Epigenetics Markers of Metastasis and HPV-Induced Tumorigenesis in Penile Cancer

Andrew Feber¹, Manit Arya²,³, Patricia de Winter², Muhammad Saqib², Raj Nigam⁴, Peter R. Malone⁴, Wei Shen Tan², Simon Rodney¹, Matthias Lechner¹, Alex Freeman⁵, Charles Jameson⁵, Asif Muneeb²,⁴, Stephan Beck¹, and John D. Kelly¹,²

Abstract

Purpose: Penile cancer is a rare malignancy in the developed world with just more than 1,600 new cases diagnosed in the United States per year; however, the incidence is much higher in developing countries. Although HPV is known to contribute to tumorigenesis, little is known about the genetic or epigenetic alterations defining penile cancer.

Experimental Design: Using high-density genome-wide methylation arrays, we have identified epigenetic alterations associated with penile cancer. QiMSP was used to validate lymph node metastasis markers in 50 cases. A total of 446 head and neck squamous cell carcinoma (HNSCC) and cervical squamous cell carcinoma (CESCC) samples were used to validate HPV-associated epigenetic alterations.

Results: We defined 6,933 methylation variable positions (MVP) between normal and tumor tissue, which includes 997 hypermethylated differentially methylated regions associated with tumor suppressor genes, including CDO1, 9R1, and WT1. Analysis of penile cancer tumors identified a 4 gene epigenetic signature which accurately predicted lymph node metastasis in an independent cohort (AUC of 89%). Finally, we explored the epigenetic alterations associated with penile cancer HPV infection and defined a 30 loci lineage-independent HPV specific epigenetic signature which predicts HPV status and survival in independent HNSCC, CESCC cohorts. Epigenetic-negative patients have a significantly worse overall survival [HNSCC P = 0.00073; 95% confidence interval (CI), 0.021–0.78; CESCC P = 0.0094; HR = 3.91, 95% CI = 0.13–0.78]. HPV epigenetic signature is a better predictor of survival than HPV status alone.

Conclusions: These data demonstrate for the first time genome-wide epigenetic events involved in an aggressive penile cancer phenotype and define the epigenetic alterations common across multiple HPV-driven malignancies. Clin Cancer Res; 21(5); 1196–206. ©2014 AACR.

Introduction

Penile cancer is relatively rare in the developed world, but represents a global health problem, showing high prevalence and posing significant morbidity and mortality in developing countries (1, 2). The age-standardized incidence of penile cancer is 0.3–1.0 per 100,000 men in European countries and the United States, equating to approximately 1,600 new cases per annum in the United States (2). In contrast, the incidence in developing nations varies from 3 to 8.3 per 100,000 (3, 4).

The presence of inguinal lymph node involvement is at present the most important prognostic indicator of unfavorable prognosis in penile cancer (5). Although, histopathologic factors, including tumor subtype, grade, stage, and the presence of lymphovascular and perineural invasion are useful predictors of inguinal lymph node metastases, they are still not accurate and if used exclusively would lead to overtreatment of a significant proportion of patients. The etiology of penile cancer is multifactorial with smoking, phimosis, poor personal hygiene, and low socioeconomic status all being risk factors for tumor development (6). In addition, there is strong evidence linking development of penile cancer to infection with high-risk HPV (HPV 16, 18), suggesting that HPV plays a significant role in the pathogenesis of at least a subset of cases. High-risk HPV infection is transformative in other tumor types, including cervical squamous cell carcinoma and head and neck squamous cell carcinoma (CESCC and HNSCC, respectively; refs. 7, 8). Contrary to cervical cancers, which appear to be almost exclusively (>90%) driven by HPV, only a proportion of penile, vulvar, anal, and oropharyngeal cancers appear to be HPV driven (9, 10). Interestingly, despite the clear oncogenic effects of HPV infection, HPV positivity appears to confer a survival benefit, this is particularly true for HNSCC, and also appears to be for penile cancer, although as yet only limited data are available (11).

Changes in DNA methylation play a key role in malignant transformation, leading to the silencing of tumor suppressor genes and overexpression of oncogenes (12). The ontogenic plasticity of DNA methylation makes epigenetic changes ideal...
Penile Epigenetics

Translational Relevance

Penile cancer is rare in the developed world, but represents a global health problem, with an incidence of up to 8.3-100,000 in developing nations. The most important predictive factor of an unfavorable prognosis in penile cancer is the presence of regional inguinal lymph node involvement. Currently, no molecular markers exist that can accurately predict the presence of lymph node metastases. Using genome-wide DNA methylation profiling, we defined the epigenetic alterations involved in penile cancer and validated an epigenetic signature that is predictive of lymph node metastasis. HPV represents a major oncogenic driver in penile cancer, we identify HPV-induced epigenetic alterations, and from these we define an epigenetic signature that is predictive of survival across multiple HPV-drivers cancers. The identification of epigenetic biomarkers of metastasis and survival may play a significant role in improving the management, treatment, and survival of penile cancer and also other HPV-driven cancers.

Biomarkers for diagnosis or as predictive and prognostic markers in cancer. However, little is known about the molecular genetics or epigenetics driving the development and progression of penile cancer. Aberrant methylation of a handful of candidate genes has previously been identified, including CDKN2A and RASSF1A (13–16). Recently, epigenetic changes in both host and virus epigenomes have been reported in other HPV-induced cancers (17–20). To date, no substantial genome-wide analysis has been performed in penile cancer and linkage between viral subtypes has not been elucidated. We have therefore sought to define the epigenetic alterations associated with penile carcinogenesis, including a subset of cases associated with high risk HPV infection. Using high density genome-wide methylation array on a panel of penile cancer and matched normal tissue, we have annotated epigenetic alterations which define penile cancer, we also interrogated these data to reveal epigenetic changes associated with disease progression and HPV infection.

Materials and Methods

Ethics approval

Ethics approval for this study was granted by the University College London (UCL)/University College London Hospital (UCLH, London, UK) Biobank for Health and Human Disease (NC06.11). Informed consent was obtained.

Patient samples and clinical data

Thirty-eight fresh penile cancers and 11 matched normal tissue samples (stored in RNAlater) from the UCL/UCLH Urology Biobank, and 50 formalin-fixed paraffin-embedded (FFPE) tissue blocks from the Department of Pathology (UCLH, London, UK) with confirmed histopathologic and clinical diagnosis of penile cancer and with >80% tumor cellularity were included and analyzed. Normal samples taken adjacent from tumor tissue and confirmed to be histologically normal in pathologic review (Supplementary Tables S1 and S2).

DNA extraction

DNA was extracted from RNA later preserved frozen tissue using the QIAamp DNA Mini Kit (Qiagen), and FFPE tissue using the QIAamp DNA FFPE Tissue Kit (Qiagen) according to the manufacturer’s instructions.

HPV assessment

All samples were assessed for the presence of low-risk HPV 6 and 11 and high-risk HPV 16, 18, and 31 viral DNA by qPCR with primers specific for each genotype (Supplementary Table S2A). The reference genes GAPDH and ACTB were used to normalize DNA input and calculate the number of HPV genomic copies present. HPV qPCR was carried out as previously described by Lechner and colleagues (22). HPV-type data for CESC and HNSCC TCGA samples were taken from Tang and colleagues, and based the expression of viral genes in RNA-seq data (21).

Methylation analysis

Five-hundred nanograms of DNA from 38 tumor and 11 matched normal RNA later-preserved samples from patients with penile cancer were bisulphite converted and hybridized to the Infinium 450 K Human Methylation array, and processed in accordance with the manufacturer’s recommendations. DNA bisulphite conversion was carried out using the EZ DNA Methylation Kit (Zymo Research) as per manufacturer’s instructions. Samples were processed in a single batch. R statistical software (version 2.14.0; ref. 22) was used for the subsequent data analysis. The ChAMP pipeline was used to extract and analyze data from iDat files, samples were normalized using BMIQ (23). Raw β values (methylation value) were subjected to a stringent quality control analysis as follows: samples showing reduced coverage were removed and only probes with detection levels above background across all samples were retained (detection P < 0.01). Differentially methylated regions (DMR) were called using the Probe Lasso algorithm (implemented in the R package ChAMP, see ref. 23) with default parameters with the exception of applying a minimum DMR size of 100 bp. As a result, all DMRs identified have a minimum of 3 significant probes, are at least 1 Kb from a neighboring DMR, and have a minimum size of 100 bp. Maximum DMR size is effectively unbounded but is dependent the genomic separation between contiguous CpG probes, which itself is contingent on the local underlying genomic and epigenomic features with larger DMRs more likely to occur in probe-poor regions (Butcher and colleagues, in press; ref. 23).

The statistical significance of methylation variable positions (MVP) enrichment in genomic and epigenomic features was calculated on the basis of the random selection of equal numbers of probes (4,935 for hypermethylated MVPs, 1,998 hypomethylated MVPs, from the overall probe set (472,655 probes) used in the analysis and repeated 10,000 times (24).

Gene set enrichment analysis was used to assess whether gene-associated DMRs are overrepresented in a particular gene set. Gene sets, categorized by gene ontology, molecular pathways, chromosomal locations, or targets of regulatory motifs and miRNAs, were derived from the Molecular Signatures Database (MSigDB). Enrichment was assessed by comparing the number of genes associated with DMRs belonging to the gene set with those that are not members. The significance of the over representation was then assessed by a Fisher exact test.
and adjusted for false discovery by the Benjamini Hochberg procedure. Genes containing multiple DMRs were counted only once to remove any bias in gene set enrichment.

Motif analysis was performed using the MEME-ChIP tool of the MEME suite; parameters were set to default except for the number of repetitions (set to “any number of repetitions”), motif width (min = 4, max = 15), and maximum number of motifs to find set to 20 (25).

Validation of methylation

Aberrant methylation was validated in the external cohort using methylation-specific qPCR (MSP, Supplementary Table S2). Genomic DNA from FFPE samples was bisulphite converted as above. Ten nanograms of converted DNA was subjected to MSP. Briefly, all reactions were carried out in a 13 μL reaction volume containing 6.5 μL 2 x SYBR Green reaction buffer, 0.3 μmol/L forward primer, and 0.3 μmol/L reverse primer with 1 ng genomic DNA (RNAinter—preserved) or 10 ng for FFPE samples. Reactions were run on an ABI 7300 RealTime PCR machine, denaturation was done for 10 minutes at 95°C, with 40 cycles of 95°C for 15 seconds and 60°C for 60 seconds. All reactions were performed in triplicate. Sensitivity and specificity of all reactions was assessed using spiked dilutions of fully methylated DNA. The methylation state of individual samples was determined using a standard curve with a range of control methylation states (0% to 100%). The absolute methylation was subsequently used to determine the association with lymph node metastasis.

Integration with publicly available CESC and HNSCC methylation data

Statistical software v2.15.1 (35) was used for preprocessing of data and for classic multidimensional scaling (MDS) using principal components analysis (PCA). HPV-specific epigenetic signature and prediction of HPV infection was determined using the shrunk centroid method implemented through the pamr bioconductor package. Survival analysis was carried using the bioconductor package, Survival (26). MDS was used to visualize HPV+ and HPV− penile cancer methylation signatures within methylation datasets obtained from an HPV-induced HNSCCs [(20) GEO accession numbers: GSE38266, GSE38268, GSE38270 and GSE38271, and TCGA samples from HNSCC (27) and CESC (28)]; Raw iDAT files were processed and normalized in line with human data as above.

RT-PCR

RNA was extracted from tissue, determined by hematoxylin and eosin (H&E) staining of frozen sections to be tumor or normal tissue from the same individuals, using an RNeasy kit according to the manufacturer’s instructions. RNA was quantified using a NanoDrop spectrophotometer and for each sample, 1 μg was reverse transcribed to cDNA in a 20 μL reaction using a Quantitect reverse transcription kit (QIAGEN), including a gDNA wipeout step. Completed reactions were diluted 10-fold with yeast tRNA 0.5 μg/mL, and 2 μL were used for qPCR using Brilliant III SYBR Green UltraFast qPCR master mix (Agilent) and with primers at 500 nmol/L each in a final reaction volume of 10 μL. Standards (10⁻¹⁻¹⁰ copies/reaction) were amplified together with samples in a Rotor-Gene Q (QIAGEN) using the following parameters: 95°C for 3 minutes followed by 40 cycles of 95°C for 5 seconds and 57°C for 10 seconds. Melting curve data were collected to confirm product identity. For all assays, efficiency was >95%, and reactions were linear over 7 log and sensitive to at least 10 copies and a single PCR product of the correct size was observed on a 2% agarose gel. Copy numbers per reaction were derived from the standard curves and normalized using the normalization factor for the three most stable reference genes identified by geNorm software: HPRT1, SDHA, YWHAZ. Data were analyzed using a paired Student t test with α at 0.05.

Results

Tumor-specific methylation events

To investigate whether penile tumors are epigenetically distinct from normal tissue, we performed genome-wide DNA methylation profiling using the 450k Illumina Infinium platform (29) to interrogate the methylation state of over 485,000 cytosine residues.

Unsupervised hierarchical clustering of β values (methylation score) revealed three distinct clusters based on histologic phenotype (Fig. 1A). Clustering of the most variable probes (n = 500) separated samples based on histopathology confirming that penile cancer and normal penile tissue are epigenetically distinct, and pointing to a hypermethylation phenotype associated with malignant transformation (Fig. 1B).

Supervised analysis, using a Wilcoxon rank sum test to assign directionality, was used to identify (MVPs) between penile cancer versus normal tissue. MVPs were selected on the basis of statistical significance (Wilcoxon P < 0.001), an additional filter of Δβ > 0.30 was applied to compensate for not taking into account the absolute difference in methylation between the groups. The cutoff is empirically defined to result in a false discovery rate (FDR) of <2%. This allowed us to reduce our candidate loci to those with largest methylation differences and therefore greatest potential for functional effect. A total of 6,933 MVPs met these requirements (4,935 hyper MVPs, 1,998 hypo MVPs), hierarchical clustering of the samples yielded three clusters: (i) normal, (ii) node positive, and (iii) node negative (Fig. 1C).

There is a clear hypermethylation profile associated with the cancer phenotype (Fig. 1C), with over 71% of MVPs being hypermethylated in tumor tissue compared with matched normal tissue (Supplementary Fig. S1). Mapping of the MVPs to gene features revealed a significant (random resampling P < 0.0001) enrichment of hypermethylated CpG islands (CpGI), 44% enrichment (Fig. 2A and B). To assess the potential functional impact of CpG methylation on gene expression we tested the association with MVPs in either promoter-associated or nonpromoter-associated CpGIs. This showed a enrichment (P < 0.0001) of MVPs in promoter associated CpGIs, and is further supported by the enrichment (P < 0.0001) of MVPs in regulatory regions, including transcription start sites (TSS2000), first exons, and 5’UTRs, which show enrichments of 8%, 7%, and 4% respectively (Fig. 2A and B).

Analysis of hypomethylated MVPs showed enrichment (P = 0.0010, 14%) of intergenic regions (IGR, Fig. 2C and D), potentially pointing to hypomethylation of repeats regions. This is confirmed by the enrichment of loci within ALU and SINE1 repeat elements.

As single MVPs are less likely to have functional effect on gene expression, we next sought to amalgamate individual MVPs into DMRs. The analysis defined 1,255 significant DMRs...
(P < 0.001) associated with the malignant phenotype (997 hyper DMRs and 258 hypo DMRs). The DMRs were associated with 367 genes, CpGIs were the predominant genomic feature associated DMRs.

**Gene set enrichment analysis**

Gene ontology (GO) analysis of genes associated with DMRs identified genes involved in DNA binding (GO: 0003677), signal transduction (GO: 0007165), and receptor activity (GO: 0004872) pathways. We also performed GSEA, assigning MVPs to their closest gene, to assess whether specific classes of genes are enriched. Interrogation of the penile cancer-associated hypermethylated genes showed significant enrichment (P = 0.000106) of genes which are targets of the PRC2 complex, including TBX5, GATA4, CDH7, and SOX14. Motif analysis of PRC2 target DMRs showed enrichment for PBX1, KLF4, and HIF1A transcription factor binding sites. Interestingly, we also see an increase in the expression of PRC2 complex members SUZ12 and EZH2 in tumors compared with normal tissue (Supplementary Fig. S2).

The high rate of CpGI methylation would suggest the potential for frequent inactivation of tumor suppressor genes (TSG). We therefore compared genes associated with both MVPs and DMRs with a list of 712 known TSGs. This revealed the enrichment of hyper-MVPs in TSGs (P = 0.00019), with 52 TSGs showing CpGI hypermethylation, these include RASSF2, WT1, and CDO1.

We also identified aberrant methylation of several potential therapeutic targets, including tyrosine kinases, EPHA5, EPHA6 along with FLT1 (VEGFR1), FLT3, and FLT4 (VEGFR3), and aberrant methylation of the androgen receptor (AR) and programmed cell death receptor 1 (PDCD1) the gene which encodes PD1, highlighting potential therapeutic targets for the treatment of penile cancer (Supplementary Fig. S3, S4A, and S4B).

To confirm the functional relevance of methylation we assessed the expression of two candidate genes (CDO1 and AR) in an independent cohort of matched penile cancer and normal tissues. This showed a significant reduction of expression in penile cancer compared with matched normal tissue (Supplementary Fig. S4C and S4D).

---

**Figure 1.**

Unsupervised clustering of MVPs in penile cancer squamous cell carcinomas. A, hierarchical clustering of penile cancer (PeCa) and normal tissue based on global epigenetic profiles, generates 3 groups: a normal (center, green), non-HPV–associated group (right, blue), and HPV–associated group (left, red). B, heatmap of methylation values of the 500 most variable loci, showing clear separation of normal and malignant disease. The DNA methylation (β) values are represented using a color scale from yellow (low DNA methylation) to blue (high DNA methylation). C, heatmap of β values for significant MVPs (n = 6,933) between normal and penile cancer tissue. The DNA methylation (β) values are represented using a color scale from yellow (low DNA methylation) to blue (high DNA methylation).
Epigenetic markers of lymph node metastasis

Unsupervised clustering of the top 500 most variable (tumor only) probes was performed to assess the association of aberrant epigenetic events with pathologic factors. This defined two clusters (Fig. 3A), which showed a significant correlation with lymph node status \( (P = 0.00017) \), with a hypermethylated lymph node–positive cluster and hypomethylated lymph node–negative cluster. No correlation was

Figure 2.
MVP canonical feature enrichment. Assessment of MVP enrichment in canonical gene features, for both hyper- (A and B) and hypo- (C and D) methylated MVPs. Shows enrichment of hypermethylated MVP in promoter-associated features (A) and CpGIs (D); hypomethylated MVP are enriched in intergenic regions (IGR; C). Genomic features with significant \( (P \leq 0.0001) \) enrichment are shown in red.

Figure 3.
Epigenetic signature of local lymphatic metastasis. Heatmap for the top 500 most variably methylated loci in penile cancers shows three pathologically defined clusters: a hypermethylated lymph node negative (right, top green bar), lymph node–positive hypomethylated group (center, top gold bar), and a HPV-associated cluster (left, bottom two bars; red = first HPV positive; blue = HPV negative; black = second HPV viral load, high HPV (>1 copy/cell), white = low (<1 copy/cell) gray, no HPV detectable). B, heatmap of methylation values for 962 significant MVPs between node-negative (green top bar), and node-positive (gold top bar).
found between these clusters and tumor grade or stage \((P > 0.05)\).

To more clearly define the epigenetic alterations associated with local metastatic spread we carried out a supervised analysis utilizing all 48,577 informative loci (Fig. 3B). This defined a small number of MVPs \((n = 112)\), which separate samples into two main groups, a hypomethylated lymph node–positive group, and a hypermethylated lymph node–negative disease group. Analysis of the enriched MVPs in canonical gene features, shows enrichment of hypomethylated MVPs within CpGIs \((P < 0.0001)\), with 72% of MVPs located in CpGIs. These data suggest that CpGI hypermethylation is associated with lower metastatic potential.

The ability to predict lymph node metastasis may have potential utility in the clinical management of patients by identifying which patients with clinically impalpable inguinal lymph nodes require an inguinal lymphadenectomy. To explore this we sought to define a minimal epigenetic signature, which could be used to predict lymph node metastasis. Using a shrunken centroids approach, we identified a minimum 54 CpG signature which in cross validation, could predict the lymphatic metastases with an accuracy of 93%. When individual MVPs were coalesced into potentially functional DMRs, we identified DMRs in four genes, HMX3, IRF4, FLI1 and PPP2R5C, to be predictive of lymph node positive disease (Fig. 4A and B, Supplementary Fig. S5). These DMRs were combined to define a final predictive methylation index for each sample (mean methylation state across DMRs). This predictive index reached an ROC of 98% (specificity 100%, sensitivity 92%) (Supplementary Fig. S5). We then tested the association of this gene panel in a validation cohort of a further 50 patients with FFPE DNA using qMSP for each DMR. In the validation cohort, the predictive lymph node metastasis signature reached an AUC of 0.89 (specificity 80%, sensitivity 93%) (Fig. 4C).

Multivariable analysis showed this minimal signature to be an independent predictor of lymph node metastasis \((P = 0.0053)\), a surrogate for disease-specific survival, there was no significant association with age, stage, or grade \((P = 1, P = 0.98, P = 0.76)\) in multivariable analysis.

Immunohistochemical analysis for FLI1 and IRF4 (available antibodies) was carried out on a tissue microarray containing the 50 penile cancer tumors. Although we observe a reduction in protein expression in samples with corresponding hypermethylation, the relationship with lymph node metastasis was not statistically significant (Supplementary Fig. S5B).

HPV-driven tumorigenesis

Unsupervised clustering points to the presence of a potential HPV-related epigenetic component (Fig. 5A). To define a HPV-induced epigenetic signature, we performed a supervised analysis and ranked probes using a Bayesian regularized t-statistics model. We identified a significant association between DNA methylation and HPV status, with 960 significant MVPs at an FDR of less than 0.01, and 5,037 at an FDR of \(< 0.05\). Of the 960 MVPs, the overwhelming majority (747, 77%) were hypo-MVPs in HPV-positive samples, compared with HPV negative, indicating that HPV infection is associated with widespread loss of DNA methylation (Fig. 5A). Analysis of the
canonical gene features in which these MVPs reside showed that over 67% are located with CpGIs, shores, and shelves, with a significant enrichment (P < 0.001) of MVPs in CpGI shores. When individual MVPs were coalesced into potentially functional DMRs, we identified DMRs in several candidate genes, including GRAMD4 and GPX5 (Fig. 5B and C). GO analysis of analysis of genes associated with penile cancer HPV DMRs identified genes involved in WNT signaling, DNA binding, signal transduction and receptor activity pathways. They also showed significant overlap with genes shown to be upregulated in nasopharyngeal tumors, which are also frequently driven by HPV. Motif analysis of penile cancer HPV DMRs showed enrichment for TCF3, MAZ, JUN, PAX4, and MYC transcription factor–binding sites.

**Lineage-independent HPV signature**

We sought to assess whether the effect of HPV infection on DNA methylation is lineage dependent by evaluating the methylation state of penile cancer HPV MVPs in HNSCC and CESC. Using all penile cancer HPV MVPs, we were able to accurately define HPV-positive from HPV-negative disease in 42 HNSCC (data not shown). We subsequently identified the overlapping loci between these two data sets, to define a lineage-independent HPV signature. Despite the apparent strong association of our penile cancer

![Figure 5.](image_url)

Penile cancer HPV induced epigenetic signature. A, heatmap of significant MVPs (P < 0.01) between HPV-positive and HPV-negative penile cancer. HPV-positive samples show a significantly lower methylation profile than HPV-negative disease. B and C, methylation profiles of candidate genes epigenetically deregulated during HPV tumorigenic transformation. Comparison of DMR profiles across canonical features for HPV-associated penile cancer (green) and non-HPV-associated penile cancer (red), for candidate epigenetically regulated genes involved in the HPV-driven penile cancer (GPX5 (B) and GRAMD4 (C)). Profiles show Feature annotation is as provided by BeadChip, and methylation values are color-coded accordingly: orange = TSS1500, [1500–200 bp upstream of the transcription start site (TSS)]; red = TSS200 [200 bp upstream of the TSS]; yellow = 5’ untranslated region (UTR), blue = gene body; black = CpG islands; darker gray = CpG shores; and light gray = CpG shelves.
HPV epigenetic signature across different tissue lineages, there is little overlap in epigenetically altered loci, with only 30 overlapping loci MVPs in both tissue types. Analysis of the methylation state of these loci reveals a distinct hypomethylated signature associated with HPV-positive disease (data not shown). For cross validation, we performed a shrunken centroid class prediction used the 30 MVPs and were able to accurately predict the HPV status of 27 of 28 HPV-positive and 57 of 58 HPV-negative samples from the combined penile cancer-HNSCC training cohort. We were also able to accurately predict the HPV status of a panel of HPV-positive and HPV-negative HNSCC cell lines (n = 6; Supplementary Fig. S6).

We subsequently applied this HPV epi-signature to an independent set of HNSCC (n = 310) and CESC (n = 136) samples. When applied to HNSCC, the HPV epi-signature predicted 40 HPV positive and 290 HPV negative (Fig. 6). When comparing those samples with a known HPV status this accurately predicted the HPV status of 299/310 HNSCC samples (4 false positives, 7 false negatives), giving an overall misclassification rate of 3.5% (Fig. 6A).

When comparing the predicted HPV status of all 310 HNSCC samples with pathological features, there was a significant association with patient overall survival, with a 5-year survival for signature negative patients of 38% compared to 81% for signature positive patients (P = 0.00073, HR = 5.6, 95% CI 10.021–0.78) (Fig. 6B) although not independent of HPV status. There was no significant association of our HPV epigenetic signature with stage, age or gender.

We also assessed an independent cohort of 136 cervical cancer samples, using the same 30 loci HPV epi-signature 66% (90) were predicted to be signature positive compared to 34% (46) predicted to be signature negative. Epi-signature negative samples had a significantly (P = 0.05) worse overall survival than signature positive samples, with a 5 year overall survival for
signature-positive patients of 77% compared with 50% for signature-negative patients. Age ($P = 0.0552$) and stage ($P = 0.035$) were also significant in multivariate analysis.

As >90% of CESCC are a result of HPV infection, using only those samples with a known HPV status ($n = 84$), we compared the predicted and actual HPV status (Fig. 6C). Of those 62 epi-signature–positive samples, 53 (85%) were HPV16 positive, compared with 9 of 62 (15%) which contained other high-risk HPV subtypes, including HPV18. Of those epi-signature negative samples only 2 of 22 (9%) contained HPV16, suggesting the possibility of a HPV16-specific epigenetic alteration signature. Of the 84 patients with a confirmed HPV genotype, 73 had confirmed outcome data. Signature-positive patients had a significantly better overall survival than signature negative (Fig. 6D; $P = 0.0094$; $HR = 3.91$; 95% CI: 0.13–0.78; adjusted for age, grade and stage). Despite correlating strongly with HPV genotype, the HPV16-signature appears a stronger predictor of CESCC patient survival than HPV genotype alone ($P = 0.07$; $HR = 2.56$; 95% CI: 0.14129–1.083; adjusted for age, grade, and stage).

**Discussion**

Penile cancer is a rare disease in the developed world; however, it represents a significant source of patient morbidity and mortality in developing nations. The results reported here represent the most comprehensive epigenetic study of penile squamous cell carcinoma to date and shed light on the epigenetic alterations involved in penile cancer. Using high-density genome-wide methylation arrays, we have revealed distinct penile cancer–associated epigenetic signatures and define an epigenetic signature which can predict local lymph node metastasis, one of the most important prognostic indicators for penile cancer survival, and, to our knowledge, this is the first study to demonstrate the existence of an HPV-mediated DNA methylation signature in HPV-positive penile cancer.

Previous studies have identified differentially methylated genes in penile cancer (14, 15). These have been targeted studies in which candidate epigenetic-regulated genes have been identified, including RAS and TFIHB1. Using the Illumina Infinium Human Methylation arrays, we defined more than 1,200 DMRs associated with the malignant phenotype and CIMP relating to 367 genes. Supervised analysis of penile cancer versus normal tissue identified penile cancer-associated hypermethylated genes with significant enrichment of genes which are targets of PRC2 complex, these include TBX5, GATA4, CDH7, and SOX14. Aberrant methylation of genes regulated by the PRC2 complex has been observed in many cancer types, including head and neck, cervical, and prostate cancer but not previously in penile cancer. However, changes in the epigenetic regulation of PRC2 target genes has been noted during the HPV16 transformation of normal foreskin keratinocytes, with HPV16 infection resulting in the increased EZH2 expression and decreased global H3K27me3 (30). Furthermore, we also see overexpression of the members of the PRC2 complex (EZH2 and SUZ12) in penile cancers. This has been reported in other tumor types and shown to result in loss of PRC2 target gene expression (31). These data would suggest that deregulation (through either aberrant methylation, altered histone code or increased PRC2 complex expression) of PRC2-regulated genes is an essential part of the oncogenic transformation of both HPV and non-HPV–related penile cancer and warrants further investigation.

The hypermethylation of TSGs is a key feature of tumorigenesis. To identify key TSGs regulated by methylation, we compared with both MVP and DMR with a list of 712 known TSGs. This included CDO1, which we also show to be differentially expressed between penile cancer and normal tissue. The inactivation of CDO1 by DNA methylation has recently been implicated in many cancers, including bladder, breast cancer, colon, and lung cancer (32–36). Cysteine deoxygenase 1 (CDO1) is integral to the biodegradation of toxic cysteine, and reduced CDO1 expression has been shown to increase cell proliferation in vitro, whereas overexpression resulted in decreased tumor growth both in vitro and in vivo (33).

We also identified aberrant methylation of several potential therapeutic targets, including the hypermethylation and epigenetic regulation of the androgen receptor (AR). The aberrant methylation of the AR is particularly intriguing. Increased AR signaling is important in hormonally driven tumors, including prostate and breast cancers. Although it is assumed increased AR expression is oncogenic in hormonally driven cancers, it has recently been shown that loss of AR in hormone-refractory prostate cancer results in the activation of STAT3 (37). STAT3 regulates gene involved in the control of cellular processes, including proliferation, survival, and immune responses (38). Persistent activation of STAT3 is oncogenic and has been implicated in the development of a wide variety of human malignancies, including leukemia and lymphoma, and solid tumors, including head and neck cancer, prostate, breast, and colon cancers (39–41). Although still to be functionally validated, these data would suggest the potential for a pivotal role for loss of the androgen receptor in the development of penile cancer.

The presence of metastatic disease in the inguinal lymph nodes is one of the most important prognostic factors in penile cancer (42). Occult nodal metastasis are present in 20% to 25% of cases at presentation (43, 44) and inguinal lymph node dissection is largely directed by clinical examination and the histopathologic features of the primary lesion. Because of the lack of biomarkers which can accurately identify or predict lymph node metastasis, all patients with ≥T1G2 disease and impalpable inguinal lymph nodes undergo inguinal lymphadenectomy (removal of the inguinal lymph nodes), which is unnecessary in 75% to 80% of patients. Lymph node metastasis is an independent predictor of survival in penile cancer and therefore may be used as a surrogate disease-specific survival (45).

Methylome analysis identified a distinct epigenetic signature associated with lymph node metastasis. This 122 CpG classifier, which in cross validation, could predict the lymphatic metastases with an accuracy of 93%. The majority of MVPs were located within DMRs in four genes, HMX3, IRAF, FLI1, and PP2R2SC and DMR methylation was also predictive of lymph node–positive disease. When combined as predictive methylation index for each sample, the predictive accuracy of this signature (90% methylation array and 89% for qMSP) to identify the presence of lymph node metastasis is at least comparable with if not better than the sensitivity of sentinel lymph node biopsy. We are currently assessing the feasibility of using the methylation state of these loci as biomarkers in “liquid” biopsy, using plasma cell free DNA to detect metastasis-specific methylation events.

Finally, we also sought to understand the relationship between epigenetic alterations and HPV and clinical pathologic factors. High-risk HPV infection is a key oncogenic driver in several
different tumor types, including, cervical cancer, head and neck squamous cell cancers along with penile cancer. It is well documented in HNSCC and cervical cancers that HPV infection results in the epigenetic reprogramming of the host cell during malignant transformation resulting in a distinct HPV-induced epigenetic phenotype (20, 46). In this cohort, we found HPV infection in 23% of samples which was lower than expected although the incidence of HPV-positive penile cancer ranges from 14% to 100% and is also dependent on prior circumcision which was not recorded in our cohort (47). Only HPV16 was detected in our cohort and HPV16 represents the predominant subtype in penile cancer and head and neck cancers (20, 45, 48). We defined a distinct, predominately hypomethylated, HPV16-associated epigenetic signature. This large probe set was able to accurately separate an independent cohort of HNSCC cases, suggesting a genetic signature. This large probe set was able to accurately separate an independent cohort of HNSCC cases, suggesting a lineage independent HPV-speciﬁc epigenetic phenotype (20).

In this cohort, the apparent synergy in epigenetic alterations associated with HPV infection, only 30 HPV-Speciﬁc MVPs were found to be overlapping between the two cohorts. We validated this minimal HPV signature, in independent HNSCC and CESC cohorts, and show it to be predictive of disease-free survival in both HNSCC and CESC, and predictive of HPV infection in HNSCC. Interestingly when applied to CESC, this signature appeared to separate by HPV subtype, speciﬁcally HPV16 versus HPV18/other HPV, supporting the postulate that we have deﬁned a HPV16 signature. While 50% to 60% of CESC are associated with HPV16 infection, a further 20% are associated with HPV18 (6, 8, 49); this contrasts with HNSCC and penile cancer in which 90% of HPVs are HPV16. We found only HPV16 in each of the two training cohorts. Although only a single CESC cohort, these data suggest the presence of speciﬁc HPV subtype epigenetic alterations, and further suggest a distinct survival advantage to HPV16-driven tumors compared with those associated with other high-risk HPVs, such as HPV18 (50). In future studies, it will be important to elucidate the functional impact of differential methylation of these genes and their role in HPV subtype–speciﬁc driven cancer development. In terms of clinical utility, this novel methylation signature can be tested as a strategy to stratify cases at high risk with the potential to direct multimodal therapy. Moreover, the encoded proteins affected by aberrant methylation may represent promising drug targets for innovative and more efﬁcient cancer therapy.

In summary, this work shows that changes in DNA methylation are key components in penile cancer. We show the utility of an epigenetic signature, which has been validated on an independent cohort, to identify occult lymph node metastasis in penile cancer with equivalent or greater sensitivity to methods in current clinical practice. In addition we deﬁne a penile cancer–speciﬁc HPV signature and a HPV-associated host epigenetic signature which is a lineage-independent predictor of disease-free survival and suggests distinct HPV subtype–speciﬁc epigenetic alterations.

Disclosure of Potential Conﬂicts of Interest
A. Muneeb reports receiving speakers bureau honoraria from Eli Lilly. No potential conﬂicts of interest were disclosed by the other authors.

Authors’ Contributions
Conception and design: A. Feber, M. Arya, M. Saqib
Development of methodology: A. Feber, M. Arya, M. Saqib
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): M. Arya, P. de Winter, M. Saqib, R. Nigam, P.R. Malone, W.S. Tan, S. Rodney, M. Lechner, A. Freeman, A. Stunnet, J.D. Kelly
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): A. Feber, P. de Winter, M. Lechner, J.D. Kelly
Writing, review, and/or revision of the manuscript: A. Feber, M. Arya, P. de Winter, M. Saqib, S. Rodney, A. Freeman, C. Jameson, A. Muneeb, S. Beck
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): M. Arya, P. de Winter, A. Muneeb
Study supervision: A. Muneeb, J.D. Kelly

Acknowledgments
The authors thank Jenny Patterson and UCL Advanced Diagnostics for the assistance with immunohistochemical analysis, along with Kerra Peirce from UCL genomics for processing the Illumina methylation arrays.

Grant Support
A. Feber is supported by the UCL/UCLH Comprehensive Biomedical Research Centre; the Rosetrees Trust and research involved in this project was supported by Orchid. M. Arya is supported by Orchid. Research in the Beck lab was supported by the Welcome Trust (WT084071, WT093855), Royal Society Wolfson Research Merit Award (WM100023), MRC (G100041), IMI-JU Onco-Track (115234) and EU-FP7 projects EPIGENESYS (257082), IDEAL (259679), and BLUEPRINT (282510). J.D. Kelly is supported by the UCL/IFIC Comprehensive Biomedical Research Programme.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received June 30, 2014; revised November 25, 2014; accepted December 1, 2014, published OnlineFirst December 23, 2014.

References
Development of methodology: A. Feber, M. Arya, M. Saqib
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): M. Arya, P. de Winter, M. Saqib, R. Nigam, P.R. Malone, W.S. Tan, S. Rodney, M. Lechner, A. Freeman, A. Stunnet, J.D. Kelly
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): A. Feber, P. de Winter, M. Lechner, J.D. Kelly
Writing, review, and/or revision of the manuscript: A. Feber, M. Arya, P. de Winter, M. Saqib, S. Rodney, A. Freeman, C. Jameson, A. Muneeb, S. Beck
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): M. Arya, P. de Winter, A. Muneeb
Study supervision: A. Muneeb, J.D. Kelly

Acknowledgments
The authors thank Jenny Patterson and UCL Advanced Diagnostics for the assistance with immunohistochemical analysis, along with Kerra Peirce from UCL genomics for processing the Illumina methylation arrays.

Grant Support
A. Feber is supported by the UCL/UCLH Comprehensive Biomedical Research Centre; the Rosetrees Trust and research involved in this project was supported by Orchid. M. Arya is supported by Orchid. Research in the Beck lab was supported by the Wellcome Trust (WT084071, WT093855), Royal Society Wolfson Research Merit Award (WM100023), MRC (G100041), IMI-JU Onco-Track (115234) and EU-FP7 projects EPIGENESYS (257082), IDEAL (259679), and BLUEPRINT (282510). J.D. Kelly is supported by the UCL/IFIC Comprehensive Biomedical Research Programme.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received June 30, 2014; revised November 25, 2014; accepted December 1, 2014, published OnlineFirst December 23, 2014.

References


