Pro-metastatic gene expression, immune evasion and an altered HPV spectrum characterize an aggressive subtype of cervical cancer.

Ankur Chakravarthy[1,], Stephen Henderson[2], Cindy Dong[3], Nerissa Kirkwood[3], Maxmilan Jeyakumar[3], Jacqueline McDermott[4], Xiaoping Su[5], Nagayasau Egawa[6], Christina S Fjeldbo[7], Mari Kyllesø Halle[8], Camilla Krakstad[8], Afschin Soleiman[9], Susanne Sprung[10], Heidi Lyng[7], Heidi Fiegl[11], Helga Salvesen[8], John Doorbar[6], Kerry Chester[4,*], Andrew Feber[12,*], Tim R Fenton[3,*].

[1] Princess Margaret Cancer Centre, University Health Network. Toronto, Ontario, Canada

[2] UCL Cancer Institute, Bill Lyons Informatics Centre, University College London, London, UK

[3] School of Biosciences, University of Kent, Canterbury, Kent, UK

[4] UCL Cancer Institute, University College London, London, UK

[5] MD Anderson Cancer Center, Houston, Texas, USA

[6] Department of Pathology, University of Cambridge, Cambridge, UK

[7] Department of Radiation Biology, Oslo University Hospital, Oslo, Norway

[8] Department of Obstetrics and Gynaecology, Haukeland University Hospital, Bergen, Norway; Centre for Cancer Biomarkers, Department of Clinical Science, University of Bergen, Norway

[9] INNPATH, Institute of Pathology, Tirol Kliniken Innsbruck, Innsbruck, Austria.

[10] Institute of Pathology, Medical University of Innsbruck, Innsbruck, Austria.

[11] Department of Obstetrics and Gynaecology, Medical University of Innsbruck, Innsbruck, Austria.

[12] Division of Surgery and Interventional Science, University College London, London, UK

*Correspondence to: k.chester@ucl.ac.uk; a.feber@ucl.ac.uk or t.fenton@kent.ac.uk

Abstract

Cervical cancer is caused by carcinogenic human papillomavirus infection and represents one of the leading causes of cancer death worldwide. Effective means of tumour classification are required for better disease understanding. We performed an integrated multi-omic analysis of 655 cervical cancers, using epigenomic and transcriptomic signatures to discover two distinct cervical cancer subtypes we named "typical" and "atypical". Typical tumours were largely HPV16-driven and frequently displayed an 'immune-hot' tumour microenvironment. Atypical tumours were associated with poor prognosis; they were more likely to be driven by HPVs from the HPV18-containing α7 clade, displayed distinct genomic aberrations, greater evidence of past immunoediting and a microenvironment associated with immune-evasion and failure of anti-PD1 checkpoint inhibition. The finding that atypical tumours encounter stronger anti-tumour immune responses during development may explain the lower frequency at which α7 HPV infected-lesions progress from pre-invasive disease. However those escaping this selection pressure evolve into aggressive tumours (independent of HPV-type) in which more intensive adjuvant treatment may be warranted.

--

Despite screening and the introduction of prophylactic human papillomavirus (HPV) vaccination in developed countries, cervical cancer continues to be one of the leading worldwide causes of cancer-related deaths in women. Prognosis for patients with metastatic disease remains poor, thus new treatments and effective molecular markers for patient stratification are urgently required. Cervical cancer is caused by at least 14 high-risk human papillomaviruses (hrHPVs), with HPV16 and HPV18 together accounting for over 70% of cases worldwide[1]. Although among the hrHPVs, HPV16 and HPV33 (both from the α9 clade) infections are associated with much higher chance of progression to

high-grade neoplasias but the presence of HPV16 in particular, has been linked to improved survival in cervical cancer and HPV+ head and neck squamous cell carcinoma (HNSCC)[1-3]. We have previously shown in an interim analysis of TCGA data that the majority of HPV16+ cervical tumours fall into a good prognosis group based on similarity at the DNA methylation level to HPV16+ HNSCC and penile cancer[4]. This apparent paradox, in which HPV16 behaves more aggressively in the context of tumour development but in which the resulting tumours are less aggressive, suggests fundamental differences in the natural history of tumours driven by different HPV types. Integrated molecular analysis of 228 cervical cancers by TCGA reported differential micro-RNA expression and transcription factor activation between tumours harbouring different HPV types[5]. However, a biological explanation for the apparent type-specific clinical differences noted in the above studies remains unexplored, largely due to a lack of sufficiently large cohorts for which HPV typing, molecular and clinical data are available.

To address this question we used data from 281 cervical tumours profiled by TCGA[5] as our discovery cohort (**Table 1, Table S1**) and 374 cervical cancers from three European centres, all with detailed clinical annotation and long term follow-up as our validation cohort[6,7] (**Table 1, Table S2**). Representing to our knowledge, the largest study of its kind in cervical cancer, we defined two cervical cancer subtypes; an "atypical" aggressive subtype defined by a lymphocyte-depleted microenvironment and evidence for epithelial-mesenchymal transition, and a "typical" subgroup comprised almost entirely of tumours harbouring α9 HPVs and associated with longer overall survival. The molecular, cellular and clinical differences identified between typical and atypical tumours also reveal new potential therapeutic options for the treatment of cervical cancers.

1 ## Results

2

3

		TCGA Training Cohort	European Validation Cohort
Histology			
	Adenocarcinoma	44	32
	Squamous cell Carcinoma	237	335
	Adenosquamous	0	7
Stage			
	I	153	98
	II	61	194
	III	43	65
	IV	17	17
	NA	7	0
Age	Median (Range)	46 (20-88)	51 (22-91)
HPV Sub Type			
	16	166	160
	18	38	140
	45	22	15
	Other	55	50
	Negative	0	11
Survival Status			
	Alive	226	275
	Dead	55	99
Cluster Assignment			
	Typical	231	301
	Atypical	50	73

4

5 **Table 1:** Summary of clinicopathological characteristics for the two cervical cancer cohorts.

Generation of HPV type-specific gene expression and DNA methylation profiles in cervical cancer

To look for associations between HPV type and overall survival while avoiding confounding from histology and advanced stage, we initially reduced our discovery cohort to 139 stage I and II squamous cell carcinomas (SCCs), which we confirmed using VirusSeq[8] to be transcript-positive for at least one of the three most common HPV types (HPV16, HPV18 and HPV45), and for which the covariates age and tumour stage were available. Multivariate Cox regression identified a significantly worse prognosis in HPV45-driven tumours relative to HPV16 (HR = 5.040, $p < 1e-3$), while HPV18+ tumours exhibited an intermediate prognosis (**Figure 1A**).

Modelling transcriptomic and epigenomic (DNA methylation) differences between HPV16 and HPV45-associated early-stage tumours identified 713 DEGs (Differentially Expressed Genes, FDR=0.01, FC > 2; **Figure 1B**, **Table S3**) and 689 MVPs (Methylation Variable Positions, delta-Beta 0.1, FDR < 0.01), (**Figure 1C, Table S4**). 10 DEGs previously shown to be aberrantly expressed in HPV-associated cancers from different anatomic sites[9] displayed greater dysregulation in HPV16+ than in HPV45+ cervical tumours, while two (*PLOD2* and *KRT18*) were more strongly upregulated in HPV45+ tumours (**Figure 1D**). Several DEGs were also differentially methylated between HPV16+ and HPV45+ tumours (**Figure S1**). These findings indicate molecular differences between cervical cancers driven by different hrHPV types which may manifest in clinical differences.

1

2 ## Development and validation of a prognostic classification.

3 Interestingly, when using either the 713 DEG or 613 MVP signatures, a minority (10% gene expression-
4 based, 11% DNA methylation-based) early-stage HPV16+ SCCs clustered with the HPV45+ tumours in
5 the discovery cohort (**Figures 1B and 1C**). We subsequently used consensus clustering based on the
6 713 DEG signature, which identified two robust clusters, which we termed "typical" and "atypical"
7 cervical cancer subtypes. 6 of 104 HPV16+ tumours co-clustered with the majority (13 of 17) of
8 HPV45+ tumours in the atypical subgroup, which also contained 28% (5/18) of the HPV18+ tumours
9 (**Figure 2A, Table S1**). To assess if these typical and atypical subgroups also exist at the epigenetic
10 level, we developed a DNA methylation Support Vector Machine (SVM) classification model, using
11 TCGA DNA methylation data reduced to 178 CpG sites at which methylation differed significantly
12 between tumours in typical versus atypical clusters (**Figure 2B,** mean delta-Beta > 0.3, FDR < 0.01,

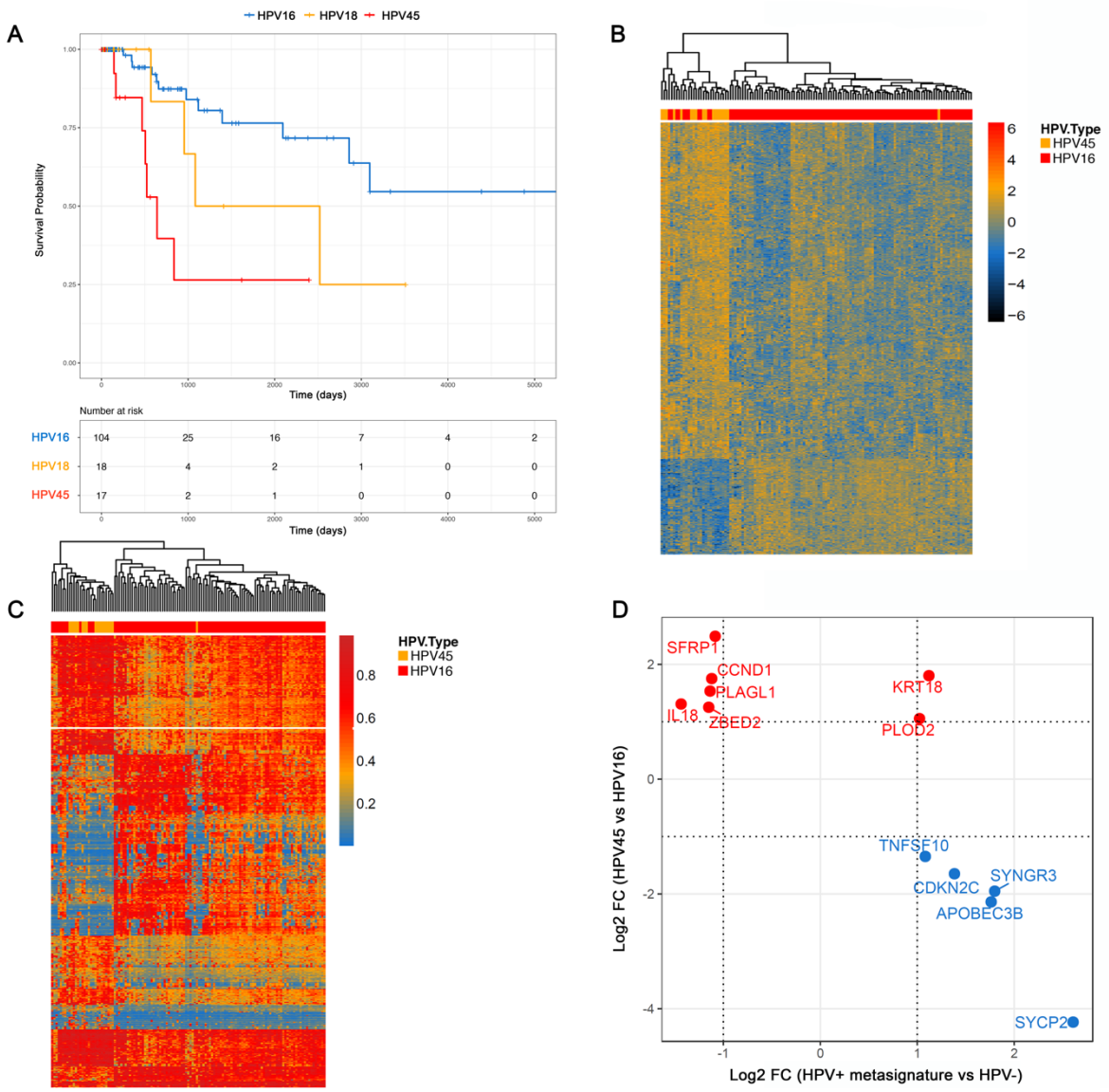


Figure 1. Clinical and molecular variation among TCGA cervical cancers driven by different HPV types. A) HPV45+ early stage Cervical Squamous Cancers display markedly worse prognosis compared to HPV16+ cancers with HPV18+ tumours showing intermediate survival. HR and p-value from Cox regression controlling for stage. **B)** Comparisons between HPV45+ and HPV16+ tumours identify large scale variation in transcriptional and **C)** epigenetic (genome-wide DNA methylation) profiles. **D)** Twelve genes from a pan-tissue signature for HPV-driven tumorigenesis show significant variation between HPV16+ and HPV45+ tumours.

Table S5). Using this signature we allocated cluster membership to a further 374 cervical cancers from our validation cohort (**Figure 2C and Table S2)**. Adenocarcinomas (12 of 32) and adenosquamous

carcinomas (5 of 7) were more likely to be classified as atypical than SCCs (56/335, p = 4.526e-09, Fisher's Exact Test) and again, the majority (141 of 160) HPV16+ tumours were designated as typical (**Figure 2C**). We could accurately predict the DEG-based cluster assignment using the MVP signature for all validation cohort tumours for which gene expression data were available (RNA-seq for Bergen samples (n = 65) and Illumina HumanHT-12 V4.0 expression beadchip arrays for Oslo samples (n = 268), **Figure S2A**), confirming that both MVP and DEG signatures classify the same tumours as typical or atypical. Single-sample gene set enrichment analysis (ssGSEA)[10] of the same tumours confirmed differential expression of the signature genes in tumours classified as typical or atypical using DNA methylation (**Figure S2B**). Having derived our typical and atypical clusters directly from the HPV45 vs HPV16 expression signature and shown that they were consistent whether assigned from gene expression or from DNA methylation data, for clarity we henceforth refer to all comparisons as atypical and typical. Integrating DNA and RNA-based HPV typing where available, we confirmed that co-infection with HPV45 or other HPV types was not responsible for the assignment of HPV16 transcript-positive samples to the atypical group. In both the validation and discovery cohorts, HPV types from the α7 clade (HPV18, 45, 59, 68, 70) were strongly enriched in atypical tumours (atypical tumours were 2.3X more likely to harbour α7 HPVs than typical tumours, p = 1.85e-14 Fisher's Exact Test).

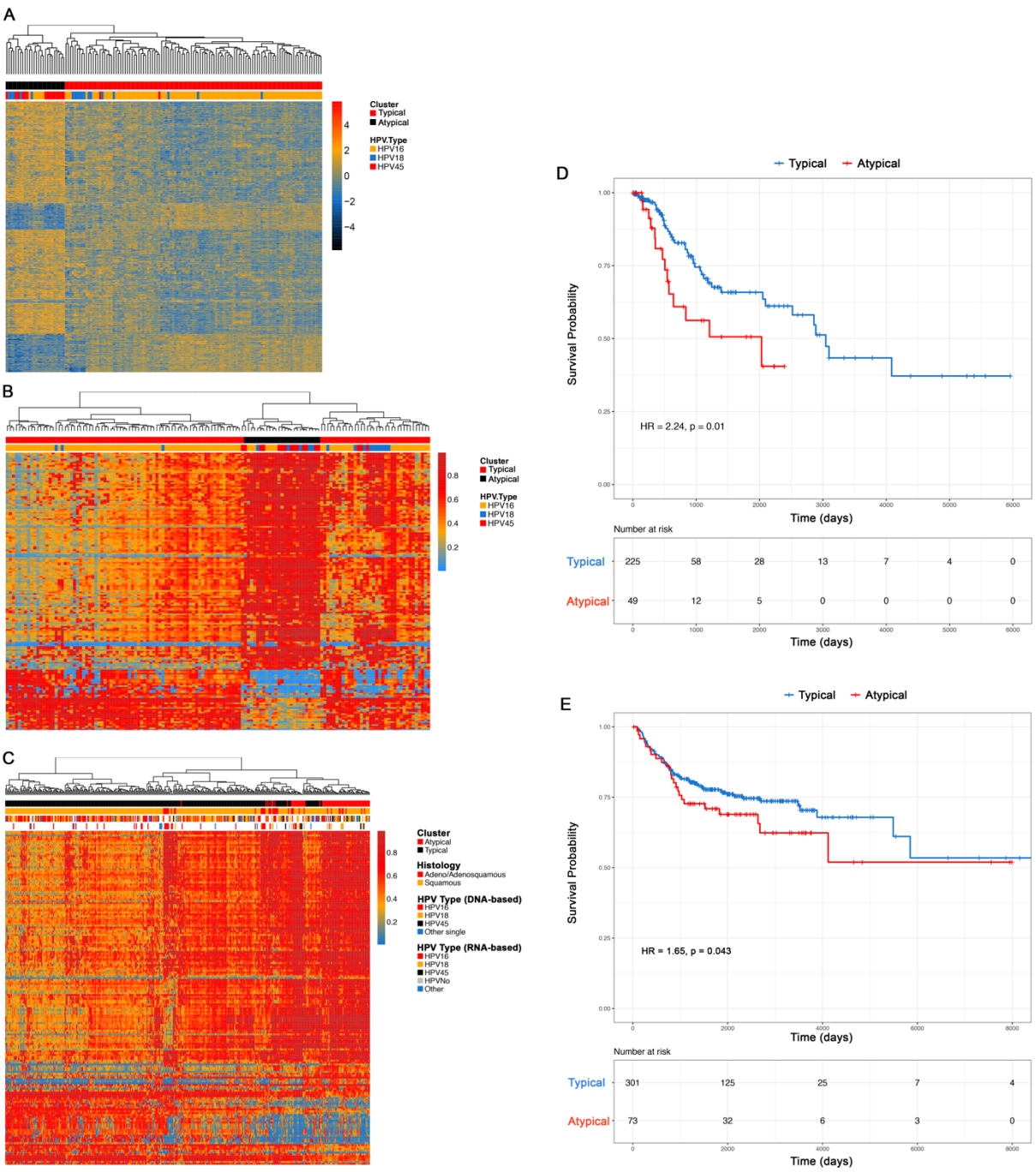


Figure 2. Derivation of two type-associated prognostic subgroups in cervical cancer and validation across independent cohorts. A) HPV45-like transcriptional profiles are also shared by small numbers of HPV16+ and HPV18+ tumours, coalescing into HPV45-like ('Atypical') and HPV16-like ('Typical') clusters. **B)** A signature of DNA methylation (dB > 0.3, FDR < 0.01) separates these groups based on consensus clustering (see methods for details). **C)** The methylation patterns are reproduced in a validation dataset from three European centres (n = 374). **D)** Survival curves and statistics from multivariate Cox regression of overall survival in TCGA cervical cancer

cohort stratified by cluster. **E)** Survival curves and statistics from multivariate Cox regression of overall survival in the European validation cohort stratified by cluster.

Multivariate analysis of survival data from the 274 TCGA tumours for which tumour stage was available identified a significant prognostic difference between the typical and atypical subgroups (**Figure 2D;** HR = 2.24, p = 0.01). This difference became even greater when restricting to stage I/II tumours (**Figure S3A**; n = 139, HR = 4.88, p = 0.0006), and was retained even upon removal of the HPV45+ tumours (**Figure S3B**; n = 122, HR = 4.91, p = 0.03). Cox regression stratifying by histology and controlling for FIGO stage and treatment (surgery alone, surgery with radio-chemotherapy and surgery with chemotherapy alone) identified typical/atypical status to be an independent predictor of overall survival in the validation cohort (n = 374, HR = 1.65 , p = 0.043**, Figure 2E)**. Again, the survival difference between the typical and atypical groups was greater when stage IV tumours were excluded (n = 357, HR = 1.73, p = 0.04) and was most pronounced in stage II tumours (n = 194, HR = 2.59, p = 0.02).

Evidence for epithelial to mesenchymal transition (EMT) in atypical tumours.

Gene set enrichment using Ingenuity Pathway Analysis 'Diseases and Functions Ontology' (**Figure S4**, **Table S6**) which identified cellular movement as the most activated pathway in atypical tumours, with 85 of 121 genes in the set expressed consistent with increased metastatic potential. Prominent pro-metastatic genes in this pathway included the transcription factor *SNAI1* (a master regulator of EMT that accompanies invasion through the basement membrane and dissemination from the primary tumour[11] , fibronectin 1, which is known to trigger EMT-associated transcriptional cascades[12], *RHOF*, a prominent player in invasion through pseudofilopodia formation[13] and *VEGFC,* involved in prometastatic lymphangiogenesis[14]. Multiple other gene sets also pertaining to cell movement were strongly enriched and associated with high activation z-scores in atypical tumours (**Figure S4, Table**

S6). Moreover, these tumours expressed high levels of Transforming Growth Factor (*TGFB1* and *TGFB2,* **Table S3**), which have been implicated as key inducers of EMT that are potentially amenable to therapeutic targeting (reviewed in[15]). Given these findings, we examined the relationship between our subtypes and the poor-prognosis cervical cancer EMT cluster defined by TCGA based on reverse phase protein array (RPPA) data[5]. 62% of atypical TCGA tumours with RPPA data available belong to the EMT cluster compared with only 20% of typical tumours. Consistent with the proteomic classification, atypical tumours displayed higher EMT gene expression scores, as defined by TCGA[5], than typical tumours (**Figure S5).** Upstream regulator analysis identified *EZH2* and *SMARCA4*, both chromatin modifiers, as the leading differentially-activated regulators in atypical tumours (**Figure S6, Table S7**). Notably *EZH2* expression has previously been linked to poor prognosis in cervical cancer[16]. Other key activated regulators include β-catenin and HIF1α, both of which have been linked to chemo or radio- resistance and poor prognosis in cervical cancer[17-19]. These analyses link the atypical expression signature to several independently discovered poor prognostic factors.

Genomic analyses of prognostic clusters

To search for genomic differences between typical and atypical cancers, we analysed mutation (WEX) and copy number data from samples with matched methylation data[20]. We first generated segmented copy number data for all tumours (combining the TCGA and validation cohort samples for which the necessary data were available for maximum statistical power), which identified 387 focal candidate copy number alterations at FDR < 0.1. Following binomial regression, we identified 12 discrete copy

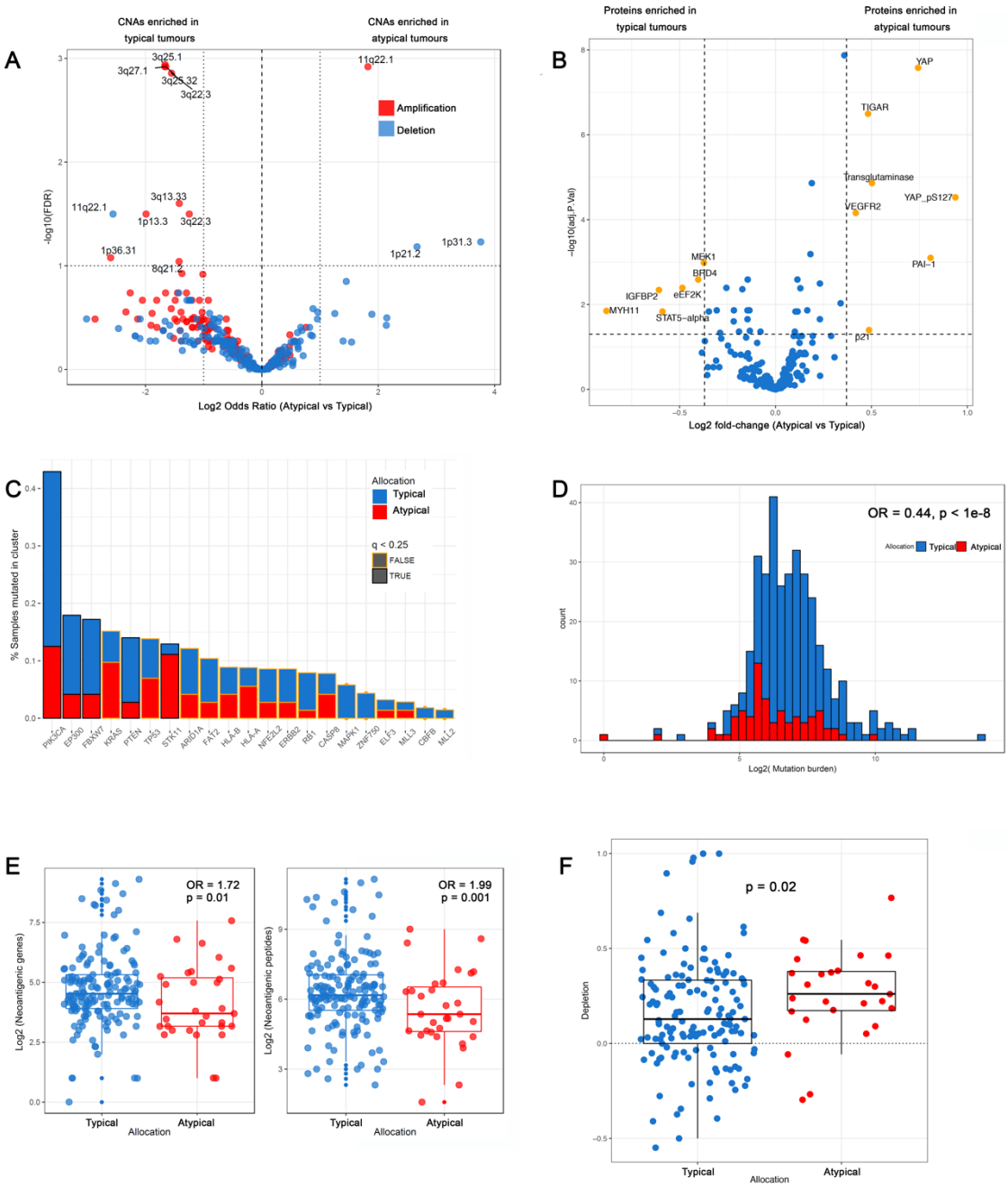


Figure 3. Genomic differences between cervical cancer subgroups. A) Volcano plot showing differences in GISTIC copy number peak frequencies between typical and atypical tumours, with -log10(FDR) on the y axis and odds ratio on the x axis. **B)** Volcano plot showing differentially abundant proteins and phospho-proteins (FDR < 0.05, FC > 1.3, represented by yellow dots) between typical and atypical TCGA tumours, as measured by Reverse Phase Protein Array. **C)** Bar chart showing mutation frequencies for candidate driver mutations in typical and atypical cancers. Y axis indicates percentage of tumours mutated within tumour subtype, and the outline colour indicates statistical significance of differences in mutation frequencies. **D)** Histograms show overall mutational

burdens are greater in HPV16-like cancers. Odds ratios and p values are from a negative binomial GLM. **E)** Neoantigen burdens are elevated in typical tumours cervical cancers (estimates and p-values from negative binomial regression). **F)** Atypical tumours display greater evidence of past immunoediting as measured by depletion of predicted neoantigens versus total mutations.

number alterations between typical and atypical clusters (**Table S8**, **Figure 3A;** FDR < 0.1, log2 (Odds Ratio) > 1). These included 1p31.3 loss and 11q22.1 gain which were more prevalent in atypical tumours and multiple 3q gains and 1p13.3 gain, and 11q22.1 loss, which were disproportionately common in typical cancers (**Figure 3A**). Notably, the 11q22.1 gain seen in atypical tumours are centred on the Yes-Associated Protein 1 (YAP1): a key transcription factor downstream of the HIPPO signalling pathway. Analysis of Reverse Phase Protein Assay (RPPA) data from TCGA also revealed significantly higher YAP1 protein expression in the atypical tumours (**Fig 3B**). We confirmed that those same cases with *YAP1* amplification (7/28 atypical tumours and 12/141 typical tumours) also showed increased YAP1 mRNA and protein expression.

We next compared the somatic mutation rates in a defined set of candidate driver genes, using a binomial regression. This identified *PIK3CA, FBXW7* and *PTEN* mutations as disproportionately more common in typical cancers and loss-of-function *STK11* mutations as more frequent in atypical tumours (FDR < 0.25, **Figure 3C),** *STK11* (LKB1) is also under-expressed in atypical tumours compared with typical tumours (**Table S1**). We observed a higher overall mutation burden in typical tumours (OR= 0.48, p = 1.4e-5, **Figure 3D**), leading us to investigate whether there is also a difference in neoantigen load between the subgroups. Fitting a negative binomial GLM to neoantigen data from TCGA (31 atypical, 157 typical for which neoantigen estimates were available from the Cancer Immunome Atlas [21]) revealed markedly more predicted neoantigens in typical tumours, at both the gene and individual MHC class 1-binding peptide level (OR = 1.72, p = 0.01 and OR = 1.99, p = 0.001 respectively, **Figure**

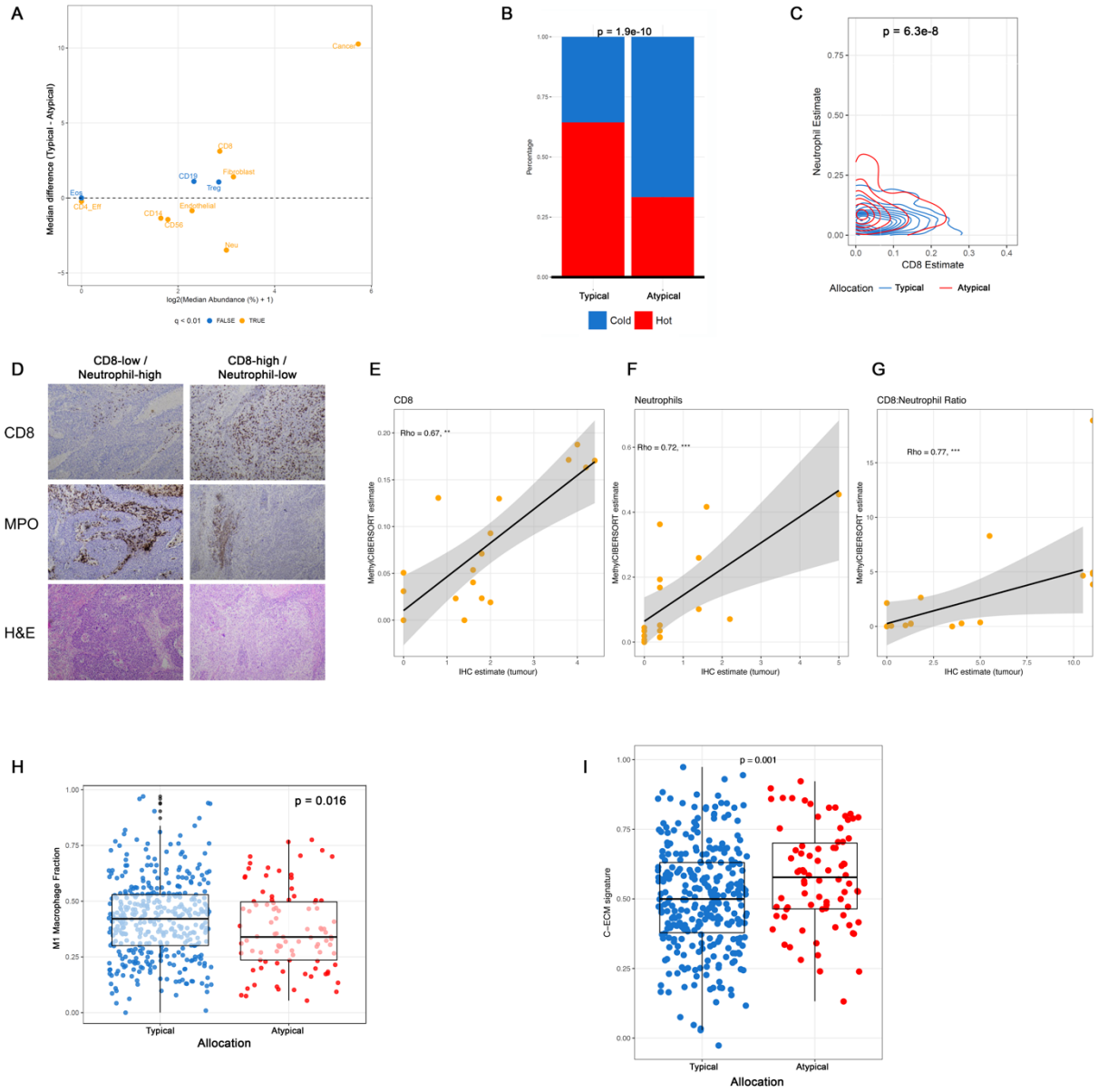
3E). Interestingly, the ratio of expected versus observed neoantigens per tumour (neoantigen depletion, see methods) is greater in atypical tumours (p = 0.02 (Wilcoxon's Rank Sum test), **Fig 3F**), suggesting more extensive immunoediting during their development and leading us to compare the tumour immune microenvironment between subgroups.

Immunological analyses of prognostic clusters implicate microenvironmental differences and highlight potential therapeutic interventions.

The nature of the tumour immune microenvironment, particularly the abundance of tumour infiltrating lymphocytes (TILs) is a strong prognostic factor in HPV-associated cancers[9,22-26]. Pathway analysis revealed activation of granulocyte and aggranulocyte adhesion along with diapedesis in atypical tumours, suggesting increased neutrophil infiltration. Other pathways activated in atypical tumours include inflammatory processes such as Acute Phase Response, TREM1 signalling, complement activation and AHR (Aryl Hydrocarbon Receptor) signalling which are often associated with macrophages (**Figure S7, Table S9**), suggesting a strong microenvironmental component may mediate type-associated pathology. This is supported by the presence of multiple cytokines and chemokines in the atypical-associated transcriptional signature, including *IL11, IL18, IL1B, IL24, IL6, IL8, CCL2, CXCL2, CXCL3, CXCL5*, *TNF* (TNF-α) and *TNFAIP6* (**Table S3**). We next used DNA methylation data to compare the cellular composition of tumours[27], observing differences in the proportions of multiple cell types between the subgroups (**Figure 4A**); most notably decreased CD8+ (cytotoxic T lymphocytes (CTL)), a marked elevation of neutrophil and natural killer (NK)-cells and lower tumour purity in atypical cancers (the latter confirmed using genomic estimates[28] (ABSOLUTE); **Figure S8**, p = 0.02 (Wilcoxon's Rank Sum test). Integrating these cell type estimates with our previously-published pan-cancer immune hot/cold classification[27], we also found significant enrichment for immune-hot

1 cancers in the typical subgroup (**Figure 4B;** p = 1.9e-10), consistent with increased CTL infiltration,

2 CTL:Treg ratios (**Figure S9**) and higher MHC class I neoantigen loads in these tumours.



3

4 **Figure 4. Differences in the immune microenvironment between cervical cancer subgroups. A)** Plot showing

5 median abundances (x-axis) and median differences (%, y-axis) for different cell types estimated using

6 MethylCIBERSORT, with significant differences in orange. **B)** Applying a pan-cancer classifier to DNA methylation

7 data from the cervical cancer samples shows typical tumours are significantly more likely to be immune-hot than

8 atypical tumours. **C)** Atypical tumours display increased neutrophil:CTL ratios as estimated using

9 MethylCIBERSORT. **D)** Representative images showing Immunohistochemistry for MPO (neutrophils) and CD8

in cervical tumour sections. **E)** Correlations between MethylCIBERSORT estimates and immunohistochemistry-based scoring for neutrophils (MPO+). **F)** Correlations between MethylCIBERSORT estimates and immunohistochemistry-based scoring for CD8+ T-cells. **G)** Correlations between MethylCIBERSORT estimates and immunohistochemistry-based scoring for CD8+ T-cell:neutrophil ratios in 17 cervical tumours from the validation cohort. **H)** Typical tumours show higher fractions of proinflammatory (M1) compared to immunosuppressive (M2/M0) macrophages (y axis = ratios, p value from Wilcoxon's Rank Sum Test). **I)** ssGSEA scores (y-axis) showing enrichment for a TGFβ-associated extracellular matrix gene expression signature in atypical tumours.

The estimation of increased neutrophil abundance in atypical tumours (**Figure 4A)** is supported by the upregulation of genes associated with granulocyte diapedesis inferred from pathway analysis (**Table S4**) and with increased ssGSEA scores for a neutrophil gene set derived from publicly-available gene expression data (see methods; p<9.2e-6 (Wilcoxon's Rank Sum test) **Figure S10**). Atypical tumours also exhibit a markedly higher neutrophil:CTL ratio (**Figure 4C**); an established adverse prognostic factor in cervical cancer[29,30]. Enumerating CTLs and neutrophils by IHC for CD8 and MPO respectively in 17 tumours from our validation cohort (representative images shown in **Figure 4D**, see methods for details) revealed strong correlations with MethylCIBERSORT estimates (**Figure 4E-G**). Given the upregulation of multiple CXCR2-ligands (CXCL1, CXCL2, CXCL3, CXCL3) as well as G-CSF, PTGS1, PTGS2, IL-1B, CCL2, IL6 (**Table S3**) and activation of TREM1 signalling (**Figure S7, Table S9**) [31-36] in atypical tumours, all of which are associated with Monocytic-Myeloid Derived Suppressor Cells or tolerogenic macrophages, we used CIBERSORT[37] to derive estimates of macrophage subpopulations for samples with expression data in our dataset. Atypical tumours displayed significantly lower pro-effector (M1) macrophage proportions relative to suppressive (M0 and M2-like) macrophages (p = 0.016 (Wilcoxon's Rank Sum test), **Figure 4H**). Finally, we observed enrichment for a TGFβ-associated extracellular matrix gene expression signature (C-ECM) in atypical tumours that we have previously linked to immune evasion and failure of PD1 blockade in melanoma and bladder cancer[27]. Although total fibroblast

content was similar between typical and atypical tumours (**Figure 4A**) enrichment for the C-ECM signature indicates increased myofibroblast trans-differentiation, which could explain the lack of CTL infiltration and the poor prognosis seen in atypical tumours.

Discussion

Of the studies indicating an association between HPV type and prognosis in cervical cancer, both HPV18 and the α7 clade to which it belongs have been linked to worse prognosis in early-stage tumours[38-40]. Our analysis of TCGA cervical cancer data agrees with these studies; among early-stage SCCs, HPV18 was associated with worse prognosis than HPV16 and for tumours harbouring HPV45 (also an α7 type), prognosis was worse still. Although the numbers are small, this observation could help to explain why in a further study, positivity for HPV16 or 18 was associated with favourable prognosis compared with other HPV types or no HPV detection[41].

Examining the molecular alterations and tumour microenvironment in a large set of tumours provided insight into the complex interplay of HPV type, histology and changes to the host genome that drive cervical carcinogenesis. Firstly, when clustering based on gene expression differences between HPV16+ and HPV45+ tumours, a small minority of HPV16+ tumours co-clustered with the majority of HPV45+ tumours, while HPV18+ tumours were found in both clusters. These clusters were robust; they could be recapitulated at the DNA methylation level and in a larger independent cohort. In both cohorts, we observed a clear difference in prognosis between the clusters, even after inclusion of tumours displaying poor prognostic features, including locally advanced tumours, adenocarcinomas and adenosquamous carcinomas[42]. Likewise, the survival differences were not driven by particularly poor outcomes among HPV45+ tumours, as they remained following exclusion of these samples (**Figure S3**) and there were very few HPV45+ samples in the validation cohort (**Table S4**). Consistent

with these observations, membership of the atypical subgroup was an independent prognostic factor in the validation cohort.

Our aggressive, atypical subgroup is enriched for genes linked to EMT and displays significant overlap with the EMT cluster defined for the subset of TCGA samples with RPPA data available; in particular increased expression of YAP1, which in most cases appears to be driven by *YAP1* gene amplification. Our finding that a subgroup of HPV16+ tumours co-cluster with tumours driven almost exclusively by α7 HPV types suggests that cervical cancers driven by different HPV types evolve along different trajectories but that HPV16+ tumours can occasionally develop via the atypical route more commonly associated with the α7 types. HPV18+ tumours can likewise evolve via either route but like the other α7 types, HPV18+ tumours are frequently atypical. We postulate that the host immune response is the key driver here: a cervical cancer developing in the presence of a stronger immune response will become an atypical tumour, hence the greater neoantigen depletion and immune-evasive/suppressive features displayed in the tumours at time of resection, including M2 macrophage polarization, increased neutrophil abundance, reduced CTL:Treg ratios and increased ECM deposition. By extension, this implies that α7 HPV types elicit more effective immunosurveillance than HPV16 and the other α9 types, however when HPV16-transformed cells do encounter an effective immune response, the resulting tumours develop down the atypical trajectory. This model could also explain why HPV16 infections are at much higher risk to progress to CIN3 than other HPV types[3]. Greater immunogenicity of α7 HPVs might also explain why they are so rarely, if ever, associated with cancers developing at sites such as the oropharynx, where HPV16 dominates but other α9 types are occasionally seen[43].

Two genomic features of atypical tumours are particularly likely to be selected for in cells under immune surveillance, *STK11* and *YAP1*. *STK11* loss-of-function drives immunosuppressive TGF-beta

signalling[44] and activates glycolysis, in turn suppressing T-cell homing and activity[45 46,47] . Consistent with this, *STK11* mutations have recently been implicated in mediating resistance to PD1/PD-L1 blockade [48]. *YAP1*, is frequently amplified in atypical tumours, has been shown to drive cervical cancer development in mice even in the absence of HPV, and strongly synergizes with HPV16 E6 and E7 to promote tumorigenesis [49,50]. It is also associated with poor prognosis, reduced lymphocyte activation and resistance to PD1/PD-L1 blockade in HNSCC [51]. Downstream of these genomic alterations, TGF-beta is known to modulate several features of the microenvironment observed in atypical cancers: it suppresses NK cell activation[52]; drives M2-polarisation in macrophages[53], induces neutrophil switching to an immunosuppressive phenotype[54] and drives immunotherapy resistance through T-cell exclusion by CAFs[55] in addition to its established role in driving metastasis. Our analysis suggests that typical cervical cancers, with their higher mutation burdens and greater T-lymphocyte infiltration will be good candidates for immunotherapy. What therapeutic strategies might be efficacious in the atypical tumours, which display worse prognosis, even when resected at early stage? Although TGF-beta inhibition for cancer treatment has thus far been limited by toxicity, inhibiting NOX4, an NADPH oxidase required for fibroblast differentiation into ECM-depositing myofibroblasts has recently been shown to permit CTL infiltration and to potentiate immunotherapy in a mouse model of HPV-associated cancer [56]. GKT137831, the NOX4 inhibitor used in this study is already approved for use in fibrosis, so could readily be trialled in patients. In those atypical cervical cancers harbouring loss-of-function *STK11* mutations, treatment with the mitochondrial inhibitor Phenformin is a possible therapeutic option[57].

In summary, by assembling the largest multi-omics cervical cancer dataset to date, we have gained novel insights into the development and progression of this disease. It has also allowed us to develop a prognostic classification that captures variation in HPV type, host genomic alterations and the tumour microenvironment and offers the potential for stratification of cervical cancer patients for

improved clinical management. Genome-wide DNA methylation profiling is already being used in the clinic for diagnosis and stratification of brain tumours[58], thus a similar strategy could readily be tested for cervical cancer.

Methods

Dataset assembly

DNA methylation (Illumina Infinium 450k array) and RNAseq data were obtained for CESC from the TCGA data portal. TCGA mutation data were obtained from the MC3 project on SAGE Synapse (syn7214402). DNA methylation (Illumina Infinium 450k array) and gene expression (Illumina HumanHT-12 V4.0 expression beadchip) data from the Oslo cohort were obtained from the Gene Expression Omnibus (GSE68339). RNAseq data were obtained for the Bergen cohort from dbGaP (phs000600/DS-CA-MDS) and were converted to fastq files using SRA-dump from the SRA Toolkit (http://ncbi.github.io/sra-tools/). Kallisto[59] was then used to quantify expression of GENCODE GrCh37 transcripts, repbase repeats and transcripts from 20 different high-risk HPV types with bias correction. Where IDAT files for 450k data were available, they were parsed using *minfi*[60] and were subjected to Functional Normalisation[61], followed by BMIQ-correction[62] for probe type distribution (which was done for all methylation data). For TCGA samples, viral type allocation was performed using VirusSeq[8].

Generation of 450k methylation profiles.

100ng DNA was bisulphite converted using the EZ DNA Methylation kit (Zymo Research) as per manufacturer's instructions. Bisulphite converted DNA hybridised to the Infinium 450K Human Methylation array, and processed in accordance with the manufacturer's recommendations.

HPV typing.

HPV16 or 18 was detected in 230 samples from the Oslo cohort by PCR, using the primers listed in[63]. The PCR products were detected by polyacrylamide gene electrophoresis or the Agilent DNA 1000 kit (Agilent Technologies Inc, Germany). Samples from the Innsbruck cohort and the remaining non-HPV16/18 samples from the Oslo cohort (n=38) were HPV-typed by DDL Diagnostic Laboratory (Netherlands) using the SPF10 assay, in which a PCR-based detection of over 50 HPV types is followed by a genotyping assay ($LIPA_{25}$) that identifies 25 HPV types (HPV 6, 11, 16, 18, 31, 33, 34, 35, 39, 40, 42, 43, 44, 45, 51, 52, 53, 54, 56, 58, 59, 66, 68/73, 70 and 74). HPV type data for the remaining samples were published previously[5,7]

Prognostic analyses and tumour clustering

Associations between HPV type and survival were tested for early stage (Stage I and II) CESC in the TCGA cohort containing either HPV16, HPV18 or HPV45. Limma-voom and limma on BMIQ and Functionally-normalised 450k data were used to identify differentially expressed genes and methylation variable positions between HPV45 and HPV16 associated tumours. Expression profiles were clustered to yield HPV45-like (atypical) and HPV16-like (typical) cancers using the clusterCons package[64]. The caret R package and limma were used to develop an SVM using 5 iterations of 5-fold Cross-Validation to allocate 450k samples to these subgroups.

Samples from our validation cohort, comprise of cases from three European centres (Bergen and Oslo in Norway and Innsbruck, Austria) were binned into these categories, along with TCGA samples not in the original training set, and taken together were used for subsequent statistical analyses to identify genomic and microenvironmental correlates. Associations between nodal dissemination and the HPV45 signature were carried out using the caret R package. A GLMnet with 5 iterations of 5-fold Cross-Validation was applied, with out-of-fold estimates used to assess performance using Affymetrix array data from GSE26511. Survival analyses of epigenetic allocations were carried out using Cox Proportional Hazards regression with stratification by histology, and with surgery, radiotherapy and

chemotherapy (given/not given) as covariates. For all clinical analyses, stages were collapsed into Stages I, II, III and IV.

Pathway analyses

Pathways were analysed with Ingenuity Pathway Analysis. Settings used were experimentally validated interactions in human models. Z-score cutoffs of 2 and FDR cutoffs of 0.05 were used to identify significant hits from Canonical Pathways, Upstream Regulator and Functions ontology analyses for plotting.

Copy number analysis

450k total intensities (Methylated and Unmethylated values) were used to generate copy number profiles with normal blood samples from Renius et al[65] as the germline reference. Functional normalisation [61] was used to regress out technical variation across the reference and tumour datasets before merging and quantile normalisation was used to normalise combined intensities followed by Circular Binary Segmentation as previously described[20]. Median density peak correction was performed to ensure centering before further analysis. GISTIC2.0[66] was then used to identify regions of significant copy number change at both arm and gene levels. Candidate copy number changes were evaluated for association with cluster using binomial GLMs.

The parameters chosen were a noise threshold of 0.1 with arm-level peel off and a confidence level of 0.95 was used to nominate genes targeted by copy number changes. Binomial regression was finally used to estimate rates of differential alteration.

Mutational analyses

For TCGA data, mutation calls were obtained from SAGE synapse as called by the MC3 project. Mutations for the Bergen cohort were obtained from[7]. Binomial GLMs were then used to estimate associations between the aggressiveness clusters and mutation frequencies.

Immunological Analyses

MHC-class I neoantigens were retrieved using The Cancer Immunome Atlas[21] for TCGA samples. Immunoediting estimates were computed based on silent mutation rates as previously described in Rooney et al [67] . Comparisons with immune infiltration in HPV+ Head and Neck Tumours were carried out using a previously published, manually-curated signature of immune-checkpoints, infiltration markers and effector molecules [9]. MethylCIBERSORT[27] was used to estimate tumour purity and abundances of seven other microenvironmental cellular fractions. Monocyte polarisation was computed using CIBERSORT on the basis of the original LM22 matrix (1000 permutations, data supplied in counts per million) provided with the software. We normalised the estimates to total monocyte fractions and estimated the fraction of proinflammatory (M1 and dendritic cells) relative to all monocytes to yield a proinflammatory monocyte fraction, which was tested for associations with prognostic cluster using Wilcoxon's Rank Sum Test.

To generate a neutrophil gene expression signature for assessing neutrophil content by ssGSEA, RNA-seq data were downloaded from the European Nucleotide Archive for the following datasets - PRJEB11844[68], GSE60424[69], and E-MTAB-2319[70] in order to derive an RNAseq dataset of immune cell types. *Kallisto*[59] was used to quantify gene expression with a reference transcriptome consisting of Gencode Grch37 assembly of protein coding and lincRNA transcripts. Data were then modelled using limma trend and overexpressed markers (3FC, FDR < 0.05) were selected for each cell subset from one versus all comparisons.

The GSVA R package was then used to compute ssGSEA scores for T-cell subsets of interest with absolute ranking, score normalisation and RNA-seq flags set to true. Enrichment scores were then normalised by cellular abundance and differences were estimated using Wilcoxon's rank sum test with Benjamini Hochberg correction for multiple testing.

Immunohistochemistry

All immunohistochemical staining was conducted by HSL-Advanced Diagnostics (London, UK) using the Leica Bond III platform with Leica Bond Polymer Refine detection as per manufacturer's recommendations. Sections from a series of 17 tumour samples from the validation cohort were stained for CD8 (mouse monoclonal 4B11, Leica Biosystems PA0183, used as supplied for 15 minutes at room temperature. HIER was performed on-board using Leica ER2 solution (high pH) for 20 minutes), CD68 (mouse monoclonal PGM1, Agilent M087601-2, used at a dilution of 1/50 for 15mins at room temperature. HIER was performed on-board using Leica ER1 solution (low pH) for 20 minutes) or MPO (rabbit polyclonal, Agilent A039829-2, used at a dilution of 1/4000 for 15 minutes at room temperature without epitope retrieval. Scoring was performed blinded to cluster membership by a histopathologist (JM) as follows: 0 = no positive cells / field (200X magnification); 1 = 1 – 10 positive cells; 2 = 11 – 100 positive cells; 3 = 101 – 200 positive cells; 4 = 201 = 300 positive cells; 5 = over 300 positive cells.

Data availability

R markdowns used to run these analyses available on request. Data generated in-house have been deposited in the Gene Expression Omnibus with the accession.(to be deposited upon publication)

Acknowledgements

AC was supported by postgraduate research scholarships from UCL and received additional research support from a Debbie Fund grant to KC. Research in TRF's lab is supported by Rosetrees Trust (M229-CD1) and Cancer Research UK (A25825). DNA methylation data were generated through funding provided by the Debbie Fund to KC and analyses were carried out in part using data generated by The Cancer Genome Atlas. AF was supported by grants from the MRC (MR/M025411/1), PCUK(MA-TR15-009), BBSRC (BB/R009295/1), TUF, Orchid and the UCLH BRC. The authors dedicate this manuscript

to the late Dr Helga Salvesen, a wonderful collaborator and colleague who played a key role in the project.

Disclosures

AC, AF, KC and TRF have filed a patent application for a prognostic methylation biomarker based on some of the results in this paper.

Contributions

TRF, AC, AF and KC conceived the project; AC and SH carried out computational analyses; AF, CD, MJ, NK, NE, JM performed experiments; HL, HF, HS, XS, CSF, MKH, JD, CK, AS, SS provided samples and data. AC and TRF and AF wrote the paper and TRF supervised the study.

References

1. Li, N., Franceschi, S., Howell-Jones, R., Snijders, P.J.F. & Clifford, G.M. Human papillomavirus type distribution in 30,848 invasive cervical cancers worldwide: Variation by geographical region, histological type and year of publication. *International Journal of Cancer* **128**, 927-935 (2011).
2. Bratman, S.V. *et al.* Human Papillomavirus Genotype Association With Survival in Head and Neck Squamous Cell Carcinoma. *JAMA Oncol* **2**, 823-6 (2016).
3. Jaisamrarn, U. *et al.* Natural history of progression of HPV infection to cervical lesion or clearance: analysis of the control arm of the large, randomised PATRICIA study. *PLoS One* **8**, e79260 (2013).
4. Feber, A. *et al.* Epigenetics markers of metastasis and HPV-induced tumorigenesis in penile cancer. *Clin Cancer Res* **21**, 1196-206 (2015).
5. Cancer Genome Atlas Research, N. *et al.* Integrated genomic and molecular characterization of cervical cancer. *Nature* **543**, 378-384 (2017).
6. Lando, M. *et al.* Interplay between promoter methylation and chromosomal loss in gene silencing at 3p11-p14 in cervical cancer. *Epigenetics* **10**, 970-80 (2015).
7. Ojesina, A.I. *et al.* Landscape of genomic alterations in cervical carcinomas. *Nature* **506**, 371-5 (2014).
8. Chen, Y. *et al.* VirusSeq: software to identify viruses and their integration sites using next-generation sequencing of human cancer tissue. *Bioinformatics* **29**, 266-7 (2013).
9. Chakravarthy, A. *et al.* Human Papillomavirus Drives Tumor Development Throughout the Head and Neck: Improved Prognosis Is Associated With an Immune Response Largely Restricted to the Oropharynx. *J Clin Oncol* **34**, 4132-4141 (2016).
10. Barbie, D.A. *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108-12 (2009).
11. Kaufhold, S. & Bonavida, B. Central role of Snail1 in the regulation of EMT and resistance in cancer: a target for therapeutic intervention. *J Exp Clin Cancer Res* **33**, 62 (2014).
12. Park, J. & Schwarzbauer, J.E. Mammary epithelial cell interactions with fibronectin stimulate epithelial-mesenchymal transition. *Oncogene* **33**, 1649-57 (2014).

13. Reymond, N., d'Agua, B.B. & Ridley, A.J. Crossing the endothelial barrier during metastasis. *Nat Rev Cancer* **13**, 858-70 (2013).
14. Skobe, M. *et al.* Induction of tumor lymphangiogenesis by VEGF-C promotes breast cancer metastasis. *Nat Med* **7**, 192-8 (2001).
15. Steeg, P.S. Targeting metastasis. *Nat Rev Cancer* **16**, 201-18 (2016).
16. Liu, Y. *et al.* Increased EZH2 expression is associated with proliferation and progression of cervical cancer and indicates a poor prognosis. *Int J Gynecol Pathol* **33**, 218-24 (2014).
17. Halle, C. *et al.* Hypoxia-induced gene expression in chemoradioresistant cervical cancer revealed by dynamic contrast-enhanced MRI. *Cancer Res* **72**, 5285-95 (2012).
18. Kim, B.W. *et al.* Prognostic assessment of hypoxia and metabolic markers in cervical cancer using automated digital image analysis of immunohistochemistry. *J Transl Med* **11**, 185 (2013).
19. Zhang, Y., Liu, B., Zhao, Q., Hou, T. & Huang, X. Nuclear localizaiton of beta-catenin is associated with poor survival and chemo-/radioresistance in human cervical squamous cell cancer. *Int J Clin Exp Pathol* **7**, 3908-17 (2014).
20. Feber, A. *et al.* Using high-density DNA methylation arrays to profile copy number alterations. *Genome Biol* **15**, R30 (2014).
21. Charoentong, P. *et al.* Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cell Rep* **18**, 248-262 (2017).
22. Gilbert, D.C. *et al.* Tumour-infiltrating lymphocyte scores effectively stratify outcomes over and above p16 post chemo-radiotherapy in anal cancer. *Br J Cancer* (2016).
23. Gooden, M.J.M., de Bock, G.H., Leffers, N., Daemen, T. & Nijman, H.W. The prognostic influence of tumour-infiltrating lymphocytes in cancer: a systematic review with meta-analysis. *British Journal Of Cancer* **105**, 93 (2011).
24. Jordanova, E.S. *et al.* Human leukocyte antigen class I, MHC class I chain-related molecule A, and CD8+/regulatory T-cell ratio: which variable determines survival of cervical cancer patients? *Clin Cancer Res* **14**, 2028-35 (2008).
25. Nedergaard, B.S., Ladekarl, M., Thomsen, H.F., Nyengaard, J.R. & Nielsen, K. Low density of CD3+, CD4+ and CD8+ cells is associated with increased risk of relapse in squamous cell cervical cancer. *British Journal Of Cancer* **97**, 1135 (2007).
26. Ward, M.J. *et al.* Tumour-infiltrating lymphocytes predict for outcome in HPV-positive oropharyngeal cancer. *Br J Cancer* **110**, 489-500 (2014).
27. Chakravarthy, A. *et al.* Pan-cancer deconvolution of tumour composition using DNA methylation. *Nat Commun* **9**, 3220 (2018).
28. Carter, S.L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol* **30**, 413-21 (2012).
29. Mizunuma, M. *et al.* The pretreatment neutrophil-to-lymphocyte ratio predicts therapeutic response to radiation therapy and concurrent chemoradiation therapy in uterine cervical cancer. *Int J Clin Oncol* **20**, 989-96 (2015).
30. Lee, Y.Y. *et al.* Pretreatment neutrophil:lymphocyte ratio as a prognostic factor in cervical carcinoma. *Anticancer Res* **32**, 1555-61 (2012).
31. Condamine, T. & Gabrilovich, D.I. Molecular mechanisms regulating myeloid-derived suppressor cell differentiation and function. *Trends Immunol* **32**, 19-25 (2011).
32. Zelenay, S. *et al.* Cyclooxygenase-Dependent Tumor Growth through Evasion of Immunity. *Cell* **162**, 1257-70 (2015).
33. Elkabets, M. *et al.* IL-1beta regulates a novel myeloid-derived suppressor cell subset that impairs NK cell development and function. *Eur J Immunol* **40**, 3347-57 (2010).
34. Chun, E. *et al.* CCL2 Promotes Colorectal Carcinogenesis by Enhancing Polymorphonuclear Myeloid-Derived Suppressor Cell Population and Function. *Cell Rep* **12**, 244-57 (2015).
35. Hartwig, T. *et al.* The TRAIL-Induced Cancer Secretome Promotes a Tumor-Supportive Immune Microenvironment via CCR2. *Mol Cell* **65**, 730-742.e5 (2017).

36. Xu, M. *et al.* Interactions between interleukin-6 and myeloid-derived suppressor cells drive the chemoresistant phenotype of hepatocellular cancer. *Exp Cell Res* **351**, 142-149 (2017).
37. Newman, A.M., Liu, C.L. & Green, M.R. Robust enumeration of cell subsets from tissue expression profiles. **12**, 453-7 (2015).
38. Lai, C.H. *et al.* Role of human papillomavirus genotype in prognosis of early-stage cervical cancer undergoing primary surgery. *J Clin Oncol* **25**, 3628-34 (2007).
39. Burger, R.A. *et al.* Human papillomavirus type 18: association with poor prognosis in early stage cervical cancer. *J Natl Cancer Inst* **88**, 1361-8 (1996).
40. Wang, C.C. *et al.* Clinical effect of human papillomavirus genotypes in patients with cervical cancer undergoing primary radiotherapy. *Int J Radiat Oncol Biol Phys* **78**, 1111-20 (2010).
41. Cuschieri, K. *et al.* Influence of HPV type on prognosis in patients diagnosed with invasive cervical cancer. *Int J Cancer* **135**, 2721-6 (2014).
42. Halle, M.K. *et al.* Clinicopathologic and molecular markers in cervical carcinoma: a prospective cohort study. *Am J Obstet Gynecol* **217**, 432 e1-432 e17 (2017).
43. Gillison, M.L., Chaturvedi, A.K., Anderson, W.F. & Fakhry, C. Epidemiology of Human Papillomavirus-Positive Head and Neck Squamous Cell Carcinoma. *J Clin Oncol* **33**, 3235-42 (2015).
44. Moren, A., Raja, E., Heldin, C.H. & Moustakas, A. Negative regulation of TGFbeta signaling by the kinase LKB1 and the scaffolding protein LIP1. *J Biol Chem* **286**, 341-53 (2011).
45. Faubert, B. *et al.* Loss of the tumor suppressor LKB1 promotes metabolic reprogramming of cancer cells via HIF-1alpha. *Proc Natl Acad Sci U S A* **111**, 2554-9 (2014).
46. Ottensmeier, C.H. *et al.* Upregulated glucose metabolism correlates inversely with CD8+ T cell infiltration and survival in squamous cell carcinoma. *Cancer Res* (2016).
47. Ngwa, V.M., Edwards, D.N., Philip, M. & Chen, J. Microenvironmental Metabolism Regulates Antitumor Immunity. *Cancer Res* **79**, 4003-4008 (2019).
48. Skoulidis, F. *et al.* STK11/LKB1 Mutations and PD-1 Inhibitor Resistance in KRAS-Mutant Lung Adenocarcinoma. *Cancer Discov* **8**, 822-835 (2018).
49. Eun, Y.G. *et al.* Clinical significance of YAP1 activation in head and neck squamous cell carcinoma. *Oncotarget* **8**, 111130-111143 (2017).
50. He, C. *et al.* A Human Papillomavirus-Independent Cervical Cancer Animal Model Reveals Unconventional Mechanisms of Cervical Carcinogenesis. *Cell Rep* **26**, 2636-2650 e5 (2019).
51. Ayers, M. *et al.* IFN-gamma-related mRNA profile predicts clinical response to PD-1 blockade. *J Clin Invest* **127**, 2930-2940 (2017).
52. Viel, S. *et al.* TGF-beta inhibits the activation and functions of NK cells by repressing the mTOR pathway. *Sci Signal* **9**, ra19 (2016).
53. Zhang, F. *et al.* TGF-beta induces M2-like macrophage polarization via SNAIL-mediated suppression of a pro-inflammatory phenotype. *Oncotarget* **7**, 52294-52306 (2016).
54. Fridlender, Z.G. *et al.* Polarization of tumor-associated neutrophil phenotype by TGF-beta: "N1" versus "N2" TAN. *Cancer Cell* **16**, 183-94 (2009).
55. Mariathasan, S. *et al.* TGFbeta attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* **554**, 544-548 (2018).
56. Ford, K. *et al.* NOX4 inhibition potentiates immunotherapy by overcoming cancer-associated fibroblast-mediated CD8 T-cell exclusion from tumours. *Cancer Res* (2020).
57. Shackelford, D.B. *et al.* LKB1 inactivation dictates therapeutic response of non-small cell lung cancer to the metabolism drug phenformin. *Cancer Cell* **23**, 143-58 (2013).
58. Capper, D. *et al.* DNA methylation-based classification of central nervous system tumours. *Nature* **555**, 469-474 (2018).
59. Bray, N.L., Pimentel, H. & Melsted, P. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**, 525-7 (2016).
60. Aryee, M.J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363-9 (2014).

61. Fortin, J.P. *et al.* Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol* **15**, 503 (2014).
62. Teschendorff, A.E. *et al.* A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* **29**, 189-96 (2013).
63. Lyng, H. *et al.* Intratumor chromosomal heterogeneity in advanced carcinomas of the uterine cervix. *Int J Cancer* **111**, 358-66 (2004).
64. Simpson, T.I., Armstrong, J.D. & Jarman, A.P. Merged consensus clustering to assess and improve class discovery with microarray data. *BMC Bioinformatics* **11**, 590 (2010).
65. Reinius, L.E. *et al.* Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS One* **7**, e41361 (2012).
66. Mermel, C.H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* **12**, R41 (2011).
67. Rooney, M.S., Shukla, S.A., Wu, C.J., Getz, G. & Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48-61 (2015).
68. De Simone, M. *et al.* Transcriptional Landscape of Human Tissue Lymphocytes Unveils Uniqueness of Tumor-Infiltrating T Regulatory Cells. *Immunity* **45**, 1135-1147 (2016).
69. Linsley, P.S., Speake, C., Whalen, E. & Chaussabel, D. Copy number loss of the interferon gene cluster in melanomas is linked to reduced T cell infiltrate and poor patient prognosis. *PLoS One* **9**, e109760 (2014).
70. Bonnal, R.J. *et al.* De novo transcriptome profiling of highly purified human lymphocytes primary cells. *Sci Data* **2**, 150051 (2015).