

RESEARCH PAPER



Secretory IgA dysfunction underlies poor prognosis in *Fusobacterium*-infected colorectal cancer

Ilseok Choi^{a*}, Kyung-A Kim^{b*}, Sang Cheol Kim^{c*}, Donghwan Park^d, Ki Taek Nam^e, Jun Hyung Cha^a, Seungbyn Baek^a, Junha Cha^a, Hye-Yeong Jo^c, Minsun Jung^e, Melody Y. Zeng^f, Irina Matei^f, Susan Bullman^g, Joong Bae Ahn^b, Yoon Dae Han^h, Han Sang Kim^{i,j}, and Insuk Lee^{a,j}

^aDepartment of Biotechnology, College of Life Science and Biotechnology, Yonsei University, Seoul, Republic of Korea; ^bYonsei Cancer Center, Division of Medical Oncology, Department of Internal Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea; ^cDivision of Healthcare and Artificial Intelligence, Department of Precision Medicine, National Institute of Health, Cheongju, Republic of Korea; ^dDepartment of Biomedical Sciences, Brain Korea 21 FOUR Project for Medical Science, Yonsei University College of Medicine, Seoul, Republic of Korea; ^eDepartment of Pathology, Yonsei University College of Medicine, Seoul, Republic of Korea; ^fChildren's Cancer and Blood Foundation Laboratories, Departments of Pediatrics, and Cell and Developmental Biology, Drukier Institute for Children's Health, Meyer Cancer Center, Weill Cornell Medicine, New York, NY, USA; ^gImmunology, James P. Allison Institute, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; ^hDivision of Colorectal Surgery, Department of Surgery, Yonsei University College of Medicine, Seoul, Republic of Korea; ⁱGraduate School of Medical Science, Brain Korea 21 FOUR Project, Yonsei University College of Medicine, Seoul, Republic of Korea; ^jDECODE BIOME Co., Ltd., Incheon, Republic of Korea

ABSTRACT

Fusobacterium nucleatum (Fn) is commonly enriched in colorectal cancer (CRC) and associated with poor outcomes, though its mechanisms remain unclear. Our study investigated how Fn affects the tumor microenvironment through single-cell transcriptomic analyses of 42 CRC patient tissues, comparing Fn-positive and Fn-negative tumors. We discovered that Fn impairs IgA plasma cell development and secretory IgA (sIgA) production by disrupting communication with tumor-associated macrophages. Additional experiments in germ-free mice, together with our re-analysis of a publicly available single-cell RNA-seq data set from a CRC mouse model with an intact gut microbiome—both models having been orally gavaged with Fn—jointly validated the causal role of Fn in impairing sIgA induction. We identified a dysregulated IgA maturation (IGAM) module in Fn-positive patients, indicating compromised mucosal immunity and increased bacterial infiltration. This IGAM signature effectively stratified Fn-positive patients, suggesting potential for targeted therapeutic approaches. Our findings reveal that Fn disrupts sIgA production, increasing tumor microbial burden and worsening prognosis through chronic inflammation in Fn-positive CRC.

ARTICLE HISTORY







Received 28 April 2025
Revised 17 June 2025
Accepted 27 June 2025

KEYWORDS


Colorectal cancer;
Fusobacterium nucleatum;
secretory IgA; plasma cell
development; tumor
bacterial burden

Introduction

Colorectal cancer (CRC) is one of the most prevalent and deadly cancers worldwide.¹ Recent studies have underscored the potential role of the gut microbiome in both the development and progression of CRC.^{2–4} Specifically, *Fusobacterium nucleatum* (Fn) sensu lato, an anaerobic bacterium typically found in the oral cavity, has been identified as a key contributor to the pathogenesis of CRC.^{5–7} Previous research has demonstrated that Fn is highly enriched in colorectal cancer tissue and is associated with poor prognosis.^{8–10} These studies also suggest that Fn may promote tumorigenesis through various mechanisms, including evasion of the immune system, induction of a pro-inflammatory microenvironment, and activation of oncogenic pathways.^{3,11,12}

CONTACT Yoon Dae Han  acylyoon@yuhs.ac  Division of Colorectal Surgery, Department of Surgery, Yonsei University College of Medicine, Seoul, Republic of Korea; Han Sang Kim  modeerfhs@yuhs.ac  Yonsei Cancer Center, Division of Medical Oncology, Department of Internal Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea; Insuk Lee  insuklee@yonsei.ac.kr  Department of Biotechnology, College of Life Science and Biotechnology, Yonsei University, Seoul, Republic of Korea

*These authors contributed equally to this work.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/19490976.2025.2528428>.

© 2025 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Despite these significant discoveries, how Fn interacts with the tumor microenvironment to influence CRC outcomes remains unclear. Unraveling this complexity requires an in-depth single-cell-level analysis of these interactions. Single-cell RNA sequencing (scRNA-seq) has emerged as a powerful tool for studying the heterogeneity of cell populations within tumors, providing high-resolution information on the gene expression profiles of individual cells.^{13,14} This method is particularly effective in identifying specific immune cell types associated with the presence of Fn in colorectal tumors, thereby illuminating the interactions between Fn and the systemic immune response.

In this study, we aimed to investigate the effects of Fn on the tumor microenvironment by analyzing scRNA-seq data from both Fn-positive and Fn-negative samples. Our single-cell transcriptome analysis, corroborated by spatial transcriptome analysis, revealed that the presence of Fn impedes the maturation of secretory IgA (sIgA) by disrupting the communication between IgA plasma cells and tumor-associated macrophages (TAMs) involved in IgA induction and dysregulating a co-regulatory gene module essential for sIgA maturation. The causal role of Fn in sIgA dysfunction was confirmed by disrupted IgA plasma cell and macrophage development in Fn-administered mice. Furthermore, we observed increased bacterial infiltration within tumors, suggesting that Fn infection compromises the protective function of sIgA, potentially impacting prognosis. Additionally, we found that the activity of co-regulatory genes for sIgA maturation can stratify Fn-positive CRC patients for survival. These findings may provide valuable clinical implications for the treatment of Fn-positive CRC patients.

Materials and methods

Sample preparation

Single-cell isolation was performed using a Human Tumor Dissociation Kit (Cat# 130-095-929, Miltenyi Biotec, USA) following the manufacturer's instructions. Fresh surgical tissues were cut into small pieces ($< 1 \text{ mm}^3$) and placed into a gentleMACS C Tube (Cat# 130-093-237, Miltenyi Biotec, USA) containing a mixture of Enzyme H, Enzyme R, and Enzyme A in RPMI 1640 medium. The tube was then processed on a gentleMACS Dissociator (Cat# 130-093-235, Miltenyi Biotec, USA) using the GentleMACS programs h_tumor_01 and h_tumor_02, each followed by a 30-minute incubation at 37°C with rotation.

The dissociated tissues were filtered through a 70 μm MACS SmartStrainer (Cat# 130-098-462, Miltenyi Biotec, USA) to achieve a single-cell suspension. The filtrate was centrifuged at $300 \times g$ for 10 minutes at 4°C. The resultant cell pellet was resuspended in 4 mL of RPMI 1640 medium (Cat# 11875093, Thermo Fisher Scientific, USA), and dead cells were removed using Ficoll-Paque PLUS (Cat# 17-1440-03, GE Healthcare, USA). The cells were subsequently washed and resuspended in PBS supplemented with 0.1% bovine serum albumin (BSA, Cat# 15561020, Invitrogen, USA). Cell viability was assessed using the LIVE/DEAD™ Viability/Cytotoxicity Kit (Cat# L3224, Invitrogen, USA), and the single-cell suspension was further enriched for viability with a Dead Cell Removal Kit (Cat# 130-090-101, Miltenyi Biotec, USA).

Single-cell RNA sequencing (scRNA-seq) data generation and analysis

We generated single-cell transcriptome profiles using the 10x Chromium Single Cell 5' Gene Expression Dual Index library. Single-cell suspensions were processed on a Chromium system (10x Genomics, USA). For the discovery cohort consisting of 24 samples, we prepared 5' scRNA-seq libraries using the Chromium Next GEM Single Cell 5' Library and Gel Bead Kit v1.1 (PN-1000165). For the validation cohort of 18 samples, 3' scRNA-seq libraries were generated using the Chromium Next GEM Single Cell 3' v3.1 Library and Gel Bead Kit v3.1 (PN-1000121). The libraries were sequenced on an Illumina NovaSeq 6000 for the 5' scRNA-seq and an Illumina HiSeq 2000 for the 3' scRNA-seq. Sequencing reads were aligned to the human reference genome (GRCh38) using Cell Ranger software (v6.0.0).

For data processing, we utilized the Seurat R package (v4.0.0).¹⁵ Initially, cells expressing fewer than 200 features were removed. We then conducted cell quality control for each sample, evaluating mitochondrial read percentage, total read count, and feature count thresholds. Doublets, which represent two or more cells captured together in a single droplet, were identified and excluded using the DoubletFinder (v2.0.3) R package.¹⁶

The count matrix underwent log-normalization for read depth normalization via the *NormalizeData()* function with default parameters. We identified 3,000 variable features using the *FindVariableFeatures()* function with the selection method set to 'vst'. T-cell receptor and immunoglobulin variable genes, known for patient-specific variances, were classified as “bad genes” and removed from the list of 3,000 variable features and all subsequent analysis stages.

The cell cycle score, which quantifies the extent to which a cell's transcriptional program aligns with cells in the G1 or S/G2/M phase, was calculated using the *CellCycleScoring()* function. We then regressed this score against mitochondrial percentage while scaling the data with the *ScaleData()* function.

Principal Component Analysis (PCA) was performed using the *RunPCA()* function, inputting the previously determined variable features. We corrected for batch effects among samples by removing patient-specific signals with the *RunHarmony()* function from the Harmony (v1.0)¹⁷ R package, using the top 50 principal components (PCs) as input. A *knn* graph was constructed using the *FindNeighbors()* function with the corrected 50 PCs. Non-linear dimensional reduction was achieved through the *RunUMAP()* function using the same 50 coordinates.

Cell clustering was performed with the *FindClusters()* function at a resolution of 0.5. To identify different cell types, we annotated each cell cluster using marker genes. To ensure a clear immune cell profile, we removed clusters annotated as epithelial cells, endothelial cells, goblet cells, fibroblasts, and myofibroblasts, and those with high mitochondrial or cell-cycle scores or lacking marker gene expression. After filtering, we repeated the analysis with the remaining immune cells to refine the immune cell profile.

For detailed analysis at the cell subtype level, each cell type was processed individually. For B-cells and plasma cells, we employed 1,000 variable genes, ten PCs, and a resolution of 0.8, removing clusters lacking marker gene expression before repeating the processing steps. We identified a cluster of dendritic cells from macrophages and monocytes using ten PCs and a resolution of 0.8, annotating clusters expressing *CD1C* and *CLEC10A* as dendritic cells.

Fusobacterium nucleatum (Fn) tumor infection analysis via 16S rRNA sequencing

Total genomic DNA from surgical tumor tissues was extracted using QIAamp® DNA Mini Kit (Qiagen, USA) following the manufacturer's protocol. DNA concentration was measured with the Qubit dsDNA HS Assay Kit and Qubit Fluorometer (Invitrogen, USA). Library preparation for 16S rRNA gene sequencing followed the Illumina protocol. Amplicon PCR targeting the bacterial 16S rRNA V3-V4 region (primers Bakt_341F-805 R) was performed under the following conditions: 95°C for 3 min; 25 cycles of 95°C for 30 sec, 55°C for 30 sec, 72°C for 30 sec; final extension at 72°C for 5 min. PCR products were verified on an Agilent Bioanalyzer (DNA 1000 chip, USA) and quantified using a Qubit Fluorometer. Amplicons were purified using AMPure XP beads (Beckman Coulter, USA), followed by index PCR with Nextera® XT Index primers (Illumina, USA) under the same cycling conditions, except for 8 cycles. Indexed libraries were purified with AMPure XP beads, and library concentration was determined using the Quant-iT™ PicoGreen™ dsDNA Assay Kit (Invitrogen, USA). Sequencing was performed on the Illumina MiSeq platform (Macrogen, South Korea), generating 2 × 300 bp paired-end reads.

Raw paired-end reads were assembled using FLASH (v1.2.11)¹⁸ to merge overlapping sequences and improve quality. Assembled reads were length filtered with CD-HIT-OUT (v4.8.1),¹⁹ retaining sequences between 400–500 bp. Redundant sequences were clustered at 100% identity using CD-HIT-DUP, and chimeric sequences were removed. Secondary clusters were merged into primary clusters, and noise sequences were filtered out based on size thresholds. Non-chimeric representative reads were clustered into OTUs at 97% identity for species-level classification using a greedy algorithm. Taxonomic assignment of OTUs was performed with QIIME against the NCBI 16S rRNA database (version 20,211,127). Taxonomic abundance ratios were calculated, and samples were classified as Fn-positive if Fn exceeded 1% relative abundance.

Cellular composition analysis using Pearson residual

To investigate proportional changes between single-cell groups (e.g., Fn-positive and Fn-negative samples) across cell types, we evaluated deviations in the observed cell count for each cell type from the expected cell count using *Pearson residual*:

$$\text{Pearson residual } (r_{ij}) = \frac{O_{ij} - E_{ij}}{\sqrt{E_{ij}}},$$

where i and j represent indices for each group and cell type, respectively, and O and E represent the observed and expected cell counts, respectively. An expected cell count (E) for a group i of a cell type j was calculated by the following equation:

$$E_{ij} = \frac{T_i}{T_{tot}} \cdot T_j,$$

where T_{tot} , T_i , T_j represent total cell count for the entire data set, total cell count for a group i , and total cell count for a subset j , respectively. The sign of the residual indicates the direction of the difference of the observed count from the expected count (i.e., positive for increase and negative for decrease compared to expected count). *Pearson residual* (r) follows an approximately normal distribution; thus, scores larger than 2.58 or smaller than -2.58 are significant by $p < 0.01$. For a highly conservative statistical test along with *Bonferroni* correction for multiple hypothesis test adjustment, we used a significance threshold of $p\text{-value} = 0.00045$ ($p\text{-value}$ of 0.01 was divided by the number of subsets). Thus, we counted only $r > 3.5$ (increase) and $r < -3.5$ (decrease) for cell types with significant changes in their abundance between groups. The goodness of fit for all subsets was also evaluated by the chi-square statistic ($p\text{-value}$).

Determination of Fn-infection status in TCGA samples

Unmapped reads from the TCGA-COAD RNA-seq bam files, which do not align to the human reference genome, are expected to potentially include tumor-infiltrating microbial reads. To minimize false positives, we used GATK PathSeq,²⁰ an alignment-based algorithm, to filter out human reads by aligning them against the human reference genome. Subsequently, the remaining reads were aligned to the Human Reference Gut Microbiome (HRGM) database²¹ to obtain a microbial profile using the *PathSeqPipelineSpark* tool with default parameters (min-clipped-read-length = 31, min-score-identity = 0.9, identity-margin = 0.02). The TCGA samples exhibiting a normalized Fn score greater than 0.001—a threshold heuristically determined to maximize the effects of Fn on survival probability – were annotated as Fn-positive.

Survival analysis for TCGA-COAD samples

The count matrix and metadata were collected from TCGA-COAD datasets using TCGAbiolinks R package (v2.25.2).²² We obtained clinical data from TCGA Pan-Cancer Clinical Data Resource (TCGA-CDR),²³ and the samples without clinical data were filtered. To effectively elicit the mechanism of Fn affecting the tumor microenvironment, we only used TCGA data obtained from the right colon (ascending colon and cecum), where Fn infection affects survival probability most robustly. Samples with microsatellite instability-high (MSI-H) were excluded based on MSI status obtained from an external database,²⁴ leaving 121 TCGA samples for analysis.

We used progress-free interval (PFI) information from TCGA-CDR as clinical data for survival analysis. Survival curves were estimated with the *survfit()* function in the survival (v3.2–7) R package using the Kaplan-Meier method.²⁵ We used the *ggsurvplot()* function in the survminer (v0.4.9) R package to plot survival curves. The $p\text{-value}$ was calculated through the log-rank test with default parameters of the *ggsurvplot()* function. In survival analysis with a combination of the Fn label with the group of each cell type, the pairwise $p\text{-value}$ was calculated through the log-rank test with samples belonging to the groups.

RNA velocity analysis and estimation of cell state transition probability

The spliced and unspliced counts in each cell from the scRNA-seq data were estimated using the *run10x()* function of the Velocyto package (v0.17.17).²⁶ RNA velocity was analyzed separately for each cell type to compare differentiation patterns between the Fn groups, utilizing the Dynamo Python package (v1.0.0).²⁷ This recently released tool addresses the fundamental limitations of traditional RNA velocity analysis by enhancing the measurement of RNA velocity and enabling the determination of transition probabilities between cell types. We adhered to the standard workflow outlined in the Dynamo user guide to estimate RNA velocity. The transition probability matrix was derived by annotating observations within the data object. To determine the transition probability to a specific cell type, we summed the transition probabilities from a given cell to all cells of that type, providing a clearer understanding of cellular dynamics across different conditions.

Cell-cell interaction estimation

The CellChat (v1.6.1)²⁸ R package was used to infer interactions between whole cell types. Subsequently, interactions between B-cells and myeloid cells were investigated. The default CellChat database was utilized, and only significant interactions in at least one Fn group were visualized; if the interactions were insignificant for a particular group, no dot was plotted on the dot plot. The biological roles of the ligand-receptor pairs involved in the IgA induction were annotated based on previous studies.^{29,30}

Spatial transcriptomic data analysis

We obtained publicly available 10x Visium data from CRC tumors located in the right colon.³¹ We aligned the raw fastq files to the same reference genome as the scRNA-seq data using 10x Genomics Space Ranger (v3.0.0). To label Fn infection status, we followed the same method described in the section of tumor microbiome profiling of TCGA samples, selecting samples with the highest detection of Fn reads as Fn-positive and those with none as Fn-negative for further analysis. We used Scanpy (v1.9.8)³² to perform the basic processing of the Visium data. The raw count matrix was normalized using *calculate_qc_metrics()* function followed by log1p transformation. To reduce noise in the Visium data, we excluded mitochondrial genes, immunoglobulin variant genes, and T cell receptor variant genes, consistent with the scRNA-seq data.

To infer cell type composition of each spot, we used Cell2location (v0.1.3)³³ to build a reference model based on our scRNA-seq data. To accurately reflect the histological characteristics within the tumor, the reference model included both immune and stromal cells from the scRNA-seq data. The spatial mapping model training followed the tutorial with recommended parameters (*N_cells_per_location* = 5, *detection_alpha* = 20, *max_epochs* = 30000). We used *q05_cell_abundance* as a conservative estimate of cell abundance in each spot. For further analysis, we labeled each spot as containing a particular cell type if the Cell2location estimate for that cell type was higher than 0.1. To measure the distance between cell types, we collected the spots containing each cell type and calculated the distance between coordinates on a spot-by-spot basis. For measuring ligand-receptor interactions in the 10x Visium data, we used COMMOT (v0.0.3)³⁴ with the optimal transport algorithm. For the ligand-receptor database, we utilized the CellChat DB.²⁸ When scoring CCI between cell types, we only used the CCI scores assigned to spots where the Cell2location estimate was higher than 0.1 for each cell type.

Trajectory-based gene clustering and differential expression analysis of B-cell

For trajectory-based analysis, B-cells and plasma cells were analyzed separately using the Slingshot R package (v2.2.0).³⁵ This tool was utilized to construct a lineage trajectory from naïve B-cell to IgA-secreting plasma cell, reflecting the IgA maturation pathway. Following lineage construction, the tradeSeq R package (v1.8.0)³⁶ was employed to identify differentially expressed genes between Fn groups associated with this IgA maturation lineage. We applied a negative binomial generalized additive model using the

fitGAM() function from *tradeSeq*, incorporating the count matrix, pseudotime, and cell weight information (*nknobs* = 6). The *Fn* labels of the cells were integrated into the model using the condition parameter of *fitGAM()* to fit a condition-specific smoother for each lineage. The *conditionTest()* function was then used to identify genes with distinct expression dynamics across *Fn* labels. Genes meeting the criteria of a *p*-value less than 0.05 and a *waldStat* value greater than 20 were selected, ensuring the exclusion of ribosomal and mitochondrial genes, often considered as ‘bad genes.’ For gene clustering, we used the *clusterExpressionPatterns()* function (*nPoints* = 20), and labeling was performed with the *primaryCluster()* function of the *clusterExperiment* R package (v2.14.0). A gene cluster containing marker genes of IgA-secreting plasma cells, including *PRDM1*—a known marker of plasma cell maturation – was selected.

We employed the *singscore* R package (v1.10.0)³⁷ for robust rank-based expression scoring of the IGAM gene set in TCGA data. TCGA samples with gene set expression scores higher than their median were classified into IGAM-high activity groups. Samples with scores below the median were classified into IGAM-low activity groups for detailed analysis.

Co-expression network analysis for prioritizing core genes of the IGAM module

Initially, we utilized the *metacell*³⁸ R package to reduce the sparsity of scRNA-seq data and construct a robust co-expression network. We divided the scRNA-seq matrix of IgA plasma cells into two separate matrices for *Fn*-negative and *Fn*-positive cells to compare the co-expression networks between the two *Fn* groups. We refined the raw matrix by removing mitochondrial genes, ribosomal genes, and previously identified ‘bad genes’ such as *NEAT1*, *TMSB4X*, and *TMSB10*.³⁸ Cells with fewer than 500 read counts were also excluded. Following a guided tutorial, we processed the data using recommended parameters to obtain the *metacell* expression matrix. We employed *mcell_gset_filter_varmean()* and *mcell_gset_filter_cov()* functions to select informative genes (*T_vm* = 0.08, *T_tot* = 100, *T_top3* = 2), *mcell_add_cgraph_from_mat_bknn()* function to construct a balanced cell graph (*K* = 100), *mcell_coclust_from_graph_resamp()* (*min_mc_size* = 20, *p_resamp* = 0.75, *n_resamp* = 500) and *mcell_mc_from_coclust_balanced()* (*K* = 30, *min_mc_size* = 30, *alpha* = 2) functions to resample and generate the co-clustering graph, and *mcell_mc_split_filt()* (*T_lfc* = 3) function to remove outlier cells within the *metacell*. After obtaining the refined *metacell* expression matrix, we constructed a co-expression network by calculating the Pearson correlation coefficient (PCC) using the *cor()* function from the R stats package.

To prioritize genes in the IGAM module, we analyzed the PCC difference between *Fn*-negative and *Fn*-positive gene pairs. Gene pairs were sorted by PCC difference, and a two-sided Wilcoxon test was used to prioritize genes with significant differences. Genes with a *p*-value lower than 0.05 were selected as the core components of the IGAM module.

Tumor microbiota analysis

To quantify the bacterial burden within tumor tissues, we analyzed read counts for tumor-infiltrating bacteria using GATK PathSeq. We obtained the number of unambiguous reads aligned to bacteria at the domain level. To adjust for sequencing depth bias, we normalized the bacterial read counts by dividing by the total number of reads in the original bam file, then multiplied by a scale factor of 1 million, and converted the results to a logarithmic scale.

For species-specific analyses, we employed a similar normalization method but increased the scale factor to 10 million to account for the typically smaller number of reads aligned to each species. This adjustment allows for a more precise comparison of tumor bacterial abundance between *Fn* groups.

To compare the IgA binding probabilities of tumor bacterial species between samples, we applied previously established binding probabilities to each species. We averaged probabilities obtained from multiple samples within each group. For consistent comparison, we adjusted the distribution of values by replacing infinite values with 20 and zero values with 0.0001, following the approach used in a previous study.

Analysis of validation cohort tumor samples

For estimating cell-cell interactions, we employed CellPhoneDB (v2.0.0),³⁹ and enhanced the analysis by integrating ligand-receptor information from SingleCellSignalR (v0.99.24)⁴⁰ and CellTalkDB (v1.0)⁴¹ to expand our custom database.

RNA velocity measurements were conducted using Dynamo, following the procedures outlined previously. Additionally, we generated bulk RNA-seq data for the same samples. The Fn infection profile was determined from annotations based on 16S rRNA-seq data.

To align our RNA-seq data with the TCGA standards, the FASTQ files were processed using the STAR aligner, adhering to the RNA-seq alignment workflow provided by the NCI Genomic Data Commons (GDC).⁴² Tumor-infiltrating bacterial profiling was then carried out using PathSeq, following the same methods previously described. This also included a comparison of read abundance using the same normalization and analytical approaches.

Cultivation of *Fusobacterium nucleatum* (Fn)

Fusobacterium nucleatum (ATCC 25586) was obtained from the American Type Culture Collection (ATCC, USA). Fn was cultured in Gifu Anaerobic Medium (GAM, Kisanbio, South Korea) broth under anaerobic conditions (20% CO₂, 5% H₂, 75% N₂) at 37°C in a Coy chamber (Coy Laboratory Products, USA). For solid cultures, Fn was grown anaerobically on Brucella agar plates with 5% sheep blood at 37°C in the Coy chamber.

Fn oral administration in germ-free (GF) mice

Six-week-old female GF C57BL/6 mice were used for all experiments. Ten mice were randomly divided into two groups (5 mice per group): a control group and an Fn group. All animal procedures were approved by the Institutional Animal Care and Use Committee (IACUC, #2022–0286) of Yonsei University College of Medicine and conducted in accordance with the Public Health Service Policy on Humane Care and Use of Laboratory Animals. Yonsei University is accredited by AAALAC International (#001071). GF mice were housed in individually ventilated cages (IVC) under controlled conditions (temperature: 22–23°C; light cycle: 12 h light/12 h dark). Sterile food and water were provided *ad libitum*.

Prior to oral gavage, GF mice were fasted for 2 hours. The Fn group received 1×10^9 colony-forming units (CFU) of Fn in 200 µL of PBS, administered orally three times per week for 5 weeks. The control group received 200 µL of PBS alone on the same schedule. Mouse body weight was measured three times weekly until sacrifice. At the end of the experiment, the mice were sacrificed, and colon tissues were collected from both the control ($n = 5$) and Fn ($n = 5$) groups.

Mouse colon and cecum tissue processing

Colon tissues, from proximal to distal sections, were cut open longitudinally and washed twice with ice-cold PBS. The tissues were then immersed in ice-cold RPMI-1640 medium for further processing and dissociation. The cecum was isolated from the gastrointestinal tract by horizontal dissection at the junctions between the cecum and the ileum, and between the cecum and the proximal colon. The cecum weight was measured for each group.

Preparation of single-cell suspensions from mouse colon tissue for scRNA-seq library

Dissected colon tissues from PBS control ($n = 5$) and Fn-treated ($n = 5$) GF mice were pooled for each group. The tissues were cut into ~5 mm fragments and washed three times with ice-cold PBS. Pre-digestion was performed twice for 20 min at 37°C in Hanks' Balanced Salt Solution (HBSS, Gibco, USA) without Ca²⁺ and Mg²⁺, supplemented with 5 mM EDTA, 5% fetal bovine serum (FBS), and 1 mM DTT, using a rotator mixer (Miltenyi Biotec, USA). The tissue fragments were then transferred

to C-tubes (Cat# 130–096–334, Miltenyi Biotec, USA) containing HBSS with Ca^{2+} and Mg^{2+} (Gibco, USA) and an enzyme mix (enzymes D, R, and A) from the Mouse Lamina Propria Dissociation Kit (Cat# 130–097–410, Miltenyi Biotec, USA). Dissociation was performed at 37°C for 30 min using a gentleMACS Octo Dissociator with heaters (Cat# 130–134–029, Miltenyi Biotec, USA). Following incubation, ice-cold PB buffer (PBS with 0.5% bovine serum albumin) was added, and the samples were passed through a 100 μm cell strainer. The cell suspensions were centrifuged at $300 \times g$ for 10 min at 4°C, and the pellet was resuspended in ice-cold PB buffer. The samples were filtered through a 70 μm cell strainer and rinsed once with PB buffer. Viable cells were purified using the Dead Cell Removal Kit (Cat# 130–090–101, Miltenyi Biotec, USA). Single-cell suspensions were prepared for scRNA-seq library construction using the 10X Chromium Next GEM Single Cell 5' RNA Library v2 kits.

Analysis of mouse-derived scRNA-seq data

The libraries were sequenced on an Illumina NovaSeq 6000, and sequencing reads were aligned to the mouse reference genome (GRCm39) using Cell Ranger (v9.0.0). The same processing steps—quality control, doublet removal, normalization, batch-effect correction, cell typing, and immune-cell filtering—were applied to both the germ-free mouse dataset and the publicly available CRC mouse dataset with a normal gut microbiome. B cells and IgA plasma cells were isolated and ordered along pseudotime using Monocle (v2.30.1),⁴³ with DDRTree (v0.1.5) used for visualization. Spliced and unspliced read counts were generated with Velocyto, and RNA velocity was calculated with Dynamo.

In the colon tissue of the germ-free model, 2,593 B cells and IgA plasma cells were analyzed. Among these, 288 IgA plasma cells were assigned to state 6, while the remaining cells were divided evenly across five pseudotime states (approximately 461 cells per state) for RNA velocity analysis. In the tumor tissue of the CRC mouse model with a normal gut microbiome, 119 GC B cells, 1,158 B cells, and 603 IgA plasma cells were divided evenly into six pseudotime states (approximately 313 cells per state).

Results

Fn infection modifies the tumor immune landscape and is associated with reduced survival in CRC

We conducted single-cell transcriptomic analysis to thoroughly examine the cellular and molecular elements of the tumor microenvironment in CRC. We compared 12 Fn-positive and 12 Fn-negative patients, with Fn-infection status determined through 16S rRNA analysis. Our discovery cohort of 24 tumor samples was then subjected to scRNA-seq to assess cellular composition and gene expression profiles (Figure 1(a), Supplementary Table S1, S2a). Given the role of immune system in bacterial infections, we focused on changes in the immune microenvironment following Fn infection. After identifying 22 immune cell types from 109,204 cells (Figure 1(b), Supplementary Figure S1a, Supplementary Table S2b), we observed an unbiased distribution across patients, ensuring data integrity through proper pre-processing and batch corrections (Supplementary Figure S1b).

We investigated the impact of Fn infection on immune cell composition. Our analysis revealed significant changes in the composition of intratumor immune cells between Fn-positive and Fn-negative groups. The majority of immune cell types showed a marked decrease in abundance ($r < -4$) in the Fn-positive group (Figure 1(c), Supplementary Table S2c). Conversely, immune cell types involved in immunoglobulin (Ig)-mediated immunity, such as naive B cells, germinal center (GC) B cells, and IgG plasma cells, showed a significant increase in abundance ($r > 4$) in the Fn-positive group. These findings suggest that Fn infection may lead to substantial shifts in the intratumor immune microenvironment in CRC.

Previous research has linked Fn infection with poorer prognosis in CRC, especially in tumors located in the right colon.^{5,9} To verify the prognostic impact of Fn, we analyzed bulk RNA sequencing data from right-sided CRC tumors in The Cancer Genome Atlas Colon Adenocarcinoma (TCGA-COAD). We determined Fn status, identifying 116 Fn-positive and 63 Fn-negative patients (Figure 1(d)). To minimize bias, we excluded microsatellite instability-high (MSI-H) samples, leaving 75 Fn-positive and 46 Fn-negative patients for survival analysis (Supplementary Table S3). Consistent with prior reports, Fn-positive tumors

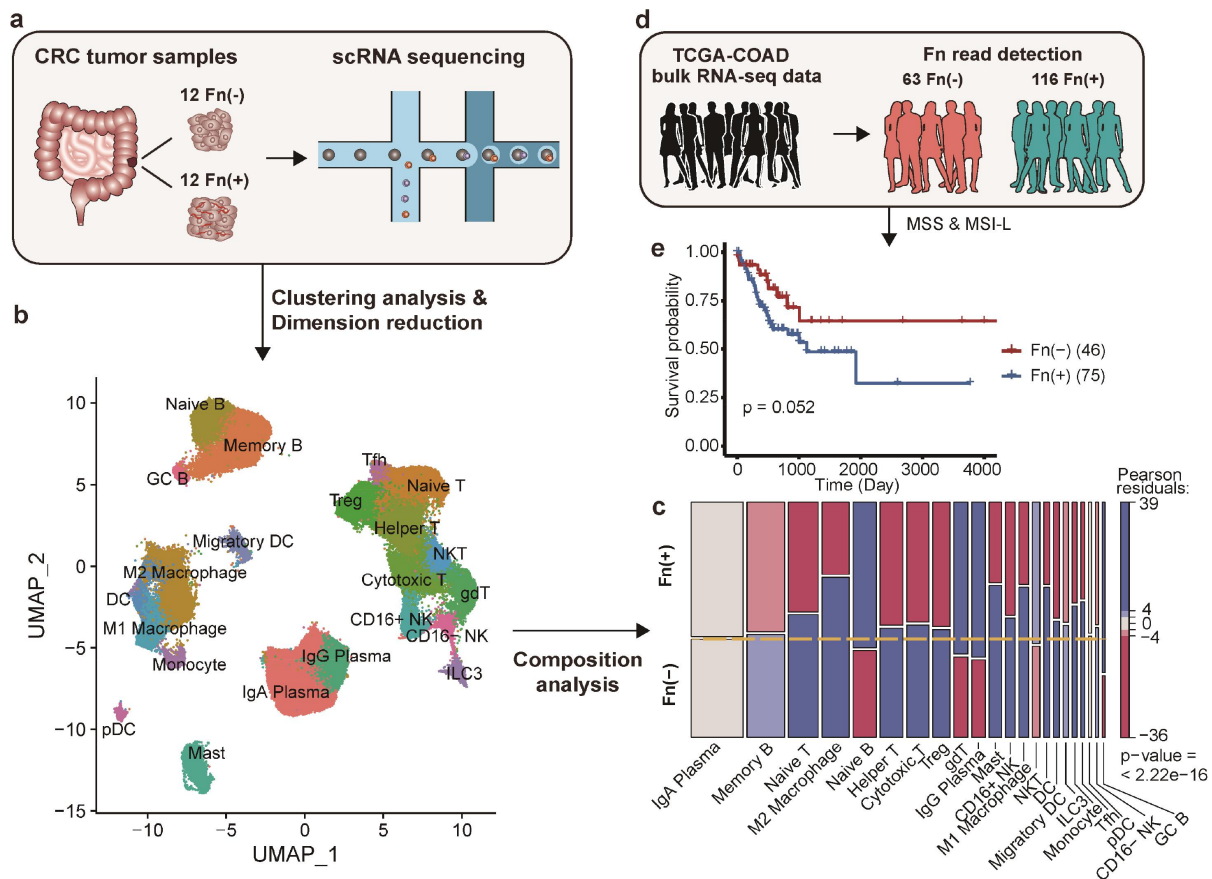


Figure 1. Single-cell RNA sequencing analysis shows Fn-associated changes in immune composition of CRC tumor. (a) Tumor tissues from 24 patients with CRC (12 Fn-positives and 12 Fn-negatives) were subjected to the single-cell RNA sequencing analysis. (b) UMAP visualization of immune cells from 24 patients after pre-processing and batch correction. A total of 22 immune cell types were obtained. (c) Mosaic plot of immune cell clusters, each of which is composed of Fn-positive and Fn-negative groups. Red indicates significant decrease in abundance, whereas blue indicates significant increase in abundance ($|\text{Pearson residual}| > 2$ for $p\text{-value} < 0.05$, $|\text{Pearson residual}| > 4$ for $p\text{-value} < 0.0001$, one-sided chi-square test). (d) Bulk RNA sequencing data were collected from TCGA-COAD cohort and their Fn-infection was determined by detection of Fn sequence reads. (e) Progression-free survival curves by Kaplan-Meier estimation using tumors located in the right colon of the TCGA-COAD patients for different Fn-infection status.

showed worse progression-free survival, though it narrowly missed statistical significance (log-rank test, $p = 0.052$, Figure 1(e)). These findings suggest that changes in intratumor immunity may influence the prognostic effects of Fn infection in CRC.

Development of IgA-secreting plasma cells is impeded in Fn-positive CRC

Our single-cell composition analysis revealed that IgA plasma cells occupy the highest proportion in CRC tumor samples. Furthermore, we noted a significant increase in naive B and GC B cells in Fn-positive CRC samples compared to those without Fn infection (Figure 1(c)). These B cells are typically found in the Peyer's patches of the gut-associated lymphoid tissue (GALT), where they may be activated to differentiate into IgA plasma cells.⁴⁴ IgA plasma cells are crucial for mucosal immunity against gut commensal bacteria, producing sIgA that binds polyreactively with the gut microbiota.⁴⁵ This binding helps prevent pathogen colonization and invasion and facilitates bacterial neutralization. Given the importance of IgA plasma cells in gut mucosal immunity, our investigation focused on these cell types. B cells within the GALT differentiate into IgA plasma cells through interactions with other immune cells, including T cells and macrophages.^{46,47}

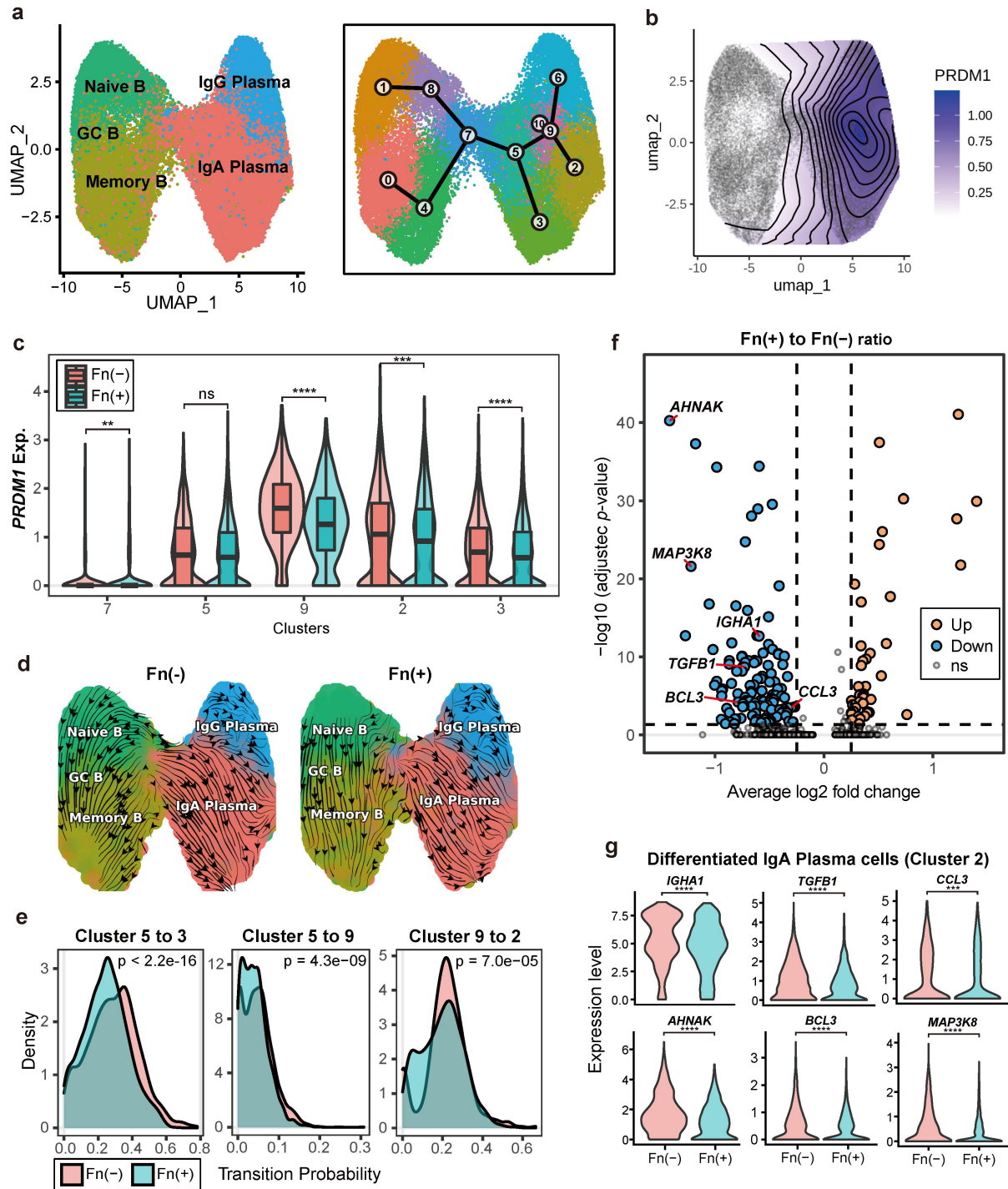


Figure 2. Impairment of IgA plasma cell development by Fn infection in CRC. (a) UMAP plot of subsets of B cells and plasma cells labeled based on cell types annotated (left), and 11 clusters obtained using the Louvain algorithm (right). Trajectory illustrating differentiation from B cells to plasma cells based on clusters. (b) Contour plot visualizing *PRDM1* expression on UMAP. (c) Boxplot comparing *PRDM1* expression in IgA plasma cells, grouped by clusters and according to the direction of differentiation, between Fn groups. The significance of differences in expression levels was estimated using a two-sided Wilcoxon rank sum test. (d) RNA velocity of B cells and plasma cells in Fn-negative (left) and Fn-positive (right) tumors. (e) Density plot of the transition probability during differentiation along the trajectory within IgA plasma cells. The statistical significance of the probability distribution was measured using the Kolmogorov-Smirnov test. (f) Volcano plot depicting the gene expression levels of differentiated IgA plasma cells (cluster 2) in Fn-negative and Fn-positive tumors. (g) Violin plots depicting differences in expression levels of genes of interest. Significance was tested using a two-sided Wilcoxon rank sum test. P -values were adjusted using the Family-wise Type 1 Error Rate (FWER). ns, $P \geq 0.05$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

We employed dimension reduction techniques to reveal distinct separations of plasma cells from B cells (Figure 2(a) left). We constructed trajectories from naive B cells to IgA plasma cells, capturing transcriptional changes during differentiation (Figure 2(a) right). Expression levels of *PRDM1* (encoding Blimp-1), a well-known marker of plasma cell maturation,^{48,49} confirmed the identity of cell subsets at different stages of differentiation (Figure 2(b)). Notably, *PRDM1* expression was significantly lower in Fn-positive CRC compared to Fn-negative CRC across all differentiation stages (Wilcoxon rank sum test, $p < 0.001$; Figure 2(c)), suggesting that Fn infection hinders IgA plasma cell development. To validate the observed retardation, we conducted RNA velocity analysis, which revealed a reduced likelihood of progression along the IgA plasma cell differentiation pathway in Fn-positive CRC (Kolmogorov-Smirnov test, $p < 0.001$; Figure 2(d–e)).

To determine whether the observed impairment in IgA plasma cell differentiation is influenced by tumor anatomical location, we stratified the discovery cohort into left- and right-sided tumors and reevaluated *PRDM1* expression and RNA velocity dynamics. In both left- and right-sided tumors, *PRDM1* expression was consistently lower in Fn-positive samples compared to Fn-negative samples (Supplementary Figure S2a–b). Similarly, the transition probability along the IgA plasma cell differentiation trajectory was significantly reduced in the Fn-positive group regardless of tumor location (Supplementary Figure S2c–d). These findings indicate that the Fn-associated impairment of IgA plasma cell development is not restricted to a particular tumor site, but rather occurs consistently across different anatomical locations in the colon.

To investigate the functional impact of Fn infection on differentiated IgA plasma cells (cluster 2), we identified differentially expressed genes (DEGs) between Fn-positive and Fn-negative groups (Figure 2(f–g)). We confirmed the hampered development of IgA plasma cells by observing reduced *IGHA1* expression in Fn-positive group. Notably, *TGFBI* expression in IgA plasma cells also decreased in Fn-positive group. Previous studies have shown that TGF- β signaling induces differentiation from B cells to IgA plasma cells and promotes IgA secretion.^{50,51} *AHNAK* and *BCL3* were identified as regulators that enhance TGF- β signaling through interaction with the Smad-dependent canonical TGF- β signaling pathway.^{52,53} *MAP3K8* has been demonstrated to be another positive regulator of TGF- β signaling pathway.⁵⁴ TGF- β , an immunosuppressive cytokine, has been shown to recruit TAMs and promote their immunosuppressive phenotypes, inducing M2 polarization.^{55,56} Additionally, *CCL3*, which encodes a chemokine involved in the recruitment and polarization of M2 macrophages,^{57,58} showed reduced expression in IgA plasma cells of Fn-positive group. These results collectively suggest that Fn infection reduces TGF- β production by IgA plasma cells, impeding the recruitment of M2 macrophages necessary for IgA plasma cell maturation.

IgA induction through macrophage interaction is disrupted in Fn-positive CRC

Since the DEG analysis suggested a reduced capability of IgA plasma cells to recruit macrophages in Fn-positive CRC, we next examined whether Fn infection also alters macrophage functions related to their interaction with IgA plasma cells. To further specify the myeloid cell subsets, we sub-clustered populations including monocytes, macrophages, and dendritic cells (Figure 3(a)). Notably, we identified subclusters of TAMs,⁵⁹ such as lipid-associated TAMs (LA-TAMs) and regulatory TAMs (Reg-TAMs), which are enriched in antigen processing and presentation pathways and canonical immunosuppressive M2-like pathways (Figure 3(b)). Observing cell transition patterns through RNA velocity, we noted a polarization toward these TAM clusters from surrounding cell clusters in Fn-negative tumors (Figure 3(c) left). However, in Fn-positive tumors, we observed a reduced probability of cell transitions, indicating a reduced M2 polarization of macrophages (Figure 3(c) right, 3(d)).

Next, we investigated whether cell-cell interactions between B cells and myeloid cells are altered by Fn infection. Among myeloid cell subsets, M2 macrophages showed most noticeable reduction in interactions with B cell subsets, including IgA plasma cells, in Fn-positive tumors (Figure 3(e)). DEG analysis of M2 macrophages revealed the reduced expression of *CCR1*, which encodes a receptor for the CCL3 chemokine involved in the recruitment of M2 macrophages^{57,58} in the Fn-positive tumors (Figure 3(f)). Previously, we observed reduced expression of *CCL3* in IgA plasma cells of Fn-positive group (Figure 2(g)). Therefore, the reduced expression of *CCL3* in IgA plasma cells and that of *CCR1* in M2 macrophages aligns with the reduced interactions between these cell types.

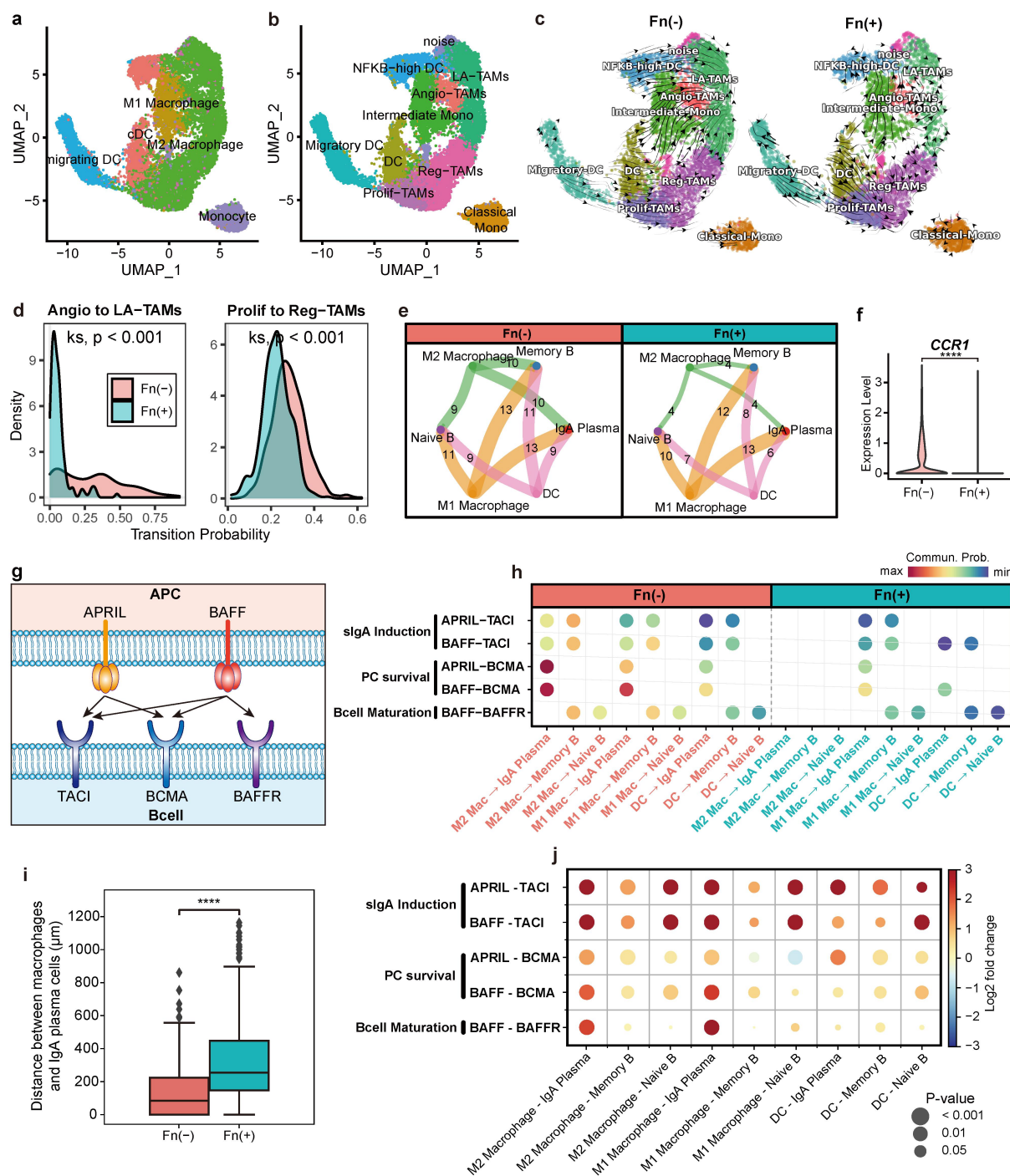


Figure 3. Disruption of cell-cell interactions for IgA induction in Fn-positive CRC. (a) UMAP plot of myeloid cell subsets labeled based on cell types. (b) UMAP plot of clusters of myeloid cells annotated to subtypes of Tumor-Associated Macrophages (TAMs). (c) RNA velocity of myeloid cells in Fn-negative (left) and Fn-positive (right) tumors. (d) Density plot showing the probability of transition toward cell subtypes involved in antigen processing and presentation (LA-TAMs, Reg-TAMs), from adjacent cell clusters (Angio-TAMs, Prolif-TAMs). The statistical significance of the probability distribution was determined using the Kolmogorov-Smirnov test. (e) Circle plot displaying the number of ligand-receptor interactions detected between each cell type. (f) Violin plots depicting differences in expression levels of *CCR1* in M2 macrophage. (g) Schematic illustrating ligand-receptor interactions involved in IgA plasma cell maturation. (h) Bubble plot representing cell-cell interactions (CCI) between myeloid cells, B cells, and IgA plasma cells. The color of the bubbles indicates the communication probability, with only significant interactions (permutation test, $P < 0.001$) being displayed. The known functions of each ligand-receptor pair are listed on the left side. (i) Boxplot comparing the distances between M2 macrophages and IgA plasma cells measured based on spatial transcriptome data. The statistical significance was determined using t-test. (j) Bubble plot representing CCI between myeloid cells, B cells, and IgA plasma cells based on spatial transcriptome data. The color of the bubbles represents the \log_2 fold change of the mean CCI scores estimated for Fn-negative over Fn-positive tumors. Positive values indicate higher scores in Fn-negative tumors. The size of the bubbles represents the statistical significance determined using a t-test. **** $P < 0.0001$.

Subsequently, we evaluated changes in ligand-receptor interactions between APRIL and BAFF, ligand molecules expressed on myeloid cells, and TACI, BCMA, and BAFFR, receptors expressed on B cells (Figure 3(g)). The BAFF-BAFFR interaction mediates B cell maturation.⁶⁰ BCMA activation by either APRIL or BAFF is essential for plasma cell survival,⁶¹ while TACI, which interacts with either APRIL or BAFF, is critical for sIgA induction.²⁹ We found that the interaction between the B-cell receptor TACI and ligands from antigen-presenting cells (APRIL, BAFF) was either absent or inhibited in the Fn-positive group (Figure 3(h)). Similarly, interactions involving BCMA, crucial for plasma cell survival, and BAFFR, involved in the maturation of B cells other than plasma cells, were also inhibited.^{62,63}

To validate whether the disruption of cell-cell communications between M2 macrophages and IgA plasma cells indeed occurs in the spatial context of tumor tissue, we analyzed 10x Visium spatial transcriptomic data for CRC tumors available from public sources (Supplementary Figure S3–4). Through deconvolution analysis using the scRNA-seq data from our study, we assigned cell types to each spot of the array. We observed that the distance between IgA plasma cell spots and M2 macrophage spots was greater in an Fn-positive sample compared to an Fn-negative sample, supporting the impairment in recruiting M2 macrophages to IgA plasma cells (Figure 3(i)). Furthermore, when examining the ligand-receptor interactions measured based on their spatial coordinates, we found a significant decrease in these interactions in Fn-positive tumors compared to Fn-negative tumors (Figure 3(j)), which is consistent with the results based solely on single-cell expression information. These results collectively suggest that Fn infection disrupts cell-cell communications between IgA plasma cells and M2 macrophages, which is required for IgA induction.

A gene network for IgA maturation is dysregulated in Fn-positive CRC

Disrupted IgA induction signaling due to Fn infection may lead to dysregulation of downstream molecular programs essential for secretory IgA (sIgA) production. Using single-cell transcriptome data, we analyzed the developmental process from naïve B cells to mature IgA-secreting plasma cells, focusing on cells from clusters 1, 8, 7, 5, 9, and 2 (Figure 2(a) right) for pseudotime analysis (Figure 4(a)). Then, we performed differential expression analysis along the pseudotime between Fn-positive and Fn-negative groups. We clustered genes displaying similar differential expression patterns between two Fn infection status. Among major clusters of differentially expressed genes, we chose a cluster of 55 genes including *PRDM1*, a marker for maturation of IgA plasma cell, which we termed the IgA maturation (IGAM) module (Figure 4(b), Supplementary Table S4), for the downstream analysis.

GO term enrichment analysis of the IGAM module revealed “protein N-linked glycosylation” as a critical pathway, significantly affecting sIgA-mediated mucosal immunity^{64,65} (Figure 4(c)). This glycosylation process is crucial for the interaction between IgA and the polymeric immunoglobulin receptor (pIgR), which facilitates the transcytosis of dimeric IgA across mucosal epithelial cells, crucial for IgA secretion into the lumen. The glycosylated structures on IgA also directly interact with both commensal and pathogenic bacteria, suggesting that Fn infection disrupts IgA glycosylation through disturbance of the IGAM module regulation, leading to reduced IgA secretion and compromised bacterial defense in the gut lumen.

Further analysis using a network biology approach showed that coregulation within the IGAM module significantly diminished in Fn-positive groups compared to Fn-negative ones (Figure 4(d)). We prioritized genes within the module based on the decrease in expression correlation in the Fn-positive group (Figure 4(e–f), Supplementary Figure S5), revealing significant dysregulation in 16 of the 55 genes ($p < 0.05$, one-sided Wilcoxon rank sum test). These include *JCHAIN*, a protein component of sIgA responsible for the monomeric IgA joining,⁶⁶ and *MEF2B*, a positive regulator of *JCHAIN* expression.⁶⁷ Significant dysregulation was also observed in *TNFRSF17* (also known as *BCMA*), the receptor involved in the T cell independent IgA induction pathway, and *MZB1*, a cochaperone crucial for proper Ig heavy chain synthesis.⁶⁸ *XBPI*, another dysregulated gene, regulates plasma cell differentiation by activating secretory machinery via unfolded protein response in endoplasmic reticulum (ER).^{69–72} Especially, *XBPI* mitigates ER stress through the upregulation of ER molecular chaperones and ER-associated protein degradation components.⁷³ *ERLEC* plays a pivotal role in ER-associated protein degradation and ER homeostasis.⁷⁴ *TXNDC5* is also a member of the protein disulfide isomerase family in ER, facilitating the cell proliferation and survival.⁷⁵

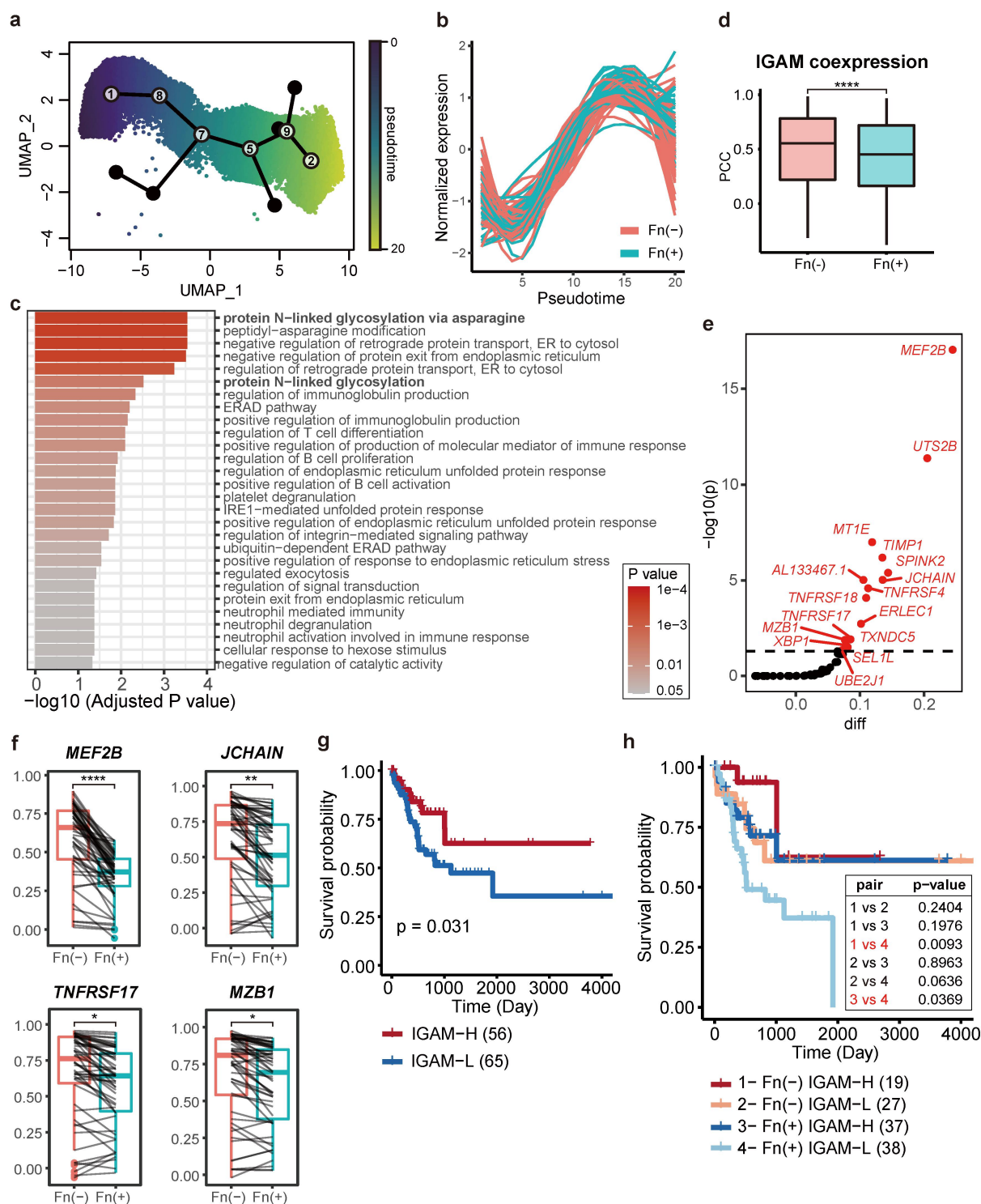


Figure 4. Dysregulation of a gene network for IgA maturation by Fn infection. (a) UMAP visualization depicting the developmental trajectory from naive B cells to mature IgA plasma cells (identified as cluster 1, 8, 7, 5, 9, and 2). Pseudotime progression is indicated by a color gradient from dark purple (early) to bright yellow (late). (b) Graph showing changes in normalized expression over pseudotime for a cluster of 55 genes that exhibit differential expression patterns along the trajectory and between Fn infection statuses. (c) Bar graph showing significantly enriched GO biological process terms ($P < 0.05$ by Fisher's exact test) for a cluster of 55 genes. (d) Boxplot comparing the Pearson correlation coefficient (PCC) of expressions within the IGAM gene set across Fn groups. Statistical significance was assessed using a two-sided Wilcoxon rank sum test. (e) Dot plot representing genes prioritized by a decrease in expression correlation with other network member genes upon Fn infection. Genes were sorted by the mean difference in PCC for each Fn group, adjusted using $-\log_{10}(p)$ from a one-sided Wilcoxon rank sum test. A p -value cutoff of 0.05 was applied. (f) Boxplots illustrating correlation changes in significantly dysregulated genes involved in sIgA maturation and function, with statistical significance

Stratification of Fn-positive CRC patients by IGAM activity predicts prognosis

The impact of Fn infection on clinical outcomes may vary due to differences in host genomes and intratumor immune environments. Therefore, we hypothesized that variations in the activity of the IGAM module could further stratify patients with Fn infection, potentially leading to more precise medical interventions for CRC. To explore this hypothesis, we utilized the TCGA-COAD cohort and conducted survival analysis based on IGAM activity levels, as determined by the mean expression of module genes (Methods). Our analysis revealed a significant separation between the survival curves of patients with high and low IGAM activity, irrespective of their Fn infection status. Patients with higher IGAM activity consistently showed better prognosis (Figure 4(g)). Furthermore, reduced IGAM activity was significantly associated with poorer survival in both univariate and multivariate Cox regression models, independent of Fn status, tumor stage, lymph node involvement, and microsatellite instability (Supplementary Table S5). This finding aligns with our earlier observation that the IGAM module is dysregulated in Fn-positive patients, who generally exhibit worse prognoses. Moreover, these results suggest that IGAM activity could serve as an independent prognostic biomarker in CRC.

Considering the potential association between the two prognostic biomarkers—Fn infection and IGAM activity—we developed a stratification strategy by combining these biomarkers. We divided the patients into four groups based on their Fn infection status and IGAM activity levels and compared their survival probabilities. Remarkably, we observed significantly different survival rates between two groups of Fn-positive patients differentiated by their IGAM activity levels (Figure 4(h)). Notably, Fn-positive patients with high IGAM activity exhibited survival rates comparable to those of Fn-negative patients with low IGAM activity. This suggests that enhancing IGAM activity in Fn-positive patients could potentially elevate their survival probabilities to levels observed in Fn-negative CRC patients, offering a promising avenue for tailored therapeutic strategies.

Bacterial burden within tumor increases in Fn-positive CRC

Our single-cell analysis indicated that Fn infection was associated with diminished production of secretory IgA (sIgA), a molecule which normally binds to commensal microorganisms, preventing their adhesion and infiltration into the mucosal layer and colon tissue. Consequently, we hypothesized that the reduced production of functional sIgA in Fn-positive CRC leads to increased colonization and infiltration of gut commensals into tumor tissue, thereby inducing chronic inflammation and resulting in poor prognosis. This hypothesis aligns with a growing body of research highlighting the clinical impact of the intratumoral microbiome.^{76,77} Specifically, Fn is known to coaggregate with other bacterial species, forming biofilms that promote host invasion.^{78,79}

To investigate this, we analyzed tumor bulk RNA sequencing data, using read counts aligned to human gut prokaryotic genomes as a proxy for the abundance of intratumor gut microbiota in the TCGA-COAD cohort. Our findings confirmed that intratumor bacterial abundance was significantly higher in Fn-positive compared to Fn-negative CRC cases (Figure 5(a)). Further analysis identified several bacterial species significantly increased (i.e., infiltrated) in Fn-infected tumors, including numerous pathogens or pathobionts (Figure 5(b), Supplementary Table S6).

This increased bacterial burden within tumors may induce chronic inflammation, contributing to the poorer prognosis observed in Fn-positive CRC. A recent study assessing IgA coating levels for thousands of gut bacterial strains in healthy humans provided a metric for the IgA binding probability of each species.⁸⁰ Our analysis found that bacterial species with increased infiltration in Fn-positive tumors also have a higher probability of binding to IgA (Figure 5(c)). These findings collectively suggest that the impairment of IgA-

determined using a two-sided Wilcoxon rank sum test. (g) Kaplan-Meier plots estimating survival based on the IgA maturation (IGAM) gene set activity score in TCGA-COAD tumors located in the right colon. (h) Paired survival analysis conducted based on groups defined by Fn infection and IGAM gene set activity score. Statistical significance for each paired group determined using Kaplan-Meier estimation; significant pairs ($p < 0.05$) are highlighted in red. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

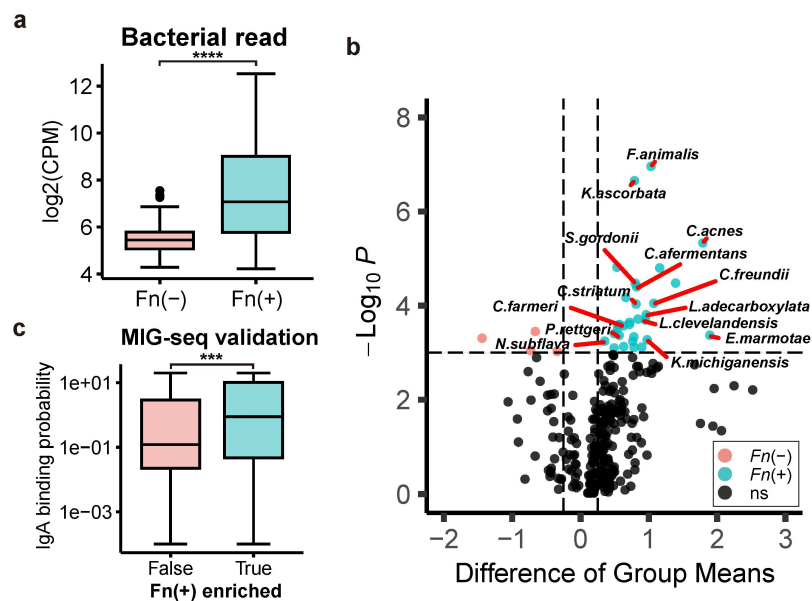


Figure 5. Intratumor bacteria burden increases in Fn-positive CRC. (a) Boxplot illustrating the distribution of bulk RNA sequencing read counts aligned to human gut bacterial genomes in TCGA-COAD tumors, grouped by Fn infection status. Read counts have been normalized to the total read depth of RNA sequencing for each sample. Statistical significance was assessed using a two-sided Wilcoxon rank sum test. (b) Volcano plot depicting species enriched in each Fn infection status group. Species reported as pathogens or opportunistic pathogens are specifically labeled. Normalized read counts from each TCGA-COAD sample were sorted, and a two-sided Wilcoxon rank sum test was used to determine enrichment for each group. P -values were adjusted using the Family-wise Type 1 Error Rate (FWER). The plot represents species based on the difference in mean normalized read counts between Fn groups, with an adjusted p -value cutoff of 0.001 and a mean difference cutoff of 0.25. (c) Boxplot illustrating the distribution of IgA binding probabilities of species in TCGA-COAD tumors, grouped by the side with the higher bacterial read abundance. Statistical significance was determined using a two-sided Wilcoxon rank sum test. *** $p < 0.001$, **** $p < 0.0001$.

mediated mucosal immunity permits species typically kept in check by IgA to infiltrate the tumor, thereby exacerbating the intratumor bacterial burden in Fn-positive CRC. This mechanism likely contributes significantly to the adverse outcomes associated with Fn infection in CRC.

Validation of Fn-associated impairment of IgA induction in an independent CRC cohort

To determine if our findings are reproducible in other CRC patients, we performed single-cell transcriptomic analysis on tumors from an independent validation cohort of 18 CRC patients (7 Fn-positives and 11 Fn-negatives) (Supplementary Table 7). This analysis identified 13 distinct immune cell types encompassing a total of 34,965 cells (Supplementary Figure S6a, Supplementary Table 8). We also examined the disruption of macrophage-mediated IgA induction by analyzing cell-cell interactions between B cells and myeloid cells (Supplementary Figure S6b). We observed notably diminished TACI-mediated interactions between B cell subsets and macrophages. Furthermore, APRIL-mediated interactions between IgA plasma cells and macrophages were reduced. These findings indicate a disruption of the interaction with macrophages necessary for IgA induction in Fn-positive tumors.

To investigate the development of IgA plasma cells further, we processed plasma cells and B cells separately (Supplementary Figure S6c). By assessing *PRDM1* expression, we distinguished mature IgA plasma cells (Supplementary Figure S6d). Using RNA velocity to measure cell transitions, we noted a significant delay in the transition to mature IgA plasma cells in the Fn-positive group, replicating our results from the discovery cohort (Supplementary Figure S6e-f). Lastly, we estimated the intratumor bacterial read count using bulk RNA-seq data from the tumor samples of the validation cohort. The analysis revealed a consistent increase in intratumor bacterial read count in the Fn-positive group compared to the Fn-negative group, although the sample size was too small for statistical significance (Supplementary Figure S6g).

In summary, single-cell and tumor microbiome analyses in an independent validation cohort reinforced our observations from the discovery cohort, confirming that the poor prognosis of Fn-positive CRC patients is linked to the impairment of IgA-mediated mucosal immunity.

Validation of Fn-associated impairment of IgA induction across mouse models

Although we identified an association between Fn infection and impaired IgA plasma cell development in an independent CRC cohort, causality was not established. To address this, we tested the causal role of Fn by infecting germ-free (GF) mice with Fn via oral gavage (Figure 6(a)). Experimental groups were established by administering Fn to GF mice (Fn group), while control groups received phosphate-buffered saline (PBS) (control group). Five weeks after oral gavage, the Fn group showed significantly lower normalized body weight ($102.11 \pm 0.89\%$, $n = 5$) compared to controls ($111.11 \pm 2.14\%$, $n = 5$) (Supplementary Figure S7a). Although GF mice typically exhibit an enlarged cecum due to mucus and undigested fibers,⁸¹ the Fn group showed reduced cecum size (Supplementary Figure S7b) and weight (Supplementary Figure S7c) compared to controls, though not statistically significant. These results indicate successful intestinal colonization by Fn.

scRNA-seq analysis generated single-cell transcriptome profiles for 7,470 mouse immune cells (Figure 6(b)). To examine IgA plasma cell differentiation, we constructed a trajectory from B cells to IgA plasma cells (Figure 6(c)), grouping cells along pseudotime into distinct developmental states (Figure 6(d)). The trajectory was validated by the expression dynamics of marker genes: early-state B cells (*Bach2*, *Ebf1*, *Foxp1*, *Pax5*)^{82–84} (Supplementary Figure S7d), mature B cells (*Cd37*, *Ce52*, *Cd74*, *Cd79a*)⁸² (Supplementary Figure S7e), and IgA plasma cells (Supplementary Figure S7f). Notably, cell density distributions shifted toward earlier pseudotime in the Fn group, indicating delayed IgA plasma cell differentiation (Figure 6(e)). RNA velocity analysis showed inhibited transition from early-state B cells to mature B cells (Figure 6f) and from monocytes to macrophages, essential for IgA induction (Figure 6(g–h)) in the Fn group. These findings suggest that Fn disrupts the development of both IgA plasma cells and macrophages, ultimately impairing IgA-mediated mucosal immunity.

To test the role of Fn under a more physiologically relevant setting—one that preserves both the tumor-immune microenvironment and the normal gut microbiome – we re-analyzed a publicly available single-cell RNA-seq data set from a CRC mouse model with an intact microbiota (GEO accession GSE172334). From 36,697 tumor-infiltrating cells collected from three mice gavaged orally with Fn and three gavaged with PBS, we identified 15 immune cell types (Supplementary Figure S8a). We then subclustered the B-lineage compartment, resolved a distinct germinal-center (GC) B-cell cluster, and reconstructed a differentiation trajectory from naïve/GC B cells to IgA plasma cells (Supplementary Figure 8b). In the Fn group, expression of key plasma-cell – differentiation genes – including *Tnfrsf13b* and *Tnfrsf13c*—was significantly reduced in B cells (Supplementary Figure S8c). When the PBS and Fn samples were visualized separately, the GC-to-IgA trajectory appeared less continuous in Fn-exposed mice (Supplementary Figure S8d). RNA-velocity analysis confirmed this impression: the transition probability from B cells to IgA plasma cells at the branch point was markedly lower in the Fn cohort (Supplementary Figure S8e–f). Thus, Fn impairs IgA-plasma-cell differentiation not only in germ-free mice but also in conventionally colonized animals.

Discussion

Our analysis revealed significant alterations in the immune cell composition within tumors due to Fn infection, including the dysregulation of pathways essential for generating IgA-secreting plasma cells. While existing studies have shown that Fn infection inhibits T-cell-driven antitumor responses,⁸⁵ our research highlights the crucial role of specific B-cell subsets in the poor prognosis of Fn-associated CRC. Notably, Fn infection hindered the development of IgA plasma cells in the colon tissue of germ-free mice and in the tumor tissue of a CRC mouse model with a conventional microbiota. Single-cell and spatial transcriptome data analysis showed disrupted ligand-receptor interactions between IgA plasma cells and M2 macrophages, resulting in impaired maturation and mucosal immunity functions of sIgA in Fn-positive CRC. Consequently, the weakened sIgA function may not properly prevent gut bacteria infiltration, leading to

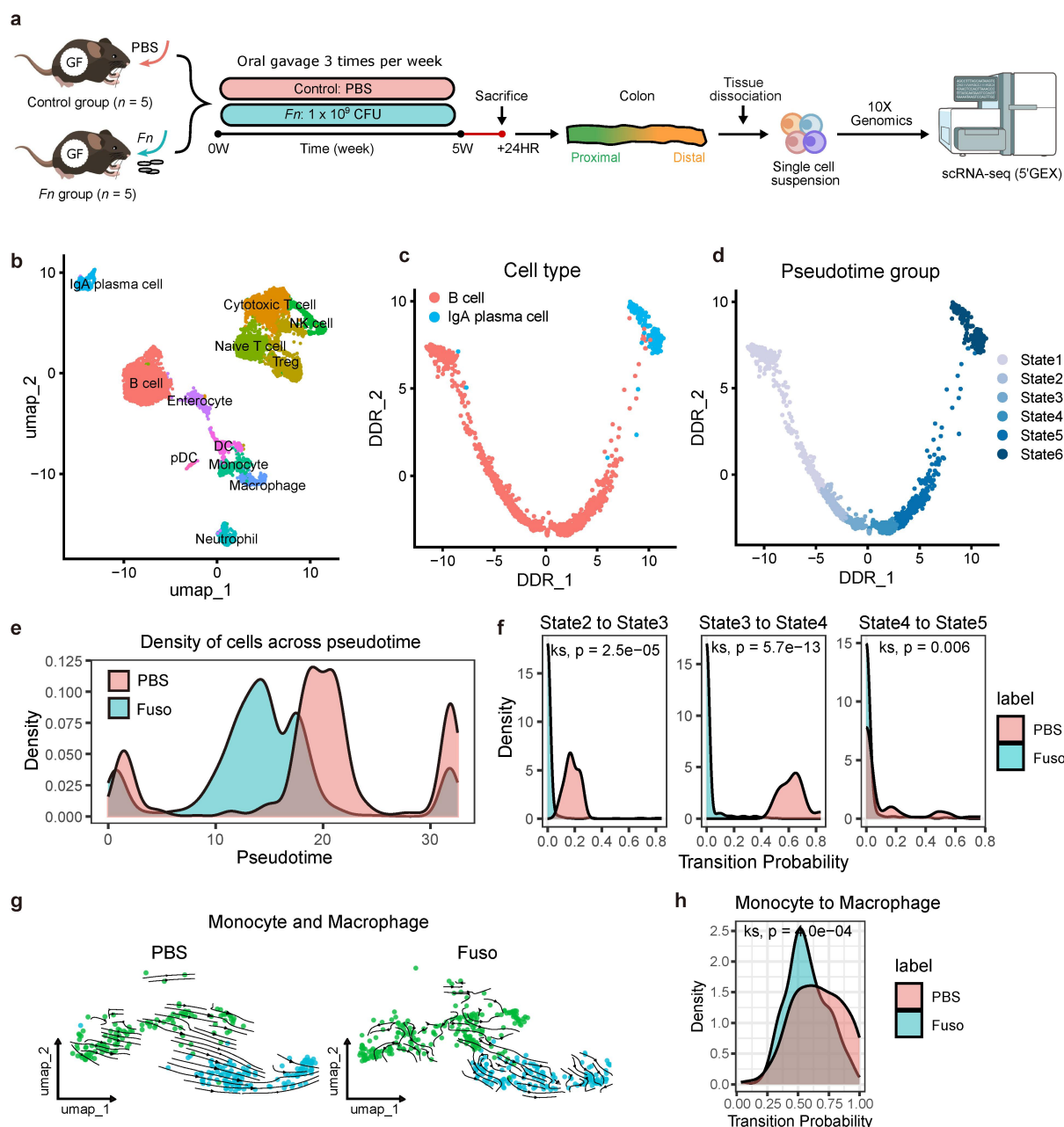


Figure 6. Disruption of IgA plasma cell and macrophage development in Fn-administered germ-free (GF) mice. (a) Experimental scheme: GF mice were administered phosphate-buffered saline (PBS) ($n = 5$) or *Fusobacterium nucleatum* (Fn) ($n = 5$) via oral gavage three times per week for five weeks, followed by sacrifice. Colon tissues were dissociated for single-cell RNA sequencing (scRNA-seq) analysis (b) UMAP visualization of immune cells, identifying 12 distinct immune cell types after preprocessing and batch correction. (c) Differentiation trajectories of B cells and plasma cells visualized using the DDRTree algorithm. (d) Six cell states along the pseudotime trajectory of B cell-to-IgA plasma cell differentiation. (e) Cell density distribution along pseudotime for IgA plasma cell differentiation in Fn (Fuso) and control (PBS) groups. (f) Density plot of transition probabilities between cell states along the IgA plasma cell trajectory (Kolmogorov-Smirnov test). (g) RNA velocity analysis of monocytes and macrophages from colon tissues in control (PBS) and Fn (Fuso) group. (h) Density plot of transition probability from monocytes to macrophages (Kolmogorov-Smirnov test).

an elevated intratumor bacterial burden, which may adversely affect patient clinical outcomes by inducing chronic inflammation. Remarkably, we identified signature genes for IgA maturation and found that their activity can stratify Fn-positive CRC patients. Those with higher gene activity exhibit survival rates comparable to Fn-negative patients. These findings illuminate the potential for developing personalized therapies for Fn-positive CRC patients by targeting the IgA development pathway.

Patients with IgA deficiency exhibit an increased risk of several types of cancer, including gastrointestinal cancer.⁸⁶ sIgA normally acts as the primary antigen-specific defense mechanism within the intestinal lumen, playing a crucial role in defending against enteric pathogens.^{45,46} Additionally, sIgA naturally binds to and encapsulates commensal bacteria in the gut, thereby playing an essential role in maintaining the balance between the host and its gut microbiota. Our results suggest that individuals with intact IgA function in colon tumors may have improved prognoses in Fn-positive cases. This implies that the role of sIgA in counteracting Fn infection could serve not only as a potential biomarker for early detection of colon cancer⁸⁷ but also as a promising leverage for therapeutic strategies aimed at inhibiting Fn-induced tumor progression. These findings underscore the need for further studies to identify the specific antigens recognized by sIgA and to elucidate how sIgA neutralizes the pro-tumoral effects of Fn infection in colon cancer.

This study has some limitations. First, while we identified the causal role of Fn infection in dysfunctional sIgA production in CRC tumors, the precise molecular mechanism remains unclear. Our spatial transcriptome analysis revealed that IgA plasma cell spots were located closest to the spots containing Fn sequence reads in situ, suggesting that Fn might impair IgA plasma cells through close contact. However, further studies are needed to identify Fn-derived molecules that negatively regulate TGF- β signaling in IgA plasma cells. Second, clinical factors such as antibiotic usage and the effects of chemotherapy were not available for the TCGA-COAD cohort and therefore not included in the multivariate survival analysis. In our independent patient cohort, however, all individuals were treatment naïve and had not received antibiotics at the time of curative-intent surgery and sample collection. Additionally, the IGAM module for stratifying Fn-positive patients was not validated using prospective cohorts in this study, which should be addressed in future follow-up studies.

In summary, our findings support the notion that Fn infection hinders macrophage-mediated sIgA induction and increases tumor bacterial burden via disruption of both IgA plasma cells and macrophage. Stratifying patients based on IGAM module activity and developing strategies to restore IgA maturation hold promises for refining treatment strategies for Fn-positive CRC. This approach offers a more tailored strategy for patient care, potentially improving therapeutic outcomes.

Resource availability

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Insuk Lee (insuklee@yonsei.ac.kr).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Single-cell RNA sequencing data have been deposited at GEO under accession number GSE273567 and will be publicly available as of the date of publication. Currently, the data can be accessed using a secure token (szaresysfxehncj).
- The corresponding raw 16S rRNA sequencing tumor microbiome data have been deposited in the Sequence Read Archive (SRA): PRJNA1142424 and will be publicly available as of the date of publication.
- Single-cell RNA sequencing data from germ-free (GF) mice have been deposited at GEO under accession number GSE288951 and will be publicly available as of the date of publication, with temporary access available via a secure token (arefyeqwllajtgn).
- Single-cell RNA sequencing data from CRC mouse with a normal gut microbiome is publicly available at GEO under accession number GSE172334.

- This paper analyzes existing, publicly available data, accessible from Broad Genome Data Analysis Center Firehose (<https://gdac.broadinstitute.org>) for the TCGA cohort dataset and the DDBJ Sequence Read Archive (DRA015288) for the spatial transcriptome dataset for CRC tumors.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study was supported in by the Post-Genome Multi-Ministerial Project grant number: [2019-NI-092-00] to S.C.K.), the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) [2022R1A2C4001879] to H.S.K., [2022M3A9F3016364], [2022R1A2C1092062] to I.L., [RS-2022-NR075480] to K.K, and [RS-2023-00247969] to J.B.A., a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea [HI22C0353] and [RS-2023-00304686] to H.S.K., the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Korean government (MSIT) [2022M3A9F3016364] to H.S.K. and I.L.), the “Hankookilbo Myung-Ho Seung” Faculty Research Assistance Program of Yonsei University College of Medicine (6-2023-0172), the Technology Innovation Program (20022947; funded by the Ministry of Trade Industry & Energy; MOTIE, Korea), Institute for Project-Y funded by Yonsei University, a research grant funded by Chong Kun Dang Bio Co. Ltd. We used ChatGPT (GPT-4o model by OpenAI) to support the refinement of English grammar and clarity during the preparation of this manuscript. All content was critically reviewed and edited by the authors, who take full responsibility for the accuracy and originality of the final submitted work.

ORCID

Ki Taek Nam  <http://orcid.org/0000-0001-5292-1280>
 Susan Bullman  <http://orcid.org/0000-0002-7713-0810>
 Han Sang Kim  <http://orcid.org/0000-0002-6504-9927>

Author contributions

S.C.K., H.S.K., and I.L. conceived and designed the study. I.C. performed the bioinformatics analysis of single-cell omics data and formulated the study hypotheses. Y.D.H., and H.S.K. collected and annotated the clinical data. K.K., S. C.K., and H-Y.J. performed the single-cell data generation. D.P. and K.K. performed GF mice experiments under supervision of K.T.N. J.C., S.B., and J.C. assisted with the bioinformatics analysis. H.S.K. and J.B.A. contributed to clinical annotations. M.Z., I.M., D.L., and S.B. provided advice on single-cell data analysis. M.J. performed pathological annotations. S.C.K. managed the project and financial support. H.S.K. and I.L. supervised the clinical and bioinformatic analyses, respectively. I.C., K.K., H.S.K., and I.L. wrote the manuscript. All authors contributed to the editing of the manuscript.

Declaration of AI-assisted technologies in the writing process

During the preparation of this work the author(s) used ChatGPT in order to improve language and readability of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Ethical approval

This study received approval from the Institutional Review Board (IRB) of Yonsei University Severance Hospital, under IRB No. 4–2019–0811. Prior to enrollment and sample collection, written informed consent was obtained from all participants at Yonsei University College of Medicine Severance Hospital. The research adhered to the principles outlined in the Declaration of Helsinki.

All animal experimental procedures were approved and conducted in accordance with the protocols of the Institutional Animal Care and Use Committee (IACUC) of Yonsei University College of Medicine, under IACUC No. 2022–0286.

References

1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. *CA Cancer J Clin.* **2024**;74(1):12–49. doi: [10.3322/caac.21820](https://doi.org/10.3322/caac.21820).
2. Nejman D, Livyatan I, Fuks G, Gavert N, Zwang Y, Geller LT, Rotter-Maskowitz A, Weiser R, Mallel G, Gigi E, 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](https://doi.org/10.1126/science.aay9189).
3. Galeano Nino JL, Wu H, LaCourse KD, Kempchinsky AG, Baryiamas A, Barber B, Futran N, Houlton J, Sather C, Sicinska E, et al. Effect of the intratumoral microbiota on spatial and cellular heterogeneity in cancer. *Nature.* **2022**;611(7937):810–817. doi: [10.1038/s41586-022-05435-0](https://doi.org/10.1038/s41586-022-05435-0).
4. Johnston CD, Bullman S. The tumour-associated microbiome. *Nat Rev Gastroenterol Hepatol.* **2022**;19(6):347–348. doi: [10.1038/s41575-022-00609-7](https://doi.org/10.1038/s41575-022-00609-7).
5. Bullman S, Pedomallu CS, Sicinska E, Clancy TE, Zhang X, Cai D, Neuberg D, Huang K, Guevara F, Nelson T, et al. Analysis of *Fusobacterium* persistence and antibiotic response in colorectal cancer. *Science.* **2017**;358(6369):1443–1448. doi: [10.1126/science.aal5240](https://doi.org/10.1126/science.aal5240).
6. Castellarin M, Warren RL, Freeman JD, Dreolini L, Krzywinski M, Strauss J, Barnes R, Watson P, Allen-Vercoe E, Moore RA, et al. *Fusobacterium nucleatum* infection is prevalent in human colorectal carcinoma. *Genome Res.* **2012**;22(2):299–306. doi: [10.1101/gr.126516.111](https://doi.org/10.1101/gr.126516.111).
7. Kostic AD, Gevers D, Pedomallu CS, Michaud M, Duke F, Earl AM, Ojesina AI, Jung J, Bass AJ, Tabernero J, et al. Genomic analysis identifies association of *Fusobacterium* with colorectal carcinoma. *Genome Res.* **2012**;22(2):292–298. doi: [10.1101/gr.126573.111](https://doi.org/10.1101/gr.126573.111).
8. Mima K, Nishihara R, Qian ZR, Cao Y, Sukawa Y, Nowak JA, Yang J, Dou R, Masugi Y, Song M, et al. *Fusobacterium nucleatum* in colorectal carcinoma tissue and patient prognosis. *Gut.* **2016**;65(12):1973–1980. doi: [10.1136/gutjnl-2015-310101](https://doi.org/10.1136/gutjnl-2015-310101).
9. Lee JB, Kim KA, Cho HY, Kim D, Kim WK, Yong D, Lee H, Yoon SS, Han DH, Han YD, et al. Association between *Fusobacterium nucleatum* and patient prognosis in metastatic colon cancer. *Sci Rep.* **2021**;11(1):20263. doi: [10.1038/s41598-021-98941-6](https://doi.org/10.1038/s41598-021-98941-6).
10. Kim HS, Kim CG, Kim WK, Kim KA, Yoo J, Min BS, Paik S, Shin SJ, Lee H, Lee K, et al. *Fusobacterium nucleatum* induces a tumor microenvironment with diminished adaptive immunity against colorectal cancers. *Front Cell Infect Microbiol.* **2023**;13:1101291. doi: [10.3389/fcimb.2023.1101291](https://doi.org/10.3389/fcimb.2023.1101291).
11. Rubinstein MR, Wang X, Liu W, Hao Y, Cai G, Han YW. *Fusobacterium nucleatum* promotes colorectal carcinogenesis by modulating E-cadherin/ β -catenin signaling via its FadA adhesin. *Cell Host & Microbe.* **2013**;14(2):195–206. doi: [10.1016/j.chom.2013.07.012](https://doi.org/10.1016/j.chom.2013.07.012).
12. Kostic AD, Chun E, Robertson L, Glickman JN, Gallini CA, Michaud M, Clancy T, Chung D, Lochhead P, Hold G, et al. *Fusobacterium nucleatum* potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host & Microbe.* **2013**;14(2):207–215. doi: [10.1016/j.chom.2013.07.007](https://doi.org/10.1016/j.chom.2013.07.007).
13. Kim MJ, Kim K, Park HJ, Kim GR, Hong KH, Oh JH, Son J, Park DJ, Kim D, Choi J-M, et al. Deletion of PD-1 destabilizes the lineage identity and metabolic fitness of tumor-infiltrating regulatory T cells. *Nat Immunol.* **2023**;24(1):148–161. doi: [10.1038/s41590-022-01373-1](https://doi.org/10.1038/s41590-022-01373-1).
14. Heumos L, Schaar AC, Lance C, Litinetskaya A, Drost F, Zappia L, Lücken MD, Strobl DC, Henao J, Curion F, et al. Best practices for single-cell analysis across modalities. *Nat Rev Genet.* **2023**;24(8):550–572. doi: [10.1038/s41576-023-00586-w](https://doi.org/10.1038/s41576-023-00586-w).
15. Hao Y, Hao S, Andersen-Nissen E, Wm M, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, et al. Integrated analysis of multimodal single-cell data. *Cell.* **2021**;184(13):3573–87.e29. doi: [10.1016/j.cell.2021.04.048](https://doi.org/10.1016/j.cell.2021.04.048).
16. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* **2019**;8(4):329–37.e4. doi: [10.1016/j.cels.2019.03.003](https://doi.org/10.1016/j.cels.2019.03.003).
17. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, Baglaenko Y, Brenner M, Loh P-R, Raychaudhuri S, et al. Fast, sensitive and accurate integration of single-cell data with harmony. *Nat Methods.* **2019**;16(12):1289–1296. doi: [10.1038/s41592-019-0619-0](https://doi.org/10.1038/s41592-019-0619-0).
18. Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics.* **2011**;27(21):2957–2963. doi: [10.1093/bioinformatics/btr507](https://doi.org/10.1093/bioinformatics/btr507).
19. Li W, Fu L, Niu B, Wu S, Wooley J. Ultrafast clustering algorithms for metagenomic sequence analysis. *Brief Bioinform.* **2012**;13(6):656–668. doi: [10.1093/bib/bbs035](https://doi.org/10.1093/bib/bbs035).
20. Walker MA, Pedomallu CS, Ojesina AI, Bullman S, Sharpe T, Whelan CW, Meyerson M, Berger B. GATK PathSeq: a customizable computational tool for the discovery and identification of microbial sequences in libraries from eukaryotic hosts. *Bioinformatics.* **2018**;34(24):4287–4289. doi: [10.1093/bioinformatics/bty501](https://doi.org/10.1093/bioinformatics/bty501).

21. Kim CY, Lee M, Yang S, Kim K, Yong D, Kim HR, Lee I. Human reference gut microbiome catalog including newly assembled genomes from under-represented Asian metagenomes. *Genome Med.* **2021**;13(1):134. doi: [10.1186/s13073-021-00950-7](https://doi.org/10.1186/s13073-021-00950-7).
22. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, Sabedot TS, Malta TM, Pagnotta SM, Castiglioni I, et al. Tcgbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **2015**;44(8):e71–e. doi: [10.1093/nar/gkv1507](https://doi.org/10.1093/nar/gkv1507).
23. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, Kovatich AJ, Benz CC, Levine DA, Lee AV, et al. An integrated TCGA Pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell.* **2018**;173(2):400–16.e11. doi: [10.1016/j.cell.2018.02.052](https://doi.org/10.1016/j.cell.2018.02.052).
24. Goldman MJ, Craft B, Hastie M, Repčič K, McDade F, Kamath A, Banerjee A, Luo Y, Rogers D, Brooks AN, et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol.* **2020**;38(6):675–678. doi: [10.1038/s41587-020-0546-8](https://doi.org/10.1038/s41587-020-0546-8).
25. Therneau TM, Lumley T. Package ‘survival’. *R Top Doc.* **2015**;128:28–33.
26. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, Lidschreiber K, Kastrić ME, Lönnerberg P, Furlan A, et al. RNA velocity of single cells. *Nature.* **2018**;560(7719):494–498. doi: [10.1038/s41586-018-0414-6](https://doi.org/10.1038/s41586-018-0414-6).
27. Qiu X, Zhang Y, Martin-Rufino JD, Weng C, Hosseinzadeh S, Yang D, Pogson AN, Hein MY, Hoi (joseph) Min K, Wang L, et al. Mapping transcriptomic vector fields of single cells. *Cell.* **2022**;185(4):690–711.e45. doi: [10.1016/j.cell.2021.12.045](https://doi.org/10.1016/j.cell.2021.12.045).
28. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan C-H, Myung P, Plikus MV, Nie Q. Inference and analysis of cell-cell communication using CellChat. *Nat Commun.* **2021**;12(1):1088. doi: [10.1038/s41467-021-21246-9](https://doi.org/10.1038/s41467-021-21246-9).
29. Grasset EK, Chorny A, Casas-Recasens S, Gutzeit C, Bongers G, Thomsen I, Chen L, He Z, Matthews DB, Oropallo MA, et al. Gut T cell-independent IgA responses to commensal bacteria require engagement of the TAC1 receptor on B cells. *Sci Immunol.* **2020**;5(49):eaat7117. doi: [10.1126/sciimmunol.aat7117](https://doi.org/10.1126/sciimmunol.aat7117).
30. Kampa M, Notas G, Stathopoulos EN, Tsapis A, Castanas E. The TNFSF members APRIL and BAFF and Their receptors TACI, BCMA, and BAFFR in oncology, with a special focus in breast cancer. *Front Oncol.* **2020**;10:827. doi: [10.3389/fonc.2020.00827](https://doi.org/10.3389/fonc.2020.00827).
31. Ozato Y, Kojima Y, Kobayashi Y, Hisamatsu Y, Toshima T, Yonemura Y, Masuda T, Kagawa K, Goto Y, Utou M, et al. Spatial and single-cell transcriptomics decipher the cellular environment containing HLA-G+ cancer cells and SPP1+ macrophages in colorectal cancer. *Cell Rep.* **2023**;42(1):42. doi: [10.1016/j.celrep.2022.111929](https://doi.org/10.1016/j.celrep.2022.111929).
32. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology.* **2018**;19(1):15. doi: [10.1186/s13059-017-1382-0](https://doi.org/10.1186/s13059-017-1382-0).
33. Kleshcheynikov V, Shmatko A, Dann E, Aivazidis A, King HW, Li T, Elmentaite R, Lomakin A, Kedlian V, Gayoso A, et al. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat Biotechnol.* **2022**;40(5):661–671. doi: [10.1038/s41587-021-01139-4](https://doi.org/10.1038/s41587-021-01139-4).
34. Cang Z, Zhao Y, Almet AA, Stabell A, Ramos R, Plikus MV, Atwood SX, Nie Q. Screening cell-cell communication in spatial transcriptomics via collective optimal transport. *Nat Methods.* **2023**;20(2):218–228. doi: [10.1038/s41592-022-01728-4](https://doi.org/10.1038/s41592-022-01728-4).
35. Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, Purdom E, Dudoit S. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics.* **2018**;19(1):477. doi: [10.1186/s12864-018-4772-0](https://doi.org/10.1186/s12864-018-4772-0).
36. Van den Berge K, Roux de Bézieux H, Street K, Saelens W, Cannoodt R, Saeys Y, Dudoit S, Clement L. Trajectory-based differential expression analysis for single-cell sequencing data. *Nat Commun.* **2020**;11(1):1201. doi: [10.1038/s41467-020-14766-3](https://doi.org/10.1038/s41467-020-14766-3).
37. Foroutan M, Bhuva DD, Lyu R, Horan K, Cursons J, Davis MJ. Single sample scoring of molecular phenotypes. *BMC Bioinform.* **2018**;19(1):404. doi: [10.1186/s12859-018-2435-4](https://doi.org/10.1186/s12859-018-2435-4).
38. Baran Y, Bercovich A, Sebe-Pedros A, Lubling Y, Giladi A, Chomsky E, Meir Z, Hoichman M, Lifshitz A, Tanay A, et al. MetaCell: analysis of single-cell RNA-seq data using K-nn graph partitions. *Genome Biology.* **2019**;20(1):206. doi: [10.1186/s13059-019-1812-2](https://doi.org/10.1186/s13059-019-1812-2).
39. Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat Protoc.* **2020**;15(4):1484–1506. doi: [10.1038/s41596-020-0292-x](https://doi.org/10.1038/s41596-020-0292-x).
40. Cabello-Aguilar S, Alame M, Kon-Sun-Tack F, Fau C, Lacroix M, Colinge J. SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. *Nucleic Acids Res.* **2020**;48(10):e55–e. doi: [10.1093/nar/gkaa183](https://doi.org/10.1093/nar/gkaa183).
41. Shao X, Liao J, Li C, Lu X, Cheng J, Fan X. CellTalkDB: a manually curated database of ligand–receptor interactions in humans and mice. *Brief Bioinform.* **2021**;22(4):22. doi: [10.1093/bib/bbaa269](https://doi.org/10.1093/bib/bbaa269).
42. Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA, Staudt LM. Toward a shared vision for cancer genomic data. *N Engl J Med.* **2016**;375(12):1109–1112. doi: [10.1056/NEJMp1607591](https://doi.org/10.1056/NEJMp1607591).
43. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, Trapnell C. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods.* **2017**;14(10):979–982. doi: [10.1038/nmeth.4402](https://doi.org/10.1038/nmeth.4402).

44. Bemark M, Pitcher MJ, Dionisi C, Spencer J. Gut-associated lymphoid tissue: a microbiota-driven hub of B cell immunity. *Trends In Immunol.* **2024**;45(3):211–223. doi: [10.1016/j.it.2024.01.006](https://doi.org/10.1016/j.it.2024.01.006).
45. Pabst O, Slack E. IgA and the intestinal microbiota: the importance of being specific. *Mucosal Immunol.* **2020**;13(1):12–21. doi: [10.1038/s41385-019-0227-4](https://doi.org/10.1038/s41385-019-0227-4).
46. Bunker JJ, Bendelac A. IgA responses to microbiota. *Immunity.* **2018**;49(2):211–224. doi: [10.1016/j.immuni.2018.08.011](https://doi.org/10.1016/j.immuni.2018.08.011).
47. Lycke NY, Bemark M. The regulation of gut mucosal IgA B-cell responses: recent developments. *Mucosal Immunol.* **2017**;10(6):1361–1374. doi: [10.1038/mi.2017.62](https://doi.org/10.1038/mi.2017.62).
48. Martins G, Calame K. Regulation and functions of blimp-1 in T and B lymphocytes. *Annu Rev Immunol.* **2008**;26(1):133–169. doi: [10.1146/annurev.immunol.26.021607.090241](https://doi.org/10.1146/annurev.immunol.26.021607.090241).
49. Turner CA, Mack DH, Davis MM. Blimp-1, a novel zinc finger-containing protein that can drive the maturation of B lymphocytes into immunoglobulin-secreting cells. *Cell.* **1994**;77(2):297–306. doi: [10.1016/0092-8674\(94\)90321-2](https://doi.org/10.1016/0092-8674(94)90321-2).
50. Cazac BB, Roes J. TGF-beta receptor controls B cell responsiveness and induction of IgA in vivo. *Immunity.* **2000**;13(4):443–451. doi: [10.1016/S1074-7613\(00\)00044-3](https://doi.org/10.1016/S1074-7613(00)00044-3).
51. Tamayo E, Alvarez P, Merino R. TGFβ superfamily members as regulators of B cell development and function-implications for autoimmunity. *Int J Mol Sci.* **2018**;19(12):3928. doi: [10.3390/ijms19123928](https://doi.org/10.3390/ijms19123928).
52. Lee IH, Sohn M, Lim HJ, Yoon S, Oh H, Shin S, Shin JH, Oh S-H, Kim J, Lee DK, et al. Ahnak functions as a tumor suppressor via modulation of TGFβ/Smad signaling pathway. *Oncogene.* **2014**;33(38):4675–4684. doi: [10.1038/onc.2014.69](https://doi.org/10.1038/onc.2014.69).
53. Chen X, Cao X, Sun X, Lei R, Chen P, Zhao Y, Jiang Y, Yin J, Chen R, Ye D, et al. Bcl-3 regulates TGFβ signaling by stabilizing Smad3 during breast cancer pulmonary metastasis. *Cell Death Dis.* **2016**;7(12):e2508–e. doi: [10.1038/cddis.2016.405](https://doi.org/10.1038/cddis.2016.405).
54. Kwon W, Choi S-K, Kim D, Kim H-G, Park J-K, Han JE, Cho G-J, Yun S, Yu W, Han S-H, et al. ZNF507 affects TGF-β signaling via TGFBR1 and MAP3K8 activation in the progression of prostate cancer to an aggressive state. *J Exp & Clin Cancer Res.* **2021**;40(1):291. doi: [10.1186/s13046-021-02094-3](https://doi.org/10.1186/s13046-021-02094-3).
55. Zhang F, Wang H, Wang X, Jiang G, Liu H, Zhang G, Wang H, Fang R, Bu X, Cai S, et al. TGF-β induces M2-like macrophage polarization via SNAIL-mediated suppression of a pro-inflammatory phenotype. *Oncotarget.* **2016**;7(32):52294–52306. doi: [10.18632/oncotarget.10561](https://doi.org/10.18632/oncotarget.10561).
56. Zhong Z, Nan K, Weng M, Yue Y, Zhou W, Wang Z, Chu Y, Liu R, Miao C. Pro- and anti- effects of immunoglobulin A- Producing B cell in tumors and its triggers. *Front Immunol.* **2021**;12:765044. doi: [10.3389/fimmu.2021.765044](https://doi.org/10.3389/fimmu.2021.765044).
57. Xuan W, Qu Q, Zheng B, Xiong S, Fan G-H. The chemotaxis of M1 and M2 macrophages is regulated by different chemokines. *J Leukoc Biol.* **2015**;97(1):61–69. doi: [10.1189/jlb.1A0314-170R](https://doi.org/10.1189/jlb.1A0314-170R).
58. Qin R, Ren W, Ya G, Wang B, He J, Ren S, Jiang L, Zhao S. Role of chemokines in the crosstalk between tumor and tumor-associated macrophages. *Clin Exp Med.* **2022**;23(5):1359–1373. doi: [10.1007/s10238-022-00888-z](https://doi.org/10.1007/s10238-022-00888-z).
59. Ma R-Y, Black A, Qian B-Z. Macrophage diversity in cancer revisited in the era of single-cell omics. *Trends Immunol.* **2022**;43(7):546–563. doi: [10.1016/j.it.2022.04.008](https://doi.org/10.1016/j.it.2022.04.008).
60. Schiemann B, Gommerman JL, Vora K, Cachero TG, Shulga-Morskaya S, Dobles M, Frew E, Scott ML. An essential role for BAFF in the normal development of B cells through a BCMA-independent pathway. *Science.* **2001**;293(5537):2111–2114. doi: [10.1126/science.1061964](https://doi.org/10.1126/science.1061964).
61. Cornelis R, Chang HD, Radbruch A. Keeping up with the stress of antibody production: BAFF and APRIL maintain memory plasma cells. *Curr Opin Immunol.* **2021**;71:97–102. doi: [10.1016/j.coi.2021.06.012](https://doi.org/10.1016/j.coi.2021.06.012).
62. Lightman SM, Utley A, Lee KP. Survival of long-lived plasma cells (LLPC): piecing together the puzzle. *Front Immunol.* **2019**;10:10. doi: [10.3389/fimmu.2019.00965](https://doi.org/10.3389/fimmu.2019.00965).
63. Smulski CR, Eibel H. BAFF and BAFF-Receptor in B cell selection and survival. *Front Immunol.* **2018**;9. doi: [10.3389/fimmu.2018.02285](https://doi.org/10.3389/fimmu.2018.02285).
64. Mathias A, Corthésy B. N-Glycans on secretory component. *Gut Microbes.* **2011**;2(5):287–293. doi: [10.4161/gmic.2.5.18269](https://doi.org/10.4161/gmic.2.5.18269).
65. Perrier C, Sprenger N, Corthésy B. Glycans on secretory component participate in innate protection against mucosal pathogens*. *J Biol Chem.* **2006**;281(20):14280–14287. doi: [10.1074/jbc.M512958200](https://doi.org/10.1074/jbc.M512958200).
66. Johansen FE, Braathen R, Brandtzaeg P. Role of J chain in secretory immunoglobulin formation. *Scand J Immunol.* **2000**;52(3):240–248. doi: [10.1046/j.1365-3083.2000.00790.x](https://doi.org/10.1046/j.1365-3083.2000.00790.x).
67. Rao S, Karray S, Gackstetter ER, Koshland ME. Myocyte enhancer factor-related B-MEF2 is developmentally expressed in B cells and regulates the immunoglobulin J chain Promoter*. *J Biol Chem.* **1998**;273(40):26123–26129. doi: [10.1074/jbc.273.40.26123](https://doi.org/10.1074/jbc.273.40.26123).
68. Rosenbaum M, Andreani V, Kapoor T, Herp S, Flach H, Duchniewicz M, Grosschedl R. MZB1 is a GRP94 cochaperone that enables proper immunoglobulin heavy chain biosynthesis upon ER stress. *Genes Dev.* **2014**;28(11):1165–1178. doi: [10.1101/gad.240762.114](https://doi.org/10.1101/gad.240762.114).
69. Shaffer A, Shapiro-Shelef M, Iwakoshi NN, Lee A-H, Qian S-B, Zhao H, Yu X, Yang L, Tan BK, Rosenwald A, et al. XBP1, downstream of blimp-1, expands the secretory apparatus and other organelles, and increases protein synthesis in plasma cell differentiation. *Immunity.* **2004**;21(1):81–93. doi: [10.1016/j.immuni.2004.06.010](https://doi.org/10.1016/j.immuni.2004.06.010).

70. Tellier J, Shi W, Minnich M, Liao Y, Crawford S, Smyth GK, Kallies A, Busslinger M, Nutt SL. Blimp-1 controls plasma cell function through the regulation of immunoglobulin secretion and the unfolded protein response. *Nat Immunol.* 2016;17(3):323–330. doi: [10.1038/ni.3348](https://doi.org/10.1038/ni.3348).
71. Yoshida H, Matsui T, Yamamoto A, Okada T, Mori K. XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. *Cell.* 2001;107(7):881–891. doi: [10.1016/S0092-8674\(01\)00611-0](https://doi.org/10.1016/S0092-8674(01)00611-0).
72. Ma Y, Shimizu Y, Mann MJ, Jin Y, Hendershot LM. Plasma cell differentiation initiates a limited ER stress response by specifically suppressing the PERK-dependent branch of the unfolded protein response. *Cell Stress Chaperones.* 2010;15(3):281–293. doi: [10.1007/s12192-009-0142-9](https://doi.org/10.1007/s12192-009-0142-9).
73. Park S-M, Kang T-I, So J-S. Roles of XBP1s in transcriptional regulation of target genes. *Biomedicines.* 2021;9(7):791. doi: [10.3390/biomedicines9070791](https://doi.org/10.3390/biomedicines9070791).
74. Cruciat C-M, Hassler C, Niehrs C. The MRH protein Erlectin is a member of the endoplasmic Reticulum Synexpression group and functions in N-Glycan Recognition*. *J Biol Chem.* 2006;281(18):12986–12993. doi: [10.1074/jbc.M511872200](https://doi.org/10.1074/jbc.M511872200).
75. Chawsheen HA, Ying Q, Jiang H, Wei Q. A critical role of the thioredoxin domain containing protein 5 (TXNDC5) in redox homeostasis and cancer development. *Genes & Dis.* 2018;5:312–322. doi: [10.1016/j.gendis.2018.09.003](https://doi.org/10.1016/j.gendis.2018.09.003).
76. Nejman D, Livyatan I, Fuks G, Gavert N, Zwang Y, Geller LT, Rotter-Maskowitz A, Weiser R, Mallel G, Gigi E, 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](https://doi.org/10.1126/science.aay9189).
77. Chen Y, Wu F-H, Wu P-Q, Xing H-Y, Ma T. The role of the tumor microbiome in tumor development and its treatment. *Front Immunol.* 2022;13. doi: [10.3389/fimmu.2022.935846](https://doi.org/10.3389/fimmu.2022.935846).
78. Metzger Z, Blasbalg J, Dotan M, Tsesis I, Weiss EI. Characterization of coaggregation of *Fusobacterium nucleatum* PK1594 with six porphyromonas gingivalis strains. *J Endod.* 2009;35(1):50–54. doi: [10.1016/j.joen.2008.09.016](https://doi.org/10.1016/j.joen.2008.09.016).
79. Saito A, Inagaki S, Kimizuka R, Okuda K, Hosaka Y, Nakagawa T, Ishihara K. *Fusobacterium nucleatum* enhances invasion of human gingival epithelial and aortic endothelial cells by porphyromonas gingivalis. *FEMS Immunol Med Microbiol.* 2008;54(3):349–355. doi: [10.1111/j.1574-695X.2008.00481.x](https://doi.org/10.1111/j.1574-695X.2008.00481.x).
80. Olm MR, Spencer SP, Takeuchi T, Silva EL, Sonnenburg JL. Metagenomic immunoglobulin sequencing reveals IgA coating of microbial strains in the healthy human gut. *Nat Microbiol.* 2025;10(1):112–125. doi: [10.1038/s41564-024-01887-4](https://doi.org/10.1038/s41564-024-01887-4).
81. Al-Asmakh M, Zadjali F. Use of germ-free animal models in microbiota-related research. *J Microbiol Biotechnol.* 2015;25(10):1583–1588. doi: [10.4014/jmb.1501.01039](https://doi.org/10.4014/jmb.1501.01039).
82. Morgan D, Tergaonkar V. Unraveling B cell trajectories at single cell resolution. *Trends In Immunol.* 2022;43(3):210–229. doi: [10.1016/j.it.2022.01.003](https://doi.org/10.1016/j.it.2022.01.003).
83. Hu H, Wang B, Borde M, Nardone J, Maika S, Allred L, Tucker PW, Rao A. Foxp1 is an essential transcriptional regulator of B cell development. *Nat Immunol.* 2006;7(8):819–826. doi: [10.1038/ni1358](https://doi.org/10.1038/ni1358).
84. Sidwell T, Kallies A. Bach2 is required for B cell and T cell memory differentiation. *Nat Immunol.* 2016;17(7):744–745. doi: [10.1038/ni.3493](https://doi.org/10.1038/ni.3493).
85. Borowsky J, Haruki K, Lau MC, Dias Costa A, Vayrynen JP, Ugai T, Arima K, da Silva A, Felt KD, Zhao M, et al. Association of *Fusobacterium nucleatum* with specific T-cell subsets in the colorectal carcinoma microenvironment. *Clin Cancer Res.* 2021;27(10):2816–2826. doi: [10.1158/1078-0432.CCR-20-4009](https://doi.org/10.1158/1078-0432.CCR-20-4009).
86. Ludvigsson JF, Neovius M, Ye W, Hammarstrom L. IgA deficiency and risk of cancer: a population-based matched cohort study. *J Clin Immunol.* 2015;35(2):182–188. doi: [10.1007/s10875-014-0124-2](https://doi.org/10.1007/s10875-014-0124-2).
87. Baik JE, Li L, Shah MA, Freedberg DE, Jin Z, Wang TC, Han YW. Circulating IgA antibodies against *Fusobacterium nucleatum* amyloid adhesin FadA are a potential biomarker for colorectal neoplasia. *Cancer Res Commun.* 2022;2(11):1497–1503. doi: [10.1158/2767-9764.CRC-22-0248](https://doi.org/10.1158/2767-9764.CRC-22-0248).