A peer-reviewed version of this preprint was published in PeerJ on 19 September 2018.

Lam KC, Vyshenska D, Hu J, Rodrigues RR, Nilsen A, Zielke RA, Brown NS, Aarnes E, Sikora AE, Shulzhenko N, Lyng H, Morgun A. 2018. Transkingdom network reveals bacterial players associated with cervical cancer gene expression program. PeerJ 6:e5590 https://doi.org/10.7717/peerj.5590

Transkingdom network reveals bacterial players associated with cervical cancer gene expression program

Khiem Chi Lam [1,2], **Dariia Vyshenska** [2], **Jialu Hu** [2,3], **Richard Rosario Rodrigues** [2], **Anja Nilsen** [4], **Ryszard A. Zielke** [2], **Nicholas Samuel Brown** [2], **Eva-Katrine Aarnes** [4], **Aleksandra E. Sikora** [2], **Natalia Shulzhenko** [5], **Heidi Lyng** [Corresp., 4], **Andrey Morgun** [Corresp. 2]

[1] National Institutes of Health, Center for Cancer Research, National Cancer Institute, Bethesda, MD, United States

[2] Oregon State University, College of Pharmacy, Corvallis, OR, United States

[3] Northwestern Polytechnical University, School of Computer Science, Xi'an, China

[4] Oslo University Hospital, Institute for Cancer Research, Norwegian Radium Hospital, Oslo, Norway

[5] Oregon State University, Carlson College of Veterinary Medicine, Corvallis, OR, United States

Corresponding Authors: Heidi Lyng, Andrey Morgun
Email address: Heidi.Lyng@rr-research.no, andriy.morgun@oregonstate.edu

Cervical cancer is the fourth most common cancer in women worldwide with human papillomavirus (HPV) being the main cause of disease. Chromosomal amplifications have been identified as a source of upregulation of cervical cancer driver genes but cannot fully explain increased expression of immune genes in invasive carcinoma. Insight into additional factors that may tip the balance from making the immune system tolerate HPV to eliminate the virus may lead to markers for better diagnosis. We investigated whether microbiota affect molecular pathways in cervical carcinogenesis by performing microbiome analysis via sequencing 16S rRNA in tumor biopsies from 121 patients. While we detected a large number of intra-tumor taxa (289 OTUs), we focused on the thirty-eight most abundantly represented microbes. To search for microbes and host genes potentially involved in the interaction, we reconstructed a transkingdom network by integrating previously discovered cervical cancer gene expression network with our bacterial co-abundance network and employed bipartite betweenness centrality (BiBC). The top ranked microbes were represented by the families Bacillaceae, Halobacteriaceae, and Prevotellaceae. While we could not define the first two families to the species level, Prevotellaceae was assigned to *Prevotella bivia*. By co-culturing a cervical cancer cell line with *P. bivia*, we confirmed that three out of ten top predicted genes in the transkingdom network (LAMP3, STAT1, TAP1), all regulators of immunological pathways, were upregulated by this microorganism. Therefore, we propose that intra-tumor microbiota might contribute to cervical carcinogenesis through the induction of immune response drivers, including the well-known cancer gene LAMP3.

Transkingdom network reveals bacterial players associated with cervical cancer gene expression program

Khiem Chi Lam [1, 3], Dariia Vyshenska [1]*, Jialu Hu [1,4], Richard Rosario Rodrigues [1], Anja Nilsen [2], Ryszard A. Zielke [1], Nicholas Samuel Brown [1], Eva-Katrine Aarnes [2], Aleksandra E. Sikora [1], Natalia Shulzhenko [1], Heidi Lyng [2†], Andrey Morgun [1†]*

[1] Oregon State University, Corvallis, OR 97330, USA

[2] Institute for Cancer Research, Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway

[3] Inflammatory Cell Dynamics Section, Cancer and Inflammation Program, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, MD 20892, USA

[4] School of Computer Science, Northwestern Polytechnical University, Xi'an, China

* - these authors contributed equally to this work

† - co-senior/co-corresponding authors:

Heidi Lyng[2]

e-mail: heidi.lyng@rr-research.no

Andrey Morgun[1]

andriy.morgun@oregonstate.edu

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26957v1 | CC BY 4.0 Open Access | rec: 24 May 2018, publ: 24 May 2018

List of abbreviations

HPV – Human papillomavirus;

hrHPV – high-risk Human papillomavirus;

OTU – Operational Taxonomic Units;

LAMP3 – lysosome-associated membrane glycoprotein 3;

STAT1 – signal transducer and activator of transcription 1;

TAP1 – transporter 1, ATP binding cassette subfamily B member;

qPCR – quantitative polymerase chain reaction;

DNA – deoxyribonucleic acid;

rRNA – ribosomal RNA;

BiBC – bipartite betweenness centrality;

GZMB – granzyme B;

CXCL10 – C-X-C motif chemokine ligand 10.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26957v1 | CC BY 4.0 Open Access | rec: 24 May 2018, publ: 24 May 2018

Abstract

Cervical cancer is the fourth most common cancer in women worldwide with human papillomavirus (HPV) being the main cause of disease. Chromosomal amplifications have been identified as a source of upregulation of cervical cancer driver genes but cannot fully explain increased expression of immune genes in invasive carcinoma. Insight into additional factors that may tip the balance from making the immune system tolerate HPV to eliminate the virus may lead to markers for better diagnosis. We investigated whether microbiota affect molecular pathways in cervical carcinogenesis by performing microbiome analysis via sequencing 16S rRNA in tumor biopsies from 121 patients. While we detected a large number of intra-tumor taxa (289 OTUs), we focused on the thirty-eight most abundantly represented microbes. To search for microbes and host genes potentially involved in the interaction, we reconstructed a transkingdom network by integrating previously discovered cervical cancer gene expression network with our bacterial co-abundance network and employed bipartite betweenness centrality (BiBC). The top ranked microbes were represented by the families Bacillaceae, Halobacteriaceae, and Prevotellaceae. While we could not define the first two families to the species level, Prevotellaceae was assigned to *Prevotella bivia*. By co-culturing a cervical cancer cell line with *P. bivia*, we confirmed that three out of ten top predicted genes in the transkingdom network (LAMP3, STAT1, TAP1), all regulators of immunological pathways, were upregulated by this microorganism. Therefore, we propose that intra-tumor microbiota might contribute to cervical carcinogenesis through the induction of immune response drivers, including the well-known cancer gene LAMP3.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26957v1 | CC BY 4.0 Open Access | rec: 24 May 2018, publ: 24 May 2018

Introduction

Cervical cancer remains the fourth most common cancer in women worldwide, and the world's second highest cause of female cancer mortality (Ginsburg et al.). Persistent infection with high-risk human papillomavirus (hrHPV) is a causative factor in cervical carcinogenesis (Walboomers et al.). The HPV vaccination is expected to decrease the cancer incidence, however, in Central and Eastern Europe and developing countries where there is a lack of systematic screening and vaccination programs, morbidity and mortality are anticipated to remain high. Thus, cervical carcinogenesis will remain a major threat to women's health. Most women who are infected with HPV never develop cancer. Hence, persistent HPV infection is necessary but may not be sufficient to trigger cancer development. A better understanding of the additional factors required for carcinogenesis would improve identification of the high-risk cases and would represent a major step towards personalized medicine in cervical cancer (Shulzhenko et al.).

According to the current model of cervical carcinogenesis, hrHPV exists in the episomal state within basal cells in the epithelium during early disease onset (Munoz et al.). The viral gene E2, while expressed, suppresses expression of E6 and E7 viral oncogenes. The longer the infection persists, more HPV integrates into the host genome. Upon integration, the open reading frame of E2 is disrupted which reduces the control of E6 and E7 and promotes cell proliferation (Shulzhenko et al.). Therefore, the key event that turns chronic HPV infection into cancer is elimination of the episomal form of virus as it is the only source of the oncosuppressor E2. The host immune system may eliminate episomal HPV, giving the cells with an integrated form of virus a growth advantage (Herdman et al.). Hence, although insufficient antiviral immunity enables persistent HPV infection at an early state, proceeding to the integrated state seems to

require activation of the woman's immune system. Chromosomal amplifications in the infected cells have recently been identified as one of the key sources of upregulation of driver genes in this disease (Mine et al.). However, these genomic alterations cannot fully explain increased expression of immune genes in invasive carcinoma. Additional factors that may tip the balance from making the immune system tolerate HPV to eliminate virus are so far unknown.

Microbiota can have a significant role in disease, contributing to development of metabolic (diabetes (Greer et al.), enteropathy associated with common variable immunodeficiency (shulzhenko et al.) etc.) and immune (IBD (Knights et al.), asthma (Ege et al. ; Stokholm et al.), allergy (Fujimura & Lynch)) disorders as well as cancer (colorectal (Geng et al. ; Marchesi et al.), gastric (Uemura et al.), lung (Hosgood et al.), pancreatic (Farrell et al.) etc.). Microbial communities have also been found to specifically contribute to virus induced carcinogenesis (Vyshenska et al.). In particular, multiple studies confirm vaginal dysbiosis (bacterial vaginosis) to be a risk factor for HPV infection (Gillet et al.) and its progression (Gillet et al. ; Oh et al.). Studies of vaginal (Champer et al.) and cervical (Vyshenska et al.) microbiota have found association between changes in microbial community, like overall diversity and abundances of particular taxa, and cancer development. Furthermore, the location of this cancer suggests a possible role of common members of cervico-vaginal microbiota in cervical carcinogenesis. While almost 25 years ago, bacteria cultured from cervical cancer biopsies were proposed to contribute to cancer progression, we are still far from understanding the role of non-viral microbes in this disease (Mikamo et al.).

In this study, we aimed to investigate whether microbiota may affect molecular pathways in cervical carcinogenesis. We reconstructed a transkingdom network that integrates microbiome and host transcriptome data in tumor samples from patients to infer key bacterial players. We

demonstrate the power of this approach by identifying several bacterial candidates and show that one of them (*Prevotella bivia)* upregulates a well-known human cancer driver, lysosomal associated membrane protein 3 (LAMP3).

Materials & Methods

Patients

Tumor specimens were retrieved from 123 patients with locally advanced squamous cell carcinoma of the uterine cervix. One to 4 biopsies were taken at different locations of the tumor at the time of diagnosis, immediately frozen in liquid nitrogen, and stored at −80°C. DNA and RNA from different biopsies of the same tumor were pooled. The clinical protocol was approved by the Regional Committee for Medical Research Ethics in southern Norway (REC no. S-01129). Written informed consent was obtained from all patients (Supplemental Form S1). DNA was isolated according to a standard protocol with proteinase K, phenol, chloroform, and isoamylalcohol (De Angelis et al.) . Purified DNA quality and concentration were assessed using the Quant-iT™PicoGreen® dsDNA Assay Kit (Life Technologies, Carlsbad, CA, USA).

Gene expression

Gene expression profiling of 123 tumors used in this study was performed using the Illumina HumanWG-6 v3 Expression BeadArrays with approximately 48,000 transcripts (Illumina Inc., San Diego, CA). All samples had more than 50% tumor cells in hematoxylin and eosin stained sections. This selection may have led to some bias in the results, but was chosen to reduce the influence of different normal cell proportion across the samples. Total RNA was isolated by the use of Trizol reagent (Life Technologies) followed by LiCl precipitation.

Hybridization, scanning, signal extraction and normalization were performed as described (Lando et al.).

Bacterial DNA quantification

Bacterial content was quantified using QuantiFast SYBR Green mix (Qiagen, Germantown, MD) and universal bacterial primers, UniF340 (5'-ACTCCTACGGGAGGCAGCAGT) and UniR514 (5'-ATTACCGCGGCTGCTGGC). Standards were created from serial dilutions of extracted DNA from bacteria grown from mouse cecum contents.

16S rRNA library preparation and sequencing

For MiSeq Illumina sequencing, total genomic DNA was subjected to PCR amplification targeting the 16S rRNA variable region 4 (V4) using the bacterial primers 515F/806R. Test reactions were performed with samples with varying amounts of bacterial DNA to determine the threshold for successful library preparation, defined as a positive band on an agarose gel. 40-250 ng of total template DNA were used for each PCR reaction, performed in triplicates per sample. Each set of triplicate PCR reactions was pooled and purified using the MinElute PCR Purification Kit (Qiagen). The pools were checked for proper band size of 382 bp by gel electrophoresis using 2% agarose pre-cast E-gels (Life Technologies). Negative controls consisted of samples without template for DNA extraction and PCR amplification. The pools were then quantified using the Qubit dsDNA BR Assay Kit (Life Technologies). Barcoded amplicons were pooled at equal volumes and concentrations (2 uL of 5 ng/uL DNA). The total pool was sequenced using the Illumina MiSeq 2000 sequencing platform at the Center for Genome Research and Biocomputing at Oregon State University (OSU) to generate pair-ended

250 nt reads. The dataset generated and analyzed during the current study is available in Sequence Read Archive under accession no. SRP131188.

Processing of raw 16S rRNA reads

Raw forward-end fastq reads from the Illumina sequencing output were quality-filtered, demultiplexed, and analyzed using quantitative insights into microbial ecology (QIIME) (Caporaso et al.) . Reads were quality filtered using default QIIME parameters; reads with a Phred quality score of <20, ambiguous base calls, and fewer than 187 nt (75% of 250 nt) of consecutive high-quality base calls were discarded. Additionally, truncation occurred on reads with three consecutive low-quality bases. The samples were demultiplexed using 12 bp barcodes, allowing for a maximum of 1.5 errors in the barcode.

Reads were clustered using UCLUST (Edgar) at 97% similarity into operational taxonomic units (OTUs) at QIIME default parameters. A representative set of sequences from each OTU were selected for taxonomic identification by selecting the cluster seeds (first read assigned to that OTU). Representative sequences for each OTU were aligned using BLAST (e-value < 0.001) to Greengenes (version 13.8) OTU reference sequences (97% similarity) to obtain taxonomy assignments. OTUs were filtered for singletons (only found in one sample) and relative abundance was quantified by dividing raw read counts by total number of reads for each sample. Alpha diversity rarefaction curves using the Shannon index and a heatmap of OTU frequencies were obtained from QIIME scripts using default parameters.

Comparing bacterial communities from different body sites

HMP data. High-quality fastq reads for region v3 and v5 were obtained from (ftp://public-ftp.ihmpdcc.org/HMQCP/seqs_v35.fna.gz). Female patients from the first visit were retained if they had samples from vagina (Posterior_fornix) and at least stool or skin (Left and Right of Antecubital fossa and Retroauricular crease) to allow multi-site comparisons for the same patient. In case there were multiple samples from the same site of a patient, we retained the sample with highest number of sequences assigned to OTUs. This resulted a total of 380 samples from 77 females.

Healthy cervix. Raw fastq data for 17 HPV neg samples was obtained from European Nucleotide Archive, Study PRJEB1872, and used as healthy cervix samples. High-quality reads (Phred score > 19) were retained using QIIME's split_libraries_fastq.py command and a Phred offset of 33.

To allow comparisons between studies, reads for the 455 samples (Table S2) were assigned to OTUs at 97% sequence similarity using UCLUST and closed reference OTU picking. The reverse strand match was enabled for the HMP data during OTU picking. The studies were merged using common OTU IDs. Singleton OTUs were removed. The combined OTU table had 455 samples and 2,516,046 sequences assigned to 444 OTUs (Table S2). The table was rarefied with a sequence threshold of 1000 sequences (Table S2) and used for diversity analysis. Beta diversity was calculated using weighted and unweighted unifrac and used for PCoA analysis. Alpha diversities were compared using Shannon diversity index.

Cervical cancer samples with greater than 1000 sequences (based on the above threshold used for rarefaction) assigned to OTUs were retained and singleton OTUs were removed. The

resulting OTU table had 432 OTUs for 52 cervical cancer samples (Table S2), was relativised and used to create a heat tree at the genus level using Metacoder (Foster et al.).

OTU collapsing

An initial quality check of *de novo* cervical cancer OTUs were made against healthy vaginal samples from the Human Microbiome Project (data not shown). OTU representative sequences, OTU table, and mapping file for 16S rRNA V3-V5 sequencing were downloaded from the Human Microbiome Project (HMP) QIIME Community Profiling datasets publicly available . Representative sequences from cervical cancer *de novo* picked OTUs were aligned to HMP representative sequences using USEARCH. Sequences with identity matches greater than 0.97 were collapsed together, keeping HMP OTU identification, taxonomy, and representative sequences as appropriate; if multiple cervical cancer OTUs matched, their read counts were summed for each sample. These new "collapsed" OTUs were used for subsequent analyses.

Transkingdom network reconstruction

Pairwise Spearman correlations were calculated for each OTU with mean relative abundance >0.5%. Spearman correlations were additionally calculated between top OTUs and 738 differentially expressed genes (quantile-normalized gene expression microarray, raw data available in Gene Expression Omnibus under accession no. GSE68339) found in previous cancer gene regulatory network (Mine et al.). Correlations with $P < 0.001$ and FDR < 0.1 were imported into igraph R package for network reconstruction. Gene-OTU correlations were integrated with OTU-OTU correlations along with the previously published gene-gene network. Multiple edges and self-loops were removed. Visualization was performed in Cytoscape.

Bipartite betweenness centrality (biBC) is the measure of probability for nodes belonging to one subnetwork to be bottlenecks in the transfer of signal to the nodes in another subnetwork, and vice versa. It was calculated as previously described (Dong et al.) between the microbial subnetwork and each differentially expressed gene (DEG) subnetwork as well as all DEGs. Relativized biBC values were calculated by dividing the value of a particular node to the sum of all nodes in that metric. Relativized biBC values were then log transformed as followed:

$$\log2((biBC*10^6)+1)$$

Prevotella OTU alignment

Representative sequences were extracted for all original cancer *de novo* OTUs that were collapsed into OTU_97.1949. These sequences were aligned to SILVA 16S rRNA sequences for *Prevotella* using USEARCH. Matches with length >200 bp and mismatch ≤40 bp were kept for quantifications. For each species, the numbers of hits for each representative sequence were summed after normalization to the total number of hits for that representative sequence. This method was used due to multiple database matches for each given representative sequence due to high similarity and low specificity of the query sequence.

Human cell culture

HeLa cervical cancer cell line was acquired from ATCC and maintained in Modifed Eagle's Medium (EMEM) with Earle's Balanced Salt Solution, L-Glutamine, and Non-Essential Amino Acids, without Calcium. EMEM was supplemented with 5% Heat-Inactivated Fetal Bovine Serum (FBS) and 1% Penicillin Streptomycin solution (PEST). Frozen cells stored in 10% DMSO in liquid nitrogen were thawed in a 37°C water bath until thawed (approx. 3 min), added to 4 mL of fresh media, and centrifuged for 3 minutes at 2500 rpm to pellet cells. The

media solution was aspirated, and cells were resuspended in 15 mL of fresh media. Cells were grown in 5% CO_2 at 37°C in 75 cm^2 treated, vented cap flat bottom culture flasks. Cells were passaged at approximately 80% confluency using 3 mL of 0.25% Trypsin supplemented with 2.21 mM EDTA without sodium bicarbonate to detach adherent cells.

Bacteria cell culture

The following bacteria were used in this study: *Prevotella bivia* (ATCC 29303) and *Lactobacillus crispatus* (ATCC 33197). All bacteria were grown on Brucella blood agar plates supplemented with hemin and vitamin K (Hardy Diagnostics) at 37°C in anaerobic conditions using the GasPak™ EZ Container System (BD Diagnostics). Cultures were stored in 25% glycerol at -80°C. Bacterial plates were restreaked once every 2-3 days as needed for experiments.

Human-bacteria co-culture

Bacterial cultures were taken from frozen stocks, plated on Brucella blood agar plates, and restreaked after 3 days. After incubation for 2 days, bacteria were suspended in 3 mL of PBS and quantified by optical density at OD600 nm; correlations between OD600 and Colony Forming Units (CFU) were made prior to the experiments. HeLa cells were resuspended in EMEM media without PEST, counted using a hemocytometer, seeded at 75,000 cells/well in 24-well flat bottom culture plates, and incubated at 37°C in 5% CO_2 24 hours prior to bacterial treatment. Bacteria were adjusted to appropriate concentration for 40 uL treatments at a multiplicity of infection (MOI) of 10 in replicates of 3 or more, using sterile PBS as the negative control. Bacteria-treated HeLa cells were incubated at 37°C in anaerobic conditions using

GasPak™ EZ Anaerobe Container System Sachets for either 24 hours, followed by on-plate lysis using RLT Lysis Buffer (Qiagen) and lysate storage at -80°C.

RT-qPCR for co-culture experiment

RNA was extracted from cell lysate using the RNeasy Mini Kit (Qiagen) with on-column DNase digestion. RNA was quantified using Qubit RNA BR Assay Kit (Life Technologies) and reverse transcribed using qScript cDNA Synthesis Kit (Quantabio, Beverly, MA). RT-qPCR was performed using PerfeCTA SYBR Green FastMix (Quantabio) and human gene primers (Table S5). RT-qPCR set up was as follows: sample was heated to 95°C, followed by 40 cycles of 95°C for 10 sec and 60°C for 30 sec. Fold change was calculated by normalizing to 18S rRNA housekeeping gene.

Results

Varying amounts of bacterial DNA were detected by quantitative PCR (qPCR) in 121 out of 123 tumor samples from cervical cancer patients (Table S1). Sufficient amounts for 16S rRNA library preparation and sequencing was found in 58 samples. After 16S rRNA sequencing and quality filtering, we obtained 3,975,755 high-quality reads with a median of 87,320 reads/sample.

As a first step of analysis, we evaluated our cancer microbiome data in relation to publicly available datasets for adjacent healthy body sites (Audirac-Chalifour et al. ; Di Paola et al. ; Gajer et al. ; Hong et al. ; Lee et al. ; Ling et al. ; Liu et al. ; Ravel et al. ; Si et al.) and sites sampled by the Human Microbiome Project . While such analysis is limited since the samples were not matched (i.e. not the same individuals), it provides general information about the

microbiota in cervical cancer compared to other tissues. Unsupervised principal coordinate analysis using common OTUs showed that the cervical cancer samples had a different microbiome composition than samples from healthy cervix, vagina, stool, or skin samples (Fig. 1a). Compared to healthy uterine cervix and vagina, the cancer microbiome showed increased alpha diversity (Fig. S1) and high abundance of genera such as *Prevotella* (27.3%), *Fusobacterium* (17%), and *Peptoniphilus* (10.8%) as well as a low abundance of *Lactobacillus* (0.5%) (Fig. 1b, c; Table S2).

To investigate whether the bacteria influence molecular pathways in the tumors, we built a transkingdom network that integrated microbial abundance and cancer gene expression networks. First, OTUs with an abundance of at least 0.5% were selected, and a bacterial co-abundance network that contained 38 nodes and 54 edges was reconstructed. For the human portion of the transkingdom network, we selected the network based on meta-analysis of transcriptomes from five cohorts of cervical cancer patients (Shulzhenko et al.). In that study, we reconstructed a gene expression meta-network, identifying three pathways with up-regulated cell cycle and proinflammatory/antiviral genes, and down-regulated epithelial cell differentiation genes as key features of cervical carcinogenesis. To integrate the microbial and human subnetworks we calculated correlations between microbial abundances and gene expression from the cancer network measured in the same tumor samples as bacteria. The network contained 21 taxa and 698 human genes connected by 19 transkingdom edges (Fig. 2a).

The inflammatory/antiviral genes are the most plausible targets of regulation by bacteria. In addition, antiviral gene expression is of special interest in cervical cancer due to its involvement in elimination of episomal HPV that leads to tumor growth (Pett et al. ; Vyshenska

et al.). Therefore, to infer bacteria that may drive changes in cancer gene expression we searched for 'bottleneck' bacterial nodes that link microbial and inflammatory/antiviral subnetworks using the bipartite betweenness centrality (BiBC) metric (Dong et al. ; Morgun et al. ; Thomas et al.). The top three microbes with highest BiBC represented OTUs from families *Bacillaceae* and *Halobacteriaceae* and the genus *Prevotella* (Fig. 2b).

Due to limitations of 16S rRNA sequencing, QIIME was unable to assign these OTUs to specific species. However, the *Prevotella* OTU had more advanced assignment than the other two top OTUs, which were assigned only to family level. An alternative bioinformatics approach was therefore used to identify bacterial species that were represented by the *Prevotella*-related OTU. We aligned representative sequences from all OTUs that were collapsed into the *Prevotella* OTU (OTU_97.1949) to the SILVA 16S rRNA database containing sequences of *Prevotella* genus (Fig. 2c). While matches for seven different species from this genus were found, *P. bivia* had the most hits, indicating that this species was the most probable bacterium connected to cancer gene expression. Review of recent literature with publicly available 16S rRNA data of healthy cervix and vagina showed that *Prevotella* had much higher abundance in our cervical cancer cohort than in five studies of healthy vagina (Gajer et al. ; Hong et al. ; Ling et al. ; Liu et al. ; Ravel et al.) and four studies of healthy cervix (Audirac-Chalifour et al. ; Di Paola et al. ; Lee et al. ; Si et al.) (Fig. 2d, Table S4). Altogether these results provided *P. bivia* as a candidate bacterium that may contribute to upregulation of proinflammatory/antiviral genes in cervical cancer.

Cervical carcinomas consist of a mixture of stromal, immune and cancer cells (Pilch et al. ; Sheu et al.). Therefore, the bacteria harbored by tumor might have direct and indirect effects on

cancer cells. To test if bacteria predicted by our analysis can drive cervical cancer gene expression program directly, we performed *in vitro* co-incubation of *P. bivia* with a cervical cancer cell line (HeLa) and measured expression of selected genes. We used *Lactobacillus crispatus* as a control because it is a common standard in co-culture studies for bacterial induction of innate immunity (Eade et al. ; Libby et al.). While *L. crispatus* was among top most abundant bacteria detected in cancer samples, it showed decreased abundance in cervical cancer compared to other body sights (Fig. S2, Table S4) and it was not connected to any cancer genes in the transkingdom network (BiBC=0) indicating no influence on cancer gene expression program. To evaluate which genes might be the most affected by bacteria we calculated BiBC between the microbial portion of the transkingdom network and inflammatory/antiviral subnetwork and selected top ten genes for further analysis (Fig. 3a). Our gene expression data was generated from whole tissue samples, therefore, it was not surprising that two (GZMB, CXCL10) out of ten selected genes were not detected in the co-incubation experiments with the cancer cell line. Indeed, GZMB is mostly expressed by T lymphocytes and NK cells (Bratke et al. ; Johnson et al.) and probable sources of CXCL10 expression are endothelial and stromal cells (Panzer et al. ; Proost et al.). Among the eight remaining genes we found that three (LAMP3, STAT1, and TAP1) were upregulated by *P. bivia* (Fig. 3b) whereas no genes were downregulated by this bacterium (Fig. 3b, Fig. S3). These results further support that intra-tumor bacteria like *Prevotella* may be involved in control of the gene expression program of cervical cancer.

Discussion

The use of transkingdom network analysis in our study provided novel insight into the host-microbiome relationship in cervical cancer. The method has previously proved to be

successful for identification of members of the microbial community responsible for a variety of pathological situations, such as enteropathy associated with common variable immunodeficiency (shulzhenko et al.), diabetes (Greer et al.), as well as microbes and microbial genes that regulate the effect of antibiotics on the intestine (Morgun et al.). Our work demonstrates a new approach where transkingdom network is used to elucidate the pathogenic role of microbiota in cancer. We combined a bacterial co-abundance network with gene expression of key pathways in cervical carcinogenesis (Mine et al.), increasing the probability of detecting host-microbiome relationships of relevance for cancer development. The results strongly support a role of microbes from the *Bacillaceae* and *Halobacteriaceae* families and the genus *Prevotella* in regulation of a pro-inflammatory pathway that is activated in cervical cancer (Mine et al.). This opens the possibility that bacteria are involved in the elimination of episomal HPV in cervical cancer and thereby play a role in carcinogenesis.

Results from the transkingdom network analysis were further tested and showed that the most promising candidate, *P. bivia,* up-regulated three genes (LAMP3, STAT1, and TAP1) in the pro-inflammatory pathway (Mine et al.). Noteworthy, LAMP3 is one of the key drivers of this pathway that controls expression of STAT1 and several other antiviral genes (Mine et al.). Overexpression of LAMP3 has been shown to promote metastasis in cervical cancer xenografts and to associate with poor treatment outcome in clinical studies (Kanao et al.). Moreover, LAMP3 is induced by hypoxic conditions (Mujcic et al. ; Nagelkerke et al.) and potentially stimulates hypoxia dependent cancer metastasis (Lu & Kang ; Nagelkerke et al.). Co-culture of a cervical cancer cell line with *P. bivia* in our work showed that this bacterium indeed upregulates LAMP3 under anaerobic conditions. It is therefore reasonable to hypothesize that *P. bivia* (and

possibly anaerobes) attracted by hypoxic environment, infiltrates the tumor and induces a pro-inflammatory gene expression program via up-regulation of LAMP3. LAMP3 itself seems to play a crucial role in cervical cancer. Similarly to its role in other cancer types (Nagelkerke et al. ; Sun et al.), it promotes metastases and, by driving expression of multiple antiviral genes, it is potentially involved in elimination of episomal HPV which leads to overexpression of the E6 and E7 HPV oncogenes and disease progression (Munger et al.).

Our study is a first step towards connecting intra-tumor microbiome to molecular pathways operating in cervical cancer. We analyzed bulk gene expression of the tumor containing a mixture of transcriptomes of different cell types. Therefore, it might not be surprising that genes such as GZMB and CXCL10, which are expressed by non-cancerous cells, were among top ones predicted to be induced by intra-tumor microbiota. Therefore, future studies may take advantage of single cell sequencing allowing for identification of indirect communication between cancer cells and bacteria, for example via infiltrating immune cells. In addition, while our results pointed to several bacteria that might be important, we were able to focus only on one of them, *P. bivia*. It is commonly believed that microbiota may affect its host through combined effect of several taxa rather than as single bacterium (Lamont & Hajishengallis ; Round & Mazmanian). Improved species identification could probably be achieved by employing shotgun together with 16S rRNA sequencing (Dong et al. ; shulzhenko et al.), and this approach should be considered in future studies.

Conclusions

A possible involvement of bacteria in the carcinogenesis of cervical cancer could have implications for women presenting with persistent HPV infection. Unraveling the crosstalk

between cervico-vaginal microbiota and woman body may allow development of personalized preventative measures against the infection through shaping the microbiome to favor fast virus elimination (Kassam et al.). It could also help to identify single bacterial or community markers of disease progression for better diagnostics. When used together with HPV testing, such markers may lead to more precise evaluation of a woman's risk of developing cancer that would be especially useful in undeveloped countries where access to the healthcare system is limited (Denny et al.). In addition, expansion of our study to reveal whether the cancer microbiome may affect treatment outcomes at later stages of the disease (Iida et al. ; Muls et al. ; Vetizou et al.), could empower identification of bacterial and host gene targets for new anticancer therapies.

Acknowledgements

We thank Center for Genome Research and Biocomputing personnel for support in sequencing and computation infrastructure.

REFERENCES

2012a. A framework for human microbiome research. *Nature* 486:215-221. 10.1038/nature11209

2012b. Structure, function and diversity of the healthy human microbiome. *Nature* 486:207-214. 10.1038/nature11234

Audirac-Chalifour A, Torres-Poveda K, Bahena-Roman M, Tellez-Sosa J, Martinez-Barnetche J, Cortina-Ceballos B, Lopez-Estrada G, Delgado-Romero K, Burguete-Garcia AI, Cantu D, Garcia-Carranca A, and Madrid-Marina V. 2016. Cervical Microbiome and Cytokine Profile at Various Stages of Cervical Cancer: A Pilot Study. *PLoS One* 11:e0153274. 10.1371/journal.pone.0153274

Bratke K, Kuepper M, Bade B, Virchow JC, Jr., and Luttmann W. 2005. Differential expression of human granzymes A, B, and K in natural killer cells and during CD8+ T cell differentiation in peripheral blood. *Eur J Immunol* 35:2608-2616. 10.1002/eji.200526122

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, and Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7:335-336. 10.1038/nmeth.f.303

Champer M, Wong AM, Champer J, Brito IL, Messer PW, Hou JY, and Wright JD. 2017. The role of the vaginal microbiome in gynaecological cancer. *Bjog*. 10.1111/1471-0528.14631

De Angelis PM, Clausen OP, Schjolberg A, and Stokke T. 1999. Chromosomal gains and losses in primary colorectal carcinomas detected by CGH and their associations with tumour DNA ploidy, genotypes and phenotypes. *Br J Cancer* 80:526-535. 10.1038/sj.bjc.6690388

Denny L, Quinn M, and Sankaranarayanan R. 2006. Chapter 8: Screening for cervical cancer in developing countries. *Vaccine* 24 Suppl 3:S3/71-77. 10.1016/j.vaccine.2006.05.121

Di Paola M, Sani C, Clemente AM, Iossa A, Perissi E, Castronovo G, Tanturli M, Rivero D, Cozzolino F, Cavalieri D, Carozzi F, De Filippo C, and Torcia MG. 2017. Characterization of cervico-vaginal microbiota in women developing persistent high-risk Human Papillomavirus infection. *Sci Rep* 7:10200. 10.1038/s41598-017-09842-6

Dong X, Shulzhenko N, Lemaitre J, Greer RL, Peremyslova K, Quamruzzaman Q, Rahman M, Hasan OS, Joya SA, Golam M, Christiani DC, Morgun A, and Kile ML. 2017. Arsenic exposure and intestinal microbiota in children from Sirajdikhan, Bangladesh. *PLoS One* 12:e0188487. 10.1371/journal.pone.0188487

Dong X, Yambartsev A, Ramsey SA, Thomas LD, Shulzhenko N, and Morgun A. 2015. Reverse enGENEering of Regulatory Networks from Big Data: A Roadmap for Biologists. *Bioinform Biol Insights* 9:61-74. 10.4137/bbi.s12467

Eade CR, Diaz C, Wood MP, Anastos K, Patterson BK, Gupta P, Cole AL, and Cole AM. 2012. Identification and characterization of bacterial vaginosis-associated pathogens using a comprehensive cervical-vaginal epithelial coculture assay. *PLoS One* 7:e50106. 10.1371/journal.pone.0050106

Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460-2461. 10.1093/bioinformatics/btq461

Ege MJ, Mayer M, Normand AC, Genuneit J, Cookson WO, Braun-Fahrlander C, Heederik D, Piarroux R, and von Mutius E. 2011. Exposure to environmental microorganisms and childhood asthma. *N Engl J Med* 364:701-709. 10.1056/NEJMoa1007302

Farrell JJ, Zhang L, Zhou H, Chia D, Elashoff D, Akin D, Paster BJ, Joshipura K, and Wong DT. 2012. Variations of oral microbiota are associated with pancreatic diseases including pancreatic cancer. *Gut* 61:582-588. 10.1136/gutjnl-2011-300784

Foster ZS, Sharpton TJ, and Grunwald NJ. 2017. Metacoder: An R package for visualization and manipulation of community taxonomic diversity data. *PLoS Comput Biol* 13:e1005404. 10.1371/journal.pcbi.1005404

Fujimura KE, and Lynch SV. 2015. Microbiota in allergy and asthma and the emerging relationship with the gut microbiome. *Cell Host Microbe* 17:592-602. 10.1016/j.chom.2015.04.007

Gajer P, Brotman RM, Bai G, Sakamoto J, Schutte UM, Zhong X, Koenig SS, Fu L, Ma ZS, Zhou X, Abdo Z, Forney LJ, and Ravel J. 2012. Temporal dynamics of the human vaginal microbiota. *Sci Transl Med* 4:132ra152. 10.1126/scitranslmed.3003605

Geng J, Fan H, Tang X, Zhai H, and Zhang Z. 2013. Diversified pattern of the human colorectal cancer microbiome. *Gut Pathog* 5:2. 10.1186/1757-4749-5-2

Gillet E, Meys JF, Verstraelen H, Bosire C, De Sutter P, Temmerman M, and Broeck DV. 2011. Bacterial vaginosis is associated with uterine cervical human papillomavirus infection: a meta-analysis. *BMC Infect Dis* 11:10. 10.1186/1471-2334-11-10

Gillet E, Meys JF, Verstraelen H, Verhelst R, De Sutter P, Temmerman M, and Vanden Broeck D. 2012. Association between bacterial vaginosis and cervical intraepithelial neoplasia: systematic review and meta-analysis. *PLoS One* 7:e45201. 10.1371/journal.pone.0045201

Ginsburg O, Bray F, Coleman MP, Vanderpuye V, Eniu A, Kotha SR, Sarker M, Huong TT, Allemani C, Dvaladze A, Gralow J, Yeates K, Taylor C, Oomman N, Krishnan S, Sullivan R, Kombe D, Blas MM, Parham G, Kassami N, and Conteh L. 2017. The global burden of women's cancers: a grand challenge in global health. *Lancet* 389:847-860. 10.1016/s0140-6736(16)31392-7

Greer RL, Dong X, Moraes AC, Zielke RA, Fernandes GR, Peremyslova E, Vasquez-Perez S, Schoenborn AA, Gomes EP, Pereira AC, Ferreira SR, Yao M, Fuss IJ, Strober W, Sikora AE, Taylor GA, Gulati AS, Morgun A, and Shulzhenko N. 2016. Akkermansia muciniphila mediates negative effects of IFNgamma on glucose metabolism. *Nat Commun* 7:13329. 10.1038/ncomms13329

Herdman MT, Pett MR, Roberts I, Alazawi WO, Teschendorff AE, Zhang XY, Stanley MA, and Coleman N. 2006. Interferon-beta treatment of cervical keratinocytes naturally infected with human papillomavirus 16 episomes promotes rapid reduction in episome numbers and emergence of latent integrants. *Carcinogenesis* 27:2341-2353. 10.1093/carcin/bgl172

Hong KH, Hong SK, Cho SI, Ra E, Han KH, Kang SB, Kim EC, Park SS, and Seong MW. 2016. Analysis of the Vaginal Microbiome by Next-Generation Sequencing and Evaluation of its Performance as a Clinical Diagnostic Tool in Vaginitis. *Ann Lab Med* 36:441-449. 10.3343/alm.2016.36.5.441

Hosgood HD, 3rd, Sapkota AR, Rothman N, Rohan T, Hu W, Xu J, Vermeulen R, He X, White JR, Wu G, Wei F, Mongodin EF, and Lan Q. 2014. The potential role of lung microbiota in lung cancer attributed to household coal burning exposures. *Environ Mol Mutagen* 55:643-651. 10.1002/em.21878

Iida N, Dzutsev A, Stewart CA, Smith L, Bouladoux N, Weingarten RA, Molina DA, Salcedo R, Back T, Cramer S, Dai RM, Kiu H, Cardone M, Naik S, Patri AK, Wang E, Marincola FM, Frank KM, Belkaid Y, Trinchieri G, and Goldszmid RS. 2013. Commensal bacteria control cancer response to therapy by modulating the tumor microenvironment. *Science* 342:967-970. 10.1126/science.1240527

Johnson BJ, Costelloe EO, Fitzpatrick DR, Haanen JB, Schumacher TN, Brown LE, and Kelso A. 2003. Single-cell perforin and granzyme expression reveals the anatomical localization of effector CD8+ T cells in influenza virus-infected mice. *Proc Natl Acad Sci U S A* 100:2657-2662. 10.1073/pnas.0538056100

Kanao H, Enomoto T, Kimura T, Fujita M, Nakashima R, Ueda Y, Ueno Y, Miyatake T, Yoshizaki T, Buzard GS, Tanigami A, Yoshino K, and Murata Y. 2005. Overexpression of LAMP3/TSC403/DC-LAMP promotes metastasis in uterine cervical cancer. *Cancer Res* 65:8640-8645. 10.1158/0008-5472.can-04-4112

Kassam Z, Hundal R, Marshall JK, and Lee CH. 2012. Fecal transplant via retention enema for refractory or recurrent Clostridium difficile infection. *Arch Intern Med* 172:191-193. 10.1001/archinte.172.2.191

Knights D, Lassen KG, and Xavier RJ. 2013. Advances in inflammatory bowel disease pathogenesis: linking host genetics and the microbiome. *Gut* 62:1505-1510. 10.1136/gutjnl-2012-303954

Lamont RJ, and Hajishengallis G. 2015. Polymicrobial synergy and dysbiosis in inflammatory disease. *Trends Mol Med* 21:172-183. 10.1016/j.molmed.2014.11.004

Lando M, Holden M, Bergersen LC, Svendsrud DH, Stokke T, Sundfor K, Glad IK, Kristensen GB, and Lyng H. 2009. Gene dosage, expression, and ontology analysis identifies driver genes in the carcinogenesis and chemoradioresistance of cervical cancer. *PLoS Genet* 5:e1000719. 10.1371/journal.pgen.1000719

Lee JE, Lee S, Lee H, Song YM, Lee K, Han MJ, Sung J, and Ko G. 2013. Association of the vaginal microbiota with human papillomavirus infection in a Korean twin cohort. *PLoS One* 8:e63514. 10.1371/journal.pone.0063514

Libby EK, Pascal KE, Mordechai E, Adelson ME, and Trama JP. 2008. Atopobium vaginae triggers an innate immune response in an in vitro model of bacterial vaginosis. *Microbes Infect* 10:439-446. 10.1016/j.micinf.2008.01.004

Ling Z, Kong J, Liu F, Zhu H, Chen X, Wang Y, Li L, Nelson KE, Xia Y, and Xiang C. 2010. Molecular analysis of the diversity of vaginal microbiota associated with bacterial vaginosis. *BMC Genomics* 11:488. 10.1186/1471-2164-11-488

Liu MB, Xu SR, He Y, Deng GH, Sheng HF, Huang XM, Ouyang CY, and Zhou HW. 2013. Diverse vaginal microbiomes in reproductive-age women with vulvovaginal candidiasis. *PLoS One* 8:e79812. 10.1371/journal.pone.0079812

Lu X, and Kang Y. 2010. Hypoxia and hypoxia-inducible factors: master regulators of metastasis. *Clin Cancer Res* 16:5928-5935. 10.1158/1078-0432.ccr-10-1360

Marchesi JR, Dutilh BE, Hall N, Peters WH, Roelofs R, Boleij A, and Tjalsma H. 2011. Towards the human colorectal cancer microbiome. *PLoS One* 6:e20447. 10.1371/journal.pone.0020447

Mikamo H, Izumi K, Ito K, Watanabe K, Ueno K, and Tamaya T. 1993. Internal bacterial flora of solid uterine cervical cancer. *Kansenshogaku Zasshi* 67:1057-1061.

Mine KL, Shulzhenko N, Yambartsev A, Rochman M, Sanson GF, Lando M, Varma S, Skinner J, Volfovsky N, Deng T, Brenna SM, Carvalho CR, Ribalta JC, Bustin M, Matzinger P, Silva ID, Lyng H, Gerbase-DeLima M, and Morgun A. 2013. Gene network reconstruction reveals cell cycle and antiviral genes as major drivers of cervical cancer. *Nat Commun* 4:1806. 10.1038/ncomms2693

Morgun A, Dzutsev A, Dong X, Greer RL, Sexton DJ, Ravel J, Schuster M, Hsiao W, Matzinger P, and Shulzhenko N. 2015. Uncovering effects of antibiotics on the host and microbiota using transkingdom gene networks. *Gut* 64:1732-1743. 10.1136/gutjnl-2014-308820

Mujcic H, Rzymski T, Rouschop KM, Koritzinsky M, Milani M, Harris AL, and Wouters BG. 2009. Hypoxic activation of the unfolded protein response (UPR) induces expression of the metastasis-associated gene LAMP3. *Radiother Oncol* 92:450-459. 10.1016/j.radonc.2009.08.017

Muls A, Andreyev J, Lalondrelle S, Taylor A, Norton C, and Hart A. 2017. Systematic Review: The Impact of Cancer Treatment on the Gut and Vaginal Microbiome in Women With a Gynecological Malignancy. *Int J Gynecol Cancer* 27:1550-1559. 10.1097/igc.0000000000000999

Munger K, Phelps WC, Bubb V, Howley PM, and Schlegel R. 1989. The E6 and E7 genes of the human papillomavirus type 16 together are necessary and sufficient for transformation of primary human keratinocytes. *J Virol* 63:4417-4421.

Munoz N, Castellsague X, de Gonzalez AB, and Gissmann L. 2006. Chapter 1: HPV in the etiology of human cancer. *Vaccine* 24 Suppl 3:S3/1-10. 10.1016/j.vaccine.2006.05.115

534 Nagelkerke A, Bussink J, Mujcic H, Wouters BG, Lehmann S, Sweep FC, and Span PN. 2013. Hypoxia
535 stimulates migration of breast cancer cells via the PERK/ATF4/LAMP3-arm of the unfolded
536 protein response. *Breast Cancer Res* 15:R2. 10.1186/bcr3373
537 Nagelkerke A, Mujcic H, Bussink J, Wouters BG, van Laarhoven HW, Sweep FC, and Span PN. 2011.
538 Hypoxic regulation and prognostic value of LAMP3 expression in breast cancer. *Cancer*
539 117:3670-3681. 10.1002/cncr.25938
540 Oh HY, Kim BS, Seo SS, Kong JS, Lee JK, Park SY, Hong KM, Kim HK, and Kim MK. 2015. The association of
541 uterine cervical microbiota with an increased risk for cervical intraepithelial neoplasia in Korea.
542 *Clin Microbiol Infect* 21:674.e671-679. 10.1016/j.cmi.2015.02.026
543 Panzer U, Steinmetz OM, Reinking RR, Meyer TN, Fehr S, Schneider A, Zahner G, Wolf G, Helmchen U,
544 Schaerli P, Stahl RA, and Thaiss F. 2006. Compartment-specific expression and function of the
545 chemokine IP-10/CXCL10 in a model of renal endothelial microvascular injury. *J Am Soc Nephrol*
546 17:454-464. 10.1681/asn.2005040364
547 Pett MR, Herdman MT, Palmer RD, Yeo GS, Shivji MK, Stanley MA, and Coleman N. 2006. Selection of
548 cervical keratinocytes containing integrated HPV16 associates with episome loss and an
549 endogenous antiviral response. *Proc Natl Acad Sci U S A* 103:3822-3827.
550 10.1073/pnas.0600078103
551 Pilch H, Schlenger K, Steiner E, Brockerhoff P, Knapstein P, and Vaupel P. 2001. Hypoxia-stimulated
552 expression of angiogenic growth factors in cervical cancer cells and cervical cancer-derived
553 fibroblasts. *Int J Gynecol Cancer* 11:137-142.
554 Proost P, Vynckier AK, Mahieu F, Put W, Grillet B, Struyf S, Wuyts A, Opdenakker G, and Van Damme J.
555 2003. Microbial Toll-like receptor ligands differentially regulate CXCL10/IP-10 expression in
556 fibroblasts and mononuclear leukocytes in synergy with IFN-gamma and provide a mechanism
557 for enhanced synovial chemokine levels in septic arthritis. *Eur J Immunol* 33:3146-3153.
558 10.1002/eji.200324136
559 Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, McCulle SL, Karlebach S, Gorle R, Russell J, Tacket CO,
560 Brotman RM, Davis CC, Ault K, Peralta L, and Forney LJ. 2011. Vaginal microbiome of
561 reproductive-age women. *Proc Natl Acad Sci U S A* 108 Suppl 1:4680-4687.
562 10.1073/pnas.1002611107
563 Round JL, and Mazmanian SK. 2009. The gut microbiota shapes intestinal immune responses during
564 health and disease. *Nat Rev Immunol* 9:313-323. 10.1038/nri2515
565 Sheu BC, Lin RH, Lien HC, Ho HN, Hsu SM, and Huang SC. 2001. Predominant Th2/Tc2 polarity of tumor-
566 infiltrating lymphocytes in human cervical cancer. *J Immunol* 167:2972-2978.
567 shulzhenko N, Dong X, Vyshenska D, Greer R, Gurung M, Vasquez-Perez S, Peremyslova K, Sosnovtsev
568 SV, Yao M, Montgomery-Recht K, Strober W, Fuss IJ, and Morgun A. 2017. Tu1884 - Low
569 Intestinal IGA Production in CVID Facilitates Pathobiont-Mediated IFN Responses and
570 Enteropathy. *Gastroenterology* 152:S997-S998. https://doi.org/10.1016/S0016-5085(17)33381-4
571 Shulzhenko N, Lyng H, Sanson GF, and Morgun A. 2014. Menage a trois: an evolutionary interplay
572 between human papillomavirus, a tumor, and a woman. *Trends Microbiol* 22:345-353.
573 10.1016/j.tim.2014.02.009
574 Si J, You HJ, Yu J, Sung J, and Ko G. 2017. Prevotella as a Hub for Vaginal Microbiota under the Influence
575 of Host Genetics and Their Association with Obesity. *Cell Host Microbe* 21:97-105.
576 10.1016/j.chom.2016.11.010
577 Stokholm J, Blaser MJ, Thorsen J, Rasmussen MA, Waage J, Vinding RK, Schoos A-MM, Kunøe A, Fink NR,
578 Chawes BL, Bønnelykke K, Brejnrod AD, Mortensen MS, Al-Soud WA, Sørensen SJ, and Bisgaard
579 H. 2018. Maturation of the gut microbiome and risk of asthma in childhood. *Nat Commun* 9:141.
580 10.1038/s41467-017-02573-2

Sun R, Wang X, Zhu H, Mei H, Wang W, Zhang S, and Huang J. 2014. Prognostic value of LAMP3 and TP53 overexpression in benign and malignant gastrointestinal tissues. *Oncotarget* 5:12398-12409. 10.18632/oncotarget.2643

Thomas LD, Vyshenska D, Shulzhenko N, Yambartsev A, and Morgun A. 2016. Differentially correlated genes in co-expression networks control phenotype transitions. *F1000Res* 5:2740. 10.12688/f1000research.9708.1

Uemura N, Okamoto S, Yamamoto S, Matsumura N, Yamaguchi S, Yamakido M, Taniyama K, Sasaki N, and Schlemper RJ. 2001. Helicobacter pylori infection and the development of gastric cancer. *N Engl J Med* 345:784-789. 10.1056/NEJMoa001999

Vetizou M, Pitt JM, Daillere R, Lepage P, Waldschmitt N, Flament C, Rusakiewicz S, Routy B, Roberti MP, Duong CP, Poirier-Colame V, Roux A, Becharef S, Formenti S, Golden E, Cording S, Eberl G, Schlitzer A, Ginhoux F, Mani S, Yamazaki T, Jacquelot N, Enot DP, Berard M, Nigou J, Opolon P, Eggermont A, Woerther PL, Chachaty E, Chaput N, Robert C, Mateus C, Kroemer G, Raoult D, Boneca IG, Carbonnel F, Chamaillard M, and Zitvogel L. 2015. Anticancer immunotherapy by CTLA-4 blockade relies on the gut microbiota. *Science* 350:1079-1084. 10.1126/science.aad1329

Vyshenska D, Lam KC, Shulzhenko N, and Morgun A. 2017. Interplay between viruses and bacterial microbiota in cancer development. *Semin Immunol* 32:14-24. 10.1016/j.smim.2017.05.003

Walboomers JM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV, Snijders PJ, Peto J, Meijer CJ, and Munoz N. 1999. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol* 189:12-19. 10.1002/(sici)1096-9896(199909)189:1<12::aid-path431>3.0.co;2-f

Figure 1(on next page)

Community composition in cervical cancer and healthy adjacent sites

(A) PCoA of unweighted unifrac comparing microbiota of samples from cervical cancer (n=52), healthy cervix (n=17), vagina (n=76), stool (n=28), and skin (n=55). (B) Bar chart of mean relative abundance of genera in cervical cancer biopsies, cytobrush from healthy ectocervical mucosa, and swab from healthy posterior vaginal fornix. Genera are arranged in descending order of mean relative abundance in cervical cancer samples from bottom to top of the bar chart. Genera with mean relative abundance <0.5% across the three sites are grouped into “Other” found at bottom of bar chart. (C) Phylogenetic tree indicating the relationship and mean relative abundance (blue color intensity) of various genera in cervical cancer samples. The size of node and its label indicate the number of OTUs belonging to that taxonomy.

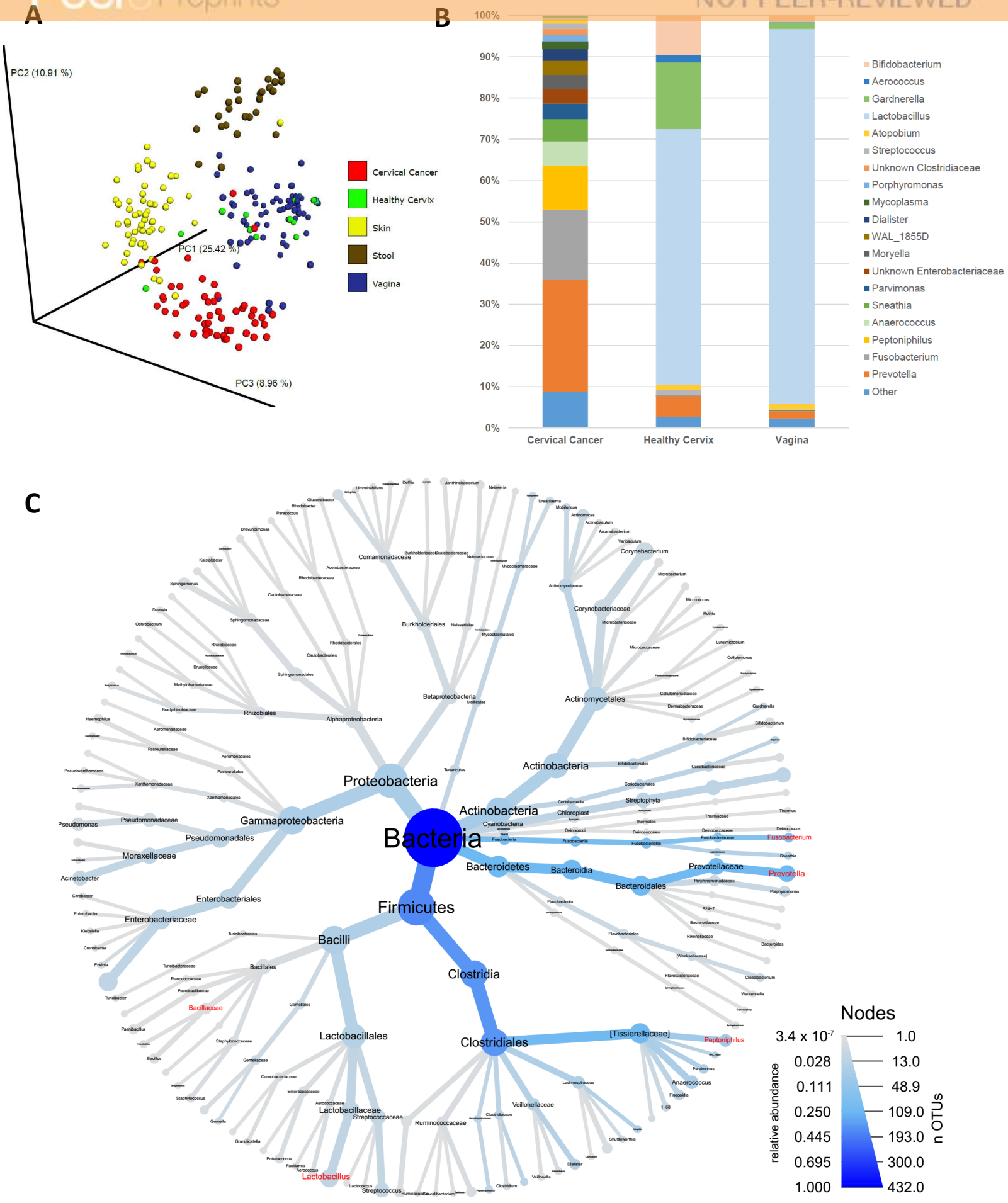
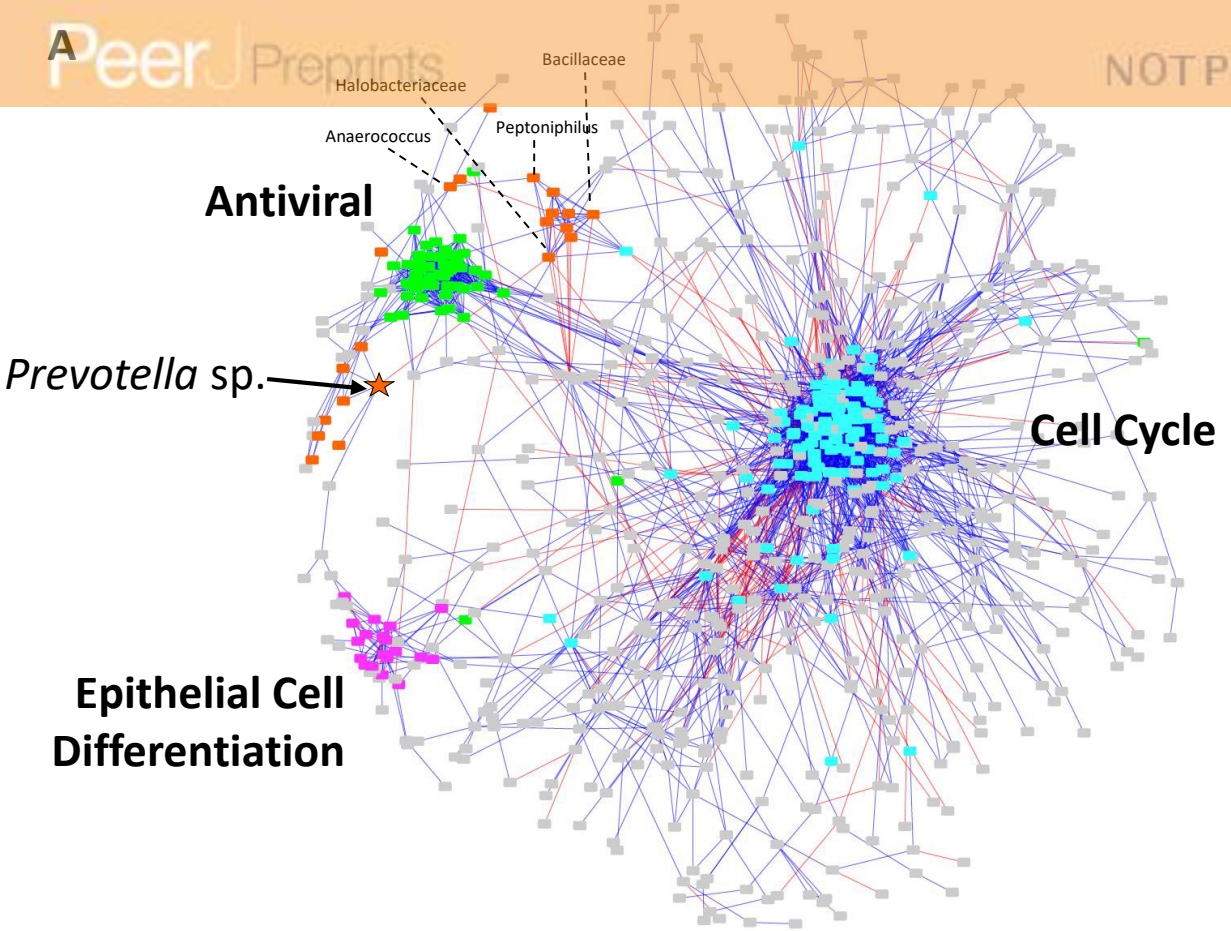
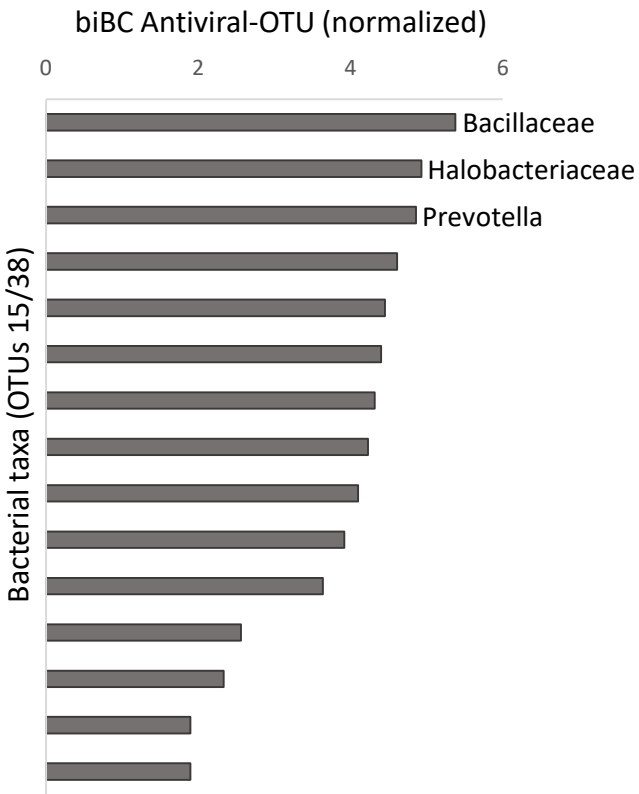


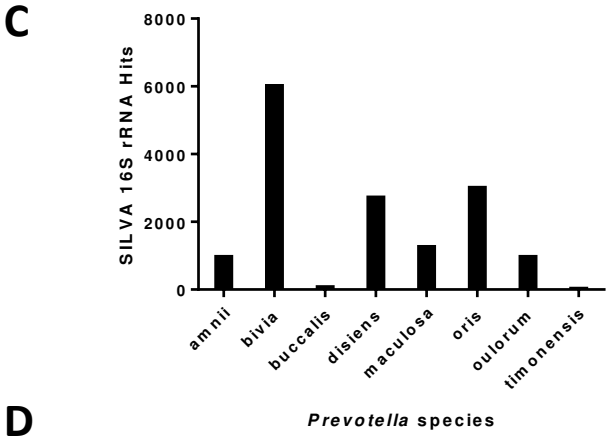
Figure 2(on next page)

Transkingdom microbe-gene regulatory network

(A) Transkingdom correlation network ($p < 0.001$; FDR < 0.1) between microbial network (21 OTUs, 50 edges) and previously described (Mine K, et al. 2013) tumor differentially expressed genes (698 DEGs, 3066 edges) connected by 19 edges. Edge-weighted spring layout was performed in Cytoscape. Nodes represent: orange – bacteria; green - antiviral genes; purple - epithelial cell differentiation genes; blue - cell cycle genes; and gray - genes not assigned to specific subnetwork or function. Lines indicate: blue - presence of correlation between nodes; red – lack of correlation between nodes. Orange star indicates Prevotella OTU with high BiBC whereas dashed lines connecting the node and its name designate the top 5 BiBC scored bacteria OTU. (B) Top BiBC OTUs (15/38) calculated between microbial subnetwork and antiviral subnetwork. (C) Top Prevotella species in SILVA 16S rRNA database matched to representative sequences assigned to OTU_97.1949 (match length >200 bp, mismatch =40 bp). (D) Prevotella mean abundance in cervical cancer (CC) compared with previous 16S studies for healthy adjacent sites: HPV negative cervix (HPV- cervix) and healthy vaginal microbiome (Healthy vagina). (PMID numbers of the source article specified for each column).







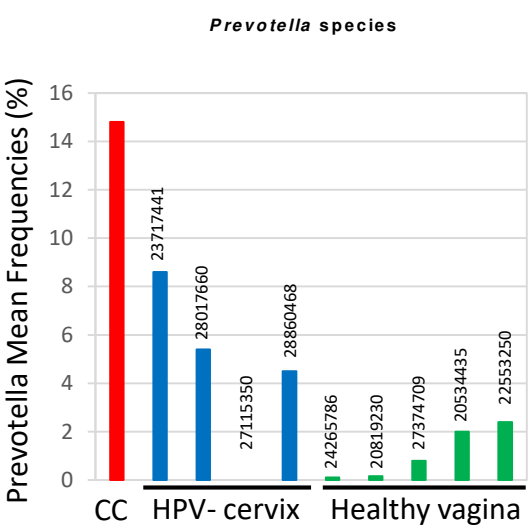
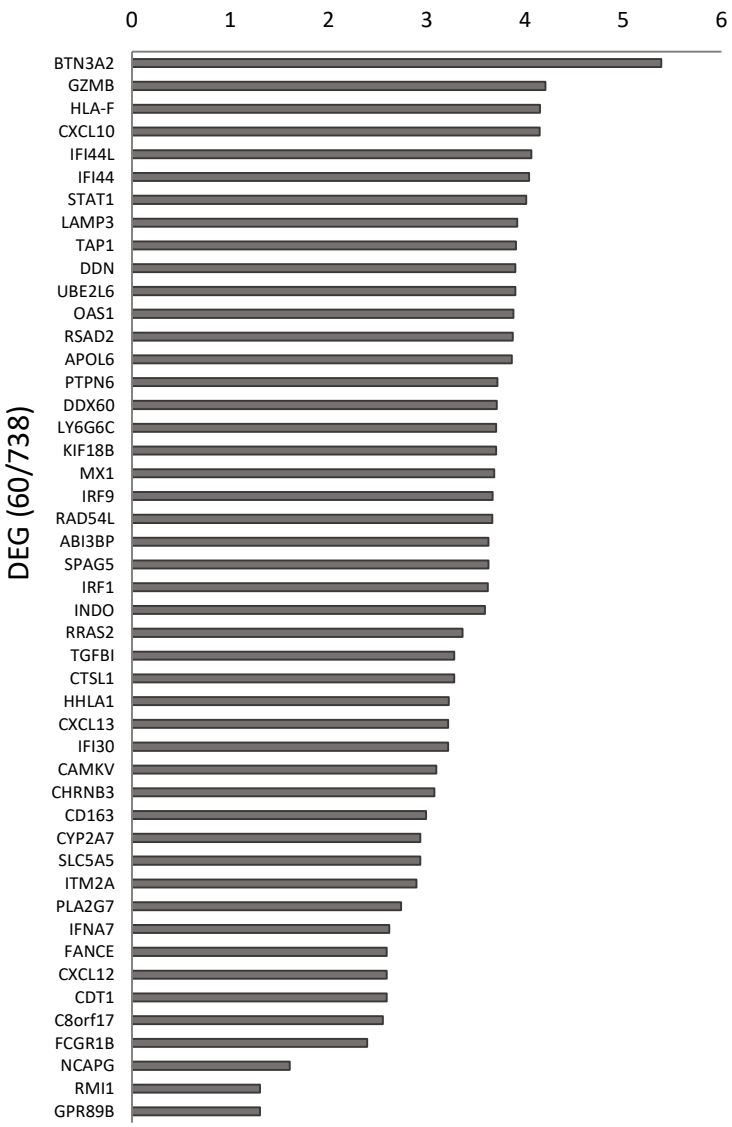


Figure 3(on next page)

Host gene expression regulated by Prevotella bivia

(A) Top DEGs (60/738) ranked by BiBC centrality calculated between bacteria and antiviral genes in transkingdom network (normalized BiBC = log2((BiBC*10^6)+1)). (B) RT-qPCR for cervical cancer top BiBC genes (LAMP3, STAT1, and TAP1), for which gene expression was upregulated in HeLa cells by P. bivia but not L. crispatus co-culture compared to negative treatment (PBS). mRNA levels were normalized to 18S rRNA gene expression. (*p-value < 0.05, one-tailed Wilcoxon matched-pairs signed rank test).

A



B