Urinary Eubacterium sp. CAG:581 promotes non-muscle-invasive bladder cancer (NMIBC) development through ECM1/MMP9 pathway

Yu-hang Zhang  
Peking University First Hospital

Wenyu Wang  
Capital Medical University

Gang Zhao  
Capital Medical University

Yimin Cui (✉ cui.pharm@pkufh.com)  
Peking University First Hospital

Research Article

Keywords: Non-muscle-invasive bladder cancer (NMIBC), Eubacterium sp. CAG:581, Extracellular matrix protein 1 (ECM1), Matrix metalloproteinase 9 (MMP9), Noninvasive urinary diagnostics.

Posted Date: November 2nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2207998/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

**Background:** Increasing evidence points to the urinary microbiota as a possible key susceptibility factor for early-stage bladder cancer (BCa) progression. However, its underlying mechanism interpretation is often insufficient, given that various environmental conditions have affected the composition of urinary microbiota. Herein, we sought to rule out confounding factors and clarify how urinary *Eubacterium sp. CAG:581* promoted non-muscle-invasive bladder cancer (NMIBC) development.

**Methods:** Differentially abundant urinary microbiota of 51 NMIBC patients and 47 healthy controls as the Cohort 1 were firstly determined by metagenomics analysis. Then we modeled the coculture of NMIBC organoids with candidate urinary *Eubacterium sp. CAG:581* in anaerobic condition and explored differentially expressed genes of NMIBC organoids by RNA-Seq. Furthermore, we dissected the mechanisms involved into *Eubacterium sp. CAG:581*-induced extracellular matrix protein 1 (ECM1) and matrix metalloproteinase 9 (MMP9) upregulation. Finally, we used multivariate Cox modeling to investigate the clinical relevance of urinary *Eubacterium sp. CAG:581* 16S ribosomal RNA (16SrRNA) levels with the prognosis of 406 NMIBC patients as the Cohort 2.

**Results:** *Eubacterium sp. CAG:581* infection accelerated the proliferation of NMIBC organoids (*P* < 0.01); ECM1 and MMP9 were the most upregulated gene induced by increased colony forming units (CFU) gradient of *Eubacterium sp. CAG:581* infection, via phosphorylating ERK1/2 in NMIBC organoids of the Cohort 1. Excluding the favorable impact of potential contributing factors, ROC curve of the Cohort 2 manifested its 3-year AUC value as 0.79 and the cut-off point of *Eubacterium sp. CAG:581* 16SrRNA as 10.3 (delta CT value).

**Conclusion:** Our evidence suggests that urinary *Eubacterium sp. CAG:581* promoted NMIBC progression through ECM1/MMP9 pathway, which may serve as the promising noninvasive diagnostic biomarker for NMIBC.

Introduction

Bladder cancer (BCa) is the second most common malignancy of the urinary tract, accounting for 4.6% of all new cancer diagnoses. It causes 400,000 new diagnoses and 160,000 deaths worldwide each year, which is age-associated and individuals aged 75–84 years accounting for the largest percentage at 30% of new cases [1]. BCa can be divided into muscle-infiltrating bladder cancer (MIBC) and non-muscle-infiltrating bladder cancer (NMIBC) according to whether it invades into the muscular layer of the bladder wall [2]. NMIBC comprises approximately 75–80% of all newly diagnosed BCa [3], which is the malignant papillary tumor of the bladder confined to the mucosal layer (TIS/CIS accounting for 10%, Ta for 70%) and lamina propria (T1 for 20%) of the bladder with no muscle infiltration [4]. After transurethral resection of bladder tumor (TURBT), the 5-year risk of recurrence for high-risk NMIBC can be as high as 80% [5]. However, if NMIBC is screened at an early stage, patients with high-risk NMIBC have much lower rate of tumor recurrence after immediate postoperative perfusion which can be cured with simple surgery [6];
improving the sensitivity and specificity of NMIBC screening is thus critical to the prognosis of NMIBC patients. Currently, the preferred method for screening patients with NMIBC is cystoscopy. Still, it is invasive and often screens to confirm for NMIBC only after the solid tumor [7]. Exploring more convenient and sensitive early-stage screening modalities is essential for NMIBC treatment.

The development of NMIBC is a complex, multifactorial and multistep pathological process, in which both intrinsic genetic factors and extrinsic environmental factors have contributed to the process [8]. Regional, ethnic and gender differences could affect the development of bladder cancer, with a high incidence at 50–70 years old. According to National Central Cancer Registry of China (NCCRC), the ratio of male to female mortality is 2.97:1 [9,10]. Smoking is the most definite and major external risk factor for NMIBC. About 50% of NMIBC patients have the history of smoking, which increases the risk of bladder cancer by 2–5 times depending on the intensity and duration of smoking [9,10]. Recent studies have shown the composition and content of the urinary flora can also impact the development and prognosis of NMIBC. Previous studies have indicated significant differences in urinary flora characteristics between BCa patients and healthy participants [11,12]. Based on their bioinformatics analysis, there may be the link between urinary flora and bladder cancer, and some different genera might be the novel biomarkers for BCa [12]. However, these findings were premature and the causal relationship between urinary flora and early-stage BCa is still not clear. It needs further validation through prospective cohort studies and further mechanistic explanation.

ECM1 is a glycoprotein expressed in epithelial organs and secreted into the extracellular matrix, which can promote the proliferation of vascular endothelial cells and angiogenesis [13]. It has been found that the expression of ECM1 may associate with BCa progression, which was confirmed with potential anti-tumor effects by siRNA gene silencing experiments [14,15]. Overexpression of ECM1 has now been incorporated into scoring models to suggest poor clinical prognosis and metastatic potential in several types of cancer patients [16]. Matrix metalloproteinases (MMPs) are critical regulators of the tumor microenvironment. Their protein hydrolytic activity initiates the degradation of extracellular matrix components so that tumors are no longer restricted by the intact basement membrane, invading surrounding tissues and ultimately leading to tumor malignancy [17]. MMPs have been shown to regulate not only epithelial-mesenchymal transition through various non-catalytic structural domains but also molecular signaling for cell growth, inflammation and angiogenesis in the non-protein hydrolytic manner [18]. Previous studies have demonstrated that ECM1 inhibited MMP9 activity through high-affinity protein/protein interactions [18,19]. In certain cancer types, the significant increase of MMP9 transcription can be observed in parallel with the upregulation of ECM1, suggesting that ECM1-MMP9 axis is of great significance for tumorigenesis and progression [18,20].

In this study, we have designed two clinical cohorts for prediction and validation, respectively. Firstly, we compared the urinary microbiota of 51 NMIBC patients and 47 healthy controls as the Cohort 1 by metagenomic analysis. Then we co-cultured NMIBC organoids with candidate urinary bacterium *Eubacterium sp. CAG:581* and found its underlying mechanisms that *Eubacterium sp. CAG:581* promoted
the growth of NMIBC organoids by upregulating the expression of ECM1-MMP9 axis, suggesting that *Eubacterium sp. CAG:581* may serve as the potential diagnostic predictor of NMIBC. To further verify whether *Eubacterium sp. CAG:581* has the similar predictive value for NMIBC in the different large-scale populations, we designed the Cohort 2 with 406 NMIBC patients and 398 healthy controls to determine the AUC and cut-off value of urinary *Eubacterium sp. CAG:581* 16SrRNA detection for potentially clinical application.

**Materials And Methods**

**Clinical Cohorts Design and Specimens**

We studied 2 cohorts of NMIBC patients from Peking University First Hospital between 2019–2022. Patients were selected based on the availability of a pre-treatment urine sample and presence of the well-modeling NMIBC organoids. Formalin-fixed paraffin-embedded tissues (FFPE) of 51 NMIBC patients and 47 healthy controls were collected as the Cohort 1. We performed metagenome sequencing studies in Cohort 1 to define which bacterium is predominant (and/or different) in the urine of NMIBC patients compared to those of healthy controls. We used the urine of 406 NMIBC patients and 398 healthy controls of the Cohort 2 as a validation dataset to determine which levels of urinary *Eubacterium sp. CAG:581* generated from Cohort 1 could be applied and validated. Thus, Cohorts 1 and 2 serve as the models for our research discovery and validation, respectively. All the patients' information could be found in Tables S1 and S2. The Ethics Committees in Peking University First Hospital approved the study protocols (Ethical number: 2019 - 138). Written informed consents were obtained from all participants in this study. The whole process of this research was carried out in accordance with the provisions of the Helsinki Declaration of 1975.

**Metagenome Sequencing**

Briefly, all urine samples were normalized to 0.2 ng/µl DNA material per library using a Quant- iT picogreen assay system (Life Technologies) on an AF2200 plate reader (Eppendorf), then fragmented and tagged via tagmentation. Amplification was performed by Veriti 96 well PCR (Applied Biosystems) followed by AMPure XP bead cleanup (Beckman Coulter). Fragment size was measured using Labchip GX Touch high-sensitivity. For cluster generation and next generation sequencing, samples were normalized to 1 nM, pooled and diluted to 8 pM. The paired-end cluster kit V4 was used and cluster generation was performed on an Illumina cBot, with pooled samples in all 8 lanes. Sequencing was performed on an Illumina HiSeq 2500 using SBS kit V4 chemistry. Median cluster densities were 908.5 for Nextera XT.

**NMIBC organoids cocultured with urinary bacterium in the 2-chamber culture system**

Coated the bottom of normoxic apical chamber with 5% (vol/vol) β-Mercaptoethanol (BME) in DMEM for 30 min at 37°C, all inserts grant apical and basal access to the cell layer, whereas standard cell culture plates allow for the greatest scalability. After the coating process, remove the supernatant and allow the
BME coating to dry in the incubator for another 15–30 min. Harvest organoids from the culture in 1 ml cold DMEM, and transfer to a 15 ml Falcon tube. Top up to 10 mL with cold DMEM. Spin for 5 min at 300g, and discard supernatant. Resuspend in expansion medium with 10 µM Rho Knise Inhibitor at a density of 1000 organoids/ml. Seed 10,000 organoids in 100 µl per well of a tissue-culture-treated normoxic apical chamber of 6.5 mm diameter. Allow cells to settle for 2 days. After 3 days or when the layer is close to confluence, equilibrate the medium in the basal chamber with anaerobic gas and subsequently sealed by inserting a plug made of butyl rubber (AsONE international, Santa Clara, CA). The oxygen concentration of the apical chamber was measured by a fiberoptic oxygen meter (PreSens. Regensburg, Germany). Then add urinary bacterium to the basal side in a medium suitable for exposure. For MOI calculations, a representative well can be harvested and cells counted as described above. Perform read-outs in analogy to 3D coculture.

**RNA-seq and Data Processing**

Following the manufacturer’s instructions, barcoded sequencing libraries were generated using Chromium Cell Reagent Kits and sequenced across 8 lanes on HiSeq 4000 platform targeting 100,000 reads per cell for the SMC dataset. RNA-seq data were then generated using the NextSeq 500 and NovaSeq 6000 systems. Sequencing data were aligned to the human reference genome (GRCh38) and processed using the Cell Ranger 2.1.0 pipeline (10x Genomics). The raw gene expression matrix from the Cell Ranger pipeline was filtered, normalized using the Seurat R package and selected according to the following criteria: cells with > 1,000 unique molecular identifier (UMI) counts; >200 genes and < 6,000 genes; and < 20% of mitochondrial gene expression in UMI counts. From the filtered cells, the gene expression matrices were normalized to the total UMI counts per cell and transformed into the natural log scale. For batch correction, we used multiple CCA implemented in Seurat v.2.3.4 and variable genes were selected as the top 1,000 highly variable genes expressed by more than 0.1% of cells in each sample. The Metagene Bicor Plot function was used to select the number of standard correlations vectors for choosing the inflection point. Then we aligned the CCA subspaces by selecting the variably per sample origin. The major cell types were annotated by comparing the canonical marker genes and the differentially expressed genes (DEGs).

**Western Blotting**

All NMIBC organoids were washed twice with cold PBS and harvested for western blot. Then they were lysed in cold radioimmunoprecipitation assay (RIPA) buffer (50 mm Tris-HCl, pH 7.4, 150 mm NaCl, 1% Triton X-100, 1% sodium deoxycholate, 0.1% SDS) containing protease inhibitors for 30 min. The lysates were then centrifuged and the supernatants were collected. Approximately 40 µg of total protein was denatured and separated by 10% SDS-PAGE, and then transferred to a polyvinylidene fluoride membrane. The membranes were blocked with 5% non-fat milk in Tris-buffered saline containing 0.1% Tween-20 (TBST) for 2 h at room temperature. The membranes were then incubated with primary antibodies overnight at 4°C. The primary antibodies were listed in the Table S3, Supporting Information. β-actin was used as a loading control. The membranes were washed five times with TBST and then incubated with
horseradish peroxidase-conjugated secondary antibodies for 1 h at room temperature. The signal was visualized using an Enhanced Chemiluminescence Detection Kit (Pierce Biotechnology, USA).

Quantitative Real-Time PCR.

Total RNA was isolated from urinary microbiome with TRIzol (Invitrogen, USA). HiScript II qRT Super Mix (Vazyme, China) was used to synthesize the first strand of cDNA. Quantitative real-time PCR was performed using the Hieff qPCR SYBR Green PCR Master Mix (Yeasen, China) with gene-specific primers. The primers of *Eubacterium sp. CAG:581* 16SrRNA, *ptpn6*, *ikzf3*, *ecm1*, *mmp9* and *gapdh* were listed in the Table S4, Supporting Information. All target gene transcripts were normalized to *GAPDH*, and the relative fold change in expression calculated using the $2^{-\Delta\Delta CT}$ method.

**Adenoviral shRNA Infection of NMIBC organoids**

The shRNA adenoviruses Ad-GFP-U6-m-ECM1-shRNA (shECM1, CCTGATATTTCCTCGGGTCTT) was purchased from Vector Biolabs. A non-specific scrambled shRNA adenovirus Ad-GFP-U6-scrambled-shRNA (Vector Biolabs, 1122N, CCGGCAACAAGATGAAGAGCACCAACTCGAGTTGGTGCT) expressing green fluorescent protein (GFP) alone was used as a control. For adenoviral infection, after centrifugation of dissociated NMIBC organoids in 3.5% BSA gradient, the resulting pellet was resuspended in serum-free DMEM medium containing Ad-GFP-U6-m-ECM1-shRNA or Ad-GFP-U6-scrambled-shRNA at a concentration of $2 \times 10^8$ pfu/ml. The mixture was then plated on normoxic basal chamber coated with poly-L-lysine/laminin and incubated at 37°C in 5% CO$_2$. Two hours after incubation, the medium was replaced with fresh supplemented DMEM medium. Three days after infection, adenovirus-infected NMIBC organoids were used for protein extraction to determine the expression of ECM1.

**Statistical Analysis**

Statistical analyses were carried out using the program R (www.r-project.org). Data from at least three independent experiments performed in triplicates are presented as the means ± SE. Error bars in the scatterplots and the bar graphs represent SE. Data were examined to determine whether they were normally distributed with the One-Sample Kolmogorov-Smirnov test. If the data were normally distributed, comparisons of measurement data between two groups were performed using independent sample t test and the comparisons among three or more groups were firstly performed by one-way ANOVA test. When the data represented skewed distribution, comparisons were performed by nonparametric test. Measurement data between two groups were performed using nonparametric Mann-Whitney test. To generate the ROC curves, patients were classified as recurrence time either longer or shorter than the median recurrence free survival, excluding patients who were alive for durations less than the median recurrence free survival at last follow-up. Single-sample gene set enrichment analysis (ssGSEA) was used to assess gene set activation scores in gene expression profiling data. ssGSEA calculates a sample level gene set score by comparing the distribution of gene expression ranks inside and outside the gene set. The ssGSEA score was calculated by Gene Set Variation Analysis (GSVA) R package. GraphPad Prism v.8.01 were also used for statistical analysis (GraphPad Software, La Jolla, CA, USA). Univariate and
multivariate Cox regression models were used to evaluate survival. The hazard ratio (HR) and 95 percent confidence interval (CI) were adopted to find genes correlated with overall survival. $P < 0.05$ indicated a difference with statistical significance unless otherwise stated.

**Results**

**Eubacterium sp. CAG:581 is clinically associated with NMIBC occurrence.** To examine the potential relationship between the urinary microbiota alteration and NMIBC development, we firstly compared the long-read metagenomics sequencing data of 51 NMIBC patients and 47 healthy controls. LEfSe algorithm was used to define their potential differential bacterium patterns. We found *Eubacterium sp. CAG:581, Bacteroides sp. 4_3_47FAA* and *Flavobacteriales* were enriched in NMIBC group as compared to healthy control group (Fig. 1A). *Flavobacteriales* belong to the taxonomy level of class, hence we further studied *Eubacterium sp. CAG:581* and *Bacteroides sp. 4_3_47FAA* for quantitative validation. The gut microbiota diversity of the above two groups was analyzed by the Chao1 index and Rank Abundance Curves for α diversity and Bray-Curtis distance and Binary-Jaccard distance for β diversity. It showed that α diversity of the urinary bacterial community of NMIBC group was lower than those of healthy control groups (Fig. 1B and 1C). β diversity of the urinary microbiota evaluated by ANOSIM was significantly different between the two groups based on the Bray-Curtis distances ($R = 0.144, P = 0.025$, Fig. 1D) and Binary-Jaccard distance ($R^2 = 0.190, P = 0.003$, Fig. 1E). Then the difference in bacterial community composition was analyzed. At the phylum, class and genus level, the predominantly abundant phyla of the NMIBC group were *Eubacterium sp. CAG:581, Bacteroides sp. 4_3_47FAA* and *Flavobacteriales*, while the enriched phyla of the healthy control group were listed as the cyan columns in Fig. 1F.

The functional composition of the urinary microbiome was compared between NMIBC patients and controls by COG and KEGG pathway analyses. Although the functional compositions of the two groups were highly similar, COG analyses indicated the clustering of metabolic modules were increased in NMIBC group including the metabolism of xenobiotics by cytochrome P450 ($E-value = 31.357, P < 0.001$), purine ($E-value = 30.835, P = 0.022$), flavone and flavonol biosynthesis ($E-value = 29.663, P = 0.029$) (Fig. 1G). KEGG pathway analysis showed that these differential metabolisms were majorly concentrated on oxidative phosphorylation, cardiac muscle contraction and carbon metabolism (Fig. 1H). These findings are consistent with the hypermetabolic activity and aggressive characteristics of NMIBC.

**Coculture of Eubacterium sp. CAG:581 promoted the growth of NMIBC organoids.** Hypoxia is essential for the growth of obligate anaerobes, but prohibitive for the maintenance of viable tumor organoids [21]. To surmount this tradeoff in oxygen demands, we developed NMIBC organoids and anaerobes coculture system, consisting of a normoxic apical chamber and a hypoxic basal chamber. The plug inserts tightly, physically blocking the influx of external oxygen, which allows maintenance of hypoxia in the basal chamber, while oxygen freely perfuses the apical chamber (Fig. 2A). NMIBC organoids model derived from NMIBC patients were established on the array chip using 50% Matrigel. Each chip had a reservoir layer on the top, a 3D implanting hole in the middle, and anaerobes culture slide underneath. The nested design allowed convenient medium exchange without disruption of the 3D organoids (Fig. 2A). It showed
that organoids size on the chip increased over time and achieved a diameter of 30 µm within 21 days. *Eubacterium sp. CAG:581* and *Bacteroides sp. 4_3_47FAA* coculture groups respectively proliferated faster by 48.5% and 37.8% when compared with control NMIBC organoids on Day 21 (Fig. 2B). On day 21, bright-field images of *Eubacterium sp. CAG:581* coculture group has presented heterogeneous morphologies with predominantly thin-walled cystic structure and solid dense structure (Fig. 2C). To gain insights into the underlying mechanisms, we randomly performed RNA-Seq to compare *Eubacterium sp. CAG:581*-cocultured NMIBC organoids (N14-N26) and control NMIBC organoids (N1-N13). It demonstrated that ECM1 and MMP9 were most significantly upregulated in *Eubacterium sp. CAG:581*-cocultured NMIBC organoids, while PTPN6 and IKZF3 were mostly downregulated in the heatmap and ssGSEA analysis (Fig. 2D and 2E), which should be further validated by molecular biological experiments.

**Eubacterium sp. CAG:581 activated ECM1/ERK1/2 phosphorylation/MMP9 of NMIBC organoids.** Based on the above RNA-Seq predictive data, we have determined the transcriptional levels of *ptpn6, ikzf3, ecm1* and *mmp9* in NMIBC organoids cocultured with *Eubacterium sp. CAG:581* and control NMIBC organoids by RT-PCR. It showed that increased colony forming units (CFU) gradient (5×10^5, 10^6, 5×10^6, 10^7, 5×10^7) of *Eubacterium sp. CAG:581* have significantly increased mRNA levels of *ecm1* and *mmp9* (Fig. 3A). Then we compared their protein expressions by the assay of western blotting, which also demonstrated increased gradient of *Eubacterium sp. CAG:581* could upregulate ECM1 and MMP9 of NMIBC organoids as compared with the control organoids. It was reported that ECM1 induces tumor growth by promoting angiogenesis or enhancing the EGF signaling in the breast cancer [22]. We thus detected their expression of ERK1/2 phosphorylation and AKT phosphorylation, which manifested the upregulated ERK1/2 phosphorylation under the exposure of increased CFU gradient of *Eubacterium sp. CAG:581* (Fig. 3B). Then we used shRNA of ECM1 to model BCa organoid[^shECM1] and control BCa organoid[^vector]. When exposed to 10^7 cfu *Eubacterium sp. CAG:581*, BCa organoid[^shECM1] was determined with decreased expression of ERK1/2 phosphorylation and MMP9. Meanwhile, BCa organoid[^shECM1] and BCa organoid[^vector] were treated with 10 µM Ravoxertinib (Rav) or Ulixertinib (Uli), both of which are inhibitors of ERK1/2 [23, 24]. It showed that both Rav and Uli could impair the expression of MMP9 (Fig. 3C). Lastly, we detected the proliferative size of Rav-treated or Uli-treated BCa organoid[^shECM1] and control BCa organoid[^vector] under the exposure of 10^7 cfu *Eubacterium sp. CAG:581* for 21 days. We found ERK1/2 inhibition or shECM1 could effectively prohibit the growth of BCa organoids (Fig. 3D), which suggested *Eubacterium sp. CAG:581* progressed NMIBC organoids by activating ECM1/ERK1/2 phosphorylation/MMP9.

**Eubacterium sp. CAG:581 was endowed with the diagnostic predictor for NMIBC.** Increasing evidence points to the urinary microbiota as a possible key susceptibility factor for early-stage bladder cancer (BCa) progression [11, 25]. However, its conclusive interpretation is often insufficient, given that various environmental conditions could significantly alter the regulation of urinary microbiota. Herein we have evaluated the relationship between the amount of *Eubacterium sp. CAG:581*, ECM1, MMP9 and other baseline features in 51 NMIBC patients and 47 healthy controls (Fig. 4A). The amount of *Eubacterium sp. CAG:581* was positively associated with the occurrence of NMIBC (HR: 4.21, 95% CI: 2.54–5.33, log-rank
The expression of ECM1 (HR: 1.87, 95% CI: 1.61–2.83, log-rank $P = 0.005$) and MMP9 (HR: 1.66, 95% CI: 1.49–1.88, log-rank $P = 0.013$) were substantially higher in NMIBC group than healthy control group. In contrast, age, sex, smoking status and alcohol consumption of this study were not statistically significant for Pearson correlation analysis (Fig. 4A). In addition, levels of *Eubacterium sp. CAG:581*, ECM1 and MMP9 with survival probability of NMIBC were also analyzed by Kaplan Meier Curve. The findings revealed that higher amount of *Eubacterium sp. CAG:581* was strongly associated with decreased survival probability of NMIBC (Fig. 4B), and the survival time was also favorably linked with the expression of ECM1 and MMP9 (Figs. 4C and 4D). These above data implied the detection of *Eubacterium sp. CAG:581* is effective at predicting the occurrence of NMIBC.

**Identification of NMIBC occurrence-associated Eubacterium sp. CAG:581 in the larger population.** To further validate whether *Eubacterium sp. CAG:581* levels had a similar prediction value in a different and larger NMIBC patients’ population, we analyzed an additional cohort with 406 NMIBC patients as the Cohort 2. Using PCA cluster analysis with 398 normal healthy controls, we identified the ability of *Eubacterium sp. CAG:581* to distinguish NMIBC urine from healthy control urine (Fig. 5A). LASSO regression analysis of *Eubacterium sp. CAG:581*-induced prognostic DEGs by RNA-Seq was also modeled to verify its fine cooperativity (Fig. 5B) and stable partial likelihood deviation from the minimum value (Fig. 5C). Based on the abundance of *Eubacterium sp. CAG:581* levels, we have also divided them into the group of high risk and low risk, which demonstrated that the prognosis model is feasible in their survival predictions (Figs. 5D and 5E). The prognostic value of increased urine ECM1 and MMP9 were also stable and accurate in larger NMIBC cohort. Decreased PTPN6 and IKZF3 were determined in the Cohort 2 as well (Fig. 5F). The total survival rate was significantly worse ($P < 0.001$) in NMIBC patients with high amounts of *Eubacterium sp. CAG:581*, according to the Kaplan-Meier survival curve (Fig. 5G). Receiver operating characteristic (ROC) curve analysis was conducted to predict the potential CRC recurrence using 1-, 2-, and 3-year NMIBC occurrence and 3-year occurrence was demonstrated with the highest AUC value of 0.79. Youden Index was used to determine the optimal cut-off point of 3-year NMIBC occurrence as 10.3 (delta CT value) that provided the best balance between the sensitivity and the specificity of *Eubacterium sp. CAG:581* to predict NMIBC occurrence (Fig. 5H). Therefore, the data in Cohort 2 not only confirm our observation in Cohort 1 but also define the potential value of the *Eubacterium sp. CAG:581* signature in predicting NMIBC occurrence.

**Discussion**

Several studies have demonstrated that 70% NMIBC patients have the disease progression for recurrence within 5 years, and 10%-20% can progress to advanced muscle-invasive disease or distant metastatic disease [26]. In case of an advanced stage, high-risk NMIBC patients are thus treated with electrodesiccation of the urinary bladder tumor and postoperative instillation of Bacillus Calmette-Guérin (BCG) [27]. Unfortunately, instillation of BCG failed in approximately half of these patients, resulting in approximately 30% persistence or early relapse for high-risk NMIBC. 50% of them will undergo radical cystectomy [28]. Despite the advent of minimally invasive surgery and robotic technology, 90-day mortality
and morbidity rates for patients undergoing cystectomy still remained as high as 3–6% and 28–64%, respectively [29]. For the past 30 years, bladder perfusion for NMIBC patients is dominated by chemotherapeutic agents in China and BCG immunotherapy in the United States. However, bladder chemotherapy instillation is an invasive drug delivery method which has led to 30% of urethral irritation such as dysuria, frequency, urgency, hematuria and cystitis. Moreover, the treatment course usually took up more than one year, resulting in poor patient compliance [30]. The standard treatment regimen of BCG, on the other hand, can trigger more adverse reactions. Not only is the prognosis relatively poor, but patients will also endure much suffering in the process. Currently, NMIBC diagnosis mainly relies on cystoscopy [31, 32]. Conventional imaging methods (CT urography, IVU or ultrasound, etc.) depend decisively on the personal experience of the physician, while ultrasonography, intravenous pyelogram, cystography, magnetic resonance imaging, and lymphography are not very sensitive. Cystoscopy is relatively sensitive, but invasive and expensive, which is difficult for large-scale screening [33]. Urine exfoliative cytology is highly specific but lacks sensitivity [34, 35]. Eurovision fluorescence in situ hybridization has high sensitivity which is widely used for routine clinical detection of NMIBC, but owned with poor sensitivity for low-grade or small tumors [36]. Although some studies suggest that methylated CpG sites in urine may be promising markers for the detection or monitoring of NMIBC [37], these assays are not fully adopted in routine clinical practice and still need to be validated in multicenter and large-scale cohorts in Asia. Novel insights for early-stage screening of NMIBC are still required with further breakthrough.

In this study, we have compared 51 patients with NMIBC and 47 healthy controls as the Cohort 1 to explore the clinical relevance of altered urinary microbiota with the development of NMIBC. Then the functional composition of the urinary microbiome was compared between NMIBC patients and controls by COG and KEGG pathway analysis. *Eubacterium sp. CAG:581* was confirmed to promote the growth of NMIBC organoids by activating ECM1/ERK1/2 phosphorylation/MMP9. Lastly, we used multivariate Cox modeling to investigate the clinical relevance of urinary *Eubacterium sp. CAG:581* abundance with the prognosis of 406 NMIBC patients as the Cohort 2, which suggested urinary *Eubacterium sp. CAG:581* could be used for early diagnosis and prognosis of NMIBC. Thereinto, phosphorylation of ERK1/2 signalling could regulate numerous transcription factors, known to activate the transcription of a broad number of MMPs. Herein we found that inhibiting ERK1/2 phosphorylation resulted in decreased MMP-9 expression, indicating that the mediation of ERK1/2 in promoting NMIBC progression. A similar study performed by Lee et al. has manifested with tandem decreases in MMP-9 mRNA levels following inhibition of ERK1/2 phosphorylation in rhabdomyosarcoma, fibrosarcoma, bladder, colon and prostate carcinoma cells [38], which is in accordance with NMIBC patients of this study. In addition, as the amount of *Eubacterium sp. CAG:581* has influenced the expression of ERK1/2 phosphorylation and MMP9, ECM1 is also positioned as potential therapeutic target for the functional transcriptional network with *Eubacterium sp. CAG:581*-induced NMIBC migration and invasion.

Normal flora of microbial colonies has developed over a long historical evolutionary process and reside on the human body’s surface and inside. Microbial flora exists in sites of the oral cavity, urinary tract and
intestinal tract. These flora have been found to have a tremendous effect on their hosts. Intestinal flora has been widely used in the early-stage diagnosis of neurodegenerative diseases, mood disorders, lung diseases, as well as the recurrence of digestive tract cancers [39–41]. In recent studies, a wealth of research on urinary flora have confirmed that bacteria in urine may play an essential role in many urinary disorders, such as urge incontinence [42]. The study of Thomas White et al. found that bacterial diversity of urinary flora was associated with high body mass index, hormonal status, and symptoms of urge incontinence [43]. Several studies of bioinformatics analysis have also shown that among NMIBC patients, the relative abundance Spingomonas, Immunobacterium, Aeromonas and Bacillus spp. in the urinary flora of the tumor group was significantly higher than in the control group [11, 44]. However, more detailed underlying validation is insufficient for these predictions. In this study, MMP9 served as the downstream executor of NMIBC growth. It belongs to the family of zinc-dependent endopeptidases with proteolytic activity against extracellular matrix components, involving in numerous physiological and pathological processes including tissue remodeling, embryonic development, tumor growth and metastasis [45,46]. Previous studies have revealed that MMP-9 got involved into the whole process of pathogenesis in bladder cancer [45,47,48]; however, the present study has only focused on the occurrence of NIMBC. We are constructing the cohort to compare NMIBC patients and MIBC patients for further validating the Eubacterium sp. CAG:581 signature in the conversion from NMIBC into MIBC.

In conclusion, understanding the mechanisms that urine bacterium contributes to NMIBC development has the potential to improve its early-stage diagnostic decisions. This study presents a novel mechanism of Eubacterium sp. CAG:581 in upregulating ECM1-MMP9 axis and identifies Eubacterium sp. CAG:581 to be clinically useful in predicting the occurrence of NMIBC.

**Declarations**

**Ethics approval and consent to participate** The Ethics Committees in Peking University First Hospital approved the study protocols (Ethical number: 2019-138). Written informed consents were obtained from all participants in this study.

**Consent for publication** Written informed consent was obtained from the patients for publication and any accompanying images. The copy of the written consent is available.

**Availability of data and material** All generated as well as analyzed data in the present research are contained in this manuscript.

**Competing interests** The authors declare that they have no competing interests.

**Funding** This project was supported by National Natural Science Foundation of China (82204515), National High Level Hospital Clinical Research Funding (Research Achievement Transformation Project of Peking University First Hospital) (2022RT04), Research Seed Fund of Peking University First Hospital (2022SF04).
Authors' contributions
The manuscript written was completed by Yuhang Zhang and Wenyu Wang. Yimin Cui has conceived the whole program. Experiment performance was done by Yuhang Zhang and Gang Zhao. Data collection was conducted by Wenyu Wang. All authors have discussed the results and reviewed the manuscript.

Acknowledgements
We thank Prof. Zu-Hua Gao at Department of Pathology, McGill University, for helpful discussions and data analysis.

References


Figures
Figure 1

Metagenomics data manifested the clinical association between *Eubacterium sp. CAG:581* and NMIBC occurrence. (A) A cladogram representation of data in 51 NMIBC patients versus 47 healthy controls by 16S rDNA sequencing. Taxa enriched in NMIBC patients (Amber) and healthy controls (Cobalt). The brightness of each dot is proportional to its effect size. (B, C) Chao 1 richness index (B) and Rank Abundance Curves (C) of urine samples from 51 NMIBC patients (Amber) versus 47 healthy controls.
Bray-Curtis’s distance (D) and Binary-Jaccard distance (E) of urine samples from 51 NMIBC patients (Amber) versus 47 healthy controls (Cobalt). Linear discriminant analysis (LDA) coupled with the effect size measurements identifies the significant abundance of data in A. Taxa enriched in NMIBC patients are indicated with negative (Cobalt) or positive (Amber) LDA scores as compared with healthy controls, respectively. Only taxa greater than LDA threshold of 3.5 are shown. DEGs enriched in the COG classification. Dots represent term enrichment with different colors based on Q values. Red color indicates high enrichment, while blue color indicates low enrichment. The sizes of the dots represent the gene number of enrichment. Histogram presentation of the KEGG pathway. A total of 148 DEGs were successfully annotated and grouped into 26 functional categories. P-values were determined using two-sided Fisher’s exact tests with Benjamini-Hochberg correction for multiple testing.
Coculture of *Eubacterium sp. CAG:581* promoted the growth of NMIBC organoids. (A) Schematic diagram of the coculture system of NMIBC organoids and anaerobic urine microbiota using the two-chambers chip. Each chip had a reservoir layer on the top, a 3D implanting hole in the middle, and anaerobes culture slide underneath. (B) Growth curve of the NMIBC organoids cocultured with *Eubacterium sp. CAG:581* (10^6 cfu per chip) and *Bacteroides sp. 4_3_47FAA* (10^6 cfu per chip). Results are presented as mean ±
SEM, n = 3. *p < 0.05, **p < 0.01, ***p < 0.001. (C) Bright-field images showing the representative NMIBC organoids size on Day 0 and Eubacterium sp. CAG:581 coculture group, Bacteroides sp. 4_3_47FAA coculture group and NMIBC control group on Day 21. Scale bars, 1000 μm (left panels) and 10 μm (right panels). (D) Heatmap of DEGs of Eubacterium sp. CAG:581 coculture group as compared with NMIBC control group on Day 21. Gene copy numbers are transformed as log2 ratios per gene (blue, decrease; red, increase). Gene expression profiling analysis was used to identify the potentially upregulated ECM1 and MMP9 as well as the downregulated PTPN6 and IKZF3 (fold change > 2, FDR < 0.001). (E) Gene Set Enrichment Analysis (GSEA) plot in Eubacterium sp. CAG:581 coculture group versus NMIBC control group. GSEA algorithm was used to evaluate the statistical significance of ECM1, MMP9, PTPN6 and IKZF3, according to the normalized enrichment scores (NES). These values were assigned to each gene and set after normalization across all analyzed gene sets.
Figure 3

*Eubacterium sp. CAG:581* activated ECM1/ERK1/2 phosphorylation/MMP9 of NMIBC organoids. (A) qRT-PCR analysis of *ptpn6, ikzf3, ecm1* and *mmp9* levels of NMIBC organoids cocultured with increased gradient (5×10^5, 10^6, 5×10^6, 10^7, 5×10^7 cfu) of *Eubacterium sp. CAG:581*. (B) Western blot analysis of PTPN6, IKZF3, ECM1, MMP9, ERK1/2 phosphorylation, and AKT phosphorylation in NMIBC organoids cocultured with increased gradient (5×10^5, 10^6, 5×10^6, 10^7, 5×10^7 cfu) of *Eubacterium sp. CAG:581*. (C) Western blot analysis of ECM1, MMP9 and ERK1/2 phosphorylation in BCa organoid^{shECM1} and BCa organoid^{vector} treated with 10 μM Ravoxertinib (Rav) or Ulixertinib (Uli). (D) Growth curve of BCa organoid^{shECM1} and BCa organoid^{vector} treated with 10 μM Ravoxertinib (Rav) or Ulixertinib (Uli). Results are presented as mean ± SEM, n = 3. *p < 0.05, **p < 0.01, ***p < 0.001.
Figure 4

_Eubacterium sp. CAG:581_ was endowed with the diagnostic predictor for NMIBC. (A) Univariate analysis of baseline information was performed in 51 NMIBC patients versus 47 healthy controls. The bars correspond to 95% confidence intervals. (B) Univariate predictive value of _Eubacterium sp. CAG:581_ was compared between 51 NMIBC patients and 47 healthy controls, Log-rank test. (C) Univariate predictive value of ECM1 was compared between 51 NMIBC patients and 47 healthy controls, Log-rank test. (D)
Univariate predictive value of MMP9 was compared between 51 NMIBC patients and 47 healthy controls, Log-rank test.

Figure 5

Identification of NMIBC occurrence-associated *Eubacterium sp. CAG:581* signature in the larger population. (A) The result of PCA analysis demonstrating the efficacy of *Eubacterium sp. CAG:581* signature in differentiating NMIBC from healthy control. (B) LASSO coefficient profiles of the prognostic DEGs. Each curve represents a coefficient, and the x-axis represents the regularization penalty parameter. As the tuning parameter (λ) changes, a coefficient that becomes non-zero enters the LASSO regression
model. (C) Cross-validation to select the optimal $\lambda$. The red dotted vertical line crosses over the optimal log $\lambda$, which corresponds to the minimum value for multivariate Cox modeling. The two dotted lines represent one standard deviation from the minimum value. (D) The distribution of risk scores is shown for the cohort 2. The dotted horizontal line indicates the cut-off level of the risk score used to stratify patients, and the dotted vertical line separates participants on the basis of low-risk (green) or high-risk (red). (E) The distribution of overall outcomes in the cohort 2. NMIBC patients with low-abundance of *Eubacterium sp. CAG:581* are shown in green, while NMIBC patients with high-abundance of *Eubacterium sp. CAG:581* are shown in red. (F) Heatmap of ECM1, MMP9, PTPN6 and IKZF3 in the cohort 2. Gene copy numbers are transformed as log2 ratios per gene (blue, decrease; red, increase). (G) Kaplan-Meier survival plots predicted to be at risk for NMIBC occurrence in the cohort 2. The number of patients remaining at a particular timepoint is shown at the bottom. (H) Time-dependent ROC curves for predicting one-year, two-year and three-year survival in the cohort 2, with the AUC value of 0.72, 0.73 and 0.79, respectively. ROC, receiver operating curve; AUC, area under the curve.

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryData.docx