

Original Article

Integrated microbiome and transcriptomic analyses link *Paenibacillus* to aggressive features of prostate adenocarcinoma

Langxi Zhou¹, Yi Zhang²

¹South Pasadena High School, South Pasadena, California, USA; ²Department of Biomedical Sciences, Samuel Oschin Comprehensive Cancer Center, Cedars-Sinai Medical Center, 8700 Beverly Blvd, Los Angeles, California 90048, USA

Received July 15, 2026; Accepted September 13, 2026; Epub October 15, 2026; Published October 30, 2026

Abstract: Objective: To identify tumor-associated microbial genera associated with aggressive clinicopathologic features in prostate adenocarcinoma (PRAD) and to characterize host transcriptomic alterations associated with the leading microbial candidate. Methods: Genus-level microbial abundance profiles from the Bacteria in Cancer resource were integrated with clinicopathologic annotations and host transcriptomic data from TCGA-PRAD. Differential microbial abundance, clinicopathologic association, differential gene-expression and pathway enrichment analyses were performed. Results: The microbiome dataset included 1,617 genera across 493 tumors and 52 adjacent normal tissues. Sixty-seven genera were nominally differentially abundant, and 15 remained significant after false discovery rate correction. *Paenibacillus* showed the strongest tumor enrichment ($\log_2FC = 3.04$, $FDR = 2.54 \times 10^{-7}$) and the most consistent associations with Gleason-related variables and pathologic stage. Tumor-normal transcriptomic analysis identified 17,555 differentially expressed genes, including 1,215 with $FDR < 0.05$ and $|\log_2FC| > 1$. Comparison of *Paenibacillus*-high and *Paenibacillus*-low tumors identified 1,079 differentially expressed genes. MYC targets, oxidative phosphorylation and DNA repair were positively enriched, whereas epithelial-mesenchymal transition, inflammatory signaling, hypoxia and androgen-response programs were negatively enriched in the *Paenibacillus*-high comparison. Conclusion: *Paenibacillus* was the most strongly tumor-enriched genus and was consistently associated with aggressive clinicopathologic features and host transcriptomic alterations in PRAD.

Keywords: Prostate adenocarcinoma, tumor-associated microbiota, *Paenibacillus*, transcriptomics, clinicopathologic features

Introduction

Prostate cancer is one of the most frequently diagnosed malignancies among men worldwide and remains a leading cause of cancer-related mortality [1, 2]. The risk of prostate cancer is strongly associated with older age, family history and inherited genetic susceptibility, and Black race [3]. Although many patients with localized prostate adenocarcinoma (PRAD) have favorable outcomes following definitive treatment, a subset develops aggressive disease despite having similar clinicopathologic characteristics at diagnosis. Aggressive or invasive disease carries increased risks of extraprostatic extension, locoregional recurrence, lymph-node involvement, and distant metastasis

[3, 4]. High-risk localized or locally advanced disease is commonly treated with radical prostatectomy or radiotherapy combined with long-term androgen-deprivation therapy (ADT), whereas metastatic disease is primarily managed with ADT intensified by androgen-receptor pathway inhibitors, with chemotherapy used in selected patients [3, 4]. These treatments improve disease control and survival; however, once distant metastasis develops, the 5-year survival rate remains approximately 37% [3]. Established factors, including Gleason score, pathologic stage, prostate-specific antigen and surgical margin status, remain fundamental for clinical risk stratification but do not fully explain the biological heterogeneity underlying disease progression [2, 5]. A better understanding

of the molecular and microenvironmental factors associated with aggressive PRAD is therefore important for improving disease characterization and identifying biological processes involved in tumor progression.

Recent advances in sequencing technologies have revealed tumor-associated microbial signals across several solid tumors. Tumor-associated microbiota may be related to cancer biology through interactions with immune regulation, inflammatory responses, cellular metabolism and the tumor microenvironment [6-10]. These findings indicate that microbial signals represent an additional dimension of cancer biology beyond tumor-intrinsic genomic and transcriptomic alterations.

Compared with the gastrointestinal tract, the prostate is generally considered a relatively low-biomass tissue, making characterization of its microbial communities technically challenging [11, 12]. Nevertheless, accumulating evidence suggests that microorganisms may contribute to multiple prostatic diseases. Chronic bacterial infection has long been implicated in prostatitis, and persistent inflammation has been proposed as a contributing factor in benign prostatic hyperplasia and prostate cancer [13, 14]. Culture-independent sequencing approaches have subsequently detected microbial communities in prostate tissue, urine and prostate cancer cohorts [15-20]. However, the clinicopathologic significance of individual microbial genera and their relationships with host transcriptomic alterations remain incompletely understood.

This study integrated genus-level microbial abundance profiles from the Bacteria in Cancer resource with clinicopathologic annotations and host transcriptomic data from the TCGA-PRAD cohort [5, 21]. We first compared tumor and normal prostate tissues to identify differentially abundant genera and then evaluated their associations with Gleason-related variables and pathologic stage. Host transcriptomic analyses were used to characterize the molecular context associated with the leading microbial candidate. This approach was intended to identify tumor-associated microbial genera linked to aggressive clinicopathologic features and to guide future biological investigation.

Materials and methods

Data sources

Genus-level microbial abundance profiles for PRAD were retrieved from the Bacteria in Cancer (BIC) resource [21]. Clinicopathologic annotations and host transcriptomic data were obtained from The Cancer Genome Atlas prostate adenocarcinoma cohort (TCGA-PRAD) [5]. The microbiome matrices comprised 493 tumor samples and 52 adjacent normal prostate samples across 1,617 bacterial genera. The 52 adjacent normal samples had corresponding tumor samples from the same TCGA participants. The host gene-expression matrices comprised 501 tumor samples and 52 normal prostate samples across 59,427 genes.

Microbial, clinicopathologic and transcriptomic datasets were harmonized using TCGA sample identifiers. Samples lacking variables required for a specific analysis were excluded from that analysis. Samples were included when the data required for the corresponding analysis were available: microbial abundance data for tumor-normal microbial comparisons, clinicopathologic annotations for clinical association analyses, and matched microbial and transcriptomic data for the *Paenibacillus*-stratified transcriptomic analysis. Samples with missing or unavailable data for the variable(s) required for a given analysis were excluded only from that specific analysis. For binary Gleason comparisons, tumors with Gleason scores ≤ 7 were classified as low/intermediate grade and those with scores ≥ 8 as high grade. The Gleason score is determined by summing the grades of the primary and secondary histologic growth patterns, with higher scores indicating less differentiated and more aggressive disease. In the present study, Gleason scores were obtained directly from the TCGA-PRAD clinicopathologic annotations rather than reassessed by the authors.

Differential microbial abundance analysis

Microbial abundance was compared between PRAD tumor and normal prostate samples at the genus level using the Wilcoxon rank-sum test. Mean and median abundance, $\log_2$ fold change, nominal *P* values and false discovery rate (FDR)-adjusted *P* values were calculated

for each genus. Multiple-testing correction was performed using the Benjamini-Hochberg procedure [22], and genera with $FDR < 0.05$ were considered significantly differentially abundant. Genera with nominal $P < 0.05$ but $FDR \geq 0.05$ were considered exploratory and are presented in [Supplementary Figure 1](#) and [Supplementary Table 1](#). Volcano plots and heatmaps were used to display tumor-normal microbial differences.

Clinicopathologic association analyses

Genera that remained significant after FDR correction in the tumor-normal comparison were evaluated for associations with clinicopathologic features. The clinicopathologic association analysis included three components: binary Gleason grade grouping (Gleason score ≤ 7 vs ≥ 8), continuous Gleason score, and pathologic stage. Microbial abundance was compared between low/intermediate-grade and high-grade tumors using the Wilcoxon rank-sum test. Associations with continuous Gleason score were evaluated using Spearman rank correlation, and differences across pathologic stage groups were assessed using the Kruskal-Wallis test. The leading microbial candidate for transcriptomic analysis was prioritized on the basis of tumor enrichment and consistent associations across Gleason-related and stage-related variables.

Differential gene-expression analysis

Normalized host gene-expression profiles were analyzed using the limma package in R [23]. Tumor-normal differential expression was evaluated across 501 PRAD tumors and 52 normal prostate tissues. Genes with $FDR < 0.05$ were considered statistically significant, and a more stringent threshold of $FDR < 0.05$ and $|\log_2 FC| > 1$ was used to summarize the principal tumor-normal expression changes and construct the corresponding volcano plot and heatmap.

For the microbial-stratified analysis, tumors with matched microbial and transcriptomic data were divided into *Paenibacillus*-high and *Paenibacillus*-low groups according to the cohort median abundance. Differential expression between the two groups was analyzed using limma. Genes with $FDR < 0.05$ and $|\log_2 FC| > 0.5$ were considered significantly associated with *Paenibacillus* abundance.

Positive $\log_2 FC$ values indicated higher expression in *Paenibacillus*-high tumors, whereas negative values indicated higher expression in *Paenibacillus*-low tumors.

Functional enrichment analysis

Gene set enrichment analysis was applied to ranked gene-expression results from both the tumor-normal and *Paenibacillus*-stratified comparisons. Hallmark gene sets were obtained from the Molecular Signatures Database, and enrichment was evaluated using established GSEA methods [24, 25]. Gene Ontology biological-process and Kyoto Encyclopedia of Genes and Genomes pathway analyses were performed for *Paenibacillus*-associated genes using clusterProfiler [26-28]. Benjamini-Hochberg-adjusted P values < 0.05 were considered significant. Normalized enrichment scores were used to determine enrichment direction, and results were displayed using dot plots and representative enrichment plots.

Statistical analysis and data visualization

Statistical analyses and data visualization were performed in R version 4.3.2. Microbial abundance is reported as mean and median values, detection frequency as number and percentage of samples, differential abundance and gene-expression effects as $\log_2$ fold change ($\log_2 FC$), and clinicopathologic correlations as Spearman correlation coefficients (ρ). Gene set enrichment results are reported as normalized enrichment scores (NES), with nominal P values and Benjamini-Hochberg-adjusted P values or false discovery rates (FDR) reported as appropriate; GO and KEGG enrichment results are additionally summarized by gene ratios and gene counts. All statistical tests were two-sided. P values < 0.05 were considered significant for exploratory clinicopathologic comparisons, whereas $FDR < 0.05$ was required for high-dimensional microbial, transcriptomic and functional enrichment analyses.

Ethics statement

This study used publicly available, de-identified data from TCGA-PRAD and the Bacteria in Cancer resource. No new participants or biological specimens were enrolled or collected, and no identifiable private information was accessed. Additional ethics committee approval

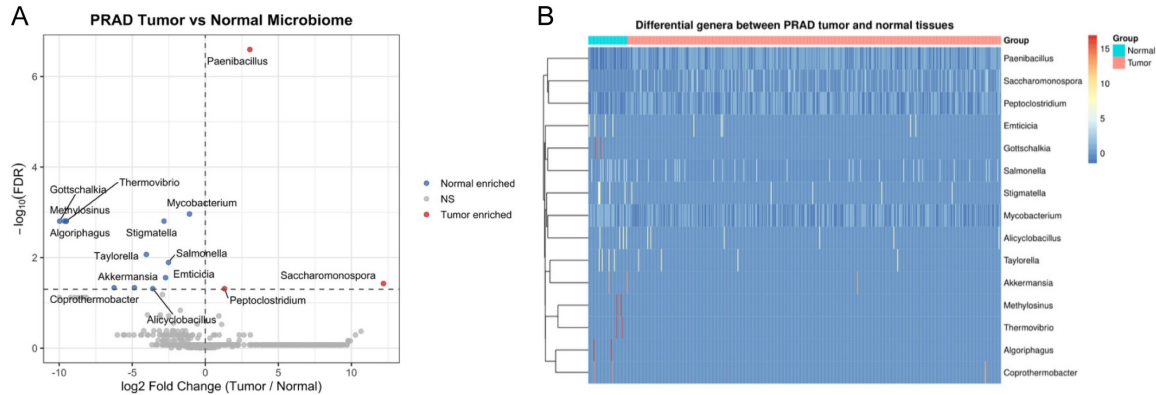


Figure 1. Prostate adenocarcinoma (PRAD) tumor-normal microbial differences. A. Differential abundance volcano plot comparing tumor and normal prostate tissues. B. Heatmap of significant tumor-normal microbial differences.

Table 1. Differentially abundant genera between PRAD tumor and normal prostate tissues (FDR < 0.05)

Bacteria	Normal mean	Tumor mean	log ₂ FC	P value	FDR
<i>Paenibacillus</i>	0.0117	0.0965	3.04	3.09×10^{-10}	2.54×10^{-7}
<i>Mycobacterium</i>	0.0305	0.0144	-1.09	2.65×10^{-6}	0.00109
<i>Algoriphagus</i>	7.95×10^{-4}	0	-9.64	1.34×10^{-5}	0.00157
<i>Gottschalkia</i>	0.001	0	-9.97	1.34×10^{-5}	0.00157
<i>Methylosinus</i>	7.40×10^{-4}	0	-9.53	1.34×10^{-5}	0.00157
<i>Stigmatella</i>	0.00157	2.19×10^{-4}	-2.83	1.19×10^{-5}	0.00157
<i>Thermovibrio</i>	7.33×10^{-4}	0	-9.52	1.34×10^{-5}	0.00157
<i>Taylorella</i>	0.00232	1.41×10^{-4}	-4.03	8.29×10^{-5}	0.00851
<i>Salmonella</i>	0.00539	9.39×10^{-4}	-2.52	1.40×10^{-4}	0.0128
<i>Emticicia</i>	0.0013	1.95×10^{-4}	-2.72	3.38×10^{-4}	0.0278
<i>Saccharomonospora</i>	0	0.00468	12.2	5.01×10^{-4}	0.0374
<i>Akkermansia</i>	0.00133	4.51×10^{-5}	-4.85	7.31×10^{-4}	0.0462
<i>Coprothermobacter</i>	7.95×10^{-4}	9.48×10^{-6}	-6.25	7.31×10^{-4}	0.0462
<i>Alicyclobacillus</i>	0.00184	1.51×10^{-4}	-3.6	8.27×10^{-4}	0.0485
<i>Peptoclostridium</i>	0.00716	0.0179	1.32	8.89×10^{-4}	0.0487

Note: PRAD, prostate adenocarcinoma; log₂FC, log₂ fold change; FDR, false discovery rate.

and informed consent were therefore not required for this study.

Results

Tumor-normal microbial abundance differences in PRAD

The genus-level microbiome dataset comprised 493 PRAD tumor samples and 52 adjacent normal prostate tissue samples, with abundance profiles available for 1,617 bacterial genera. Differential abundance analysis identified 67 genera with *P* values < 0.05, of which 15 remained significant after Benjamini-Hochberg correction (FDR < 0.05) (**Figure 1**; **Table 1**).

Among the FDR-significant genera, *Paenibacillus* exhibited the strongest tumor enrichment. Its mean abundance increased from 0.0117 in normal prostate tissues to 0.0965 in PRAD tumors, corresponding to a log₂FC of 3.04 (*P* = 3.09×10^{-10} , FDR = 2.54×10^{-7}). Median abundance was 0 in normal tissues and 0.0339 in tumors. A detectable *Paenibacillus* signal was present in 358 of 493 tumors (72.6%) and 16 of 52 normal tissues (30.8%). *Saccharomonospora* and *Peptoclostridium* were also enriched in tumors, whereas the remaining FDR-significant genera showed lower abundance in tumors than in normal tissues. Complete results are presented in **Table**

Paenibacillus in prostate adenocarcinoma

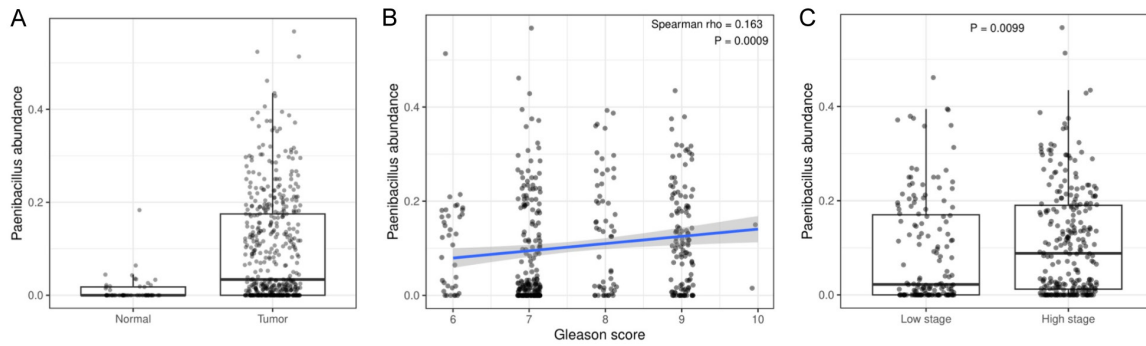


Figure 2. Clinical relevance of *Paenibacillus* in PRAD. A. *Paenibacillus* abundance in tumor and adjacent normal prostate tissues. B. Relationship between *Paenibacillus* abundance and Gleason score. C. *Paenibacillus* abundance across pathologic stage categories.

Table 2. Clinical association summary for the 15 significant PRAD genera

Bacteria	TN FDR	GS binary P	Gleason rho	Gleason P	Stage P
<i>Paenibacillus</i>	2.54×10^{-7}	1.89×10^{-5}	0.163	9.06×10^{-4}	0.00985
<i>Mycobacterium</i>	0.00109	0.749	0.0482	0.329	0.971
<i>Algoriphagus</i>	0.00157	NA	NA	NA	NA
<i>Gottschalkia</i>	0.00157	NA	NA	NA	NA
<i>Methylosinus</i>	0.00157	NA	NA	NA	NA
<i>Stigmatella</i>	0.00157	0.916	0.0166	0.737	0.874
<i>Thermovibrio</i>	0.00157	NA	NA	NA	NA
<i>Taylorella</i>	0.00851	0.398	0.057	0.248	0.914
<i>Salmonella</i>	0.0128	0.659	0.0296	0.548	0.832
<i>Emticia</i>	0.0278	0.0194	0.0916	0.0628	0.128
<i>Saccharomonospora</i>	0.0374	0.149	0.0771	0.118	0.501
<i>Akkermansia</i>	0.0462	NA	NA	NA	NA
<i>Coprothermobacter</i>	0.0462	0.394	-0.031	0.529	0.189
<i>Alicyclobacillus</i>	0.0485	0.766	0.0336	0.496	0.633
<i>Peptoclostridium</i>	0.0487	0.159	0.0334	0.499	0.24

Note: PRAD, prostate adenocarcinoma; TN, tumor-normal; FDR, false discovery rate; GS, Gleason score; p, Spearman correlation coefficient; NA, not available. NA indicates that the corresponding clinicopathologic association statistic could not be estimated because the genus showed no or insufficient variation in abundance among tumor samples with available clinicopathologic data. These genera were absent or extremely sparsely detected in the analyzable tumor samples.

1, and genera significant only at the nominal level are shown in [Supplementary Figure 1](#) and [Supplementary Table 1](#).

Paenibacillus shows the most consistent clinicopathologic associations

The 15 genera that remained significant after FDR correction were examined for associations with clinicopathologic characteristics. *Paenibacillus* showed the most consistent pattern across tumor enrichment, Gleason-related variables and pathologic stage. Its abundance differed between tumors with Gleason scores

≤ 7 and those with scores ≥ 8 ($P = 1.89 \times 10^{-5}$). In the continuous analysis, *Paenibacillus* abundance was positively correlated with Gleason score, and the magnitude of the correlation was modest (Spearman $\rho = 0.163$, $P = 9.06 \times 10^{-4}$). Its abundance also differed across pathologic stage groups ($P = 9.85 \times 10^{-3}$) (**Figure 2; Table 2**).

Emticia showed a nominal association with binary Gleason grouping ($P = 0.0194$) but was not significantly associated with continuous Gleason score or pathologic stage. The remaining FDR-significant genera did not show a

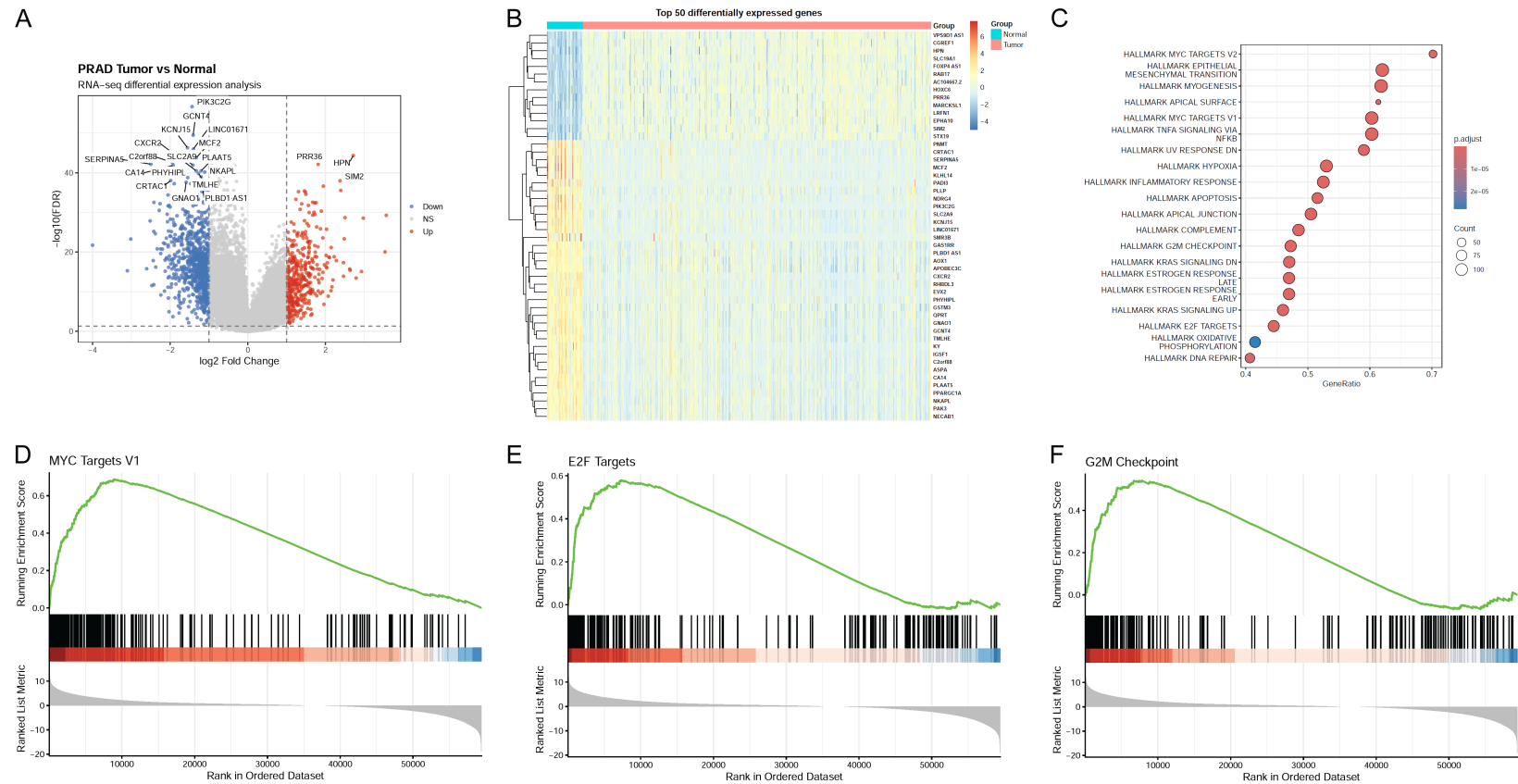


Figure 3. PRAD tumor-normal transcriptomic programs. A. Differential-expression volcano plot. B. Heatmap of tumor-normal differentially expressed genes. C. Hallmark enrichment dotplot. D. MYC targets V1 gene set enrichment analysis (GSEA) plot. E. E2F targets GSEA plot. F. G2M checkpoint GSEA plot.

Table 3. Top positively enriched Hallmark pathways in PRAD tumor-normal analysis

Description	setSize	NES	P value	Adjusted P	rank
MYC targets V1	199	2.74	1.00×10^{-10}	5.00×10^{-10}	9×10^3
MYC targets V2	57	2.57	1.00×10^{-10}	5.00×10^{-10}	7.39×10^3
E2F targets	200	2.30	1.00×10^{-10}	5.00×10^{-10}	7.42×10^3
G2M checkpoint	199	2.16	1.00×10^{-10}	5.00×10^{-10}	7.8×10^3
DNA repair	150	1.84	1.36×10^{-6}	3.57×10^{-6}	8.02×10^3
Unfolded protein response	112	1.74	1.08×10^{-4}	2.34×10^{-4}	7.17×10^3
Oxidative phosphorylation	200	1.65	1.11×10^{-5}	2.78×10^{-5}	9.22×10^3
Androgen response	100	1.64	3.90×10^{-4}	8.12×10^{-4}	4.99×10^3

Note: PRAD, prostate adenocarcinoma; NES, normalized enrichment score. Rank indicates the position in the ranked gene list at which the running enrichment score reaches its maximum deviation from zero.

similarly consistent clinicopathologic pattern. These findings supported selection of *Paenibacillus* for subsequent host transcriptomic analyses.

Tumor-normal transcriptomic changes characterize established PRAD biology

The host transcriptomic dataset included 501 PRAD tumor samples and 52 normal prostate tissue samples, with expression measurements available for 59,427 genes. Tumor-normal differential expression analysis identified 17,555 genes with $FDR < 0.05$. Application of the more stringent threshold of $FDR < 0.05$ and $|\log_2 FC| > 1$ identified 1,215 differentially expressed genes, including 390 genes with higher expression and 825 genes with lower expression in tumor tissues (**Figure 3A, 3B**). Genes with prominent tumor-associated increases included DLX1, PCA3, HPN and AMACR. Additional tumor-normal differential-expression results are provided in [Supplementary Figure 2](#) and [Supplementary Tables 2, 3, 4](#).

Hallmark gene set enrichment analysis demonstrated positive enrichment of MYC targets V1 (NES = 2.74, adjusted P = 5.00×10^{-10}), MYC targets V2 (NES = 2.57, adjusted P = 5.00×10^{-10}), E2F targets (NES = 2.30, adjusted P = 5.00×10^{-10}) and the G2M checkpoint (NES = 2.16, adjusted P = 5.00×10^{-10}). Additional positively enriched pathways included DNA repair, the unfolded protein response, oxidative phosphorylation and androgen response. Negative enrichment was observed for myogenesis, epithelial-mesenchymal transition, apical junction, hypoxia, TNF- α signaling, KRAS signaling, complement, inflammatory response and TGF- β signaling (**Figure 3C-F; Table 3**). The com-

plete tumor-normal Hallmark enrichment profile is provided in [Supplementary Figure 3](#) and [Supplementary Table 5](#).

Paenibacillus-stratified analysis identifies a distinct host transcriptional state

Comparison of *Paenibacillus*-high and *Paenibacillus*-low tumors identified 1,079 differentially expressed genes meeting the thresholds of $FDR < 0.05$ and $|\log_2 FC| > 0.5$. Among these genes, 567 had positive $\log_2 FC$ values and 512 had negative $\log_2 FC$ values in the *Paenibacillus*-high comparison (**Figure 4A, 4B**). Genes with prominent positive fold changes included SNORD104, MT-ATP8, SNORD14E, SNORD3A, FTH1, PAQR6 and GNG5. Genes with prominent negative fold changes included LTF, OLFM4, AOC1, NEFH, SCGB1A1, CD38 and CD177. Additional *Paenibacillus*-associated differential-expression results are provided in [Supplementary Figure 4](#) and [Supplementary Table 6](#).

Hallmark analysis identified three positively enriched pathways in *Paenibacillus*-high tumors: MYC targets V2 (NES = 1.75, adjusted P = 1.24×10^{-3}), oxidative phosphorylation (NES = 1.57, adjusted P = 4.69×10^{-4}) and DNA repair (NES = 1.39, adjusted P = 1.33×10^{-2}) (**Figure 4C-E; Table 4**). In contrast, epithelial-mesenchymal transition, KRAS signaling, TGF- β signaling, inflammatory response, complement, TNF- α signaling via NF- κ B, androgen response, apoptosis, IL2-STAT5 signaling, apical junction and hypoxia showed negative enrichment in the *Paenibacillus*-high comparison (**Table 5**). The complete *Paenibacillus*-stratified Hallmark enrichment profile is provided in [Supplementary Figure 5](#) and [Supplementary Table 7](#).

Paenibacillus in prostate adenocarcinoma

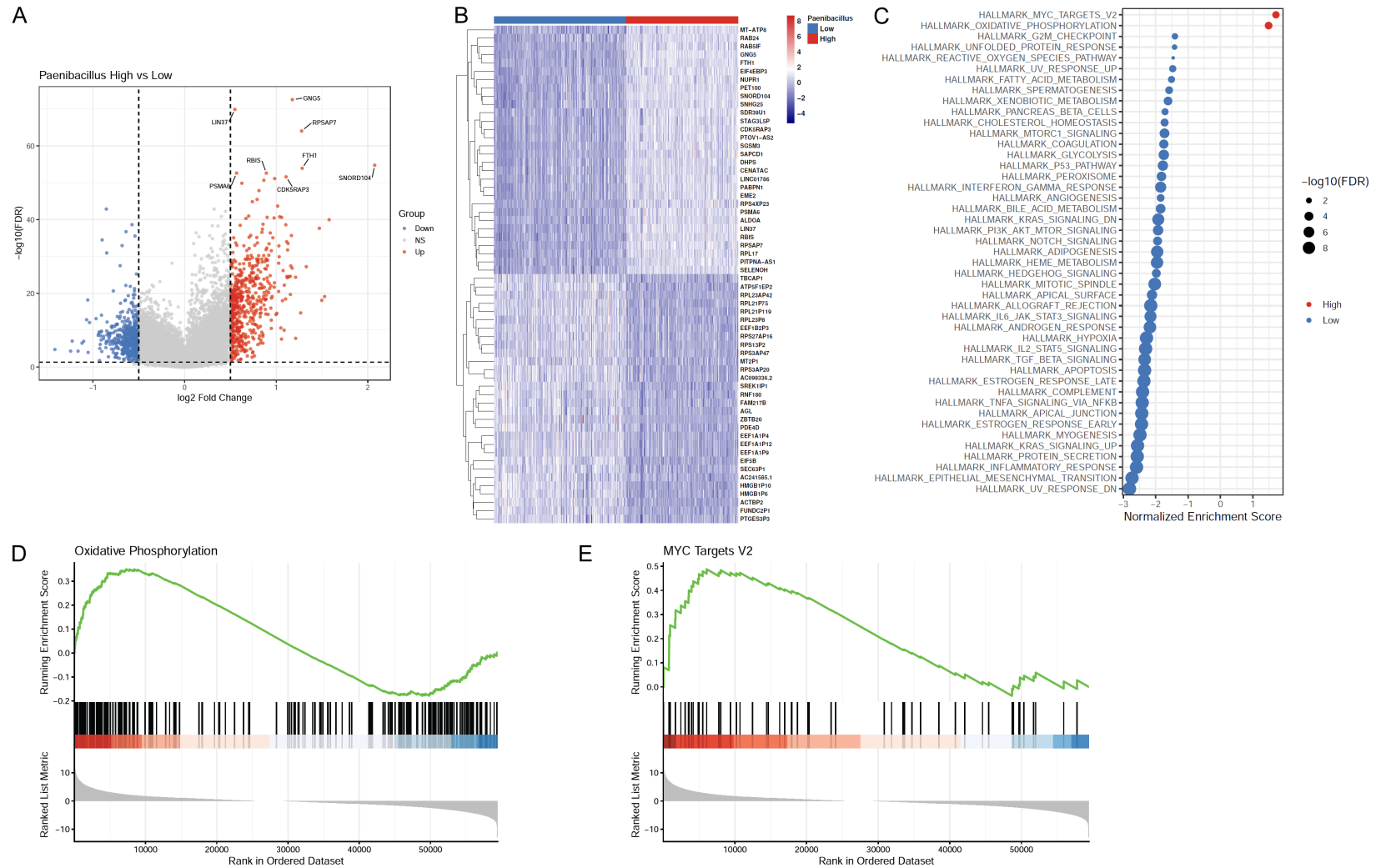


Figure 4. Host transcriptomic features associated with *Paenibacillus* abundance. A. Differential expression between *Paenibacillus*-high and *Paenibacillus*-low tumors. B. Heatmap of significant *Paenibacillus*-associated genes. C. Hallmark enrichment dotplot. D. Oxidative phosphorylation GSEA plot. E. MYC targets V2 GSEA plot.

Table 4. Top positively enriched Hallmark pathways in the *Paenibacillus*-high versus low analysis

Description	setSize	NES	P value	Adjusted P	rank
MYC targets V2	57	1.75	8.18×10^{-4}	0.00124	1.04×10^4
Oxidative phosphorylation	200	1.57	2.81×10^{-4}	4.69×10^{-4}	7.38×10^3
DNA repair	150	1.39	0.0112	0.0133	6.93×10^3

Note: NES, normalized enrichment score. Rank indicates the position in the ranked gene list at which the running enrichment score reaches its maximum deviation from zero.

Table 5. Top negatively enriched Hallmark pathways in the *Paenibacillus*-high versus low analysis

Description	setSize	NES	P value	Adjusted P	rank
UV response down	144	-2.83	1.00×10^{-10}	4.17×10^{-10}	1.01×10^4
Protein secretion	96	-2.66	1.00×10^{-10}	4.17×10^{-10}	8.88×10^3
Epithelial-mesenchymal transition	200	-2.46	1.00×10^{-10}	4.17×10^{-10}	1.31×10^4
KRAS signaling up	200	-2.44	1.00×10^{-10}	4.17×10^{-10}	1.13×10^4
TGF-beta signaling	54	-2.36	2.13×10^{-9}	7.61×10^{-9}	9.79×10^3
Estrogen response early	198	-2.35	1.00×10^{-10}	4.17×10^{-10}	1.46×10^4

Note: NES, normalized enrichment score. Rank indicates the position in the ranked gene list at which the running enrichment score reaches its maximum deviation from zero.

Functional enrichment implicates tissue architecture and microenvironment-related processes

Gene Ontology biological-process analysis of *Paenibacillus*-associated differentially expressed genes identified enrichment of muscle contraction (adjusted $P = 5.94 \times 10^{-8}$), muscle system processes (adjusted $P = 6.45 \times 10^{-7}$), muscle tissue development (adjusted $P = 1.06 \times 10^{-4}$) and smooth muscle contraction (adjusted $P = 2.64 \times 10^{-4}$). Additional enriched terms included tissue migration, responses to decreased oxygen levels and hypoxia, vascular processes, prostate gland morphogenesis, cell-substrate adhesion, mesenchyme development and extracellular matrix organization.

KEGG analysis similarly highlighted cytoskeleton in muscle cells, focal adhesion, vascular smooth muscle contraction and related cell-matrix interaction pathways (**Figure 5**). Detailed Gene Ontology and KEGG enrichment results are provided in [Supplementary Tables 8](#) and [9](#), respectively. These enrichment patterns showed that the host transcriptional differences associated with *Paenibacillus* abundance included processes related to cytoskeletal organization, smooth muscle biology, adhesion, tissue architecture and the local tumor microenvironment.

Discussion

This study integrated genus-level microbial abundance profiles with clinicopathologic and transcriptomic data from PRAD. Among the 1,617 genera examined, *Paenibacillus* showed the strongest tumor enrichment after correction for multiple comparisons. Its abundance was also associated with Gleason-related variables and pathologic stage. Transcriptomic analysis showed that tumors with higher *Paenibacillus* abundance had differences in MYC-related activity, oxidative phosphorylation, DNA repair, inflammatory signaling, androgen response and pathways related to tissue structure and cell-matrix interactions. These results identify *Paenibacillus* as the genus with the most consistent microbial, clinical and transcriptomic associations in this cohort.

Microorganisms have been investigated in several prostatic diseases. Bacterial infection is a recognized cause of some forms of prostatitis, while chronic inflammation has also been proposed to contribute to benign prostatic hyperplasia and prostate carcinogenesis [13, 14]. Sequencing-based studies have detected microbial signals in prostate tissue, urine and prostate cancer cohorts [15-20]. However, the reported taxa have differed considerably among studies. The prominence of *Paenibacillus* in our cohort should therefore be interpreted in

Paenibacillus in prostate adenocarcinoma

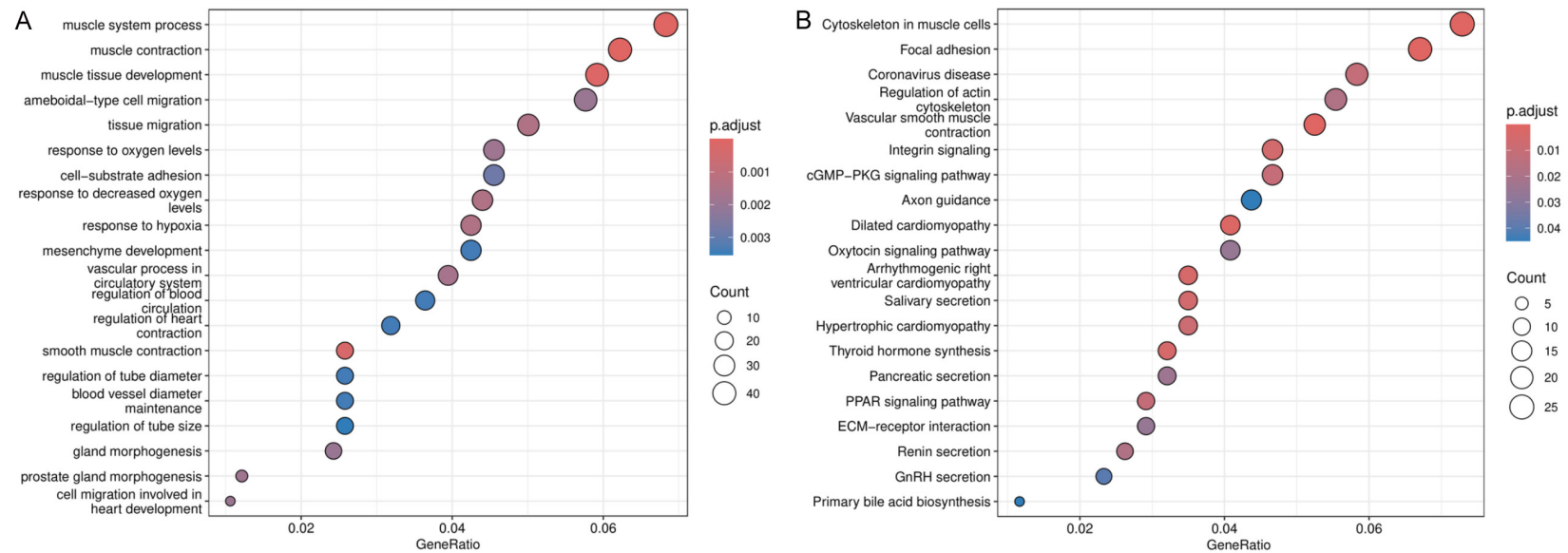


Figure 5. Functional enrichment of *Paenibacillus*-associated genes. A. Gene Ontology (GO) biological-process enrichment. B. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment.

the context of the substantial between-study heterogeneity in microbial signatures reported in previous prostate cancer studies [15-20]. Differences in specimen type, patient population, sequencing platform, microbial reference database, sample processing and contamination-control procedures may contribute to this variation [11, 12]. Experimental evidence that microbial metabolism can influence androgen availability and endocrine treatment resistance further supports the relevance of host-microbe interactions in prostate cancer [29].

Paenibacillus showed the strongest tumor enrichment in the present analysis and was more frequently detected in tumor tissues than in normal prostate tissues. Direct evidence linking this genus to cancer biology remains limited. The present study examined its relationships with clinicopathologic variables and host gene expression.

Higher *Paenibacillus* abundance was associated with Gleason score when Gleason grade was analyzed as both a categorical and a continuous variable, and it also differed across pathologic stage groups. The continuous correlation was more consistent than that observed for the other FDR-significant genera. This consistency across tumor enrichment and multiple clinicopathologic measures supported selection of *Paenibacillus* for downstream transcriptomic analysis.

The tumor-normal transcriptomic analysis recovered established features of prostate cancer, including enrichment of MYC targets, E2F targets, the G2M checkpoint, DNA repair, oxidative phosphorylation and androgen-response pathways [5, 30-32]. This provided a biological reference for the microbial-stratified analysis. When tumors were divided according to *Paenibacillus* abundance, 1,079 genes differed between the high- and low-abundance groups. Pathway analysis linked the *Paenibacillus*-high group with MYC targets V2, oxidative phosphorylation and DNA repair. These transcriptional patterns are broadly consistent with previously reported roles of MYC signaling, altered energy metabolism, DNA-repair pathways and androgen signaling in prostate cancer [30-32], while the additional inflammatory and microenvironment-related associations observed here may reflect biological features associated with *Paenibacillus* abundance.

Other associated pathways included epithelial-mesenchymal transition, TGF- β signaling, inflammatory response, TNF- α signaling through NF- κ B, hypoxia, complement and androgen response. At the individual-gene level, several transcripts with prominent expression differences have recognized links to cancer-related biology, although the evidence is stronger for some genes than for others. *SNORD104*, *SNORD14E* and *SNORD3A* are small nucleolar RNAs involved in ribosomal RNA processing; dysregulated snoRNA expression has been broadly implicated in tumor biology, but the specific roles of these three snoRNAs in PRAD remain largely undefined [33]. *MT-ATP8* encodes a mitochondrial ATP synthase subunit, and its increased expression is consistent with the oxidative phosphorylation signal observed in the pathway analysis. *FTH1* regulates intracellular iron storage and oxidative stress, processes frequently altered in cancer [34]. *PAQR6* and *GNG5* participate in membrane steroid and G-protein-associated signaling, respectively, although their relevance to PRAD requires further study. Among the transcripts with lower expression, *LTF*, *CD38* and *CD177* are related to antimicrobial defense, immune metabolism and myeloid-cell states; *OLFM4* has established roles in epithelial stem-cell and cancer biology, whereas *AOC1* connects histamine and polyamine metabolism with pathways relevant to tumor growth [35-37]. *NEFH* and *SCGB1A1* may be more indicative of differences in cellular differentiation or tissue composition.

GO and KEGG analyses identified terms related to smooth muscle contraction, actin-cytoskeleton regulation, focal adhesion, extracellular matrix organization, tissue migration, vascular processes and responses to hypoxia. The prostate contains epithelial, stromal, vascular and smooth muscle components, and differences in these pathways may reflect variation in tissue composition or local microenvironmental features. Focal adhesion and cytoskeletal pathways may also indicate differences in cell-matrix interaction and structural organization. These analyses provide biological context for the *Paenibacillus*-associated transcriptional pattern.

Limitations

This study has several limitations. First, the analyses were based on publicly available

TCGA-PRAD and BIC datasets, which limited our ability to control sample collection, processing, and other pre-analytical factors. Second, because prostate tissue has low microbial biomass, microbial signals may be particularly susceptible to reagent or laboratory contamination, and contamination cannot be fully excluded in this retrospective analysis [11, 12]. Third, unmeasured or incompletely captured clinical factors may confound the observed associations between microbial abundance and clinicopathologic features. Fourth, the findings were not validated in an independent patient cohort or by experimental studies. Finally, the observational nature of these analyses precludes conclusions regarding whether *Paenibacillus* contributes causally to aggressive prostate cancer or instead reflects other biological or clinical differences between tumors. Future studies should validate these findings in independent cohorts using rigorous contamination-controlled microbial profiling and determine through experimental studies whether *Paenibacillus* directly influences prostate cancer progression or treatment response.

Conclusion

Integrated microbiome, clinicopathologic and transcriptomic analyses identified *Paenibacillus* as the most strongly tumor-enriched genus and the genus with the most consistent associations with Gleason-related variables and pathologic stage in PRAD. Higher *Paenibacillus* abundance was associated with MYC targets, oxidative phosphorylation, DNA repair and several inflammatory, hypoxic, stromal and androgen-response pathways. Functional enrichment additionally implicated cytoskeletal organization, focal adhesion, smooth muscle biology and tissue architecture. These findings provide new direction for investigating mechanisms of aggressive PRAD and support further investigation of *Paenibacillus* in prostate cancer.

Disclosure of conflict of interest

None.

Address correspondence to: Dr. Yi Zhang, Department of Biomedical Sciences, Samuel Oschin Comprehensive Cancer Center, Cedars-Sinai Medical Center, 8700 Beverly Blvd, Los Angeles, California 90048, USA. ORCID: 0000-0001-8151-9397; E-mail: Yi.Zhang@csmc.edu

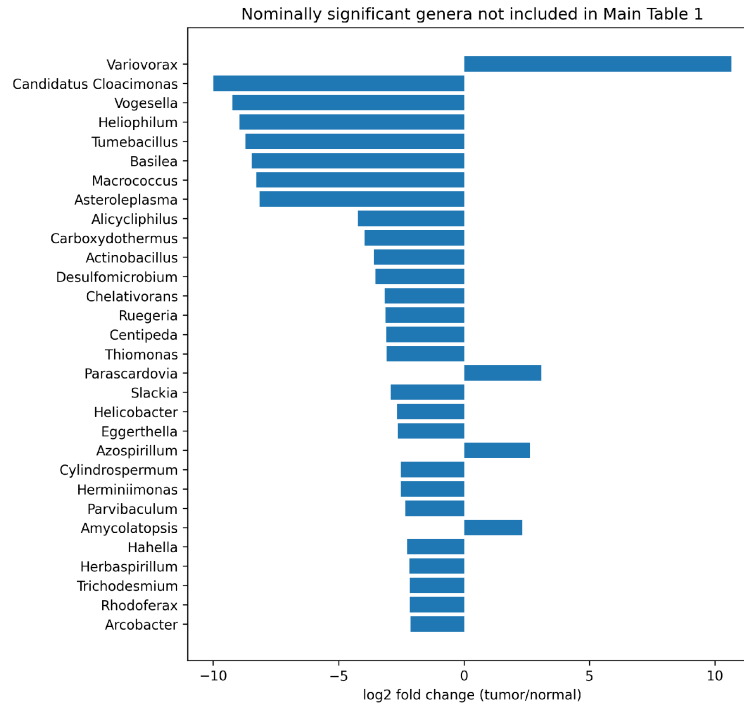
References

- [1] Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I and Jemal A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2024; 74: 229-263.
- [2] Siegel RL, Kratzer TB, Giaquinto AN, Sung H and Jemal A. Cancer statistics, 2025. *CA Cancer J Clin* 2025; 75: 10-45.
- [3] Raychaudhuri R, Lin DW and Montgomery RB. Prostate cancer: a review. *JAMA* 2025; 333: 1433-1446.
- [4] Rooney MK, El Kouzi Z, Kouzy R, Mohamad O and Hoffman KE. Debate 1: high-risk localized prostate cancer: why combination hormone therapy and radiotherapy is the optimal treatment strategy for most men. *Semin Radiat Oncol* 2025; 35: 393-401.
- [5] The Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. *Cell* 2015; 163: 1011-1025.
- [6] de Visser KE and Joyce JA. The evolving tumor microenvironment: from cancer initiation to metastatic outgrowth. *Cancer Cell* 2023; 41: 374-403.
- [7] Nejman D, Livyatan I, Fuks G, Gavert N, Zwang Y, Geller LT, Rotter-Maskowitz A, Weiser R, Mal-lal G, Gigi E, Meltser A, Douglas GM, Kamer I, Gopalakrishnan V, Dadosh T, Levin-Zaidman S, Avnet S, Atlan T, Cooper ZA, Arora R, Cogdill AP, Khan MAW, Ologun G, Bussi Y, Weinberger A, Lotan-Pompan M, Golani O, Perry G, Rokah M, Bahar-Shany K, Rozeman EA, Blank CU, Ronai A, Shaoul R, Amit A, Dorfman T, Kremer R, Cohen ZR, Harnof S, Siegal T, Yehuda-Shnaidman E, Gal-Yam EN, Shapira H, Baldini N, Langille MGI, Ben-Nun A, Kaufman B, Nissan A, Golan T, Dadiani M, Levanon K, Bar J, Yust-Katz S, Barshack I, Peeper DS, Raz DJ, Segal E, Wargo JA, Sandbank J, Shental N and Straussman R. The human tumor microbiome is composed of tumor type-specific intracellular bacteria. *Science* 2020; 368: 973-980.
- [8] Sepich-Poore GD, Zitvogel L, Straussman R, Hasty J, Wargo JA and Knight R. The microbiome and human cancer. *Science* 2021; 371: eabc4552.
- [9] Park EM, Chelvanambi M, Bhutiani N, Kroemer G, Zitvogel L and Wargo JA. Targeting the gut and tumor microbiota in cancer. *Nat Med* 2022; 28: 690-703.
- [10] Battaglia TW, Mimpfen IL, Traets JHH, van Hoeck A, Zevenijn LJ, Geurts BS, de Wit GF, Noë M, Hoffland I, Vos JL, Cornelissen S, Alkemade M, Broeks A, Zuur CL, Cuppen E, Wessels L, van de Haar J and Voest E. A pan-cancer analysis of the microbiome in metastatic cancer. *Cell* 2024; 187: 2324-2335, e19.

- [11] Eisenhofer R, Minich JJ, Marotz C, Cooper A, Knight R and Weyrich LS. Contamination in low microbial biomass microbiome studies: issues and recommendations. *Trends Microbiol* 2019; 27: 105-117.
- [12] Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ and Walker AW. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* 2014; 12: 87.
- [13] Sfanos KS, Yegnasubramanian S, Nelson WG and De Marzo AM. The inflammatory microenvironment and microbiome in prostate cancer development. *Nat Rev Urol* 2018; 15: 11-24.
- [14] De Marzo AM, Platz EA, Sutcliffe S, Xu J, Grönberg H, Drake CG, Nakai Y, Isaacs WB and Nelson WG. Inflammation in prostate carcinogenesis. *Nat Rev Cancer* 2007; 7: 256-269.
- [15] Cavarretta I, Ferrarese R, Cazzaniga W, Saita D, Lucianò R, Ceresola ER, Locatelli I, Visconti L, Lavorgna G, Briganti A, Nebuloni M, Doglioni C, Clementi M, Montorsi F, Canducci F and Salonia A. The microbiome of the prostate tumor microenvironment. *Eur Urol* 2017; 72: 625-631.
- [16] Yow MA, Tabrizi SN, Severi G, Bolton DM, Pedersen J; Australian Prostate Cancer BioResource; Giles GG and Southey MC. Characterisation of microbial communities within aggressive prostate cancer tissues. *Infect Agent Cancer* 2017; 12: 4.
- [17] Feng Y, Ramnarine VR, Bell R, Volik S, Davicioni E, Hayes VM, Ren S and Collins CC. Metagenomic and metatranscriptomic analysis of human prostate microbiota from patients with prostate cancer. *BMC Genomics* 2019; 20: 146.
- [18] Banerjee S, Alwine JC, Wei Z, Tian T, Shih N, Sperling C, Guzzo T, Feldman MD and Robertson ES. Microbiome signatures in prostate cancer. *Carcinogenesis* 2019; 40: 749-764.
- [19] Ma J, Gnanasekar A, Lee A, Li WT, Haas M, Wang-Rodriguez J, Chang EY, Rajasekaran M and Ongkeko WM. Influence of intratumor microbiome on clinical outcome and immune processes in prostate cancer. *Cancers (Basel)* 2020; 12: 2524.
- [20] Salachan PV, Rasmussen M, Fredsøe J, Ulhøi B, Borre M and Sørensen KD. Microbiota of the prostate tumor environment investigated by whole-transcriptome profiling. *Genome Med* 2022; 14: 9.
- [21] Chen KP, Hsu CL, Oyang YJ, Huang HC and Juan HF. BIC: a database for the transcriptional landscape of bacteria in cancer. *Nucleic Acids Res* 2023; 51: D1205-D1211.
- [22] Benjamini Y and Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Methodol* 1995; 57: 289-300.
- [23] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W and Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015; 43: e47.
- [24] Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP and Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst* 2015; 1: 417-425.
- [25] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES and Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005; 102: 15545-15550.
- [26] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM and Sherlock G. Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; 25: 25-29.
- [27] Kanehisa M, Furumichi M, Tanabe M, Sato Y and Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 2017; 45: D353-D361.
- [28] Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, Fu X, Liu S, Bo X and Yu G. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation (Camb)* 2021; 2: 100141.
- [29] Pernigoni N, Zagato E, Calcinotto A, Troiani M, Mestre RP, Calì B, Attanasio G, Troisi J, Minini M, Mosole S, Revandkar A, Pasquini E, Elia AR, Bossi D, Rinaldi A, Rescigno P, Flohr P, Hunt J, Neeb A, Buroni L, Guo C, Welti J, Ferrari M, Grioni M, Gauthier J, Gharaibeh RZ, Palmisano A, Lucchini GM, D'Antonio E, Merler S, Bolis M, Grassi F, Esposito A, Bellone M, Briganti A, Rescigno M, Theurillat JP, Jobin C, Gillissen S, de Bono J and Alimonti A. Commensal bacteria promote endocrine resistance in prostate cancer through androgen biosynthesis. *Science* 2021; 374: 216-224.
- [30] Koh CM, Bieberich CJ, Dang CV, Nelson WG, Yegnasubramanian S and De Marzo AM. MYC and prostate cancer. *Genes Cancer* 2010; 1: 617-628.
- [31] Bader DA and McGuire SE. Tumour metabolism and its unique properties in prostate adenocarcinoma. *Nat Rev Urol* 2020; 17: 214-231.
- [32] Polkinghorn WR, Parker JS, Lee MX, Kass EM, Spratt DE, Iaquinta PJ, Arora VK, Yen WF, Cai L, Zheng D, Carver BS, Chen Y, Watson PA, Shah NP, Fujisawa S, Goglia AG, Gopalan A, Hierony-

- mus H, Wongvipat J, Scardino PT, Zelefsky MJ, Jasin M, Chaudhuri J, Powell SN and Sawyers CL. Androgen receptor signaling regulates DNA repair in prostate cancers. *Cancer Discov* 2013; 3: 1245-1253.
- [33] Williams GT and Farzaneh F. Are snoRNAs and snoRNA host genes new players in cancer? *Nat Rev Cancer* 2012; 12: 84-88.
- [34] Torti SV and Torti FM. Iron and cancer: more ore to be mined. *Nat Rev Cancer* 2013; 13: 342-355.
- [35] Liu W and Rodgers GP. Olfactomedin 4 expression and functions in innate immunity, inflammation, and cancer. *Cancer Metastasis Rev* 2016; 35: 201-212.
- [36] Casero RA Jr, Murray Stewart T and Pegg AE. Polyamine metabolism and cancer: treatments, challenges and opportunities. *Nat Rev Cancer* 2018; 18: 681-695.
- [37] Konen JM, Fradette JJ and Gibbons DL. The good, the bad and the unknown of CD38 in the metabolic microenvironment and immune cell functionality of solid tumors. *Cells* 2019; 9: 52.

Paenibacillus in prostate adenocarcinoma



Supplementary Figure 1. Additional nominally significant tumor-normal differential genera not shown in Main Table 1. Genera with nominal $P < 0.05$ but false discovery rate (FDR) ≥ 0.05 are shown, sorted by absolute log₂ fold change (log₂FC). Main Table 1 contains the 15 FDR-significant genera.

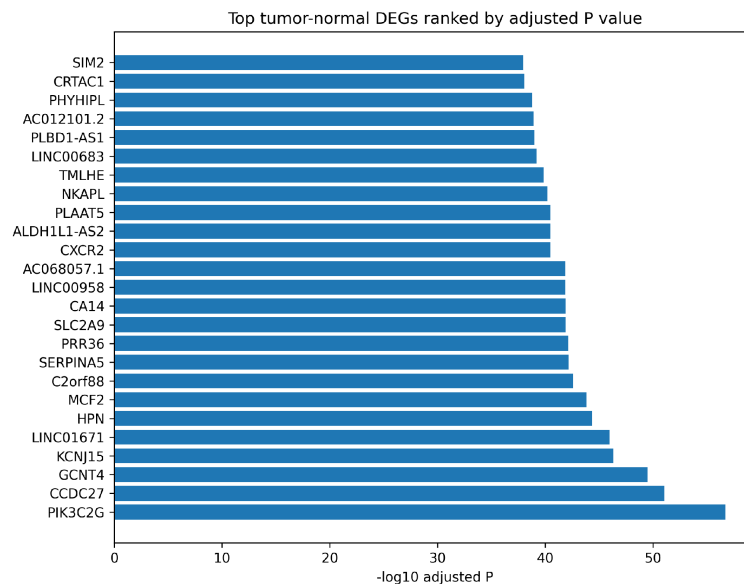
Supplementary Table 1. Additional nominally significant differential genera not listed in Main Table 1

Bacteria	log ₂ FC	P value	FDR
Slackia	-2.933	0.001276	0.06546
Asteroleplasma	-8.149	0.00212	0.07569
Basilea	-8.454	0.00212	0.07569
Candidatus Cloacimonas	-9.985	0.00212	0.07569
Heliophilum	-8.949	0.00212	0.07569
Macrococcus	-8.282	0.00212	0.07569
Tumebacillus	-8.711	0.00212	0.07569
Vogesella	-9.234	0.00212	0.07569
Faecalibacterium	-1.709	0.004265	0.1459
Carboxydotherrmus	-3.959	0.005691	0.1832
Centipeda	-3.09	0.005802	0.1832
Herminiimonas	-2.511	0.006523	0.1939
Azotobacter	0.926	0.006613	0.1939
Thiomonas	-3.084	0.007484	0.2119
Herbaspirillum	-2.184	0.01022	0.2796
Burkholderia	-0.5536	0.01105	0.2928
Prevotella	1.115	0.01165	0.2989
Sulfobacillus	-1.564	0.01689	0.413
Klebsiella	-1.349	0.01912	0.413
Helicobacter	-2.665	0.01924	0.413
Alicyclophilus	-4.22	0.01924	0.413
Desulfomicrobium	-3.526	0.01953	0.413
Trichodesmium	-2.16	0.01983	0.413
Hahella	-2.266	0.01992	0.413

Paenibacillus in prostate adenocarcinoma

Chelativorans	-3.152	0.02012	0.413
Hyphomicrobium	-1.894	0.02126	0.4238
Variovorax	10.66	0.02168	0.4238
Rhodopseudomonas	-1.523	0.02648	0.5056
Treponema	-1.967	0.0278	0.5113
Starkeya	-0.8728	0.02892	0.5113
Rhodoferax	-2.151	0.03094	0.5113
Azospirillum	2.634	0.031	0.5113
Corynebacterium	-0.4207	0.03146	0.5113
Mannheimia	-1.434	0.03184	0.5113
Micrococcus	0.6644	0.03242	0.5113
Escherichia	-1.259	0.0325	0.5113
Parascardovia	3.079	0.03286	0.5113
Gemella	0.9554	0.03721	0.5113
Streptococcus	0.5685	0.03735	0.5113
Cylindrospermum	-2.513	0.03938	0.5113
Achromobacter	-1.461	0.03981	0.5113
Sorangium	-1.401	0.04195	0.5113
Arcobacter	-2.129	0.04238	0.5113
Actinobacillus	-3.583	0.04435	0.5113
Ruegeria	-3.139	0.04435	0.5113
Anaeromyxobacter	-1.228	0.04531	0.5113
Parvibaculum	-2.338	0.04544	0.5113
Eggerthella	-2.646	0.04544	0.5113
Candidatus Phaeomarinobacter	-1.74	0.04656	0.5113
Amycolatopsis	2.316	0.04772	0.5113
Opitutus	-1.775	0.04791	0.5113
Leptolyngbya	-0.8966	0.04926	0.5113

Note: This table includes tumor-normal genera with nominal $P < 0.05$ but $FDR \geq 0.05$. $\log_2 FC$, $\log_2$ fold change; FDR, false discovery rate.



Supplementary Figure 2. Top tumor-normal differentially expressed genes (DEGs) ranked by adjusted P value. Top genes from the complete tumor-normal differential-expression output are displayed as $-\log_{10}$ adjusted P values.

Paenibacillus in prostate adenocarcinoma

Supplementary Table 2. Top 50 genes with higher expression in PRAD tumors than in normal prostate tissues

Gene	Normal_Mean	Tumor_Mean	log ₂ FC	P value	Adjusted P value
HPN	46.41	251.8	2.714	4.38e-49	4.33e-45
PRR36	7.333	26.19	1.81	1.22e-46	7.27e-43
SIM2	4.492	22.72	2.374	4.53e-42	1.08e-38
CGREF1	6.763	24.96	1.944	1.42e-40	2.28e-37
HOXC6	2.539	12.77	2.392	1.88e-39	2.80e-36
RAB17	10.64	30.15	1.456	2.73e-39	3.86e-36
SLC19A1	6.309	16.25	1.304	4.30e-39	5.80e-36
FOXP4-AS1	1.352	5.802	1.484	5.66e-39	7.31e-36
MARCKSL1	417.1	1.06e+03	1.495	2.50e-38	2.65e-35
LRFN1	3.607	9.629	1.237	5.48e-37	5.01e-34
VPS9D1-AS1	4.736	17	1.667	3.27e-36	2.70e-33
STX19	13.17	36.74	1.76	4.43e-36	3.51e-33
AC104667.2	1.782	8.093	1.718	4.99e-36	3.90e-33
EPHA10	1.213	5.148	1.387	9.37e-36	7.05e-33
TRIM36	11.78	31.21	1.558	2.42e-35	1.71e-32
SBK1	6.934	17.73	1.346	2.82e-34	1.73e-31
AC074141.1	3.492	17.01	1.892	8.39e-34	4.75e-31
TRPM4	97.47	287.3	1.702	1.77e-33	9.50e-31
PCAT7	0.6447	4.471	1.579	2.09e-33	1.07e-30
SNHG4	2.496	11.35	1.728	6.57e-33	3.23e-30
PODXL2	85.85	215.2	1.671	1.03e-32	4.86e-30
DLX1	4.847	39.36	3.564	1.13e-32	5.28e-30
TRGC1	146.8	540.4	2.502	4.81e-32	2.10e-29
RPL7P16	90.57	390.1	2.973	6.48e-32	2.77e-29
SNHG3	8.556	25	1.458	3.34e-31	1.23e-28
NME1	7.425	18.26	1.117	7.71e-31	2.71e-28
ARHGEF38	6.749	18.79	1.526	1.94e-30	6.44e-28
EEF1AKMT4	5.224	13.49	1.183	2.68e-30	8.53e-28
SAMD10	12.73	24.74	0.9248	3.75e-30	1.17e-27
C19orf48	152.4	375.5	1.565	8.24e-30	2.44e-27
AC010326.4	2.744	5.605	0.8051	1.17e-29	3.41e-27
H2AW	45.5	146.2	1.772	1.62e-29	4.56e-27
HOXC4	1.629	7.941	1.871	5.37e-29	1.41e-26
GOLM1	371.8	1.20e+03	2.026	6.68e-29	1.73e-26
PPP1R14B	51.22	122.5	1.196	2.28e-28	5.61e-26
ARHGEF26	23.01	66.86	1.484	2.97e-28	7.12e-26
AL031123.2	0.5392	2.312	1.036	4.39e-28	1.02e-25
PYCR1	65.81	162.1	1.404	5.55e-28	1.26e-25
BIK	53.58	106.1	1.292	5.76e-28	1.31e-25
ULK3	32.71	62.78	0.9411	6.21e-28	1.40e-25
TLCD1	18.11	44.62	1.266	1.41e-27	2.96e-25
AC006042.1	6.38	18.09	1.429	1.87e-27	3.83e-25
LINC01509	1.566	6.694	1.492	2.24e-27	4.49e-25
UCN	1.728	7.053	1.374	3.46e-27	6.69e-25
LINC00665	6.08	12.9	0.9614	4.20e-27	8.02e-25
POLR2H	32.49	57.15	0.7897	4.36e-27	8.30e-25

Paenibacillus in prostate adenocarcinoma

DBNDD1	10.83	27.72	1.275	4.64e-27	8.82e-25
CNPY2	2.543	4.537	0.6231	4.81e-27	9.10e-25
SNORD88A	2.083	10.88	1.931	1.57e-26	2.72e-24
AL353622.1	1.596	6.579	1.424	1.62e-26	2.78e-24

Supplementary Table 3. Top 50 genes with lower expression in PRAD tumors than in normal prostate tissues

Gene	Normal_Mean	Tumor_Mean	log ₂ FC	P value	Adjusted P value
PIK3C2G	4.028	0.2719	-1.437	3.38e-62	2.01e-57
CCDC27	0.6124	0.1225	-0.4715	3.15e-56	9.36e-52
GCNT4	5.154	1.09	-1.409	1.65e-54	3.27e-50
KCNJ15	5.343	0.7254	-1.556	3.28e-51	4.87e-47
LINC01671	3.831	0.5537	-1.405	9.15e-51	1.09e-46
MCF2	6.832	0.2871	-1.329	1.77e-48	1.50e-44
C2orf88	12.05	2.201	-1.94	3.40e-47	2.53e-43
SERPINA5	97.37	1.525	-2.497	1.03e-46	6.80e-43
SLC2A9	8.454	1.76	-1.415	2.33e-46	1.26e-42
CA14	10.69	2.004	-1.923	2.72e-46	1.35e-42
LINC00958	1.675	0.2861	-0.9224	3.25e-46	1.39e-42
AC068057.1	0.3428	0.0535	-0.3358	3.27e-46	1.39e-42
CXCR2	3.525	0.622	-1.33	8.84e-45	3.30e-41
ALDH1L1-AS2	1.13	0.2844	-0.7097	9.36e-45	3.30e-41
PLAAT5	3.438	0.8375	-1.205	9.44e-45	3.30e-41
NKAPL	3.282	0.9333	-1.113	1.96e-44	6.48e-41
TMLHE	5.046	1.521	-1.273	4.63e-44	1.45e-40
LINC00683	1.094	0.3278	-0.6246	2.19e-43	6.50e-40
PLBD1-AS1	3.394	0.9373	-1.202	3.74e-43	1.06e-39
AC012101.2	1.351	0.3467	-0.7461	4.45e-43	1.20e-39
PHYHIP1	5.959	1.379	-1.549	6.44e-43	1.66e-39
CRTAC1	26.18	1.975	-1.987	3.63e-42	8.98e-39
KRT222	1.241	0.2275	-0.7797	5.22e-42	1.16e-38
AC108718.1	1.187	0.2188	-0.7109	5.26e-42	1.16e-38
AC040168.1	0.3874	0.1216	-0.2941	6.65e-42	1.41e-38
TRIM61	1.57	0.4695	-0.7356	7.31e-42	1.50e-38
GNAO1	9.712	2.082	-1.589	1.23e-41	2.43e-38
KLHL4	0.8681	0.2178	-0.5498	1.58e-41	3.03e-38
ASPA	6.011	1.258	-1.469	3.07e-41	5.64e-38
AL109976.1	1.88	0.5252	-0.8908	3.22e-41	5.64e-38
QPRT	23.71	5.203	-1.897	3.23e-41	5.64e-38
LINC01778	1.642	0.4707	-0.7582	4.65e-41	7.90e-38
PON3	1.939	0.2663	-0.8834	1.36e-40	2.24e-37
AC092851.1	0.9064	0.09566	-0.614	1.37e-39	2.14e-36
TRIM9	0.8823	0.2149	-0.5285	1.47e-39	2.24e-36
IL21-AS1	0.4209	0.07639	-0.3655	2.21e-39	3.20e-36
PAK3	3.944	1.126	-1.183	4.21e-39	5.80e-36
RHBDL3	2.773	0.8041	-1.027	5.38e-39	7.10e-36
RBMS3-AS3	0.7775	0.2409	-0.4911	8.39e-39	1.06e-35
KY	2.71	0.5261	-1.108	8.72e-39	1.08e-35

Paenibacillus in prostate adenocarcinoma

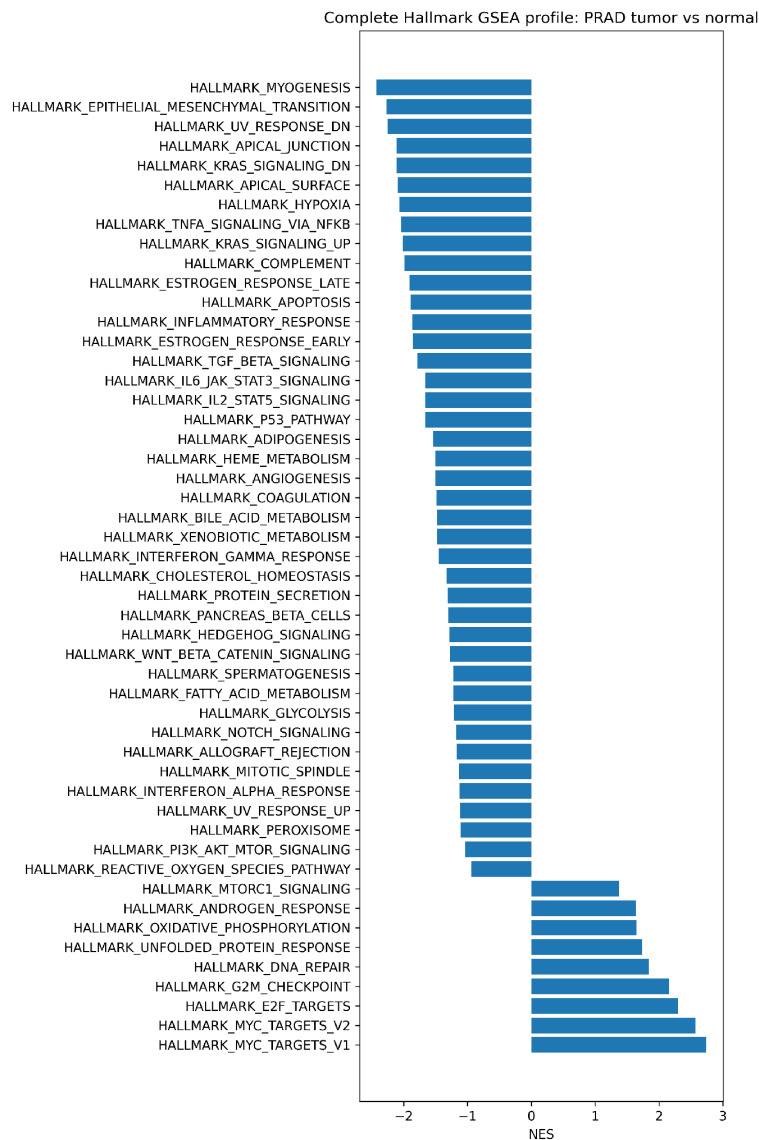
AC026469.1	0.7487	0.1226	-0.5395	1.01e-38	1.22e-35
AL161945.1	1.683	0.1408	-0.769	1.10e-38	1.31e-35
SCN5A	0.9359	0.2818	-0.5725	1.69e-38	1.97e-35
PADI3	6.701	0.4054	-1.516	2.05e-38	2.32e-35
IGSF1	4.217	0.8365	-1.368	2.07e-38	2.32e-35
KLHL14	8.828	0.7349	-1.348	2.26e-38	2.48e-35
NECAB1	5.852	1.742	-1.315	2.47e-38	2.65e-35
GSTM3	84.51	20.18	-1.874	2.80e-38	2.92e-35
PNMT	28.38	1.556	-2.057	3.87e-38	3.97e-35
PAK5	1.492	0.4352	-0.7965	1.26e-37	1.27e-34

Supplementary Table 4. Top 50 tumor-normal DEGs ranked by adjusted *P* value

Gene	Normal_Mean	Tumor_Mean	log ₂ FC	<i>P</i> value	Adjusted <i>P</i> value
PIK3C2G	4.028	0.2719	-1.437	3.38e-62	2.01e-57
CCDC27	0.6124	0.1225	-0.4715	3.15e-56	9.36e-52
GCNT4	5.154	1.09	-1.409	1.65e-54	3.27e-50
KCNJ15	5.343	0.7254	-1.556	3.28e-51	4.87e-47
LINC01671	3.831	0.5537	-1.405	9.15e-51	1.09e-46
HPN	46.41	251.8	2.714	4.38e-49	4.33e-45
MCF2	6.832	0.2871	-1.329	1.77e-48	1.50e-44
C2orf88	12.05	2.201	-1.94	3.40e-47	2.53e-43
SERPINA5	97.37	1.525	-2.497	1.03e-46	6.80e-43
PRR36	7.333	26.19	1.81	1.22e-46	7.27e-43
SLC2A9	8.454	1.76	-1.415	2.33e-46	1.26e-42
CA14	10.69	2.004	-1.923	2.72e-46	1.35e-42
LINC00958	1.675	0.2861	-0.9224	3.25e-46	1.39e-42
AC068057.1	0.3428	0.0535	-0.3358	3.27e-46	1.39e-42
CXCR2	3.525	0.622	-1.33	8.84e-45	3.30e-41
ALDH1L1-AS2	1.13	0.2844	-0.7097	9.36e-45	3.30e-41
PLAAT5	3.438	0.8375	-1.205	9.44e-45	3.30e-41
NKAPL	3.282	0.9333	-1.113	1.96e-44	6.48e-41
TMLHE	5.046	1.521	-1.273	4.63e-44	1.45e-40
LINC00683	1.094	0.3278	-0.6246	2.19e-43	6.50e-40
PLBD1-AS1	3.394	0.9373	-1.202	3.74e-43	1.06e-39
AC012101.2	1.351	0.3467	-0.7461	4.45e-43	1.20e-39
PHYHIPL	5.959	1.379	-1.549	6.44e-43	1.66e-39
CRTAC1	26.18	1.975	-1.987	3.63e-42	8.98e-39
SIM2	4.492	22.72	2.374	4.53e-42	1.08e-38
KRT222	1.241	0.2275	-0.7797	5.22e-42	1.16e-38
AC108718.1	1.187	0.2188	-0.7109	5.26e-42	1.16e-38
AC040168.1	0.3874	0.1216	-0.2941	6.65e-42	1.41e-38
TRIM61	1.57	0.4695	-0.7356	7.31e-42	1.50e-38
GNAO1	9.712	2.082	-1.589	1.23e-41	2.43e-38
KLHL4	0.8681	0.2178	-0.5498	1.58e-41	3.03e-38
ASPA	6.011	1.258	-1.469	3.07e-41	5.64e-38
AL109976.1	1.88	0.5252	-0.8908	3.22e-41	5.64e-38
QPRT	23.71	5.203	-1.897	3.23e-41	5.64e-38
LINC01778	1.642	0.4707	-0.7582	4.65e-41	7.90e-38
PON3	1.939	0.2663	-0.8834	1.36e-40	2.24e-37
CGREF1	6.763	24.96	1.944	1.42e-40	2.28e-37

Paenibacillus in prostate adenocarcinoma

AC092851.1	0.9064	0.09566	-0.614	1.37e-39	2.14e-36
TRIM9	0.8823	0.2149	-0.5285	1.47e-39	2.24e-36
HOXC6	2.539	12.77	2.392	1.88e-39	2.80e-36
IL21-AS1	0.4209	0.07639	-0.3655	2.21e-39	3.20e-36
RAB17	10.64	30.15	1.456	2.73e-39	3.86e-36
PAK3	3.944	1.126	-1.183	4.21e-39	5.80e-36
SLC19A1	6.309	16.25	1.304	4.30e-39	5.80e-36
RHBDL3	2.773	0.8041	-1.027	5.38e-39	7.10e-36
FOXP4-AS1	1.352	5.802	1.484	5.66e-39	7.31e-36
RBMS3-AS3	0.7775	0.2409	-0.4911	8.39e-39	1.06e-35
KY	2.71	0.5261	-1.108	8.72e-39	1.08e-35
AC026469.1	0.7487	0.1226	-0.5395	1.01e-38	1.22e-35
AL161945.1	1.683	0.1408	-0.769	1.10e-38	1.31e-35



Supplementary Figure 3. Complete Hallmark gene set enrichment analysis (GSEA) profile for the prostate adenocarcinoma (PRAD) tumor-normal comparison. All 50 Hallmark gene sets are shown according to normalized enrichment score (NES). The main text highlights selected leading pathways.

Paenibacillus in prostate adenocarcinoma

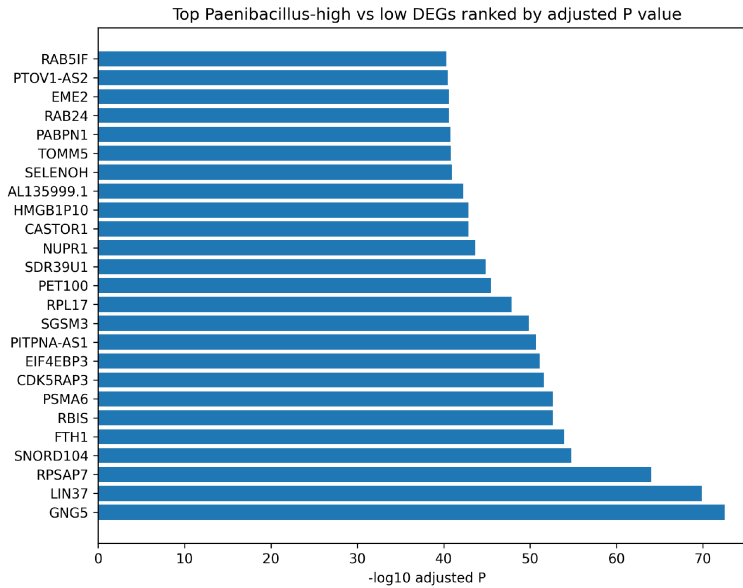
Supplementary Table 5. Complete Hallmark GSEA results for the PRAD tumor-normal comparison

Description	setSize	NES	P value	Adjusted P value	q value	rank	leading_edge
HALLMARK_MYC_TARGETS_V1	199	2.743	1.00e-10	5.00e-10	1.68e-10	9.00e+03	tags = 60%, list = 15%, signal = 51%
HALLMARK_MYC_TARGETS_V2	57	2.573	1.00e-10	5.00e-10	1.68e-10	7.39e+03	tags = 70%, list = 12%, signal = 62%
HALLMARK_MYOGENESIS	199	-2.426	1.00e-10	5.00e-10	1.68e-10	8.69e+03	tags = 62%, list = 15%, signal = 53%
HALLMARK_E2F_TARGETS	200	2.302	1.00e-10	5.00e-10	1.68e-10	7.42e+03	tags = 44%, list = 12%, signal = 39%
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	200	-2.266	1.00e-10	5.00e-10	1.68e-10	9.89e+03	tags = 62%, list = 17%, signal = 52%
HALLMARK_UV_RESPONSE_DN	144	-2.247	1.00e-10	5.00e-10	1.68e-10	9.17e+03	tags = 59%, list = 15%, signal = 50%
HALLMARK_G2M_CHECKPOINT	199	2.158	1.00e-10	5.00e-10	1.68e-10	7.80e+03	tags = 47%, list = 13%, signal = 41%
HALLMARK_APICAL_JUNCTION	198	-2.109	1.00e-10	5.00e-10	1.68e-10	9.20e+03	tags = 51%, list = 15%, signal = 43%
HALLMARK_KRAS_SIGNALING_DN	200	-2.108	1.00e-10	5.00e-10	1.68e-10	8.58e+03	tags = 47%, list = 14%, signal = 40%
HALLMARK_HYPOXIA	200	-2.067	1.00e-10	5.00e-10	1.68e-10	9.97e+03	tags = 53%, list = 17%, signal = 44%
HALLMARK_TNFA_SIGNALING_VIA_NFKB	199	-2.038	1.12e-10	5.09e-10	1.71e-10	1.15e+04	tags = 60%, list = 19%, signal = 49%
HALLMARK_KRAS_SIGNALING_UP	200	-2.009	2.13e-10	8.89e-10	2.99e-10	8.51e+03	tags = 46%, list = 14%, signal = 40%
HALLMARK_COMPLEMENT	200	-1.983	6.16e-10	2.37e-09	7.98e-10	1.13e+04	tags = 48%, list = 19%, signal = 39%
HALLMARK_ESTROGEN_RESPONSE_LATE	198	-1.906	1.01e-08	3.61e-08	1.22e-08	8.10e+03	tags = 47%, list = 14%, signal = 41%
HALLMARK_INFLAMMATORY_RESPONSE	200	-1.861	5.54e-08	1.85e-07	6.22e-08	1.18e+04	tags = 52%, list = 20%, signal = 42%
HALLMARK_ESTROGEN_RESPONSE_EARLY	198	-1.856	6.07e-08	1.90e-07	6.39e-08	8.00e+03	tags = 47%, list = 13%, signal = 41%
HALLMARK_APOPTOSIS	161	-1.893	3.38e-07	9.94e-07	3.35e-07	1.07e+04	tags = 52%, list = 18%, signal = 42%
HALLMARK_APICAL_SURFACE	44	-2.092	1.27e-06	3.52e-06	1.19e-06	9.15e+03	tags = 61%, list = 15%, signal = 52%
HALLMARK_DNA_REPAIR	150	1.842	1.36e-06	3.57e-06	1.20e-06	8.02e+03	tags = 41%, list = 14%, signal = 35%
HALLMARK_OXIDATIVE_PHOSPHORYLATION	200	1.647	1.11e-05	2.78e-05	9.35e-06	9.22e+03	tags = 42%, list = 16%, signal = 35%
HALLMARK_IL2_STAT5_SIGNALING	198	-1.663	1.99e-05	4.75e-05	1.60e-05	8.58e+03	tags = 40%, list = 14%, signal = 35%
HALLMARK_P53_PATHWAY	199	-1.659	6.59e-05	1.50e-04	5.05e-05	8.63e+03	tags = 45%, list = 15%, signal = 38%
HALLMARK_UNFOLDED_PROTEIN_RESPONSE	112	1.736	1.08e-04	2.34e-04	7.89e-05	7.17e+03	tags = 38%, list = 12%, signal = 33%
HALLMARK_ANDROGEN_RESPONSE	100	1.645	3.90e-04	8.12e-04	2.74e-04	4.99e+03	tags = 32%, list = 8%, signal = 29%
HALLMARK_ADIPOGENESIS	200	-1.538	5.09e-04	0.001017	3.43e-04	7.60e+03	tags = 40%, list = 13%, signal = 35%
HALLMARK_TGF_BETA_SIGNALING	54	-1.785	5.90e-04	0.001135	3.82e-04	1.14e+04	tags = 56%, list = 19%, signal = 45%
HALLMARK_HEME_METABOLISM	200	-1.504	9.11e-04	0.001686	5.68e-04	8.52e+03	tags = 36%, list = 14%, signal = 31%
HALLMARK_IL6_JAK_STAT3_SIGNALING	87	-1.665	0.001313	0.002344	7.90e-04	1.10e+04	tags = 46%, list = 18%, signal = 38%
HALLMARK_INTERFERON_GAMMA_RESPONSE	198	-1.453	0.002216	0.00382	0.001287	1.08e+04	tags = 38%, list = 18%, signal = 32%
HALLMARK_XENOBIOTIC_METABOLISM	199	-1.474	0.002789	0.004648	0.001566	8.74e+03	tags = 34%, list = 15%, signal = 29%
HALLMARK_COAGULATION	138	-1.482	0.004458	0.00719	0.002422	1.19e+04	tags = 46%, list = 20%, signal = 37%
HALLMARK_BILE_ACID_METABOLISM	112	-1.474	0.008154	0.01274	0.004291	7.69e+03	tags = 36%, list = 13%, signal = 31%

Paenibacillus in prostate adenocarcinoma

HALLMARK_MTORC1_SIGNALING	199	1.376	0.008875	0.01345	0.004529	7.43e+03	tags = 30%, list = 13%, signal = 26%
HALLMARK_ANGIOGENESIS	36	-1.499	0.02257	0.03318	0.01118	8.28e+03	tags = 39%, list = 14%, signal = 33%
HALLMARK_CHOLESTEROL_HOMEOSTASIS	74	-1.325	0.05972	0.08398	0.02829	9.12e+03	tags = 46%, list = 15%, signal = 39%
HALLMARK_PROTEIN_SECRETION	96	-1.31	0.06047	0.08398	0.02829	1.33e+04	tags = 46%, list = 22%, signal = 36%
HALLMARK_GLYCOLYSIS	200	-1.216	0.07703	0.1041	0.03506	9.94e+03	tags = 38%, list = 17%, signal = 31%
HALLMARK_PANCREAS_BETA_CELLS	40	-1.3	0.09968	0.1306	0.04398	6.85e+03	tags = 25%, list = 12%, signal = 22%
HALLMARK_FATTY_ACID_METABOLISM	157	-1.219	0.1018	0.1306	0.04398	1.05e+04	tags = 37%, list = 18%, signal = 30%
HALLMARK_HEDGEHOG_SIGNALING	36	-1.284	0.11	0.1341	0.04519	8.30e+03	tags = 42%, list = 14%, signal = 36%
HALLMARK_WNT_BETA_CATENIN_SIGNALING	42	-1.277	0.1084	0.1341	0.04519	8.88e+03	tags = 48%, list = 15%, signal = 41%
HALLMARK_SPERMATOGENESIS	134	-1.22	0.116	0.1381	0.04653	9.51e+03	tags = 34%, list = 16%, signal = 28%
HALLMARK_ALLOGRAFT_REJECTION	200	-1.171	0.123	0.143	0.04817	1.13e+04	tags = 35%, list = 19%, signal = 28%
HALLMARK_MITOTIC_SPINDLE	199	-1.133	0.1737	0.1974	0.06648	9.34e+03	tags = 35%, list = 16%, signal = 29%
HALLMARK_NOTCH_SIGNALING	32	-1.177	0.2101	0.2334	0.07863	1.19e+04	tags = 53%, list = 20%, signal = 43%
HALLMARK_UV_RESPONSE_UP	158	-1.115	0.2192	0.2383	0.08027	8.30e+03	tags = 35%, list = 14%, signal = 31%
HALLMARK_INTERFERON_ALPHA_RESPONSE	97	-1.125	0.2277	0.2422	0.08159	1.09e+04	tags = 39%, list = 18%, signal = 32%
HALLMARK_PEROXISOME	104	-1.103	0.253	0.2636	0.08879	7.70e+03	tags = 28%, list = 13%, signal = 24%
HALLMARK_PI3K_AKT_MTOR_SIGNALING	105	-1.041	0.3649	0.3723	0.1254	9.72e+03	tags = 33%, list = 16%, signal = 28%
HALLMARK_REACTIVE_OXYGEN_SPECIES_PATHWAY	49	-0.9397	0.5846	0.5846	0.1969	7.32e+03	tags = 27%, list = 12%, signal = 23%

Paenibacillus in prostate adenocarcinoma



Supplementary Figure 4. Top differentially expressed genes (DEGs) between *Paenibacillus*-high and *Paenibacillus*-low tumors, ranked by adjusted P value. Top genes associated with *Paenibacillus* stratification are displayed as $-\log_{10}$ adjusted P values.

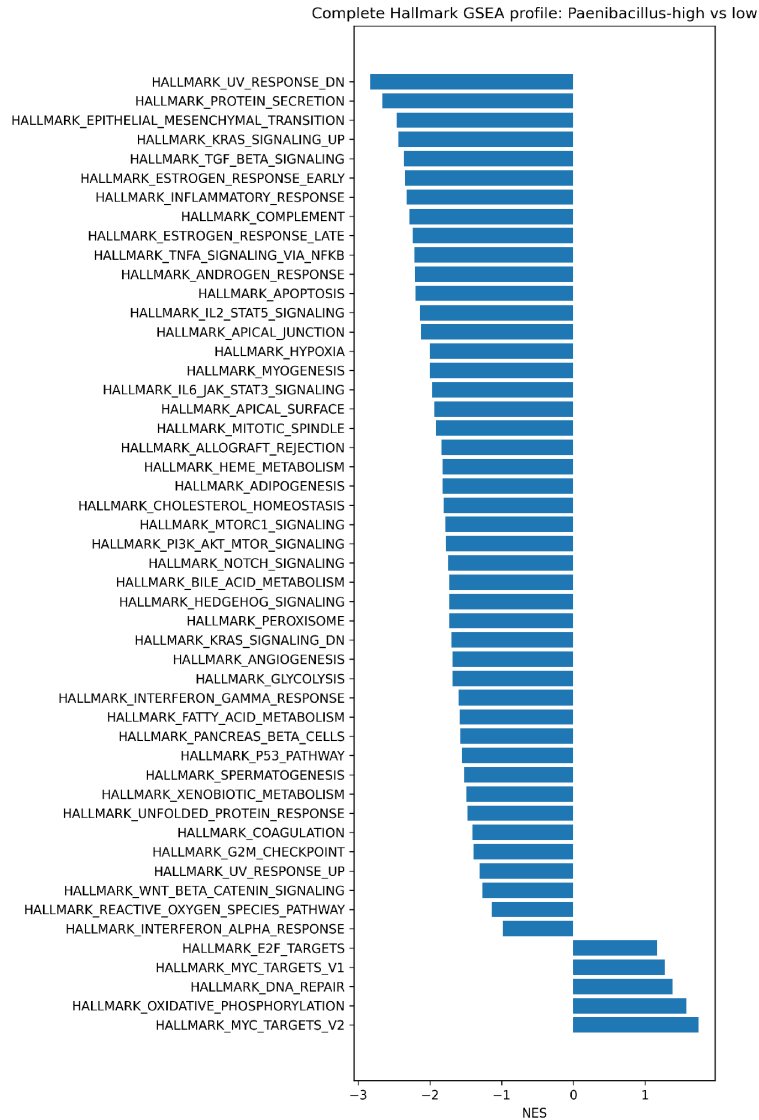
Supplementary Table 6. Top 50 significant DEGs between *Paenibacillus*-high and *Paenibacillus*-low PRAD tumors

$\log_2$ FC	AveExpr	t	P value	Adjusted P value	B	Gene
1.176	3.667	22.94	4.91e-78	2.92e-73	167.2	GNG5
0.5491	0.8852	22.3	4.43e-75	1.32e-70	160.4	LIN37
1.278	1.473	21.01	4.69e-69	9.29e-65	146.6	RPSAP7
2.075	5.834	18.98	1.20e-59	1.78e-55	125	SNORD104
1.284	6.194	18.78	1.04e-58	1.24e-54	122.8	FTH1
0.8926	3.716	18.47	2.60e-57	2.50e-53	119.6	RBIS
0.5678	1.29	18.46	2.94e-57	2.50e-53	119.5	PSMA6
1.108	1.896	18.23	3.42e-56	2.54e-52	117.1	CDK5RAP3
0.9834	2.398	18.12	1.14e-55	7.51e-52	115.9	EIF4EBP3
0.8654	2	18.01	3.62e-55	2.15e-51	114.7	PITPNA-AS1
0.6253	0.9642	17.82	2.62e-54	1.41e-50	112.7	SGSM3
0.812	2.324	17.38	2.80e-52	1.39e-48	108.1	RPL17
0.7916	1.975	16.84	7.48e-50	3.42e-46	102.5	PET100
0.7399	2.442	16.7	3.27e-49	1.39e-45	101.1	SDR39U1
1.017	3.823	16.43	5.49e-48	2.18e-44	98.24	NUPR1
0.3224	0.5344	16.24	3.77e-47	1.40e-43	96.32	CASTOR1
-0.8542	3.048	-16.24	4.04e-47	1.41e-43	96.25	HMGB1P10
0.4465	0.539	16.09	1.73e-46	5.70e-43	94.81	AL135999.1
0.7	3.657	15.8	3.65e-45	1.14e-41	91.77	SELENOH
0.4016	1.118	15.77	4.98e-45	1.48e-41	91.46	TOMM5
1.037	3.341	15.75	6.14e-45	1.74e-41	91.25	PABPN1
0.9009	2.867	15.71	9.13e-45	2.47e-41	90.86	RAB24
1.06	2.917	15.7	9.59e-45	2.48e-41	90.81	EME2
0.977	2.713	15.66	1.46e-44	3.61e-41	90.39	PTOV1-AS2

Paenibacillus in prostate adenocarcinoma

0.6699	1.108	15.63	2.08e-44	4.96e-41	90.03	RAB5IF
1.578	11.49	15.55	4.78e-44	1.09e-40	89.21	MT-ATP8
0.8562	1.753	15.27	8.23e-43	1.81e-39	86.37	DHPS
-0.5736	1.106	-15.23	1.16e-42	2.47e-39	86.03	ACTBP2
1.105	2.158	15.23	1.24e-42	2.54e-39	85.97	STAG3L5P
0.6074	0.7479	15.19	1.84e-42	3.64e-39	85.58	SAPCD1
0.3032	0.3554	15.15	2.58e-42	4.95e-39	85.24	MSH5-SAPCD1
0.364	0.4784	15.14	3.00e-42	5.57e-39	85.09	TEN1-CDK3
1.471	3.259	14.99	1.32e-41	2.37e-38	83.62	SNHG25
0.8499	2.528	14.94	2.21e-41	3.87e-38	83.1	ALDOA
-0.6391	5.996	-14.78	1.12e-40	1.90e-37	81.48	EIF5B
0.8441	0.8144	14.74	1.71e-40	2.82e-37	81.07	RPS4XP23
0.1259	0.1756	14.7	2.54e-40	4.08e-37	80.67	IPO4
0.9352	1.962	14.69	2.66e-40	4.16e-37	80.62	LINC01786
0.4441	1.85	14.62	5.32e-40	8.11e-37	79.93	MRPS18C
0.7219	2.99	14.53	1.35e-39	2.00e-36	79.01	CENATAC
1.167	3.696	14.52	1.52e-39	2.21e-36	78.89	KIFC2
0.2733	0.5287	14.51	1.62e-39	2.29e-36	78.83	ABCB6
-0.5415	3.452	-14.48	2.15e-39	2.98e-36	78.54	SREK1IP1
0.422	0.8922	14.48	2.24e-39	3.02e-36	78.5	ALKBH6
0.1582	0.1753	14.43	3.62e-39	4.78e-36	78.03	AC139530.1
0.7043	3.737	14.4	4.84e-39	6.25e-36	77.74	POLD4
0.7882	0.961	14.39	5.67e-39	7.17e-36	77.58	LMNTD2-AS1
0.9638	1.601	14.37	7.07e-39	8.76e-36	77.36	AC092171.4
0.5361	1.034	14.36	7.44e-39	9.02e-36	77.31	RPP21
0.304	0.5071	14.31	1.27e-38	1.50e-35	76.78	CCDC189

Paenibacillus in prostate adenocarcinoma



Supplementary Figure 5. Complete Hallmark gene set enrichment analysis (GSEA) profile for the Paenibacillus-high versus Paenibacillus-low comparison. All 50 Hallmark gene sets are shown according to normalized enrichment score (NES).

Paenibacillus in prostate adenocarcinoma

Supplementary Table 7. Complete Hallmark GSEA results for the *Paenibacillus*-high versus *Paenibacillus*-low comparison

Description	setSize	NES	P value	Adjusted P value	q value	rank	leading_edge
HALLMARK_UV_RESPONSE_DN	144	-2.829	1.00e-10	4.17e-10	4.39e-11	1.01e+04	tags = 68%, list = 17%, signal = 57%
HALLMARK_PROTEIN_SECRETION	96	-2.659	1.00e-10	4.17e-10	4.39e-11	8.88e+03	tags = 70%, list = 15%, signal = 59%
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	200	-2.46	1.00e-10	4.17e-10	4.39e-11	1.31e+04	tags = 57%, list = 22%, signal = 45%
HALLMARK_KRAS_SIGNALING_UP	200	-2.44	1.00e-10	4.17e-10	4.39e-11	1.13e+04	tags = 54%, list = 19%, signal = 44%
HALLMARK_ESTROGEN_RESPONSE_EARLY	198	-2.346	1.00e-10	4.17e-10	4.39e-11	1.46e+04	tags = 59%, list = 25%, signal = 45%
HALLMARK_INFLAMMATORY_RESPONSE	200	-2.322	1.00e-10	4.17e-10	4.39e-11	1.45e+04	tags = 58%, list = 24%, signal = 44%
HALLMARK_COMPLEMENT	200	-2.289	1.00e-10	4.17e-10	4.39e-11	1.32e+04	tags = 59%, list = 22%, signal = 46%
HALLMARK_ESTROGEN_RESPONSE_LATE	198	-2.235	1.00e-10	4.17e-10	4.39e-11	1.05e+04	tags = 46%, list = 18%, signal = 38%
HALLMARK_TNFA_SIGNALING_VIA_NFKB	199	-2.218	1.00e-10	4.17e-10	4.39e-11	1.40e+04	tags = 53%, list = 24%, signal = 41%
HALLMARK_APOPTOSIS	161	-2.2	1.00e-10	4.17e-10	4.39e-11	1.38e+04	tags = 58%, list = 23%, signal = 45%
HALLMARK_IL2_STAT5_SIGNALING	198	-2.135	1.00e-10	4.17e-10	4.39e-11	1.38e+04	tags = 58%, list = 23%, signal = 44%
HALLMARK_APICAL_JUNCTION	198	-2.123	1.00e-10	4.17e-10	4.39e-11	1.13e+04	tags = 53%, list = 19%, signal = 43%
HALLMARK_HYPOXIA	200	-2.005	1.35e-09	5.19e-09	5.46e-10	1.13e+04	tags = 48%, list = 19%, signal = 39%
HALLMARK_TGF_BETA_SIGNALING	54	-2.36	2.13e-09	7.61e-09	8.01e-10	9.79e+03	tags = 61%, list = 16%, signal = 51%
HALLMARK_MYOGENESIS	199	-2	2.34e-09	7.81e-09	8.22e-10	1.12e+04	tags = 42%, list = 19%, signal = 34%
HALLMARK_ANDROGEN_RESPONSE	100	-2.211	8.15e-09	2.55e-08	2.68e-09	1.21e+04	tags = 58%, list = 20%, signal = 46%
HALLMARK_MITOTIC_SPINDLE	199	-1.916	3.72e-08	1.10e-07	1.15e-08	1.06e+04	tags = 48%, list = 18%, signal = 39%
HALLMARK_ALLOGRAFT_REJECTION	200	-1.836	4.33e-07	1.20e-06	1.27e-07	1.51e+04	tags = 51%, list = 25%, signal = 38%
HALLMARK_HEME_METABOLISM	200	-1.827	5.24e-07	1.37e-06	1.44e-07	1.06e+04	tags = 42%, list = 18%, signal = 34%
HALLMARK_ADIPOGENESIS	200	-1.826	5.49e-07	1.37e-06	1.44e-07	9.30e+03	tags = 42%, list = 16%, signal = 36%
HALLMARK_MTORC1_SIGNALING	199	-1.785	1.89e-06	4.49e-06	4.73e-07	1.04e+04	tags = 46%, list = 18%, signal = 38%
HALLMARK_IL6_JAK_STAT3_SIGNALING	87	-1.966	1.75e-05	3.91e-05	4.12e-06	1.23e+04	tags = 48%, list = 21%, signal = 38%
HALLMARK_KRAS_SIGNALING_DN	200	-1.7	1.80e-05	3.91e-05	4.12e-06	1.33e+04	tags = 42%, list = 22%, signal = 33%
HALLMARK_GLYCOLYSIS	200	-1.68	3.34e-05	6.97e-05	7.33e-06	8.81e+03	tags = 38%, list = 15%, signal = 32%
HALLMARK_APICAL_SURFACE	44	-1.936	1.02e-04	2.04e-04	2.15e-05	1.27e+04	tags = 70%, list = 21%, signal = 55%
HALLMARK_PI3K_AKT_MTOR_SIGNALING	105	-1.778	1.38e-04	2.66e-04	2.80e-05	9.64e+03	tags = 45%, list = 16%, signal = 38%
HALLMARK_BILE_ACID_METABOLISM	112	-1.734	1.62e-04	3.01e-04	3.17e-05	1.24e+04	tags = 46%, list = 21%, signal = 37%
HALLMARK_INTERFERON_GAMMA_RESPONSE	198	-1.595	1.76e-04	3.14e-04	3.30e-05	1.30e+04	tags = 41%, list = 22%, signal = 32%
HALLMARK_PEROXISOME	104	-1.729	2.39e-04	4.12e-04	4.33e-05	8.05e+03	tags = 33%, list = 14%, signal = 28%
HALLMARK_OXIDATIVE_PHOSPHORYLATION	200	1.573	2.81e-04	4.69e-04	4.94e-05	7.38e+03	tags = 30%, list = 12%, signal = 27%
HALLMARK_P53_PATHWAY	199	-1.549	3.61e-04	5.82e-04	6.12e-05	9.71e+03	tags = 36%, list = 16%, signal = 30%
HALLMARK_CHOLESTEROL_HOMEOSTASIS	74	-1.809	3.77e-04	5.89e-04	6.20e-05	7.20e+03	tags = 35%, list = 12%, signal = 31%

Paenibacillus in prostate adenocarcinoma

HALLMARK_MYC_TARGETS_V2	57	1.748	8.18e-04	0.001239	1.30e-04	1.04e+04	tags = 44%, list = 18%, signal = 36%
HALLMARK_FATTY_ACID_METABOLISM	157	-1.586	8.90e-04	0.001308	1.38e-04	9.57e+03	tags = 38%, list = 16%, signal = 32%
HALLMARK_XENOBIOTIC_METABOLISM	199	-1.494	9.74e-04	0.001391	1.46e-04	1.03e+04	tags = 36%, list = 17%, signal = 30%
HALLMARK_HEDGEHOG_SIGNALING	36	-1.729	0.002681	0.003724	3.92e-04	1.36e+04	tags = 53%, list = 23%, signal = 41%
HALLMARK_NOTCH_SIGNALING	32	-1.747	0.002918	0.003943	4.15e-04	9.77e+03	tags = 56%, list = 16%, signal = 47%
HALLMARK_SPERMATOGENESIS	134	-1.519	0.003139	0.004131	4.35e-04	1.16e+04	tags = 37%, list = 20%, signal = 29%
HALLMARK_ANGIOGENESIS	36	-1.685	0.00387	0.004961	5.22e-04	1.19e+04	tags = 44%, list = 20%, signal = 36%
HALLMARK_G2M_CHECKPOINT	199	-1.388	0.006389	0.007986	8.41e-04	9.57e+03	tags = 29%, list = 16%, signal = 24%
HALLMARK_UNFOLDED_PROTEIN_RESPONSE	112	-1.473	0.007414	0.009042	9.52e-04	9.33e+03	tags = 34%, list = 16%, signal = 29%
HALLMARK_DNA_REPAIR	150	1.388	0.01119	0.01332	0.001402	6.93e+03	tags = 28%, list = 12%, signal = 25%
HALLMARK_COAGULATION	138	-1.409	0.0159	0.01849	0.001946	1.00e+04	tags = 38%, list = 17%, signal = 31%
HALLMARK_PANCREAS_BETA_CELLS	40	-1.573	0.01641	0.01865	0.001963	1.49e+04	tags = 45%, list = 25%, signal = 34%
HALLMARK_UV_RESPONSE_UP	158	-1.31	0.0293	0.03256	0.003427	1.31e+04	tags = 39%, list = 22%, signal = 31%
HALLMARK_MYC_TARGETS_V1	199	1.276	0.05112	0.05557	0.00585	7.38e+03	tags = 24%, list = 12%, signal = 21%
HALLMARK_WNT_BETA_CATENIN_SIGNALING	42	-1.27	0.1203	0.128	0.01347	1.47e+04	tags = 55%, list = 25%, signal = 41%
HALLMARK_E2F_TARGETS	200	1.167	0.1475	0.1537	0.01618	7.27e+03	tags = 28%, list = 12%, signal = 24%
HALLMARK_REACTIVE_OXYGEN_SPECIES_PATHWAY	49	-1.137	0.2549	0.2601	0.02738	1.06e+04	tags = 39%, list = 18%, signal = 32%
HALLMARK_INTERFERON_ALPHA_RESPONSE	97	-0.985	0.4833	0.4833	0.05087	1.14e+04	tags = 31%, list = 19%, signal = 25%

Supplementary Table 8. Top 50 gene ontology biological-process enrichment terms associated with *Paenibacillus*-stratified genes

ID	Description	GeneRatio	BgRatio	P value	Adjusted P value	q value	Count
GO:0006936	muscle contraction	41/659	351/18870	1.28e-11	5.94e-08	5.18e-08	41
GO:0003012	muscle system process	45/659	452/18870	2.78e-10	6.45e-07	5.62e-07	45
GO:0060537	muscle tissue development	39/659	433/18870	6.86e-08	1.06e-04	9.25e-05	39
GO:0006939	smooth muscle contraction	17/659	109/18870	2.28e-07	2.64e-04	2.30e-04	17
GO:0090130	tissue migration	33/659	380/18870	1.59e-06	0.001429	0.001246	33
GO:0036293	response to decreased oxygen levels	29/659	315/18870	2.15e-06	0.001429	0.001246	29
GO:0001666	response to hypoxia	28/659	298/18870	2.16e-06	0.001429	0.001246	28
GO:0003018	vascular process in circulatory system	26/659	269/18870	2.95e-06	0.00171	0.001491	26
GO:0060512	prostate gland morphogenesis	8/659	28/18870	3.55e-06	0.001828	0.001594	8
GO:0070482	response to oxygen levels	30/659	343/18870	4.08e-06	0.001895	0.001652	30
GO:0060973	cell migration involved in heart development	7/659	21/18870	4.65e-06	0.001962	0.001711	7
GO:0022612	gland morphogenesis	16/659	122/18870	5.40e-06	0.002013	0.001755	16
GO:0001667	ameboidal-type cell migration	38/659	497/18870	5.64e-06	0.002013	0.001755	38
GO:0031589	cell-substrate adhesion	30/659	356/18870	8.54e-06	0.002831	0.002468	30

Paenibacillus in prostate adenocarcinoma

GO:0060485	mesenchyme development	28/659	327/18870	1.26e-05	0.003408	0.002971	28
GO:1903522	regulation of blood circulation	24/659	258/18870	1.34e-05	0.003408	0.002971	24
GO:0008016	regulation of heart contraction	21/659	208/18870	1.35e-05	0.003408	0.002971	21
GO:0035296	regulation of tube diameter	17/659	146/18870	1.40e-05	0.003408	0.002971	17
GO:0097746	blood vessel diameter maintenance	17/659	146/18870	1.40e-05	0.003408	0.002971	17
GO:0035150	regulation of tube size	17/659	147/18870	1.53e-05	0.003544	0.00309	17
GO:1904062	regulation of monoatomic cation transmembrane transport	27/659	315/18870	1.77e-05	0.003737	0.003258	27
GO:0050678	regulation of epithelial cell proliferation	32/659	407/18870	1.77e-05	0.003737	0.003258	32
GO:0001655	urogenital system development	11/659	67/18870	1.88e-05	0.00379	0.003304	11
GO:0043270	positive regulation of monoatomic ion transport	21/659	214/18870	2.07e-05	0.004006	0.003492	21
GO:0062012	regulation of small molecule metabolic process	28/659	337/18870	2.19e-05	0.004033	0.003516	28
GO:0060740	prostate gland epithelium morphogenesis	7/659	26/18870	2.26e-05	0.004033	0.003516	7
GO:0010959	regulation of metal ion transport	31/659	398/18870	2.86e-05	0.004823	0.004205	31
GO:0006941	striated muscle contraction	19/659	186/18870	2.91e-05	0.004823	0.004205	19
GO:1904064	positive regulation of cation transmembrane transport	16/659	141/18870	3.42e-05	0.005472	0.004771	16
GO:0034767	positive regulation of monoatomic ion transmembrane transport	17/659	158/18870	3.91e-05	0.006053	0.005277	17
GO:0060047	heart contraction	22/659	243/18870	4.60e-05	0.006886	0.006003	22
GO:0010631	epithelial cell migration	29/659	372/18870	5.07e-05	0.007148	0.006231	29
GO:0030850	prostate gland development	9/659	50/18870	5.08e-05	0.007148	0.006231	9
GO:0007517	muscle organ development	28/659	354/18870	5.28e-05	0.007203	0.00628	28
GO:0014706	striated muscle tissue development	23/659	264/18870	5.67e-05	0.007473	0.006515	23
GO:0090132	epithelium migration	29/659	375/18870	5.85e-05	0.007473	0.006515	29
GO:0009612	response to mechanical stimulus	20/659	213/18870	6.06e-05	0.007473	0.006515	20
GO:0060445	branching involved in salivary gland morphogenesis	6/659	21/18870	6.14e-05	0.007473	0.006515	6
GO:0034765	regulation of monoatomic ion transmembrane transport	33/659	454/18870	6.28e-05	0.007473	0.006515	33
GO:0070252	actin-mediated cell contraction	13/659	105/18870	7.45e-05	0.008516	0.007425	13
GO:0042692	muscle cell differentiation	31/659	419/18870	7.53e-05	0.008516	0.007425	31
GO:0050673	epithelial cell proliferation	34/659	480/18870	8.11e-05	0.008963	0.007814	34
GO:0048732	gland development	32/659	443/18870	9.09e-05	0.009802	0.008546	32
GO:0003015	heart process	22/659	255/18870	9.43e-05	0.00994	0.008665	22
GO:2001257	regulation of cation channel activity	14/659	123/18870	1.01e-04	0.0104	0.009071	14
GO:0060048	cardiac muscle contraction	15/659	139/18870	1.06e-04	0.01072	0.009348	15
GO:0009636	response to toxic substance	22/659	258/18870	1.12e-04	0.01104	0.009623	22
GO:0043268	positive regulation of potassium ion transport	8/659	44/18870	1.23e-04	0.01193	0.0104	8
GO:0071453	cellular response to oxygen levels	17/659	174/18870	1.31e-04	0.01238	0.0108	17
GO:0050918	positive chemotaxis	10/659	69/18870	1.33e-04	0.01238	0.0108	10

Paenibacillus in prostate adenocarcinoma

Supplementary Table 9. KEGG enrichment terms associated with *Paenibacillus*-stratified genes

ID	Description	GeneRatio	BgRatio	P value	Adjusted P value	q value	Count
hsa04820	Cytoskeleton in muscle cells	25/343	233/9410	1.19e-06	1.63e-04	1.40e-04	25
hsa04510	Focal adhesion	23/343	203/9410	1.25e-06	1.63e-04	1.40e-04	23
hsa04270	Vascular smooth muscle contraction	18/343	134/9410	1.60e-06	1.63e-04	1.40e-04	18
hsa05414	Dilated cardiomyopathy	14/343	105/9410	2.54e-05	0.001938	0.001665	14
hsa05412	Arrhythmogenic right ventricular cardiomyopathy	12/343	86/9410	6.11e-05	0.003728	0.003204	12
hsa04918	Thyroid hormone synthesis	11/343	75/9410	7.78e-05	0.003953	0.003397	11
hsa04970	Salivary secretion	12/343	93/9410	1.32e-04	0.005733	0.004927	12
hsa04518	Integrin signaling	16/343	154/9410	1.50e-04	0.005733	0.004927	16
hsa05410	Hypertrophic cardiomyopathy	12/343	99/9410	2.42e-04	0.008189	0.007037	12
hsa04022	cGMP-PKG signaling pathway	16/343	166/9410	3.57e-04	0.01089	0.009356	16
hsa03320	PPAR signaling pathway	10/343	76/9410	4.07e-04	0.01129	0.0097	10
hsa05171	Coronavirus disease	20/343	239/9410	4.53e-04	0.0115	0.009886	20
hsa04810	Regulation of actin cytoskeleton	19/343	232/9410	8.13e-04	0.01831	0.01574	19
hsa04924	Renin secretion	9/343	69/9410	8.41e-04	0.01831	0.01574	9
hsa04972	Pancreatic secretion	11/343	102/9410	0.001178	0.02396	0.02059	11
hsa04512	ECM-receptor interaction	10/343	89/9410	0.001428	0.02616	0.02248	10
hsa04921	Oxytocin signaling pathway	14/343	154/9410	0.001458	0.02616	0.02248	14
hsa04929	GnRH secretion	8/343	65/9410	0.002364	0.04006	0.03443	8
hsa00120	Primary bile acid biosynthesis	4/343	17/9410	0.002832	0.04487	0.03856	4
hsa04360	Axon guidance	15/343	184/9410	0.002942	0.04487	0.03856	15
hsa04726	Serotonergic synapse	11/343	115/9410	0.003094	0.04494	0.03862	11
hsa04151	PI3K-Akt signaling pathway	24/343	362/9410	0.00334	0.04631	0.0398	24