (Supplementary Figure 1). Surprisingly, we discovered that the epithelial-specific IFNκ, and IRF1 that controls type 1 IFNs, were up-regulated by HPV16 in differentiated keratinocytes. Previously, HPV16 E7 or HPV38 E6E7 were shown to inhibit IRF1 expression (32, 33), while HPV16 E6 was shown to repress IFNk transcription through promoter methylation (53). However, these studies used overexpression of the viral oncoproteins. The levels of E6 or E7 proteins may be much lower in differentiated keratinocytes compared to that in the undifferentiated epithelial cells or cervical cancer cells used in these studies. In contrast to E6 and E7, E5 can stimulate IRF1 expression in HaCaT cells (54). Changes due to expression of the entire virus genome may be more complex and quite different to that seen with expression of individual viral proteins. Up-regulation of IFITs corresponded with the observed activation of the type 1 interferon response. Only IFIT1 has been shown to inhibit HPV replication (55, 56) therefore, the roles of other IFITs in inhibiting HPV infection remain to be determined. APOBEC3B was up-regulated however, we found no changes in expression of APOBEC3A, a known HPV restriction factor, but its expression may be differentially regulated only in less differentiated keratinocytes (57). The observed up-regulation of CXCR6 and CXCL12 is in agreement with CXCL12 detection in HPV-induced lesions and its role in the productive HPV life cycle (58). We also detected changes in some SERPINs (e.g. SERPINB3) that are involved in the inflammatory/immune response (59). We did not detected changes in STAT1 that has been shown to be controlled by E6 and E7 (60). It is possible that it undergoes changes of less than the cut-off of >3.5-fold considered here. However, STAT1 controls IRF1 expression, which was upregulated 4-fold and STAT1 was a

central node connecting gene pathways regulated by HPV16 (Supplementary Figure 1). Our data reveal that HPV suppression of intrinsic and innate immunity takes place not only in infected basal epithelial cells (17) but also in keratinocytes harbouring late events in the HPV life cycle, and that a differentiation stage-specific set of events may be relevant to this life cycle stage. The stimulation of the IFN response and viral restriction factors in differentiated HPV-infected cells requires further study. Production of progeny viral genomes and virions may stimulate the IFN response and lead to apoptosis and this could aid release and dissemination of virus particles.

The E5, E6, E7 and E2 proteins of HPV 16 are known to control cellular gene expression. E6 and E7 control keratinocyte cell cycle and apoptosis and abrogate differentiation. Many of the changes in gene expression we have observed can be attributed to these functions of the viral oncoproteins. These changes are clearly important for the replicative life cycle of HPV16 but could also contribute to HPV persistence and development of neoplasia (6). Similar to data from one overexpression study of HPV16 E6 in human foreskin keratinocytes (49), the differentiation marker involucrin, vimentin that is expressed upon epithelial stress, and signal transduction proteins MEST and H19, were up-regulated in our analysis. However, we detected none of the other changes affecting cell cycle, proliferation, DNA damage, metabolism or signalling that have previously been reported (49). We discovered only 7 genes (Semaphorin 5A (SEMA5A), CXCL1, ENTPOT, Follistatir (FST), Cytochrome P450 (CYP) 24A1, Pleckstrin homology-like domain A1 (PHLDA1) and ribosomal proteins S27-like (RPS27L)) out of a total of 99 altered in another study using siRNA depletion of E6 in HPV-positive tumour cells (46). Compared to a study of W12 cells with integrated HPV16 genomes expressing different levels of E6 and E7, we detected E6-regulated loricrin (LOR) and cytochrome P450 (CYP) 1B1, and E7-regulated FABP4, SERPINA3, SLURP1 out of the top 20 genes up-regulated by each protein (61). Only one out of 12 master regulators of E6 or E7 function defined by Smith et al (61) was in common with our study. This was downregulation of PRDM1 (BLIMP-1) which acts as a repressor of IFN-β gene expression.

E5 overexpression in HaCaT keratinocytes yielded 61 mRNAs with significant changes (21) but only two of these (Keratin 8, MMP16) were in common with our RNASeq data. In a microarray study of E2 overexpression in U2OS cancer cells where 74 genes were found to be regulated, only 3 of these (heterotrimeric G-complex protein 11 (GNG11) involved in cell signalling**,** histamine N-methyltranferase (HNMT) involved in methylation of histamine and SERPINA3 which is up-regulated in response to decreased transglutaminase activity) were altered in our study. Increased viral oncoprotein expression levels in HPV-positive cancer cells, or in cells overexpressing viral proteins, compared to the model we have used, i.e. keratinocytes supporting expression of all viral proteins from the intact HPV16 genome where expression levels are much lower (3), could explain the fact that we did not detect many of these changes. Moreover, we have only considered expression changes >3.5-fold, while these other studies considered 2-fold changes. RNASeq analysis of the W12 tumour progression series (62) would help to delineate infection versus cancer-related changes.

Liquid based cytology samples (LBCs, Pap smear samples) contain cells scraped from the top of the cervical epithelium and thus contain HPV-infected differentiated keratinocytes. Therefore, some of the mRNA changes we have detailed could be biomarkers of cervical disease. Very high levels of IL1B mRNA were detected in all patient samples, regardless of disease stage, likely due to inflammation commonly observed in diseased cervix. Statistically significant changes in KDR and SERPINB3 expression, like the known biomarker KRT17, indicate their potential in identifying high grade cervical disease. DSG1 was significantly increased between no detectable disease (NDD) and low grade disease but significantly decreased between low grade and high grade disease. This is in contrast to the clear down-regulation of DSG1 expression due to HPV16 infection of NIKS and suggests either that NIKS16 cells may not represent a low grade HPV16-positive lesion or that the levels of DSG1 in cervical keratinocytes is very different to that in foreskin keratinocytes.

In conclusion, we report for the first time RNASeq analysis of changes in the keratinocyte transcriptome caused by HR-HPV infection. Infection caused massive changes in epithelial gene expression. These changes showed mainly a profile expected of viral infection, rather than tumour progression. The large dataset we have developed opens up the possibility of a deeper understanding of late events in the HPV replication cycle in response to keratinocyte differentiation. As well as shedding light on late events during the HPV16 life cycle, the RNASeq data could uncover potential biomarkers of HPV-associated anogenital disease progression. From our analysis, DSG1, KDR and SERPINB3 expression may have potential as robust markers that can risk-stratify cervical disease, i.e. identify cervical disease cases that have a high probability of regression, and this would be of significant clinical value. However, further longitudinal studies where biomarker status is linked to clinical outcomes would be required to validate any biomarkers for such an application.

Materials and Methods

Clinical sample panel underlying pathology and HPV status

Anonymised, cervical liquid based cytology samples were obtained from the Scottish National HPV archive which holds Generic Scotland A Research Ethics Committee approval for Research Tissue banks (REC Ref 11/AL/0174) for provision of samples for HPV related research after approval from an independent steering committee. The Scottish HPV Archive also comes under the auspice of the NHS Lothian Bioresource. The panel comprised HPV negative/cytology negatives samples (no disease, n=7) samples with low-grade cytological abnormalities with histological confirmation of cervical intraepithelial neoplasia (CIN) 1 (low-grade disease, n=10) and samples with high-grade cytological abnormalities with histological confirmation of CIN2 or worse, including cancer (high-grade disease, n=10). Cytology grades were reported according to the British Society for Clinical Cytopathology (BSCC) classification (63-65). HPV testing was performed by the Optiplex HPV genotyping Assay (Diamex, Heidelberg, Germany) according to manufacturer's instructions. The Optiplex test is a PCR based assay which uses a luminex platform for the detection of 24 individual HPV types including all established as high-risk according to the International Agency on Research on Cancer. For the purposes of this panel, the main function of the genotyping was for the annotation of no disease "controls". Women with negative cytology and HPV negative status are at a very low risk of underlying disease (negative predictive value for a high grade lesion of >95% (66)) All experiments were performed in compliance with relevant laws and institutional guidelines and in accordance with the ethical standards of the Declaration of Helsinki.

Cell lines

W12E (24), NIKS (42), and NIKS16 cells (22) were co-cultured in E-medium with mitomycin C-treated J2 3T3 fibroblast feeder cells as previously described (24). 3T3 cells were grown in DMEM with 10% donor calf serum. Prior to harvesting, 3T3 cells were removed by

trypsinisation and cells layers washed twice with PBS. All cells were maintained under humidified 5% CO_2 95% air at 37°C.

RNA isolation – cell lines

Protocols followed the manufacturer's instructions. Total RNA was prepared using Qiagen RNeasy kit. RNA was quantified by measuring the ratio of absorbance at 260 and 280 nm using a Nanodrop ND-1000 spectrophotometer (ThermoScientific). Polyadenylated RNA was prepared using an oligo-dT-based mRNA extraction kit (Oligotex, Qiagen).

RNA isolation – clinical samples

LBC cells in 4ml in PreservCyt collection medium (Cytyc Corporation) were pelleted by centrifugation in a Beckman GPR bench top centrifuge at 1500g for 10 min. The cell pellet washed with sterile PBS. RNA extraction was carried out using the RNeasy miRNA preparation kit (Qiagen). RNA was quantified and purity assessed by measuring the ratio of absorbance at 260 and 280 nm using a Nanodrop ND-1000 spectrophotometer.

qRT-PCR

For cell line and clinical samples, DNA was removed using Maxima DNase and treated RNA was reverse transcribed using Maxima First Strand cDNA synthesis kit according to manufacturer's instructions (ThermoScientific). Standard curves were generated as recommended (Applied Biosystems instruction manual). Triplicate amplification reactions containing 100 ng cDNA each were carried out. Supplementary Table 1 lists primers and probes. GAPDH and β-actin were used as the internal standard controls. Reaction mixes (25 µl) contained 1x Mastermix (Stratagene), 900 nM primers, 100 nM probe, 300 nM reference dye (Stratagene). qPCR reactions were performed and analysed on an Applied Biosystems 7500 Fast System. Graphing and statistical analyses were performed using GraphPad Prism 7. Statistical analysis (all three groups were compared to each other) was performed by

469 Kruskal-Wallis test and data analysed by one way ANOVA with Tukey's post-test. A
470 significance level of $p<0.05$ was used.

471

472 ***Illumina sequencing***

473 Integrity of RNAs was assessed using an Agilent 2100 Bioanalyser. cDNA was synthesised
474 using reagents from the TruSeq RNA Sample Preparation kit (Illumina) according to the
475 manufacturer's instructions. cDNA libraries were sequenced with a 73 base single-end read
476 on an Illumina Genome Analyser IIx at the Glasgow Polyomics facility at the University of
477 Glasgow. Samples have been submitted to SRA@ncbi.nih.gov (SUB2567956). STUDY:
478 PRJNA379358 (SRP104232). SAMPLE: NIKS16_D11_Mar17 (SRS2131727),
479 EXPERIMENT: Differentiated NIKS16 cells (SRX2745325) RUN:
480 NIKS_HPV16_D11_Mar17.fq.gz (SRR5457256). SAMPLE: NIKS16_D5_Mar17
481 (SRS2131728), EXPERIMENT: Undifferentiated NIKS16 cells (SRX2745326), RUN:
482 NIKS_HPV16_D5_Mar17.fq.gz (SRR5457258). SAMPLE: NIKS_D11_Mar17
483 (SRS2131729), EXPERIMENT: Differentiated NIKS cells (SRX2745327), RUN:
484 NIKS_D11_Mar17.fq.gz (SRR5457259). SAMPLE: NIKS_D5_Mar17 (SRS2131730),
485 EXPERIMENT: Undifferentiated NIKS cells (SRX2745328), RUN: NIKS_D5_Mar17.fq.gz
486 (SRR5457260)

487

488 ***Computational analysis***

489 Datasets were cleaned of reads with runs > 12Ns. Alignment to the human cDNA set
490 (145,786 cDNAs – downloaded on 28th November 2011) was performed using Bowtie
491 version 0.12.7. Further alignment to an updated human cDNA set (180,654 cDNAs
492 downloaded April 30th 2012) was carried out using BWA 0.7.12-r1039 (Supplementary Table
493 2). DESeq implemented in BioConductor (67) was used to select cellular genes whose
494 expression was up or down-regulated by HPV in NIKS16 compared to NIKS cells

implemented in the R environment. The raw read counts were normalised using (RPKM). DESeq uses a negative binomial error distribution to model transcript abundance and determine the differential expression. The significance of differential expression was estimated for each gene and then corrected for multiple comparisons (Padj). The top 1000 differentially expressed genes based on log-fold change (Log_2FoldChange) of >1.8 (3.5-fold change) are listed in Supplementary Table 3.

Functional analysis of differentially expressed genes

GO (68) and KEGG (69) enrichment analyses were performed using Cytoscape (http://cytoscape.org/) with ClueGO (Version 2.3.2) (70). The statistical test used for the enrichment was based on a two-sided hypergeometric option with a Bonferroni step-down correction, a *P*-value less than 0.05 and a kappa score of 0.4.

Acknowledgements

We thank Dr John Doorbar (University of Cambridge) for providing the NIKS and NIKS16 cells lines and Prof Margaret Stanley (University of Cambridge) and Prof Paul Lambert (University of Wisconsin) for the W12 cell line. This work was funded by a grant to SVG from the Wellcome Trust (088848/2/09/2) and to TK and SVG from Tenovus Scotland (S11/33). We acknowledge funding from the Medical Research Council as core funding for the MRC University of Glasgow Centre for Virus Research.

Declaration of Conflicts of Interest

We declare no conflicts of interest.

519 Table 1. Changes in expression of immune regulatory molecules and viral restriction factors.

Gene ID	Category	Negative fold change	Positive fold change
TLR7	PRR	6-fold	
NLRP3	Inflammasome component	7-fold	
IL1A	Cytokine	7-fold	
IL1B	Cytokine	4-fold	
IL32G	Cytokine	17-fold	
IL36B	Cytokine	6-fold	
CCL20	Chemokine	7-fold	
CCL28	Chemokine		5-fold
CXCL12	Chemokine		4-fold
CX3CL1	Chemokine		32-fold
APOBEC3C	Restriction factor		4-fold
IFIT1	Restriction factor		6-fold
IFIT2	Restriction factor		7-fold
IFIT3	Restriction factor		7-fold
IFIT5	Restriction factor		4-fold
CD74	Restriction factor		4-fold
AOSL	Restriction factor		4-fold
RTP4	Restriction factor		13-fold
IRF1	IFN regulatory transcription factor		4-fold
IFNκ	Interferon kappa		8-fold

520

521

522 PRR: pattern recognition receptor. IFN, interferon.

523

524

525 Table 2. RNASeq expression changes in mRNAs of statistical significance ($p<0.025$) verified

526 by qPCR

Gene ID	**Padj**	**NIKS16/NIKS-fold change RNASeq**	**NIKS16/NIKS-fold change qPCR**	**Gene Function**
DSG1	2.05×10^{-5}	-19.95	-4.20	Desmoglein1: calcium-binding desmosome regulator
IL1B	1	-8.68	-5.65	Interleukin 1b: inflammatory response regulator
SERPINB3	0.008	-8.40	-4.28	Intracellular protease inhibitor, inhibits active inflammatory response
KRT10	0.021	-7.07	-10.26	Keratin10: epithelial cytofilament
KDR	0.025	10.21	10.10	VEGFR-2, tyrosine kinase receptor
VTCN1	1.4×10^{-9}	46.12	8.94	V-set domain-containing T-cell activation inhibitor-1
AZGP1	2.05×10^{-5}	12.64	7.49	Zinc alpha-2 glycoprotein: lipid metabolism
GAPDH		1	1	Glyceraldehyde-3-phosphate dehydrogenase (control)
Beta-actin		1	1	Actin (control)

527

528

Reference List

1. **zur Hausen H.** 2009. Papillomaviruses in the causation of human cancers - a brief historical account. Virol **384:**260-265.
2. **Stanley MA.** 2012. Epithelial cell responses to infection with human papillomavirus. Clin Microbiol Rev **25:**215-222.
3. **Doorbar J, Egawa N, Griffin H, Kranjec C, Murakami I.** 2015. Human papillomavirus molecular biology and disease association. Rev Med Virol **25:**2-23.
4. **Wakeham K, Kavanagh K.** 2014. The burden of HPV-associated anogenital cancers. Curr Oncol Rep **16:**1-11.
5. **Gillison ML, Chaturvedi AK, Anderson WF, Fakhry C.** 2015. Epidemiology of human papillomavirus–positive head and neck squamous cell carcinoma. J Clin Oncol **33:**3235-3242.
6. **Egawa N, Egawa K, Griffin H, Doorbar J.** 2015. Human papillomaviruses; Epithelial tropisms, and the development of neoplasia. Viruses **7:**2802.
7. **Pett M, Coleman N.** 2007. Integration of high-risk human papillomvirus: a key event in crvical carcinogenesis? J Pathol **212:**356-367.
8. **Fuchs E, Byrne C.** 1994. The epidermis: rising to the surface. Curr Opin Gen Dev **4:**725-736.
9. **Doorbar J.** 2005. The papillomavirus life cycle. J Clin Virol **32S:**S7-S15.
10. **Graham SV.** 2006. Late events in the life cycle of human papillomaviruses, p 193-212. *In* Campo MS (ed), Papillomavirus research: from natural history to vaccines and beyond, 1st ed. Caister Academic Press, Wymondham, Norfolk.
11. **Matsui T, Amagai M.** 2015. Dissecting the formation, structure and barrier function of the stratum corneum. International Immunology **27:**269-280.
12. **Partidos CD, Muller S.** 2005. Decision-making at the surface of the intact or barrier disrupted skin: potential applications for vaccination or therapy. Cell Mol Life Sci CMLS **62:**1418-1424.

13. **Westrich JA, Warren CJ, Pyeon D.** 2017. Evasion of host immune defenses by human papillomavirus. Virus Res **231:**21-33.

14. **Chang YE, Laimins LA.** 2000. Microarray analysis indentifies interferon-inducible genes and Stat-1 as major transcriptional targets of human papillomavirus type 31. J Virol **74:**4174-4182.

15. **Karstensen B, Poppelreuther S, Bonin M, Walter M, Iftner T, Stubenrauch F.** 2006. Gene expression profiles reveal an upregulation of E2F and downregulation of interferon targets by HPV18 but no changes between keratinocytes with integrated or episomal viral genomes. Virology **353:**200-209.

16. **Ruutu M, Peitsaro P, Johansson B, Syrjänen S.** 2002. Transcriptional profiling of a human papillomavirus 33–positive squamous epithelial cell line which acquired a selective growth advantage after viral integration. Int J Cancer **100:**318-326.

17. **Karim R, Meyers C, Backendorf C, Ludigs K, Offringa R, van Ommen G-JB, Melief CJM, van der Burg SH, Boer JM.** 2011. Human papillomavirus deregulates the response of a cellular network comprising of chemotactic and proinflammatory genes. PLoS ONE **6:**e17848.

18. **Gauson EJ, Windle B, Donaldson MM, Caffarel MM, Dornan ES, Coleman N, Herzyk P, Henderson SC, Wang X, Morgan IM.** 2014. Regulation of human genome expression and RNA splicing by human papillomavirus 16 E2 protein. Virology **468–470:**10-18.

19. **Vande Pol SB, Klingelhutz AJ.** 2013. Papillomavirus E6 oncoproteins. Virology **445:**115-137.

20. **Roman A, Munger K.** 2013. The papillomavirus E7 proteins. Virology **445:**138-168.

21. **Kivi N, Greco D, Auvinen P, Auvinen E.** 2007. Genes involved in cell adhesion, cell motility and mitogenic signaling are altered due to HPV 16 E5 protein expression. Oncogene **27:**2532-2541.

22. **Weschler EI, Wang Q, Roberts I, Pagliarulo E, Jackson D, Untersperger C, Coleman N, Griffin H, Doorbar J.** 2012. Reconstruction of human papillomavirus

type 16-mediated early-stage neoplasia implicated E6/E7 deeregualtion and the loss of contact inhibition in neoplastic progression. J Virol **86:**6358-6364.

23. **Stanley MA, Browne HM, Appleby M, Minson AC.** 1989. Properties of a non-tumorigenic human cervical keratinocyte cell line. Int J Cancer J Int Cancer **43**.

24. **Jeon S, Allen-Hoffman BL, Lambert PF.** 1995. Integration of human papillomavirus type 16 into the human genome correlates with a selective growth advantage of cells. J Virol **69:**2989-2997.

25. **Milligan SG, Veerapraditsin T, Ahamat B, Mole S, Graham SV.** 2007. Analysis of novel human papillomavirus type 16 late mRNAs in differentiated W12 cervical epithelial cells. Virology **360:**172-181.

26. **Bragulla HH, Homberger DG.** 2009. Structure and functions of keratin proteins in simple, stratified, keratinized and cornified epithelia. J Anat **214**.

27. **Bazzi H, Christiano AM.** 2007. Broken hearts, woolly hair, and tattered skin: when desmosomal adhesion goes awry. Curr Opin Cell Biol **19:**515-520.

28. **Laird D.** 2006. Life cycle of connexins in health and disease. Biochem J **394:**527-543.

29. **Findley MK, Koval M.** 2009. Regulation and roles for claudin-family tight junction proteins. IUBMB Life **61:**431-437.

30. **Carregaro F, Stefanini ACB, Henrique T, Tajara EH.** 2013. Study of small proline-rich proteins (SPRRs) in health and disease: a review of the literature. Arc Dermatol Res **305:**857-866.

31. **Bikle DD, Oda Y, Xie Z.** 2004. Calcium and 1,25(OH)2D: interacting drivers of epidermal differentiation. J Steroid Biochem Mol Biol **89–90:**355-360.

32. **Park J-S, Kim E-J, Kwon H-J, Hwang E-S, Namkoong S-E, Um S-J.** 2000. Inactivation of interferon regulatory factor-1 tumor suppressor protein by HPV E7 oncoprotein: Implications for the E7-mediated immuna evasion mechanism in cervical carcinogenesis. J Biol Chem **275:**6764-6769.

612 33. **Cordano P, Gillan V, Bratlie S, Bouvard V, Banks L, Tommasino M, Campo MS.**
613 2008. The E6E7 oncoproteins of cutaneous human papillomavirus type 38 interfere
614 with the interferon pathway. Virology **377:**408-418.
615 34. **Ronco LV, Karpova AY, Vidal M, Howley PM.** 1998. Human papillomavirus 16 E6
616 oncoprotein binds to interferon regulatory factor-3 and inhibits its
617 transcriptional activity. Genes Dev **12:**2061-2072.
618 35. **Li ZJ, Sohn K-C, Choi D-K, Shi G, Hong D, Lee H-E, Whang KU, Lee YH, Im M,**
619 **Lee Y, Seo Y-J, Kim CD, Lee J-H.** 2013. Roles of TLR7 in activation of NF-κB
620 signaling of keratinocytes by imiquimod. PLoS ONE **8:**e77159.
621 36. **Man X-Y, Yang X-H, Cai S-Q, Yao Y-G, Zheng M.** 2006. Immunolocalization and
622 expression of vascular endothelial growth factor receptors (VEGFRs) and neuropilins
623 (NRPs) on keratinocytes in human epidermis. Mol Med **12:**127-136.
624 37. **Brysk MMB, T,; Hoida, C.; Tyring, S.K.; Rajaraman, S.** 1991. Interferon-gamma
625 modulates terminal differentation and the expression of desquamin in cultured
626 keratinocytes. Exp Cell Res **197:**7.
627 38. **Peh W, L., Middleton K, Christensen N, Nicholls P, Egawa K, Sotlar K,**
628 **Brandsma J, Percival A, Lewis J, Liu WJ, Doorbar J.** 2002. Life cycle
629 heterogeneity in animal models of human papillomavirus-associated disease. J Virol
630 **76:**10411-10416.
631 39. **Bustin SA, Beaulieu J-F, Huggett J, Jaggi R, Kibenge FSB, Olsvik PA, Penning**
632 **LC, Toegel S.** 2010. MIQE précis: Practical implementation of minimum standard
633 guidelines for fluorescence-based quantitative real-time PCR experiments. BMC Mol
634 Biol **11:**74.
635 40. **Pfaffl MW.** 2001. A new mathematical model for relative quantification in real-time
636 RT–PCR. Nucleic Acids Research **29:**e45-e45.
637 41. **Escobar-Hoyos LF, Yang J, Zhu J, Cavallo J-A, Zhai H, Burke S, Koller A, Chen**
638 **EI, Shroyer KR.** 2014. Keratin 17 in premalignant and malignant squamous lesions

of the cervix: proteomic discovery and immunohistochemical validation as a diagnostic and prognostic biomarker. Mod Pathol **27:**621-630.

42. **Allen-Hoffmann BL, Schlosser SJ, Ivarie CAR, Meisner LF, O'Connor SL, Sattler CA.** 2000. Normal growth and differentiation in a spontaneously immortalized near-diploid human keratinocyte cell Line, NIKS. J Invest Dermatol **114:**444-455.

43. **Chen J, Xue Y, Poidinger M, Lim T, Chew SH, Pang CL, Abastado J-P, Thierry F.** 2014. Mapping of HPV transcripts in four human cervical lesions using RNAseq suggests quantitative rearrangements during carcinogenic progression. Virology **462–463:**14-24.

44. **Nees M, Geoghegan JM, Munson P, Prabhu V, Liu Y, Androphy E, Woodworth CD.** 2000. Human papillomavirus type 16 E6 and E7 proteins inhibit differentiation-dependent expression of transforming growth factor-β2 in cervical keratinocytes. Cancer Res **60:**4289-4298.

45. **Nees M, Geoghegan JM, Hyman T, Frank S, Miller L, Woodworth CD.** 2001. Papillomavirus type 16 oncogenes downregulate expression of interferon-responsive genes and upregulate proliferation-associated and NF-κB-responsive genes in cervical keratinocytes. J Virol **75:**4283-4296.

46. **Kelley ML, Keiger KE, Lee CJ, Huibregtse JM.** 2005. The global transcriptional effects of the human papillomavirus E6 protein in cervical carcinoma cell lines are mediated by the E6AP ubiquitin ligase. J Virol **79:**3737-3747.

47. **Sopov I, Sörensen T, Magbagbeolu M, Jansen L, Beer K, Kühne-Heid R, Kirchmayr R, Schneider A, Dürst M.** 2004. Detection of cancer-related gene expression profiles in severe cervical neoplasia. Int J Cancer **112:**33-43.

48. **Chen Y, Miller C, Mosher R, Zhao X, Deeds J, Morrissey M, Bryant B, Yang D, Meyer R, Cronin F, Gostout BS, Smith-McCune K, Schlegel R.** 2003. Identification of cervical cancer markers by cDNA and tissue microarrays. Cancer Res **63:**1927-1935.

49. **Duffy CL, Phillips SL, Klingelhutz AJ.** 2003. Microarray analysis identifies differentiation-associated genes regulated by human papillomavirus type 16 E6. Virology **314:**196-205.

50. **Wong Y-F, Cheung T-H, Tsao GSW, Lo KWK, Yim S-F, Wang VW, Heung MMS, Chan SCS, Chan LKY, Ho TWF, Wong KWY, Li C, Guo Y, Chung TKH, Smith DI.** 2006. Genome-wide gene expression profiling of cervical cancer in Hong Kong women by oligonucleotide microarray. Int J Cancer **118:**2461-2469.

51. **Woodworth CD, S. C, S. S, Hamacher L, Chow LT, T.R. B, DiPaolo JA.** 1992. Recombinant retroviruses encoding human papillomavirus type 18 E6 and E7 genes stimulate proliferation and delay differentiation of human keratinocytes early after infection. Oncogene **7:**8.

52. **Richards KH, Wasson CW, Watherston O, Doble R, Eric Blair G, Wittmann M, Macdonald A.** 2015. The human papillomavirus (HPV) E7 protein antagonises an Imiquimod-induced inflammatory pathway in primary human keratinocytes. Sci Rep **5:**12922.

53. **Rincon-Orozco B, Halec G, Rosenberger S, Muschik D, Nindl I, Bachmann A, Ritter TM, Dondog B, Ly R, Bosch FX, Zawatzky R, Rösl F.** 2009. Epigenetic silencing of interferon-κ in human papillomavirus type 16–positive cells. Cancer Res **69:**8718-8725.

54. **Muto V, Stellacci E, Lamberti AG, Perrotti E, Carrabba A, Matera G, Sgarbanti M, Battistini A, Liberto MC, Focà A.** 2011. Human papillomavirus type 16 E5 protein induces expression of beta interferon through interferon regulatory factor 1 in human keratinocytes. J Virol **85:**5070-5080.

55. **Terenzi F, Saikia P, Sen GC.** 2008. Interferon-inducible protein, P56, inhibits HPV DNA replication by binding to the viral protein E1. EMBO J **27:**3311-3321.

56. **Saikia P, Fensterl V, Sen GC.** 2010. The inhibitory action of P56 on select functions of E1 mediates interferon's effect on human papillomavirus DNA replication. J Virol **84:**13036-13039.

57. **Warren CJ, Xu T, Guo K, Griffin LM, Westrich JA, Lee D, Lambert PF, Santiago ML, Pyeon D.** 2015. APOBEC3A functions as a restriction factor of human papillomavirus. J Virol **89:**688-702.

58. **Meuris F, Carthagena L, Jaracz-Ros A, Gaudin F, Cutolo P, Deback C, Xue Y, Thierry F, Doorbar J, Bachelerie F.** 2016. The CXCL12/CXCR4 signaling pathway: A new susceptibility factor in human papillomavirus pathogenesis. PLoS Path **12:**e1006039.

59. **Sivaprasad U, Kinker KG, Ericksen MB, Lindsey M, Gibson AM, Bass SA, Hershey NS, Deng J, Medvedovic M, Khurana Hershey GK.** 2015. SERPINB3/B4 contributes to early inflammation and barrier dysfunction in an experimental murine model of atopic dermatitis. J Invest Dermatol **135:**160-169.

60. **Hong S, Mehta KP, Laimins LA.** 2011. Suppression of STAT-1 expression by human papillomaviruses is necessary for differentiation-dependent genome amplification and plasmid maintenance. J Virol **85:**9486-9494.

61. **Smith SP, Scarpini CG, Groves IJ, Odle RI, Coleman N.** 2016. Identification of host transcriptional networks showing concentration-dependent regulation by HPV16 E6 and E7 proteins in basal cervical squamous epithelial cells. Sci Rep **6:**29832.

62. **Gray E, Pett M, Ward D, Winder DM, Stanley MA, Roberts I, Scarpini CG, Coleman N.** 2010. *In vitro* progression of human papillomavirus 16 episome-associated cervical neoplasia displays fundamental similarities to integrant-associated carcinogenesis. Cancer Res **70:**4081-4091.

63. **Smith JHF, Patnick J.** 2013. Achievable standards, benchmarks for reporting and criteria for conducting cervical cytopathology. (2nd Edition). Sheffield. NHS Cancers Screening Programmes – available from www.cancerscreening.nhs.uk/cervical/publications/ nhscsp01.html

64. **Luesley, D. Leeson, S.** 2010. Colposcopy and Programme Management. NHS Cancer Screening Programmes. (2nd Edition) Sheffield. NHS Cancers Screening

Programmes – available from www.cancerscreening.nhs.uk/cervical/publications/nhscsp20.html

65. **Hirschowitz L.** 2012. Histopathology reporting in cervical screening-an integrated approach. (2nd Edition) Sheffield. NHS Cancers Screening Programmes – available from http://www.cancerscreening.nhs.uk/cervical/publications/cc-04.html

66. **Arbyn M, Ronco G, Anttila A, Meijer CJLM, Poljak M, Ogilvie G, Koliopoulos G, Naucler P, Sankaranarayanan R, Peto J.** 2012. Evidence regarding human papillomavirus testing in secondary prevention of cervical cancer. Vaccine **30, Supplement 5:**F88-F99.

67. **Anders S, Huber W.** 2010. Differential expression analysis for sequence count data. Genome Biol **11:**1-12.

68. **The Gene Ontology Consortium, Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G.** 2000. Gene Ontology: tool for the unification of biology. Nat Genet **25:**25-29.

69. **Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M.** 1999. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res **27:**29-34.

70. **Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman W-H, Pagès F, Trajanoski Z, Galon J.** 2009. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics **25:**1091-1093.

Figure Legends

Figure 1. HPV16 infection induces massive changes in the keratinocyte transcriptome. A. Expression levels of keratinocyte differentiation markers and viral L1 protein in undifferentiated (U = monolayer culture for 5 days) and differentiated (D = monolayer culture for 13 days) W12 and NIKS16 cells. GAPDH is shown as a loading control. B. Expression levels of viral E2 and E4 proteins at 8 (mid differentiation phase) and 13 (differentiated) days of a time course of NIKS16 differentiation in monolayer culture. C. Time course of involucrin expression over a 13 day differentiation period (monolayer cells are mostly undifferentiated after 5 days culture and fully differentiated after 13 days of culture) for NIKS and NIKS16 cells. invol, involucrin. D. mRNA numbers expressed versus the level of expression of each individual mRNA in undifferentiated (U) versus differentiated (D) NIKS cells. E. mRNA numbers expressed versus the level of expression of each individual mRNA in undifferentiated (U) versus differentiated (D) NIKS16 cells. Invol, involucrin. K10, Keratin 10.

Figure 2. Keratinocyte differentiation and epithelial barrier function is altered by HPV infection. Significant changes in expression ($>\log_2=1.8$; 3.5-fold) of proteins involved in keratinocyte differentiation and epithelial barrier function comparing HPV16-infected, differentiated NIKS keratinocytes to uninfected, differentiated NIKS keratinocytes. These are the mean values from three separate RNASeq experiments A. Markers of differentiation (filaggrin, loricrin, involucrin and transglutaminase (TGM1)). B. Keratins (K). C. Desomosomal proteins, desmogleins (DSG) 1 and 4 and desmocoilin (DSC). D. Gap junction proteins, connexins (Cx) 26, 30.2 and 32. E. Claudins. F. Cadherins. G. small proline rich proteins (SPRRs).

Figure 3. ClueGO analysis of significantly up-and down-regulated genes in HPV16-infected, differentiated NIKS keratinocytes compared to uninfected, differentiated NIKS keratinocytes. We used CluePedia, which extends ClueGO (70) functionality down to

genes, and visualizes the statistical dependencies (correlation) for markers of interest from the experimental data. A. Gene ontology (GO) pathway terms specific for up-regulated genes. B. GO pathway terms specific for down-regulated genes. The bars represent the numbers of genes associated with the term on the left hand side. The percentage of altered genes is shown above each bar. Red asterisks refer to significance. C. Functionally grouped network for up-regulated genes. D. Functionally grouped networks for down-regulated genes. Only the label of the most significant term per group is shown. The size of the nodes reflects the degree of enrichment of the terms. The network was automatically laid out using the organic layout algorithm in Cytoscape. Only functional groups represented by their most significant term were visualized in the network. Padj< 0.05 changes were analysed.

Figure 4. Western blot analysis of proteins levels encoded by selected, significantly altered mRNAs (Table 2). Protein extracts were prepared from undifferentiated and differentiated HPV-negative NIKS and HPV16-positive NIKS16 cell populations. Much greater levels of involucrin (invol) were detected in the differentiated, compared to the undifferentiated cell populations indicating differentiation was achieved. GAPDH was used a protein loading control. A. Protein levels corresponding to significantly up-regulated mRNAs. B. Protein levels corresponding to significantly down-regulated mRNAs. U, undifferentiated. D, differentiated.

Figure 5. Expression levels of selected, significantly altered mRNAs in different grades of HPV-associated pre-neoplastic cervical disease. mRNA expression levels were calculated from qPCR data using GAPDH and beta-actin as the internal controls and expressed relative to levels in a single sample of differentiated, HPV16-positive W12 cell RNA that was included in every PCR run. NDD, no detectable disease/borderline, all HPV-negative. Low grade, cervical intraepithelial neoplasia 1 (CIN1), all HPV-positive. High grade disease, cervical intraepithelial neoplasia 1 (CIN3), all HPV-positive (Supplementary Table 4).

Supplementary Figure 1. Interactome of positively and negatively changed genes comparing differentiated NIKS with differentiated NIKS16 cells.

A. interactome of genes linked through statistical correlation of up-regulated genes from the experimental data (p-value <0.05). B. interactome of genes linked through statistical correlation of down-regulated genes from the experimental data (p-value <0.05). Grey lettering and an open blue diamond indicates genes identified in the RNASeq data set. Black lettering indicates linked genes. Nodes for genes identified in the data set are indicated by red boxes. Dots/lines surrounding nodes indicate the numbers of linked pathways. The pathways analysis was produced using Cluepedia (http://apps.cytoscape.org/apps/cluepedia) [24].

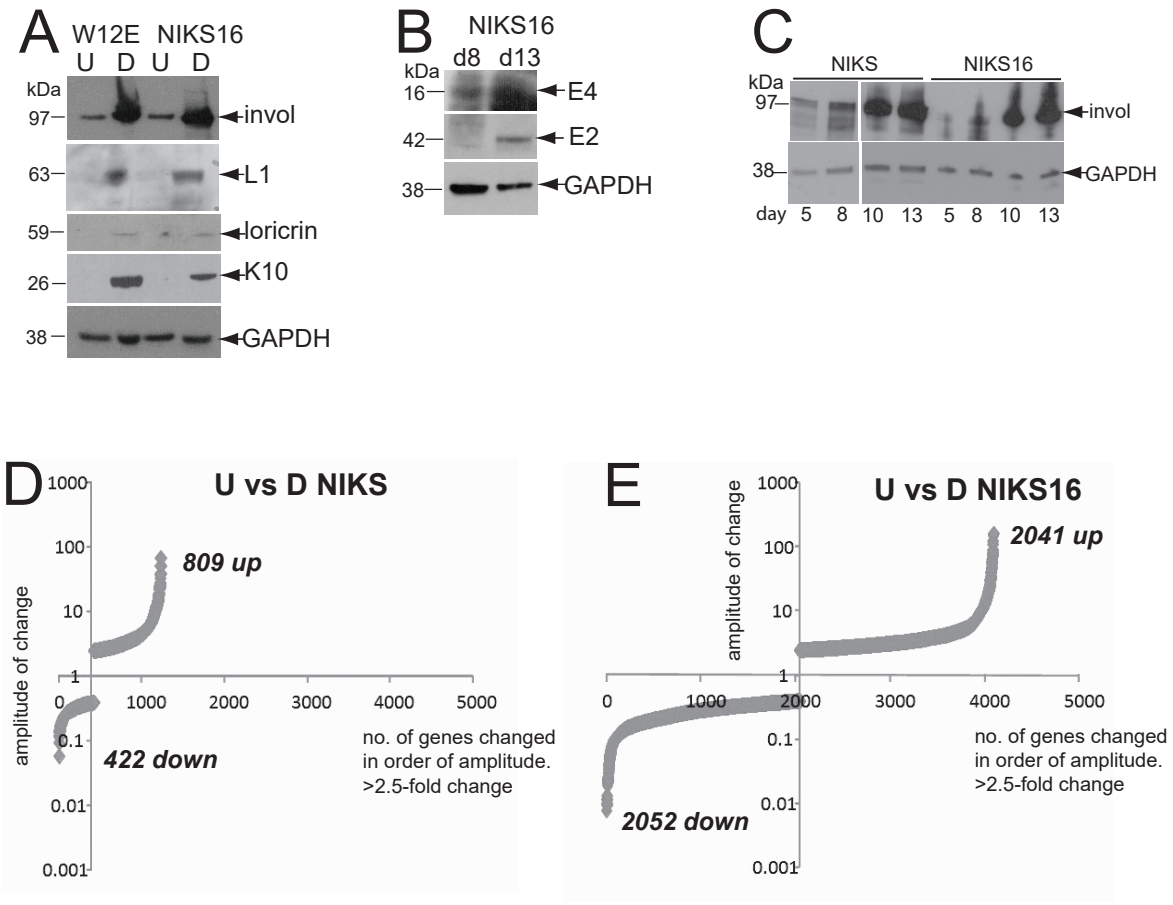

Figure 1. HPV16 infection induces massive changes in the keratinocyte transcriptome. A. Expression levels of keratinocyte differentiation markers and viral L1 protein in undifferentiated (U = monolayer culture for 5 days) and differentiated (D = monolayer culture for 13 days) W12 and NIKS16 cells. GAPDH is shown as a loading control. B. Expression levels of viral E2 and E4 proteins at 8 (mid differentiation phase) and 13 (differentiated) days of a time course of NIKS16 differentiation in monolayer culture. C. Time course of involucrin expression over a 13 day differentiation period (monolayer cells are mostly undifferentiated after 5 days culture and fully differentiated after 13 days of culture) for NIKS and NIKS16 cells. invol, involucrin. D. mRNA numbers expressed versus the level of expression of each individual mRNA in undifferentiated (U) versus differentiated (D) NIKS cells. E. mRNA numbers expressed versus the level of expression of each individual mRNA in undifferentiated (U) versus differentiated (D) NIKS16 cells. Invol, involucrin. K10, Keratin 10.

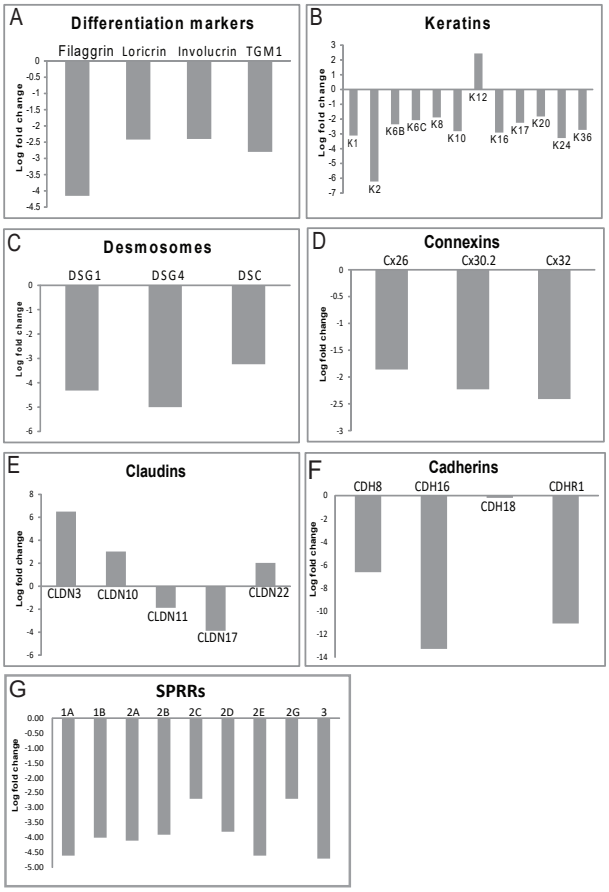

Figure 2. Keratinocyte differentiation and epithelial barrier function is altered by HPV infection. Significant changes in expression (>log2=1.8; 3.5-fold) of proteins involved in keratinocyte differentiation and epithelial barrier function comparing HPV16-infected, differentiated NIKS keratinocytes to uninfected, differentiated NIKS keratinocytes. These are the mean values from three separate RNASeq experiments A. Markers of differentiation (filaggrin, loricrin, involucrin and transglutaminase (TGM1)). B. Keratins (K). C. Desomosomal proteins, desmogleins (DSG) 1 and 4 and desmocoilin (DSC). D. Gap junction proteins, connexins (Cx) 26, 30.2 and 32. E. Claudins. F. Cadherins. G. small proline rich proteins (SPRRs).

Figure 3 Klymenko et al.

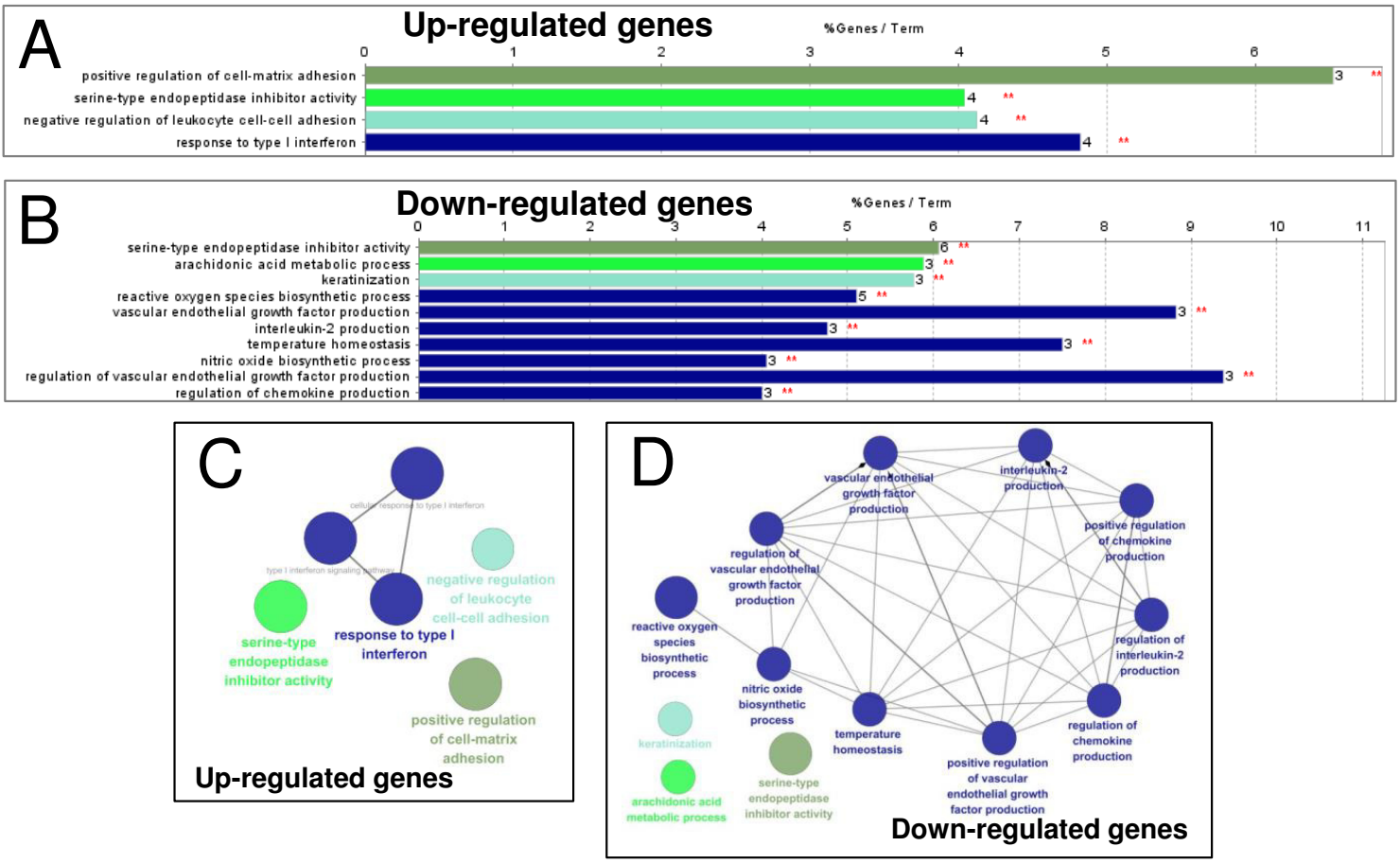

Figure 3. ClueGO analysis of significantly up-and down-regulated genes in HPV16-infected, differentiated NIKS keratinocytes compared to uninfected, differentiated NIKS keratinocytes. We used CluePedia, which extends ClueGO (70) functionality down to genes, and visualizes the statistical dependencies (correlation) for markers of interest from the experimental data. A. Gene ontology (GO) pathway terms specific for up-regulated genes. B. GO pathway terms specific for down-regulated genes. The bars represent the numbers of genes associated with the term on the left hand side. The percentage of altered genes is shown above each bar. Red asterisks refer to significance. C. Functionally grouped network for up-regulated genes. D. Functionally grouped networks for down-regulated genes. Only the label of the most significant term per group is shown. The size of the nodes reflects the degree of enrichment of the terms. The network was automatically laid out using the organic layout algorithm in Cytoscape. Only functional groups represented by their most significant term were visualized in the network. Padj< 0.05 changes were analysed

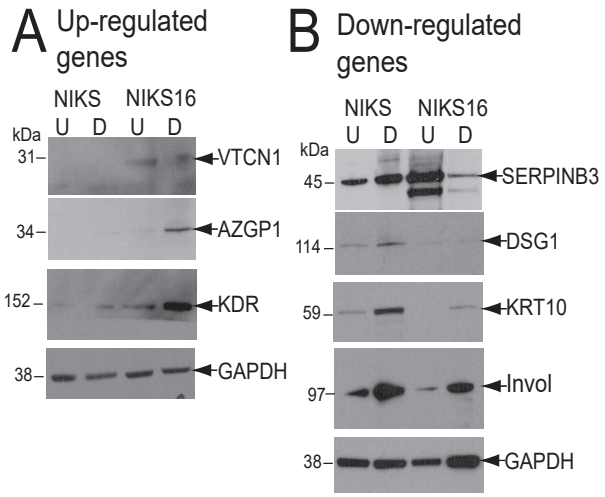

Figure 4. Western blot analysis of proteins levels encoded by selected, significantly altered mRNAs (Table 2). Protein extracts were prepared from undifferentiated and differentiated HPV-negative NIKS and HPV16-positive NIKS16 cell populations. Much greater levels of involucrin (invol) were detected in the differentiated, compared to the undifferentiated cell populations indicating differentiation was achieved. GAPDH was used a protein loading control. A. Protein levels corresponding to significantly up-regulated mRNAs. B. Protein levels corresponding to significantly down-regulated mRNAs. U, undifferentiated. D, differentiated.

Klymenko et al. Figure 5

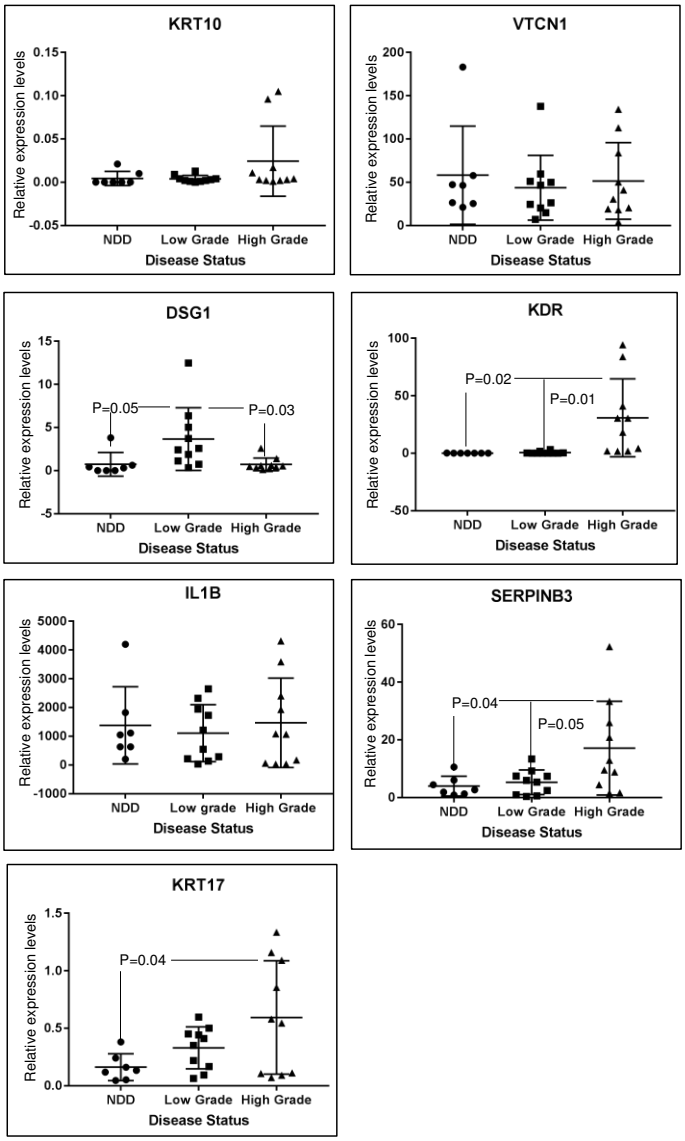

Figure 5. Expression levels of selected, significantly altered mRNAs in different grades of HPV-associated pre-neoplastic cervical disease. mRNA expression levels were calculated from qPCR data using GAPDH and beta-actin as the internal controls and expressed relative to levels in a single sample of differentiated, HPV16-positive W12 cell RNA that was included in every PCR run. NDD, no detectable disease/borderline, all HPV-negative. Low grade, cervical intraepithelial neoplasia 1 (CIN1), all HPV-positive. High grade disease, cervical intraepithelial neoplasia 1 (CIN3), all HPV-positive (Supplementary Table 4).