Intratumor bacteria is associated with prognosis in clear-cell renal cell carcinoma

Running title: ITB predicts survival in ccRCC

Yuqing Li, M.D.[1#], Dengwei Zhang, Ph.D.[2,3#], Linyi Tan, M.D.[1], Junyao Xu, M.D.[4], Ting Guo, M.D.[5], Yang Sun, Ph.D.[6], Rui Zhang, Ph.D.[7], Yao Cheng, Ph.D.[7], Haowen Jiang, M.D.[1]*, Wei Zhai, M.D.[4]*, Yong-xin Li, Ph.D.[2,3]* and Chenchen Feng, M.D.[1]*

[1]Department of Urology, Huashan Hospital, Fudan University, Shanghai 200040, China

[2]Department of Chemistry and The Swire Institute of Marine Science, The University of Hong Kong, Pokfulam Road, Hong Kong SAR, China.

[3]Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou, China

[4]State Key Laboratory of Oncogenes and Related Genes, Department of Urology, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200127, China

[5]Department of Gynecology, Obstetrics and Gynecology Hospital of Fudan University, Shanghai 200000, China

[6] Cancer Institute, Xuzhou Medical University, 209 Tongshan Road, Xuzhou, Jiangsu，221004, China

[7]Shanghai KR Pharmtech, Inc., Ltd, Shanghai, China

[#]Equal contributors

*To whom correspondence may be addressed.

Emails: urology_hs@163.com (HJ); jacky_zw2002@hotmail.com (WZ); yxpli@hku.hk (YxL); fengchenchen@fudan.edu.cn (CF)

Word count: 5265

ABSTRACT

Background

Intratumor bacteria (ITB) plays a role in various cancer types. Its role in clear-cell renal cell carcinoma (ccRCC) remains elusive due to small sample size and inadequate decontamination in relevant studies.

Objective

To establish common and reproducible ITB-associated biomarkers in ccRCC.

Design, setting, and participants

This retrospective study comprised seven bulk RNA sequencing datasets from six publicly available cohorts and one in-house Chinese cohort (Renji), one 16S rRNA sequencing dataset from an original Chinese cohort (Huashan), and one publicly available single-cell RNA sequencing dataset. All of these datasets included ccRCC cases.

Outcome measurements and statistical analysis

Composition was presented by relative abundance. Overall and progression-free survival were primary outcomes profiled by putative ITB load and risk score, respectively. Potential host interaction was exploratorily analyzed using gene set enrichment analysis and Sparse CCA.

Results and limitations

Nine cohorts encompassing a total of 1049 ccRCC cases and 130 paired normal tissues were initially analyzed and underwent decontamination. Surprisingly, neither diversity nor composition was differentially distributed between normal and cancer tissue. High putative bacterial load was associated with better overall survival. Notably, a 7-genera dichotomized ITB risk score was indicative of overall survival and a 13-genera dichotomized ITB risk score was predictive of progression-free survival, respectively. *Actinomyces*, *Rothia* and *Bifidobacterium* showed a protective role while *Exiguobacterium* was a risk factor. A limitation is lack of causation analyses.

Conclusions

ITB exists in ccRCC. High ITB loads and ITB-risk score predicts better ccRCC survival regardless of sequencing tech, sample processing or racial disparity.

KEY WORDS Clear-cell renal cell carcinoma; Intratumor bacteria; Biomarker; Prognosis

Patient Summary

In this report, we explored the role of intratumor bacteria (ITB) in renal clear-cell carcinoma (ccRCC) in patients with different race and sequencing platforms. Putative ITB load and a 7-genera ITB risk score were associated with overall survival. A 13-genera ITB risk score was predictive of progression-free survival. We conclude that certain ITB features are universally pathogenic to ccRCC.

INTRODUCTION

Intratumor microorganisms, especially intratumor bacteria (ITB) signature has shown prognostic effect in a variety of cancer types including gastric cancer[1], colorectal cancer[2], hepatocellular cancer[3], etc., establishing microbiome as a novel omics or "second genome" of cancer[4]. However, ITB may vary to vast extents that renders intratumor microbe findings in some cases, hardly reproducible[5]. Amongst all confounders, biomass of subject[6,7], race[8,9], sequencing tech[10]., and contamination[11] play the most pivotal roles.

Despite the pitfalls present in ITB studies, true bacterial signatures could be still identified by imperfect sequencing technologies and decontamination processes, which has been demonstrated through experimental validation[12,13]. Furthermore, when consistent findings emerge from multiple cohorts, the influence of these pitfalls can be minimized, leading to more reliable and robust conclusions.

Clear-cell renal cell carcinoma (ccRCC) is the most common type of malignancy in kidney that is conventionally accepted as sterile organ and ccRCC is expected to harbor a low biomass of ITB. To date, the existing literature on ITB in ccRCC is limited to three full papers[14-16]and one meeting proceeding[17]. These studies suffer from small sample sizes, lack of racial diversity, use of a single sequencing technology, and inadequate decontamination process. Consequently, the precise composition and significance of ITB in ccRCC remain elusive. Therefore, a multi-cohort study focusing on ccRCC is urgently required.

While we highly concur that ITB exists in most, if not all solid tumors including ccRCC, we aim to answer whether common ITB composition exists in ccRCC and whether ITB signature is prognostic, regardless of demographic, racial and sequencing differences. To achieve this, we incorporated various reports on decontamination and multiple ccRCC cohorts encompassing over 1000 cases that vary in region, race, batch, sequencing tech, etc. We aim

to identify inherent ITB signatures and explore its prognostication in ccRCC in the current study.

METHODS

Study Population

For 16S rRNA sequencing, we retrospectively collected the 217 tumor and 27 normal samples from 217 patients histologically diagnosed clear cell renal carcinoma who underwent partial or radical nephrectomy in Huashan hospital (Shanghai, China) between May 2013 and Oct 2022 under reasonable inclusion criteria (**Figure 1**). The samples were formalin-fixed and paraffin-embedded. We also included 10 negative controls using sliced paraffin from the margin of the block, sampling paraffin only without tissue. Tumor stages were stratified according to the 8th American Joint Committee on Cancer staging system (AJCC)[18] No subjects received preoperative treatments, including immunotherapies or molecular targeted therapies.

Six cohorts with the RNA-Seq data available were included in our study. EGAD00001000597[19] as an integrated molecular study of ccRCC and consists of 100 tumor samples. EGAD00001006029 (CheckMate 025; NCT01668784)[20] was a prospective clinical trials of the anti-PD-1 antibody nivolumab in advanced clear cell renal cell carcinoma, and 53 FFPE tumor tissues were obtained prior to initial therapy for patients enrolled in this study, including 15 patients with the objective response record of immunotherapy. Data from above two datasets (EGAD00001000597 and EGAD00001006029) were requested from the principal strictly via European Genome-phenome Archive (EGA, https://ega-archive.org/) according to the clinical data transfer agreement. GSE102101[21], GSE126964[22], GSE151419[23] were studies concerning on the renal cell carcinoma by organizations located in Singapore, China and Poland, respectively, and the raw sequencing data of tumor and paired normal samples were downloaded from the gene expression omnibus (GEO) data repository (https://www.ncbi.nlm.nih.gov/geo/). Besides, 27 fresh tumor samples with RNA sequencing were supplied by the Renji hospital (Shanghai, China).

TCGA was a cancer genomics program which molecularly characterized primary cancer and matched normal samples including clear cell renal cell carcinoma (KIRC). However, due to the limit of access to the level 1 or 2 data of TCGA hosted at the Genome Data Commons (GDC) website, the microbiome data processed by by Poore et al. [24] and by Salzberg et al. [25] respectively were directly adopted for downstream bioinformatic analysis in this study. Poore et al. derived the microbiome data from both WGS and RNA-Seq data, and we used the normalized and batch effect-corrected data of 532 tumor and 72 paired normal samples. The author performed decontamination in several degrees and got 5 microbial communities including data with non-contamination removed (NR), data with likely contaminants removed (LR), data with putative contaminants removed (PR), data with contaminants removed by sequencing "plate–center" combinations (CR), and data with mostly stringent filtering (SR). Salzberg et al. also used the TCGA data but only took WGS data into consideration and shared us with the data including 40 tumor and 35 paired normal samples. Demographics of all cohorts were demonstrated in **Table 1**.

16S rRNA gene amplicon sequencing

In preparation for the 16S rRNA gene sequencing, samples were sectioned from the paraffin-embedded tissue blocks, which accepted quality testing, purification and nested amplification. To meet the requirement of sufficient DNA for sequencing, the amplified products were detected by DNA electrophoresis, and the eligible samples were kept for further study. 16S rRNA gene sequencing was conducted at the Nonogene Co., Ltd. In brief, genomic DNA was extracted from the tissue samples using the CTAB/SDS method. The 16S rDNA V4 region was amplified through PCR employing a primer pair (515F: 5'-GTGCCAGCMGCCGCGGTAA-3' and 806R: 5'-GGACTACHVGGGTWTCTAAT-3') with a barcode. Sequencing libraries were prepared using the TruSeq® DNA PCR-Free Sample Preparation Kit (Illumina, USA) following the manufacturer's instructions. The libraries were subsequently sequenced on the Illumina NovaSeq platform, yielding 250 bp paired-end reads.

16S rRNA sequencing data processing and analysis

The raw sequencing data of 16S rRNA in FASTQ format underwent processing with QIIME 2 version

2023.2. Quality filtering, denoising, and chimera removal were performed using DADA2, resulting in high-quality sequences that were assigned to amplicon sequence variants (ASVs). The feature table was constructed using these ASVs, and taxonomic information was annotated using a Naive Bayes classifier trained with the SILVA 138 SSURef NR99 database. ASVs that could not be confidently assigned at the phylum level, as well as non-bacterial ASVs, were excluded from further analysis.

To minimize the potential impact of contaminants, we employed a previously established decontamination process. This involved three steps: Firstly, we used the isNotContaminant function in the "decontam" algorithm (ref) to identify possible contaminants. This prediction was based on the difference in ASV prevalence between FFPE samples and tissue samples. Secondly, ASVs with a relative abundance greater than 0.5% in the FFPE samples were removed. Lastly, ASVs that appeared in less than 5% of the tissue samples were further eliminated to avoid contingency. Only ASVs that met these criteria were retained for downstream analysis.

Bulk RNA sequencing data processing

Raw RNA sequencing data from tissue samples obtained from six cohorts were acquired online. Sequencing reads were quality-controlled using fastp v0.21.1, with parameters “-l 50 -5 3 -3 3”. Filtered reads that were shorter than 50 bp were discarded. To quantify human gene expression, the clean reads were aligned to the human reference genome, GRCh38.p13, available in the GENCODE database using HISAT2 v2.2.1. The gene expression values were quantified in transcripts per million (TPM) using StringTie v2.2.1.

For profiling the intratumor bacteria from bulk RNA-Seq data, clean reads were initially aligned against an indexed database to remove host or contaminant reads. This alignment was performed using bowtie2 v 2.4.5 with a “--very-sensitive” model. The indexed database included 9 mammalian genomes (hg38, felCat9, canFam6, mm39, rn7, rheMac10, susScr11, galGal6, bosTau9; University of California– Santa Cruz Genome Browser), 2145438 complete bacterial plasmids (PLSDB databse, v.2021_06_23_v2), 13705 mitochondrial genomes (NCBI RefSeq database, accessed on Aug 15, 2022), 9443 plastid sequences (NCBI RefSeq database, accessed on Aug 15, 2022), 6093 UniVec sequences (NCBI RefSeq database, accessed on Aug 15, 2022), which were considered potential sources of human habitat- or laboratory-associated or extrachromosomal sequence contaminants for taxonomic classification of microbial metagenomic sequences[26]. Unmapped paired reads were then subjected to KrakenUniq v 1.0.4 for taxonomic assignment using a pre-built database. This database includes complete microbial genomes from RefSeq, comprising

46,711 bacterial genomes, 13,011 viral genomes, and 604 archaeal genomes. Additionally, the database contains 246 eukaryotic pathogens, the UniVec set of standard laboratory vectors, and the GRCh38 human genome. The abundance of bacteria was evaluated at the genus level, which was deemed more accurate than the species level.

To ensure the removal of potential false positive assignments, the bacterial genera underwent further filtration based on the following criteria: (1) the genus must contain more than 5 reads; (2) number of duplicated kmer must be larger than half of assigned read counts; (3) genome coverage must be larger than 1e-5. (1) the genus must have a read count greater than 5; (2) the number of duplicated k-mers must exceed half of the assigned read counts; and (3) the genome coverage must be larger than 1e-5. Additionally, efforts were made to distinguish the potential host of the identified genera in order to eliminate non-human-associated genera that are likely to be contaminants. To accomplish this, information regarding the isolation sources of bacteria deposited in NCBI (https://www.ncbi.nlm.nih.gov/genome/browse#!/prokaryotes/), IMG/M (https://img.jgi.doe.gov/cgi-bin/m/main.cgi), GOLD (https://gold.jgi.doe.gov/downloads), and BV-BRC (https://www.bv-brc.org/docs/quick_references/ftp.html) was gathered. For identified genera not isolated in these four databases, potential hosts were obtained through a literature search on Google Scholar. Based on the available host information, the identified genera were classified into three groups: non-human (genera not isolated from human), human-exclusive (genera exclusively associated with the human host), and mixed (genera derived from either human or other environments). Non-human-associated genera were subsequently excluded from further downstream analysis.

Microbial analysis

For 16S rRNA sequencing data, feature table, taxonomy, and phylogenetic tree after decontamination were combined into a Phyloseq object for downstream processing. To estimate alpha diversity and beta diversity, all samples were rarefied to 2000 sequencing reads. The statistical significance of differences in alpha diversity was assessed by stat_compare_means function in R package “ggpubr”. Difference in microbial compositions was tested using Permutational multivariate analysis of variance (PERMANOVA).

For the bulk RNA-Seq data, the counts of genera were converted to relative abundance for analysis. Due to the ununiform sequencing depth that would skew the measure of alpha diversity, we did not examine alpha diversity among RNA-Seq data. Rather, we compared the bacterial read counts per million reads, which could provide an indication of bacterial load. The Bray-Curtis dissimilarity among the samples was

calculated using the vegdist function in the R package "vegan" and subjected to Principal Coordinates Analysis (PCoA). The statistical significance of the findings was evaluated using PERMANOVA analysis with the adonis2 function.

To evaluate the impact of clinical factors on intratumor microbial communities, a PERMANOVA analysis with 999 permutations was conducted based on Bray-Curtis dissimilarity. To account for multiple comparisons, all P-values were adjusted using the false discovery rate (FDR) method. To explore the relationship between the overall microbial community and overall survival or progression-free survival, dimensionality reduction was employed to reduce the complexity of the microbial data. Principal Component Analysis (PCA) was performed using the PCA function in the "FactoMineR" R package. The first five principal components (PCs) of the intratumor microbiome PCA were retained to represent the overall intratumor microbiome. Cox proportional hazard regression models were employed to examine the association between each PC and overall survival or progression-free survival. This analysis was conducted using the coxph function in the "survival" package. P-values were adjusted for multiple comparisons using FDR methods.

Identifying diagnosis-related microbiome

For differential abundance testing between tumour and normal tissues in ccRCC, we used relative abundance and counts per million reads (CPM) respectively. We performed Wilcoxon rank-sum tests for each feature in genus level, and corrected the resulting p-values with the BH method. To exclude the bias caused by the sample number imbalance, we incorporated only the matched specimen and finally got 24 pairs in Huashan cohort, 10 pairs in GSE102101 (Cohort 3), 11 pairs in GSE126964 (Cohort 4), 13 pairs in GSE151419 (Cohort 5).

We also used the Random Forest algorithm to further identify the potential features distinguishing the paired samples using randomForest function in the R package " randomForest"[27]. Ten-fold cross-validation and five repetitions were adopted to help select a specific number of features, whose importance were measured by accuracy and Gini index.

Identifying prognosis-related microbiome

Difference in microbial compositions was first tested between population with long term survival (LTS) and short term survival (STS). Due to the inconsistent following months, we used the median survival time in each cohort as the cutoff. Permutational multivariate analysis of variance (PERMANOVA) was used for

testing difference in microbial compositions as above mentioned.

Univariate cox was performed to identify the genera whose abundance associated with overall survival and progression free survival. The HR (Hazard Ratio)>1 indicated that the feature was a risk factor for the prognosis, while HR<1 indicated that the protect factor. A cluster of genera were preliminarily screened as the input for the least absolute shrinkage and selection operator (LASSO) to exclude the features with potential multi-collinearity. The glmnet function in the R package "glmnet"[28] was used and the family was set as "cox" while the other parameters were set default. Finally, we constructed the cox model using coxph function in the R package " survival". To fit the model more reasonably, we took the stepwise regression method to help select a formula-based model by Akaike information criterion (AIC). The OS-related risk cox model consist of 7 genera, including Abiotrophia, Actinomyces, Bifidobacterium, Dolosigranulum, Faecalibacterium, Kocuria, and Prevotella. The PFS-related risk cox model contained 13 genera, including Acinetobacter, Brachybacterium, Exiguobacterium, Faecalibacterium, Finegoldia, Haemophilus, Kocuria, Lactococcus, Moraxella, Porphyromonas, Prevotella, Rhodococcus, Rothia. The genera with coef>0 in the models were considered risk factors, while those with coef<0 were considered protect factors. Kaplan-Meier survival curves were plotted to report the association between the survival probability and the abundance of specific genera. The strategy for grouping included dichotomization of abundance measured by CPM and the presence or not. The significance was examined by log-rank test and two stage hazard rate comparison. Combined with the clinical covariate such as sex, age, tumor stage and grade, the risk score was tested using univariate and multivariate cox to determine whether our risk score of microbial features could acted as an independent prognostic factor.

We attempted to determine the centrality among the genera involved in the cox model and to find the hub genera. The estimateNetwork function in the R package "bootnet"[29] as used and the correlation between the features were visualized with the network plot. The influence of each genus was also measured by the indexes including “Strength”, “Closeness”, “Betweenness” and “ExpectedInfluence”.

Mapping interaction between genera and host gene

We previously got the gene expression of 6 cohorts. To filter genes non-related to protein coding, we mapped the gene list to the human genome profile named ‘Homo_sapiens.GRCh38.109.chr.gtf.gz’ downloaded from the ENSEMBL website (http://asia.ensembl.org/index.html) and 19142 genes finally remained. The gene expression of TCGA was downloaded from the GDC portal (https://portal.gdc.cancer.gov/) and the data

format was transformed to TPM.

To figure out the molecular change, especially the signaling pathway differentiation between the sub-group stratified by the risk score determined by selected microbial features in cox model, we performed the gene set enrichment analysis (GSEA). We used the GSEA function in the R package "clusterProfiler"[30], and the KEGG, PID and REACTOME database were all included using R package "msigdbr"[31]. The p-values were corrected with the BH method.

We took the mantel test to characterize the correlation between interest genera and interest molecular pathways. The mantel_test function in the R package "linkET"[32] (ref) was used. We dichotomized the genera into two clusters labeled as risk genera and protect genera. To score the immune related function, the single-sample gene set enrichment analysis (ssGSEA) method in the R package "GSVA"[33] (ref) was used. The immune cell infiltration was assessed by the quantiseq method using deconvo_tme function in the R package "IOBR"[34]. As there were 15 patients who received the nivolumab immunotherapy and were recorded the objective response rate in the EGAD00001006029 (Cohort 2), we compared the differential genera between two groups, that were CB and NCB, using the Chi-Squared test. The prediction ability was adjudged by the area under curve (AUC).

To macroscopically evaluate the association between tumor microbiome composition and host gene expression, we performed Procrustes analysis. BC dissimilarity was calculated and then the nonmetric multidimensional scaling (NMDS) was used for dimension reduction. The reduced two dimensions or axes was input for the rotations and statistical testing in Procrustes analysis. Furthermore, we took the sparse CCA to identify group level correlations between paired host gene expression and microbiome data using the CCA function in the R package "PMA"[35]. The parameters were set as default. We processed the data before the analysis. The genus whose relative abundance was higher than 0.001 in at least 10% samples were kept, and the data was transformed to the centered log ratio (CLR) format for downstream analysis. We kept the genes whose expression was greater than 0 in half of the samples and then filtered out genes with low variance, using 25% quantile of variance across samples in each disease cohort as cut-off. These filtering resulted in a unique microbiome abundance matrix and host gene expression matrix per cohort for downstream analysis, including 12477 gene × 54 taxa in the EGAD00001000597 cohort, 11817 gene × 28 taxa in the EGAD00001006029 cohort, 12633 gene × 26 taxa in the GSE126964 cohort, 11406 gene × 26 taxa in the GSE151419 cohort, and 11492 gene × 60 taxa in the Renji cohort. As the number of tumor samples in

GSE102101 was small, we didn't performed sparse CCA in this cohort. After the sparse CCA, we got paired genus and genes clusters with significant correlation, and they were classified into a component. The genes in each component were implemented with pathway enrichment analysis. The significance was determined by Fisher's exact test and BH method used for adjustment.

Statistical analysis

All data analyses were conducted via RStudio software unless otherwise specified. Visualizations were performed using ggplot2 R package. Two group comparisons were done using Wilcoxon rank-sum test. Spearman's correlations were calculated using cor.test function. The heatmap was created using Heatmap in "ComplexHeatmap" R package. In this paper, we used the following notation to indicate the significance levels of P-values: NS ($P > 0.05$), *$0.05 < P < 0.01$, **$0.01 < P < 0.001$, and *** $P < 0.001$.

Transmission Electron Microscopy (EM)

A total of 20 ccRCC tissue blocks were subject to EM. Fresh tissues were carefully handled immediately after surgical removal. Blocks sliced 1mm³ in size were placed in a culture dish containing an electron microscope fixation solution. Samples were rinsed in 0.1M phosphate buffer (PB, pH 7.4). Samples were then placed at room temperature for 2 hours using 1% osmium tetroxide prepared in 0.1M phosphate buffer (PB, pH 7.4). Gradual dehydration was applied, and infiltration was conducted in a mixture of propylene oxide and Epon 812 resin (1:1) at 37°C overnight. Samples were inserted into an embedding mold filled with pure Epon 812 resin. The embedding mold underwent polymerization in a 60°C oven for 48 hours. Ultrathin sections (70nm) were cut from resin blocks using an ultramicrotome and placed on 200 mesh Formvar-coated copper grids. Copper grids with sections were stained in a 2% uranyl acetate-saturated alcoholic solution for 15 minutes. Following three rinses with ultrapure water, sections were stained with a lead citrate solution for 10 minutes. Copper grid sections were air-dried at room temperature overnight in a copper grid box. The grids were observed under a transmission electron microscope (HITACHI, HT7800).

16S rRNA staining

We performed 16S rRNA staining in 178 samples mounted on a tissue microarray (TMA) chip from the Huashan cohort with an established protocol reported by our group previously[19]. Briefly, thorough sterilization of hood, blades, and relevant instruments was carried out. Deparaffinized sections were dehydrated, and protease K was applied at room temperature. 100 μM of EUB338-cy5 probes (sequence: 5′–GCTGCCTCCCGTAGGAGT–3′) diluted in 1 μM working solution were applied and samples were

315 finalized with DAPI (1:500) staining.

316

317

RESULTS

ccRCC has low-biomass and most ITB are contaminants

As most cohorts in this study were sequenced by bulk RNA-seq, we applied a tailored decontamination algorithm (**Fig 2A**). Analysis of these datasets revealed diverse bacteria present in ccRCC samples (**Fig S1**). Raw ITB reads took up ~(1/2.00E-06) of total sequencing reads (**Fig S2A**) and showed a positive correlation with total reads (**Fig S2B**). 327 out of 545 genera survived after decontamination (**Fig S2C**). Our passes not only managed to filter out nonhuman reads (**Fig S2D**) but also showed an increase in proportion of common contaminants after decontamination, indicating that some bacteria, previously accepted as contaminants could be indwelling in ccRCC (**Fig S2E**). Relative abundance of non-human associated bacteria dropped consistently in all cohorts following our decontamination (**Fig S2F**). Common genera across cohorts after decontamination remained comparable either grouped by dataset or by sample (**Fig S3A-D**), whereas similar trend for bacterial read drop was noticed in cancer and normal tissue, respectively, further authenticating the remaining reads were true ITB in ccRCC (**Fig S3E-F**). Compositional atlas demonstrated by relative abundance, as expected, varied drastically across cohorts (**Figure S4A-D**). Despite so, two phyla, Proteobacteria and Firmicutes were present in all bulk-sequenced cohorts (**Fig S4**) and in scRNA-sequenced samples (**Fig S5, Table S1**). They were putatively present in diverse cells such as tumor cells (**Fig S5**). Furthermore, we then applied 16S rRNA-targeted FISH probe and EM imaging to 20 ccRCC tissue blocks, validating ITB existence in ccRCC (**Fig 2B-C**). We also attempted to culture 5 tissue blocks in aerobic and anaerobic conditions, but no bacterial growth was noted, supporting low biomass feature of ccRCC (data not shown). We then cross-referenced top-20 abundant genera in all cohorts and found 11 genera were present in ≥ 5 cohorts (**Fig 2D, Fig S5B, Table S2**). Interestingly, three genera including *Cutibacterium* were also present in TCGA cohort processed by both approaches (**Fig S5C**). Here, we concluded ccRCC harbored

a low biomass of ITB and identified presence and composition of ITB which was possibly extracellular in ccRCC.

ITB does not differ between adjacent normal and cancer tissue in ccRCC

Using decontaminated reads (**Table S3**), we next probed clinical associations of ITB in three profiles: putative ITB load, ITB signature and individual ITB feature(s). Putative ITB load did not differ between paired normal and cancer tissue in all cohorts (**Fig 2E, Fig S6A**). Whereas cohorts that underwent RNA-seq could not be processed for alpha-diversity, we did not observe a difference in alpha-diversity in the Huashan cohort (**Fig 2F, Fig S6B-C**). Surprisingly, no differences in beta-diversity between normal and cancer tissue were observed in all cohorts (**Fig 2G, Fig S6D**). The only exception was TCGA_P cohort, which was challenged for its overinflated ITB reads (**Fig S6E**) and the alleged corrected version, TCGA_S cohort, again showed no difference (**Fig S6F**). We thus pursued whether individual ITB feature(s) was differentially distributed and was reproducible. Consistent with barren result of comparison between tumor and normal samples using Wilcoxon Test (**Fig S7A-C**), although the Random Forest identified 10 candidate differential ITB, this machine learning failed to validate those features with satisfactory predicting efficacy across the cohorts (**Fig S7D-E**). Again, the 10 features showed inconsistent trends in TCGA_P cohort and none was significantly different in TCGA_S cohort (**Fig S8A-B**). Here we show astonishingly that, contrary to most studies, differential ITB between adjacent normal and cancer tissue could very well be not present in ccRCC. Our findings highly suggested that most ITB in ccRCC could be inherent intra-tissue bacteria residing in kidney and only individual ITB features altered in abundance in cancer environment, supporting a passenger role of ITB in tumorigenesis stage of the disease.

Putative ITB load and risk score predict survival in ccRCC

As expected, ccRCC could not be subtyped by ITB signature based on survival (**Fig S9A-B**). Indeed, ITB signature on the whole was not associated with any major clinicopathological

parameters across cohorts (**Fig S10**). However, higher putative loads were associated with better overall survival (OS) in three cohorts available with OS profile (**Fig 3A**). In TCGA_S cohort that encompassed a small sample size, higher loads conferred a numerical better survival whereas TCGA_P cohort showed no difference, further questioning data processing in TCGA_P cohort (**Fig S11A**). Higher putative loads were solely associated with a better progression-free survival (PFS) in two cohorts, not reproducible in one of our original cohorts (cohort 6, Renji) and played a marginally protective role in TCGA_S cohort (**Fig S11B-C**). We then identified the compositional differences between patients with long and short survival, and the genera that coexisted and possessed consistent risk in univariate cox across three cohorts was used as input for LASSO and Cox model constructing (Fig S12A-B). The model identified a 7-genera ITB risk score predictive of OS in all three cohorts (**Fig 3B, Fig S12C-D, Table S4**) but not in either TCGA cohort (**Fig S12E, Table S5**). Specifically, *Actinomyces* and *Bifidobacterium* were protective ITB in ccRCC (**Fig S13**). Similar methodology was applied to PFS probing and a 13-genera risk score was generated (**Fig S14, Table S6**). Higher score predicted worsened PFS in all cohorts (**Fig 3C**) in which *Exiguobacterium* was a risk factor and *Rothia* was protective (**Fig S15**). Likewise, the results were not reproducible in either TCGA_P or TCGA_S cohort (**Fig S16, Table S7**). Whereas TCGA_P was problematic and TCGA_S consisted of only WGS samples, we here provided solid evidence that both ITB loads and features played a role in prognosis. This encouraged us to further investigate host interactions and treatment response.

ITB is immune-related in ccRCC

Of exploratory interest, we investigated interactions between prognosis-related ITB (**Fig S12D, Fig S14C**) and found *Actinomyces* and *Rothia* being consistent hub ITB features across cohorts (**Fig S17**). When host interactions were incorporated, we found the immune response to be the sole consistently enriched program in ITB risk score-stratified patients across all

cohorts (**Fig 4A, FigS18A-B, Table S8-11**). In reminiscence of inter-ITB interactions, ITB genera were associated with antigen presenting cell functions (co-inhibition and co-stimulation). (**Fig 4B, Fig S18C-D**). The risk score ITB showed in general negative correlation with pro-cancer immune infiltrates (**Fig 4C**). Specifically, absence of protective ITB *Actinomyces*, *Rothia* and *Bifidobacterium* were associated with M2 polarization of macrophages (**Fig 4D, Fig S18E-G**). Nonetheless, those three features were not associated with response to immune checkpoint inhibition and we identified *Anaerococcus* and *Corynebacterium* enriched in ccRCC with complete response (CR) to Nivolumab (**Fig 4E-F**). Lastly, we profiled host interaction using Sparce CCA and three out of five cohorts showed significant host gene-ITB interaction (**Fig S19A**). Besides immune, we also noted Ribosome signaling was associated with some microbiota across all cohorts (**Fig S19B-G**). Here, we showed ITB was associated with host immune response in particular protective ITB that were associated with decreased immune escape.

DISCUSSION

Our study encompassed thus far the largest number of ccRCC cases subject to ITB detection. In comparison to previous smaller studies[14-17], several ITB features appeared to be ubiquitously present at high relative abundance including Proteobacteria and Firmicutes at phylum level and *Pseudomonas*, *Acinetobacter* and *Staphylococcus* at genus level. Lack of difference in ITB loads, composition or diversity between normal and cancer tissue was one of our major findings. Though it was previously reported by Wang et al, we initially considered it to be a result of lack of any decontamination in their study[16]. Given that ITB features associated with prognosis were not amongst the top abundant ones, we speculate that ITB could be sporadic and commensal, not just in ccRCC but also in kidney.

Though our 7-genera panel appeared to perform consistently in all cohorts, we are yet to conclude a pathogenic mechanism regarding a single ITB. Like in genetic association studies, prognostic panel composed of multiple genes serves as a biomarker simply because none of the individual gene is statistically powerful enough to generate a reproducible survival difference and any attribution of a single element should be supported by mechanistic analyses by cell or animal modeling. Likewise, our ITB panel solely represents the prognostication of the microbial community. Moreover, our ITB panel was only aggregated at genus rather than species level, further against overinterpretation.

The causation between ITB and renal tumorigenesis remains unknown[36]. Whether those prognostic ITB are still commensal or, playing driving roles alongside tumor progression depends on human microbiota-associated murine models (HMAMMs) and microbe-phenotype triangulation (MPT)[36]. Unfortunately, there are currently no transgenic murine models for ccRCC[37] and culturomics from animal models is therefore inapplicable[38]. It was surprising that most prognosis-associated ITB features were protective and so were high putative loads, contrary to many oncobiome studies. We did not evaluate absolute ITB loads in our own

cohorts as loading could not be accurately calculated in transcriptome datasets. However, given that recent study points out that absolute, rather than relative abundance plays more important role in microbiome study[39], and that load is prognostic in nasopharyngeal cancer (NPC)[40], we are now setting up a new line to evaluate association between absolute ITB loads and prognosis.

We did not put much effort into imaging ITB. For low-biomass cancer, both LPS and FISH staining could harbor magnified signals from extra-tumor bacterial contamination[41]. We consider multiple sequencing platforms together with FISH signaling adequate to prove the existence of ITB. Of note, we did not identify any intracellular bacteria either by EM or scRNA-seq. This could either be inherent biology of ccRCC or be a result of extreme low biomass of kidney as we successfully identified ITB in all 10 samples of bladder cancer undergoing scRNA-seq in another companion project (data not shown).

Recent debate over the landmark cancer microbiome study by Poore et al[1] has drawn much attention in the oncobiome community. In their recent report[25], Salzberg's team reasoned two major points that Poore's data should be interpreted with caution, including contamination of human reads into microbial signaling and overinflation of microbial reads by machine learning. We owe great thanks to the Salzberg team for providing us KIRC WGS data processed with their protocol for reproduction and validation of our own findings. Even with the very limited sample size, our model showed a numerical OS prediction. The reason Salzberg's team did not process RNA-seq samples was that they considered poly(A)-based transcriptomes could not capture microbial signals. However, half of our cohorts were poly(A)-based transcriptomic datasets and we were able to retrieve effective reads therein. In fact, most ITB studies using scRNA-seq were also able to capture effective reads given the very few cells compared with bulk sequencing. The "poly(A)" problem in the intratumor microbiome has also been thoroughly discussed[42,43] and our findings undoubtedly further supported the notion.

Last but importantly, we show that certain ITB feature is associated with cancer immunity

and response to Nivolumab in ccRCC, in reminiscence of recent trial modulation gut microbiome in metastatic ccRCC patients receiving Nivolumab plus ipilimumab therapy[44]. The protective ITB in our findings are closely related to decreased immune escape, e.g., inhibition of antigen presentation and decreased M2 polarization, both showing pro-inflammatory effects. Interestingly, ITB with different clinical associations seldom overlap and we have not identified such an "omnipotent" ITB in ccRCC. Despite so, *Corynebacterium* is of interest as its abundance ranks top 20 in most cohorts and is associated with Nivolumab response. *Bifidobacterium* supplement has been shown in trial that augments ICI response in metastatic ccRCC patients and our findings that intratumor *Bifidobacterium* was protective shed light on the thus far elusive mechanism of this gut-tumor asix[44]. We did not analyze ITB in ccRCC patients treated with angiogenesis-targeting therapy though there are a handful of datasets available, as angiogenesis was not amongst the MAMPs we identified (**Fig S19B**). Given that combination therapy has become the mainstay of metastatic ccRCC treatment, we are now setting up an ITB analysis in such samples.

CONCLUSION

ITB exists in ccRCC. High ITB loads predicted better survival. We also developed a robust ITB score predictive of prognosis regardless of sequencing tech, sample processing or racial disparity. Those parameters and panels serve as novel biomarkers for ccRCC.

REFERENCES

1. Oosterlinck B, Ceuleers H, Arras W, et al. Mucin-microbiome signatures shape the tumor microenvironment in gastric cancer. *Microbiome*. Apr 21 2023;11(1):86. doi:10.1186/s40168-023-01534-w
2. Mouradov D, Greenfield P, Li S, et al. Oncomicrobial Community Profiling Identifies Clinicomolecular and Prognostic Subtypes of Colorectal Cancer. *Gastroenterology*. 2023;165(1):104-120. doi:10.1053/j.gastro.2023.03.205
3. Sun L, Ke X, Guan A, et al. Intratumoural microbiome can predict the prognosis of hepatocellular carcinoma after surgery. *Clinical and Translational Medicine*. 2023;13(7)doi:10.1002/ctm2.1331
4. Sepich-Poore GD, Guccione C, Laplane L, Pradeu T, Curtius K, Knight R. Cancer's second genome: Microbial cancer diagnostics and redefining clonal evolution as a multispecies process. *BioEssays*. 2022;44(5)doi:10.1002/bies.202100252
5. Aykut B, Pushalkar S, Chen R, et al. The fungal mycobiome promotes pancreatic oncogenesis via activation of MBL. *Nature*. 2019;574(7777):264-267. doi:10.1038/s41586-019-1608-2
6. Nejman D, Livyatan I, Fuks G, et al. The human tumor microbiome is composed of tumor type–specific intracellular bacteria. *Science*. 2020;368(6494):973-980. doi:10.1126/science.aay9189
7. Fu A, Yao B, Dong T, et al. Tumor-resident intracellular microbiota promotes metastatic colonization in breast cancer. *Cell*. 2022;185(8):1356-1372.e26. doi:10.1016/j.cell.2022.02.027
8. Luo M, Liu Y, Hermida LC, et al. Race is a key determinant of the human intratumor microbiome. *Cancer Cell*. 2022;40(9):901-902. doi:10.1016/j.ccell.2022.08.007
9. Byrd D, Wolf P. The microbiome as a determinant of racial and ethnic cancer disparities. *Nat Rev Cancer*. Oct 23 2023;doi:10.1038/s41568-023-00638-7
10. Galeano Niño JL, Wu H, LaCourse KD, et al. INVADEseq to identify cell-adherent or invasive bacteria and the associated host transcriptome at single-cell-level resolution. *Nature Protocols*. 2023;doi:10.1038/s41596-023-00888-7
11. Dohlman AB, Arguijo Mendoza D, Ding S, et al. The cancer microbiome atlas: a pan-cancer comparative analysis to distinguish tissue-resident microbiota from contaminants. *Cell Host & Microbe*. 2021;29(2):281-298.e5. doi:10.1016/j.chom.2020.12.001
12. Colbert LE, El Alam MB, Wang R, et al. Tumor-resident Lactobacillus iners confer chemoradiation resistance through lactate-induced metabolic rewiring. *Cancer Cell*. 2023;41(11):1945-1962.e11. doi:10.1016/j.ccell.2023.09.012
13. Zhang Z, Gao Q, Ren X, et al. Characterization of intratumor microbiome in cancer immunotherapy. *The Innovation*. 2023;4(5)doi:10.1016/j.xinn.2023.100482
14. Heidler S, Lusuardi L, Madersbacher S, Freibauer C. The Microbiome in Benign Renal Tissue and in Renal Cell Carcinoma. *Urologia Internationalis*. 2020;104(3-4):247-252. doi:10.1159/000504029
15. Liss MA, Chen Y, Rodriguez R, et al. Microbiome within Primary Tumor Tissue from Renal Cell Carcinoma May Be Associated with PD-L1 Expression of the Venous Tumor Thrombus. *Advances in Urology*. 2020;2020:1-6. doi:10.1155/2020/9068068
16. Wang J, Li X, Wu X, et al. Uncovering the microbiota in renal cell carcinoma tissue using 16S rRNA gene sequencing. *Journal of Cancer Research and Clinical Oncology*. 2020;147(2):481-491. doi:10.1007/s00432-020-03462-w
17. Wheeler C, Yang Y, Spakowicz D, Hoyd R, Li M. 942 The tumor microbiome correlates with response to immune checkpoint inhibitors in renal cell carcinoma. *Journal for ImmunoTherapy of Cancer*. 2021;9(Suppl 2):A988-A989. doi:10.1136/jitc-2021-SITC2021.942
18. Amin MB, Greene FL, Edge SB, et al. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA: A Cancer Journal for Clinicians*. 2017;67(2):93-99. doi:10.3322/caac.21388
19. Sato Y, Yoshizato T, Shiraishi Y, et al. Integrated molecular analysis of clear-cell renal cell carcinoma. *Nature Genetics*. 2013;45(8):860-867. doi:10.1038/ng.2699
20. Motzer RJ, Tannir NM, McDermott DF, et al. Nivolumab plus Ipilimumab versus Sunitinib in Advanced Renal-Cell Carcinoma. *New England Journal of Medicine*. 2018;378(14):1277-1290. doi:10.1056/NEJMoa1712126
21. Yao X, Tan J, Lim KJ, et al. VHL Deficiency Drives Enhancer Activation of Oncogenes in Clear Cell Renal Cell Carcinoma. *Cancer Discov*. Nov 2017;7(11):1284-1305. doi:10.1158/2159-8290.CD-17-0375
22. Zhao Q, Xue J, Hong B, et al. Transcriptomic characterization and innovative molecular classification of clear cell renal cell carcinoma in the Chinese population. *Cancer Cell Int*. 2020;20:461. doi:10.1186/s12935-020-

533 01552-w
534 23. Kajdasz A, Majer W, Kluzek K, et al. Identification of RCC Subtype-Specific microRNAs-Meta-Analysis of
535 High-Throughput RCC Tumor microRNA Expression Data. *Cancers (Basel)*. Feb 1
536 2021;13(3)doi:10.3390/cancers13030548
537 24. Poore GD, Kopylova E, Zhu Q, et al. Microbiome analyses of blood and tissues suggest cancer diagnostic
538 approach. *Nature*. 2020;579(7800):567-574. doi:10.1038/s41586-020-2095-1
539 25. Gihawi A, Ge Y, Lu J, et al. Major data analysis errors invalidate cancer microbiome findings. *mBio*.
540 2023;doi:10.1128/mbio.01607-23
541 26. Nakatsu G, Zhou H, Wu WKK, et al. Alterations in Enteric Virome Are Associated With Colorectal Cancer
542 and Survival Outcomes. *Gastroenterology*. Aug 2018;155(2):529-541 e5. doi:10.1053/j.gastro.2018.04.018
543 27. Liaw, A. and Wiener, M. (2002) Classification and Regression by Randomforest. R News, 2, 18-22.
544 http://CRAN.R-project.org/doc/Rnews/
545 28. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate
546 Descent. *Journal of Statistical Software*. 2010;33(1)doi:10.18637/jss.v033.i01
547 29. Epskamp S, Borsboom D, Fried EI. Estimating psychological networks and their accuracy: A tutorial
548 paper. *Behav Res Methods*. Feb 2018;50(1):195-212. doi:10.3758/s13428-017-0862-1
549 30. Wu T, Hu E, Xu S, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data.
550 *Innovation (Camb)*. Aug 28 2021;2(3):100141. doi:10.1016/j.xinn.2021.100141
551 31. Dolgalev I (2022). _msigdbr: MSigDB Gene Sets for Multiple Organisms in a Tidy Data Format_. R
552 package version 7.5.1
553 32. Houyun Huang(2021). linkET: Everything is Linkable. R package version 0.0.7.4.
554 33. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data.
555 *BMC Bioinformatics*. Jan 16 2013;14:7. doi:10.1186/1471-2105-14-7
556 34. Zeng D, Ye Z, Shen R, Xiong Y (2023). _IOBR: Immune Oncology Biological Research_. R package version
557 0.99.9.
558 35. Witten D, Tibshirani R (2020). _PMA: Penalized Multivariate Analysis_. R package version 1.2.1
559 36. Lv B-M, Quan Y, Zhang H-Y. Causal Inference in Microbiome Medicine: Principles and Applications.
560 *Trends in Microbiology*. 2021;29(8):736-746. doi:10.1016/j.tim.2021.03.015
561 37. van der Mijn JC, Laursen KB, Fu L, et al. Novel genetically engineered mouse models for clear cell renal
562 cell carcinoma. *Scientific Reports*. 2023;13(1)doi:10.1038/s41598-023-35106-7
563 38. Huang Y, Sheth RU, Zhao S, et al. High-throughput microbial culturomics using automation and machine
564 learning. *Nat Biotechnol*. Feb 20 2023;doi:10.1038/s41587-023-01674-2
565 39. Maghini DG, Dvorak M, Dahlen A, Roos M, Kuersten S, Bhatt AS. Quantifying bias introduced by sample
566 collection in relative and absolute microbiome measurements. *Nature Biotechnology*.
567 2023;doi:10.1038/s41587-023-01754-3
568 40. Qiao H, Tan X-R, Li H, et al. Association of Intratumoral Microbiota With Prognosis in Patients With
569 Nasopharyngeal Carcinoma From 2 Hospitals in China. *JAMA Oncology*.
570 2022;8(9)doi:10.1001/jamaoncol.2022.2810
571 41. de Miranda NFCC, Smit VT, van der Ploeg M, Wesseling J, Neefjes J.
572 2023;doi:10.1101/2023.08.28.555057
573 42. Ghaddar B, Biswas A, Harris C, et al. Tumor microbiome links cellular programs and immunity in
574 pancreatic cancer. *Cancer Cell*. 2022;40(10):1240-1253.e5. doi:10.1016/j.ccell.2022.09.009
575 43. Hu X, Haas JG, Lathe R. The electronic tree of life (eToL): a net of long probes to characterize the
576 microbiome from RNA-seq data. *BMC Microbiology*. 2022;22(1)doi:10.1186/s12866-022-02671-2
577 44. Dizman N, Meza L, Bergerot P, et al. Nivolumab plus ipilimumab with or without live bacterial
578 supplementation in metastatic renal cell carcinoma: a randomized phase 1 trial. *Nature Medicine*.
579 2022;28(4):704-712. doi:10.1038/s41591-022-01694-6
580

581

582

DECLARATION

Ethics approval and consent to participate

Informed consent was obtained for all patients and the study was approved by Huashan Institutional Review Board (HIRB2011-009; HIRB2023-908) and Renji Hospital, School of Medicine, Shanghai Jiao Tong University (KY2023-049-B).

Data availability statement

Read counts of un-decontaminated ITB have been deposited at China National Center for Bioinformation (GSA: CRA011414) that are publicly accessible at https://ngdc.cncb.ac.cn/gsa. Request for TCGA-KIRC data processed by Salzberg et al should be addressed to Prof. Steven Salzberg (steven.salzberg@gmail.com).

Conflict of interest

None.

Authors' contributions

Conceptualization: CF, WZ, YxL, HJ; Methodology: CF, LT, YL, DZ, YxL; Validation: LT, YL, DZ, HJ; Investigation: CF, LT, YL, DZ, RZ, YS, TG; Original Draft: CF, YL, YC

Acknowledgments

This study was sponsored in part by the National Natural Science Foundation of China (Grant No. 81874123 and 82273248). We owe great thanks to the Salzberg team for sharing their data and grant for our use for publication. This study makes use of data generated by the Department of Pathology and Tumor Biology, Kyoto University, and Dr. Seishi Ogawa was highly appreciated. We acknowledge Bristol-Myers Squibb Company (BMS) as the source of the EGAD00001006029 cohort data.

606 **Figure legends**

607 **Figure 1. Study Design**

608 **Figure 2. The presence of intratumor bacteria in ccRCC**

609 (A) Flowchart illustrating the process of analyzing bulk RNA-seq data to identify intratumor
610 bacteria. The analysis involves using bulk RNA-seq data from normal and tumor tissues. To
611 track potential microbial sources, source annotations from bacterial genomes in databases such
612 as NCBI, IMG/M, GOLD, and BV-BRC, as well as literature search, are retrieved. Bacteria
613 associated with the human host are retained for constructing the intratumor bacteria matrix. (B)
614 Representative images of fluorescence in situ hybridization (FISH) staining of 16S rRNA in
615 tumor tissues of ccRCC. (C) Representative images of the presence of bacteria in the tumor
616 tissues captured by transmission electron microscopy based on a total of 20 ccRCC tissue
617 blocks. The red arrow indicates the object. (D) Stacked bar plot showing the proportion of
618 genera present in at least five cohorts among the seven cohorts. Box plot showing the difference
619 of (E) putative load (bacterial counts per million reads) and (F) Shannon index between 24
620 tumor and paired normal samples for comparison in Huashan cohort. The statistically
621 significant difference was given by paired Wilcoxon rank-sum test. (G) Principal coordinate
622 analysis (PCoA) for 24 paired tumor and normal samples in Huashan cohort, based on the
623 Bray–Curtis dissimilarity. The *P* values were tested by Permutational multivariate analysis of
624 variance (PERMANOVA).

625 **Figure 3. Putative ITB loads and risk score predict survival in ccRCC**

626 Kaplan–Meier curves showing the overall survival probability for Huashan, Cohort 1, and
627 Cohort 2 stratified by (A) putative loads and (B) risk score. (C) Kaplan–Meier curves showing
628 the progression-free survival probability for Huashan, Cohort 2, and Cohort 6 stratified by risk
629 score. *P* values were calculated using an unadjusted Log-Rank test.

630 **Figure 4. ITB is immune-related in ccRCC**

(A) The density curves represent the distribution of the immune-related pathways that were significantly enriched between the two stratified groups using gene set enrichment analysis. The horizontal axis indicated the NES of the GSEA result. The stratification was the same as the previous result, that is the overall survival-related risk group in Cohort 1 and 2, and progression-free survival-related risk group in Cohort 2 and 6 from top to bottom. (B) The result of the Mantel test showing the interaction between genera community and potential immune function and the Spearman method was used. The thickness of the curve indicated the absolute value of the spearman rho, and the significant connection was yellow colored. Each block represented the correlation among the immune functions, and a redder color meant a greater rho. (C) Heatmap showing the correlation between specific genus and infiltration scores of immune cells in Cohort 1, Cohort 2, and Cohort 6. (D) Box plot exhibiting the level of M2 macrophage polarization in the presence or absence of *Actinomyces*, *Rothia*, *Bifidobacterium* from left to right in Cohort 6. The Wilcoxon Test was used for comparing the relative abundance between tumor and normal. (E) The heatmap at left showed the relative abundance of the differential genera in abundance in the patients with clinical benefit (CB) and non-clinical benefit (NCB) using Chi-Squared Test. The heatmap at the right indicated the mRNA expression of genes *PDCD1*, *CD274*, and *CTLA4*. (F) The ability of the abundance of *Anaerococcus* and *Corynebacterium* to predict the clinical benefit was visualized by the receiver operating characteristic curve and measured using the area under the curve (AUC).

Figure 1

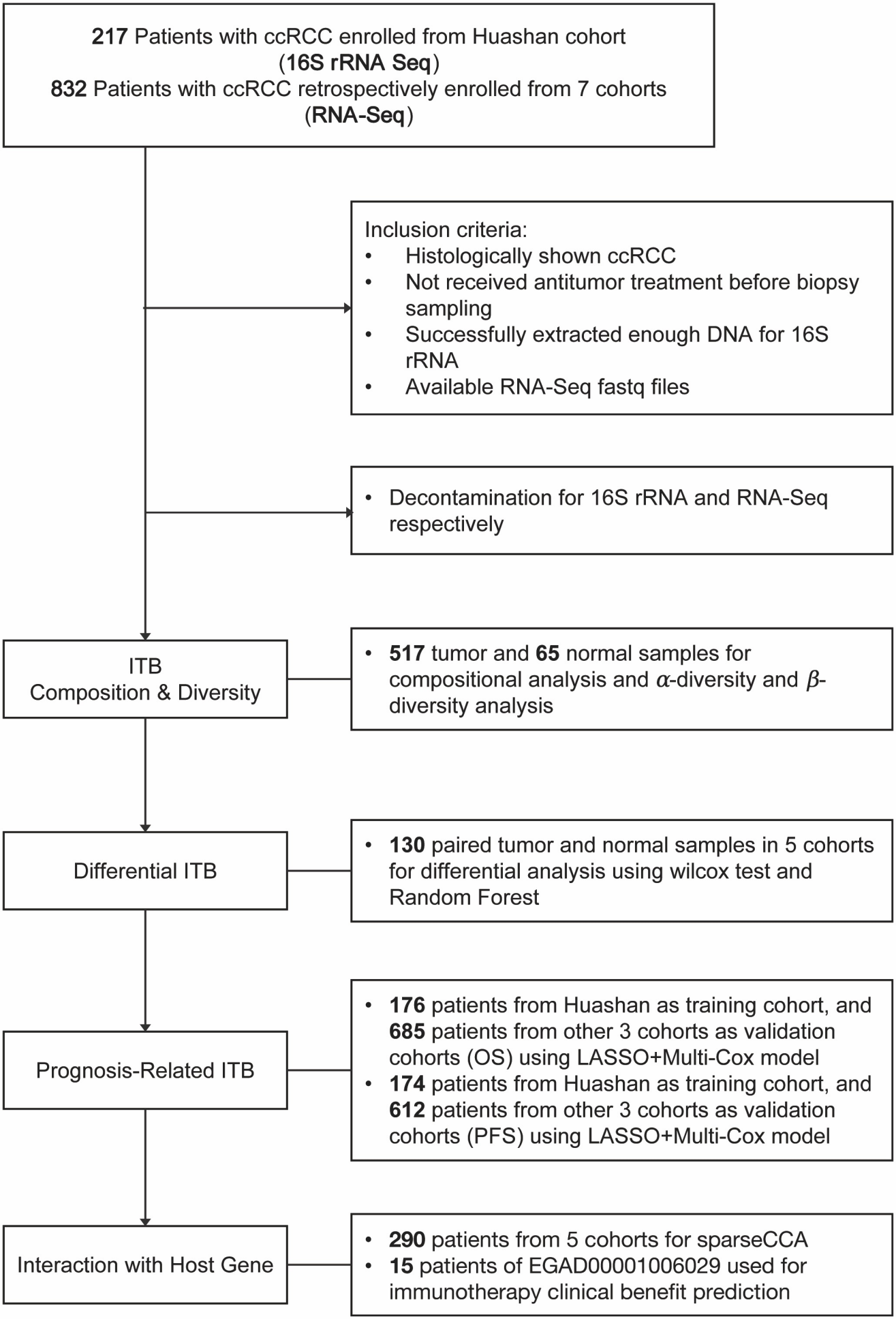
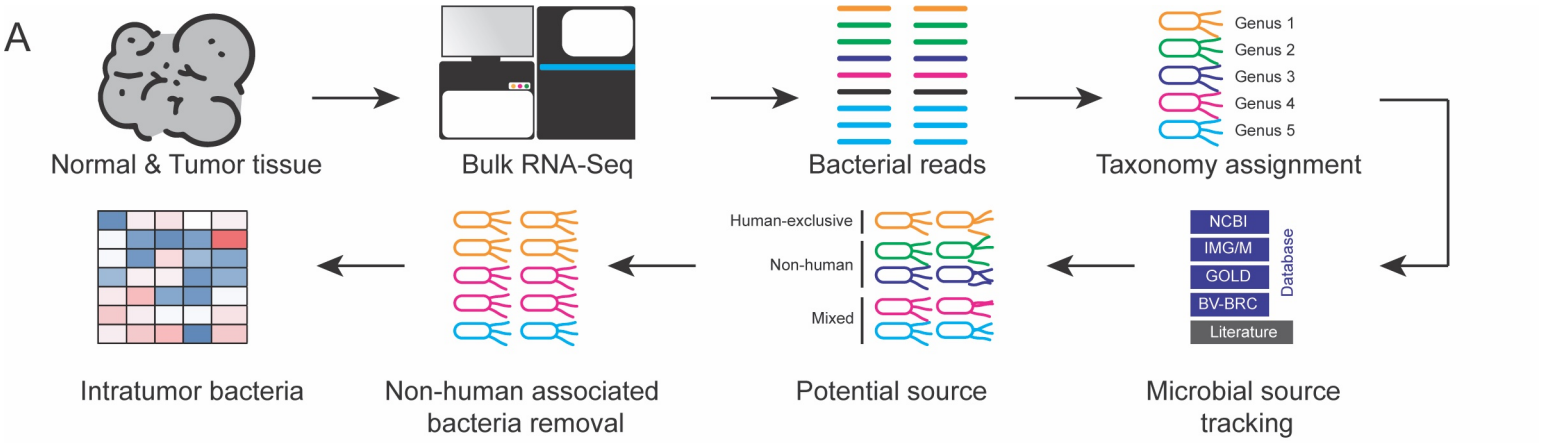
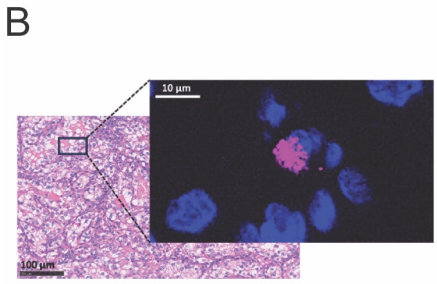
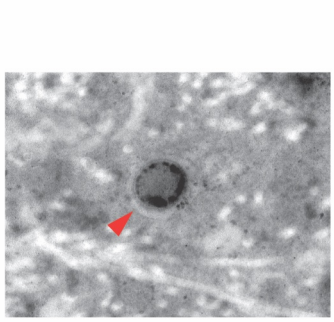
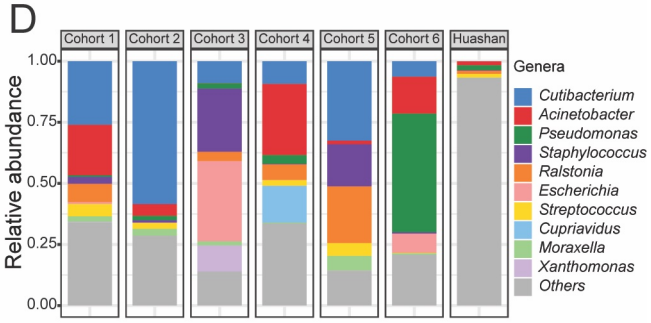


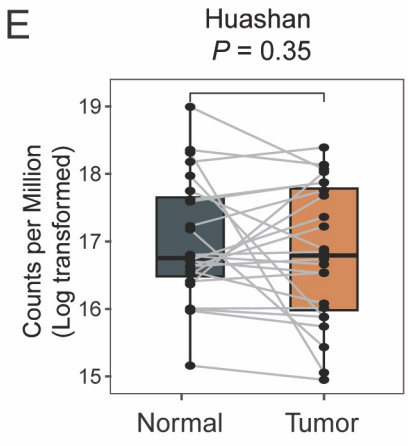
Figure 2

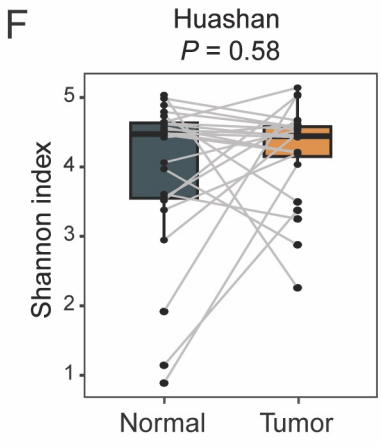












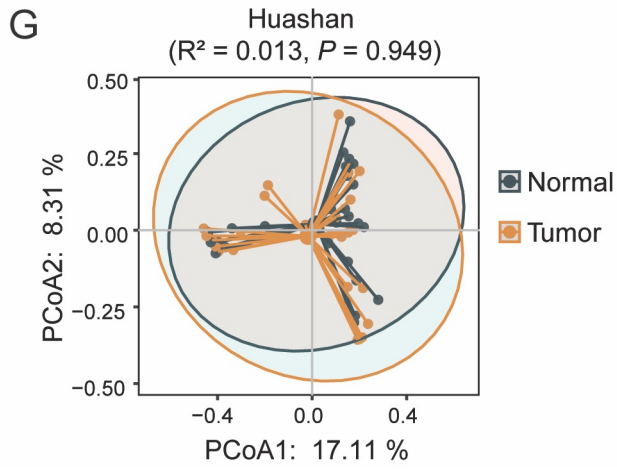


Figure 3

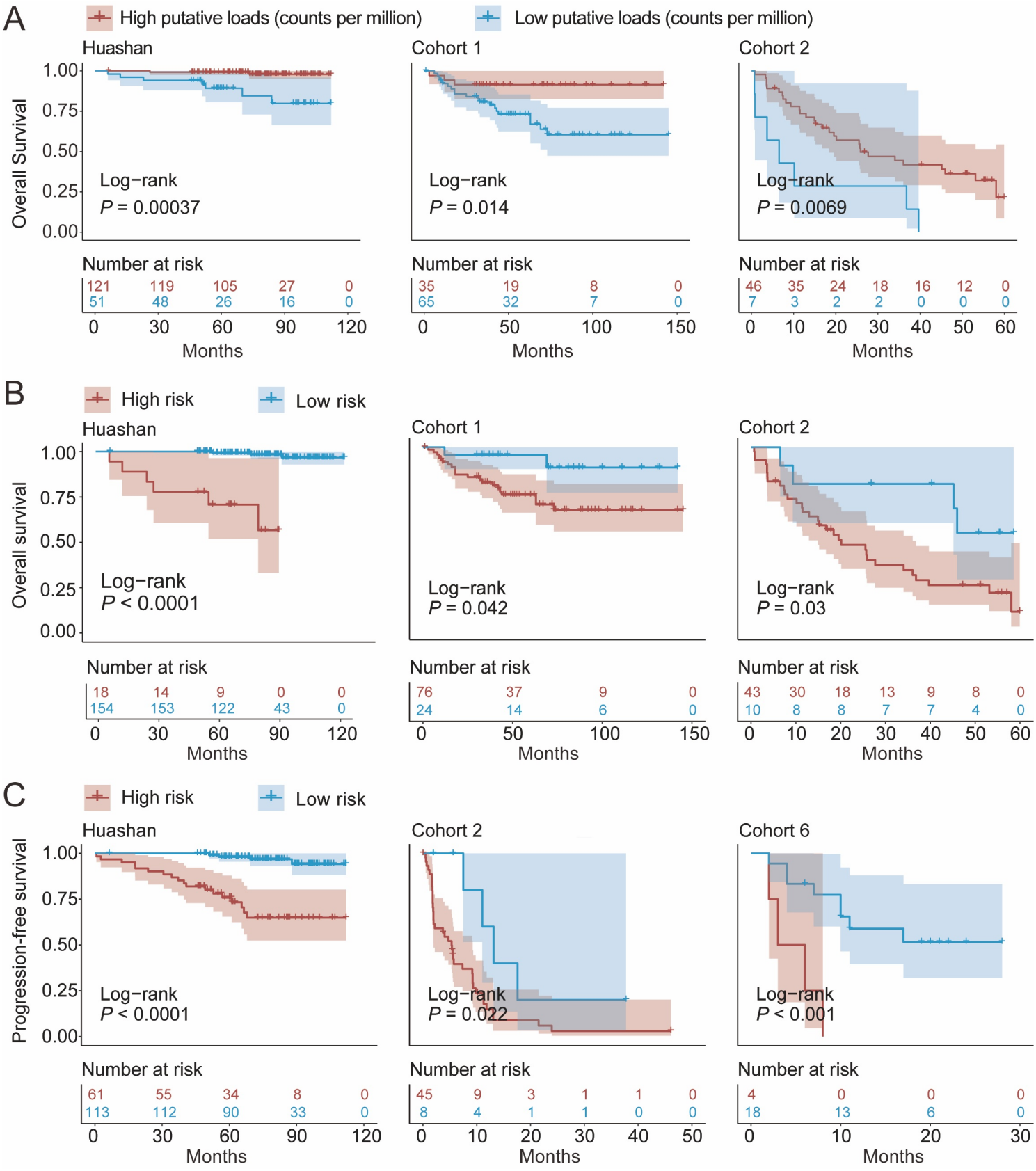


Figure 4