Integration of Bioinformatics and Machine Learning to characterize *Fusobacterium nucleatum*'s pathogenicity

Zihan Tian[1,2] and Pietro Liò[2]
[1]Dept of Biology, John Hopkins University,
[2]Department of Computer Science and Technology,
University of Cambridge, Cambridge, UK
ztian12@jh.edu pl219@cam.ac.uk

Abstract

Fusobacterium nucleatum has been found to be associated with cancer lesions in both oral and colon cancers. Although important studies have dissected the clinical aspects of its remarkable pathogenicity, there is a lack of molecular studies. This study aimed to computationally predict potential pathogenicity islands in *F. nucleatum* ATCC 25586, with candidate functional relationships among PAI-encoded proteins, and generate testable hypotheses regarding iron-dependent virulence mechanisms. We employed an integrative bioinformatics pipeline combining genomic island prediction (IslandViewer), promoter analysis (PePPER), codon adaptation index calculation, protein interaction prediction (STRING), co-expression network inference (bnlearn), structural modeling (AlphaFold-Multimer), and genome-scale metabolic modeling (CarveMe/COBRApy). Computational predication was integrated with published literature to formulate mechanistic hypotheses. Our analysis identified three candidate genomic islands, with Region 2 (PAI2; 1,496,613 to 1,523,855 bp) exhibiting the characteristics most consistent with a functional pathogenicity island, including predicted mobile genetic elements, putative toxin-antitoxin systems, and strong promoter motifs. We propose a mechanistic hypothesis linking hemolysin-mediated iron acquisition to cancer promotion through oxidative stress and Hippo pathway modulation. Our work has two immediate and important benefits: the improved understanding of the biological processes that shape the pathogenicity and evolution of *Fusobacterium nucleatum* at the molecular level and the improved ability to integrate and automate the state-of-the-art bioinformatics tools and machine learning approaches in the inference of the mechanistic interpretability of a pathogenic phenotype.

Introduction

Various studies indicate that microbial dysbiosis is associated with oral and systemic diseases (Han, 2015). *Fusobacterium nucleatum* (*F.nucleatum*), a Gram-negative anaerobe naturally found in the oral cavity of humans, contributes to periodontal disease, as well as oral and colorectal cancers (Patel et al., 2022; Queen et al., 2025). The oral cavity is full of circulating saliva that contains a high level of dissolved oxygen, and oral microbes, such as *F. nucleatum*, are subject to oxidative stress in the form of reactive oxygen species. An adaptive response that allows *F. nucleatum* to withstand oxidative stress enables the bacterium to play a prominent role in shaping the oral biofilm (McGregor and Wolthers, 2025). The association of *F. nucleatum* with oral squamous cell carcinoma (OSCC) highlights its role via oncogenic signaling pathways and has metabolic adaptability in the tumor microenvironments (Lim et al., 2025).

Beside being a pathogen, *F. nucleatum* could induce resistance to antitumor therapies, tumor recurrence, and be associated with a poor prognosis in patients. Chemotherapy resistance is one of the major contributors to the poor prognosis in colorectal cancer. Mechanistic studies have shown that F. nucleatum can inhibit pyroptosis, a form of programmed cell death induced by chemotherapeutic agents, by down-regulating the Hippo signaling pathway and up-regulating antiapoptotic BCL2 (Han et al., 2020; Wang et al., 2024). In addition, growing evidence suggests that removal of tumor-associated *F. nucleatum* can induce the host immune response to recognize tumor cells, suggesting potential future antibacteria / tumor vaccine applications (Chen and Huang, 2024).

F.nucleatum can coexist with more than 500 other species, and its elongated shape allows it to bridge other bacteria, maintaining the development of biofilm and facilitating the spreading of periodontal disease in humans. *F. nucleatum* enhances the invasive potential of other microorganisms, such as *Porphyromonas gingivalis*. Together, these bacteria form a pathogenic alliance that can significantly alter host responses and promote cancer progression (How et al., 2016; Mohammed et al., 2020; Patel et al., 2022).

Despite *F. nucleatum*'s association with various cancers in many different tissues and its contribution to therapeutic resistance, the genomic, proteomic, and metabolic characteristics of its specific pathogenic function remain poorly characterized (Ma et al., 2023). To address this gap, this study focuses on a specific strain of *F. nucleatum* ATCC 25586 that is naturally present in the human oral cavity. An integrated bioinformatics and machine learning framework is used to predict potential pathogenicity islands (PAIs). Protein and metabolic pathways that may be associated with cancer are identified. Structural predictions (AlphaFold, TMHMM, and String) are used to evaluate protein-protein interactions and potential transmembrane regions. In parallel, genome-scale metabolic modeling using COBRApy and flux balance analysis was used to explore iron dependency and pH tolerance. This study aims to better understand what contributes to *F. nucleatum* pathogenicity and offers insights into its connection to cancer progression and therapeutic resistance.

Notably, the term *F. nucleatum* can be defined in different ways, often referring to a subspecies group (*F. nucleatum subsp. animalis*, *F. nucleatum subsp. nucleatum*, *F. nucleatum subsp. polymorphum*, *F. nucleatum subsp. vincentii*) or specific lineages. In this study, all *F. nucleatum* used in computational results refers to *F. nucleatum subsp. nucleatum* (ATCC 25586, GenBank: AE009951.2), while the term is used in a broader way when referring to literature analysis (Zepeda-Rivera et al., 2025).

We emphasize that our findings are computational predictions that generate testable hypotheses for future experimental validation.

Methodology: Building a Bioinformatics-Machine Learning integrated workflow

In this section we provide a brief description of the methods and software we have used. The analysis focused on *Fusobacterium nucleatum* subsp. nucleatum strain ATCC 25586, a critical periodontal pathogen with established links to various oral diseases and colorectal cancers.

The complete genome sequence of *Fusobacterium nucleatum* subsp. nucleatum strain ATCC 25586 (GenBank accession: AE009951.2) was retrieved from the NCBI Genome database. Genome annotation was performed using Prokka v1.14 (Seemann, 2014) to generate a detailed catalog of its genetic components, including coding sequences, rRNAs, and tRNAs. Additional functional annotation was performed using eggNOG-mapper v2.0 (Huerta-Cepas et al., 2018).

To identify potential pathogenic islands (PAI) in bacteria of interest, GC skew analysis, homology-based functional analysis (UniProt + BLAST), and IslandViewer (Bertelli et al., 2017) genomic island predictions were used.

Promoter predictions were generated using PePPER (Prokaryote Promoter Prediction) v2.0 (de Jong et al., 2012) and integrated with gene annotations from NCBI GenBank to evaluate open reading frames.

Synonymous codon usage bias was assessed using the Codon Adaptation Index (CAI). The ribosomal protein genes served as the reference set, and the CAI is calculated based on the API Reference (Bagnoli and Liò, 1995; Lee, 2019). Genes were classified as high-CAI (top 20%, CAI ≥ 0.771) or low-CAI (bottom 20%, CAI ≤ 0.670) based on the genome-wide distribution. Promoter motif analysis was performed using WebLogo Crooks et al. (2004) on 8-mer sequences extracted from promoter regions.

BLASTp (Altschul et al., 1990) is used to search protein sequences against the UniProt database. Homologous proteins are identified and assigned corresponding functional annotations. For virulence factor prediction, we retrieved functional annotation from DAVID, a tool that integrates multiple databases, including KEGG pathway maps, to identify enriched biological themes and map genes to metabolic pathways. Biomni, a machine-learning tool, was applied to cross-reference existing literature and the output of DAVID to label potential pathogenic genes.

Protein-protein interactions were predicted using the STRING database v12.0 (Szklarczyk et al.,

2024) with a minimum confidence score threshold of 0.4. Network topology was analyzed using NetworkX to calculate clustering coefficients and network density. Community detection was performed to identify functional modules within the interaction network.

Transmembrane domain predictions were generated using both TMHMM v2.0 (Krogh et al., 2001) and DeepTMHMM (Hallgren et al., 2022) to identify membrane-associated proteins. For selected protein pairs with predicted interactions, AlphaFold-Multimer was used to model complex structures (Jumper et al., 2021).

Expression data for *F. nucleatum* subsp. nucleatum ATCC 25586 was obtained from GEO dataset GSE161360 (Ponath et al., 2021), comprising 18 samples across three growth phases (exponential, mid-log, and stationary) and two treatment conditions. Bayesian network structure learning was performed using the bnlearn R package with the hill-climbing algorithm and BIC scoring. Network edges were validated using 500 bootstrap iterations, with edges retained if bootstrap strength > 0.5 and direction probability > 0.5. Co-expression relationships were compared with STRING-predicted interactions to identify consensus functional associations.

We constructed a genome-scale metabolic model (GEM) from the Prokka-annotated genome using CarveMe (Machado et al., 2018). The final model included 1,914 reactions, 1,339 metabolites, and 582 gene-protein-reaction associations. We then performed Flux Balance Analysis (FBA) with COBRApy (Ebrahim et al., 2013) to simulate growth across a range of nutrient conditions, specifically exploring iron dependency (70 iron-related reactions, 40 metabolites). We built the proposed Hemolysin-Iron-ROS-Hippo pathway by combining our computational predictions with published experimental data. We systematically adjusted iron uptake constraints to characterize how iron availability influences predicted growth rate. We also carried out essential nutrient analysis by setting individual exchange reaction bounds to zero and quantifying the corresponding changes in biomass flux.

We built the proposed Hemolysin-Iron-ROS-Hippo pathway by combining our computational predictions with published experimental data. Key components of the pathway were identified through genome annotation, and their functional interactions were verified against peer-reviewed studies focused on F. nucleatum pathogenicity and its association with cancer.

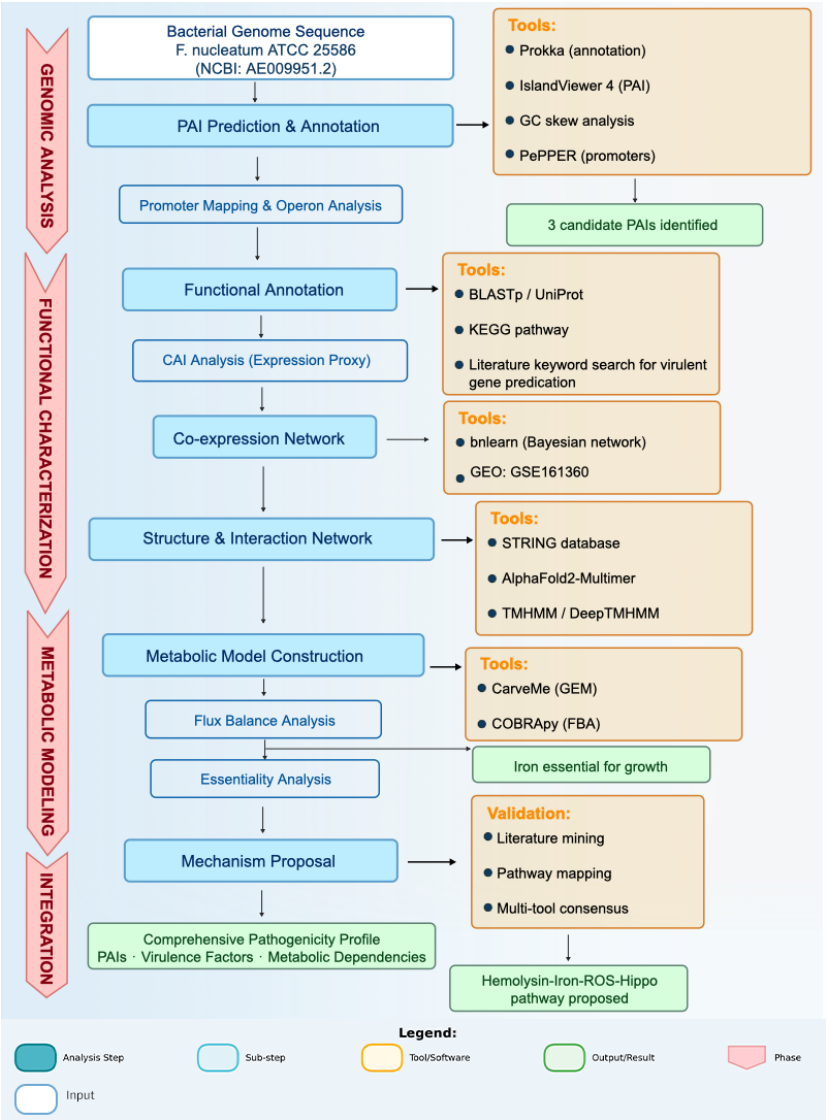

Figure 1: Integrated bioinformatics pipeline for analyzing bacterial pathogenicity. Overview of the computational workflow of four phases: 1) Genomic Analysis, including genome annotation with Prokka, GC skew analysis, and PAI prediction; 2) Functional characterization, covering homology searches, CAI analysis, co-expression network construction with bnlearn, and protein–protein interaction prediction via the STRING database; 3) Metabolic Modeling, including genome-scale model construction with CarveMe and flux balance analysis using COBRApy; and 4) Combination of computational predictions with existing literature evidence to propose a mechanistic hypothesis. Color coding: White boxes are the input. Blue boxes indicate main analysis steps; light blue boxes are sub-steps; yellow boxes indicate tools and software used; green boxes indicate outputs.

Results and Discussion

Genome Analysis

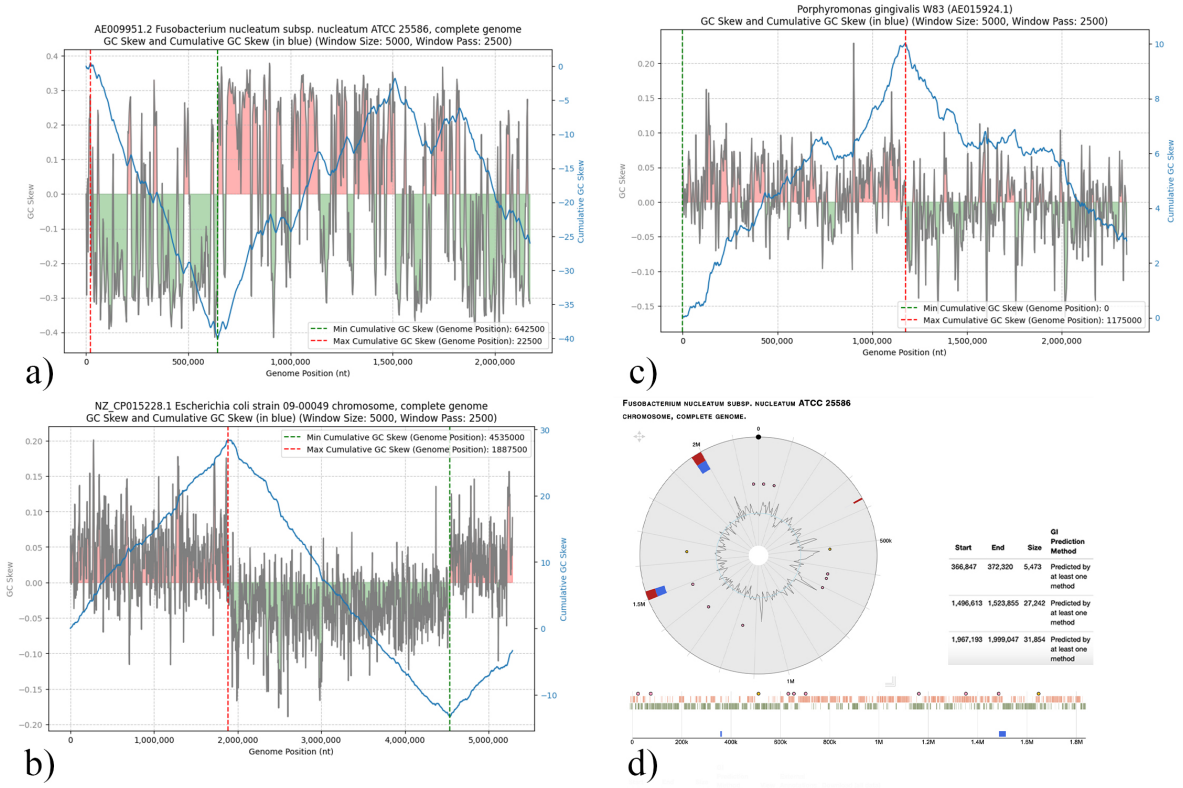

Figure 2: Panels a–c show GC Skew (in grey, where positive GC skew values have a red pattern and negative ones have a green one) and Cumulative GC Skew (blue) analyses of the complete genomes of Fusobacterium nucleatum subsp. nucleatum ATCC 25586 (GenBank: AE009951.2), Escherichia coli strain 09-00049 (GenBank: NZ_CP015228.1), and Porphyromonas gingivalis W83 (GenBank: AE015924.1), respectively. These plots were generated using the GC_skew-plot tool by Sergi Vallvé and highlight nucleotide composition asymmetries across each genome (Vallvé, 2023). Panel d presents genomic island predictions in F. nucleatum ATCC 25586 using IslandViewer, identifying three potential pathogenicity islands (PAIs) in Fusobacterium nucleatum subsp. nucleatum ATCC 25586. These regions were predicted by at least one of the integrated methods: IslandPick, IslandPath-DIMOB, or SIGI-HMM (Bertelli et al., 2017).

As shown in Figure 2, notable fluctuations in GC skew were observed between the 1.6 Mb and 1.9 Mb region, suggesting atypical nucleotide composition that may indicate genomic anomalies such as prophages or integrated mobile genetic elements (MGEs). In addition, genomic island predictions were performed using IslandViewer Bertelli et al. (2017), a computational tool that integrates multiple prediction methods, including IslandPick, IslandPath-DIMOB, and SIGI-HMM. Three regions (366,847-372,320, 1,496,613-1,523,855, 1,967,193-1,999,047) that have the potential to be PAI have been identified.

To characterize the genetic content of the predicted PAIs, gene sequences were downloaded from NCBI (GFF3 format), and corresponding protein sequences were extracted. Since a large number of hypothetical proteins were initially identified, protein sequences were queried against the UniProt database for id mapping and using BLAST to identify the top three potential identities for each protein (Altschul et al., 1990). The result, combined with annotations from NCBI and IslandViewer, was cross-referenced to make a best-fit prediction for each protein's identity.

The functional profile for each region was done through a combined analysis of manual review, homology searches, and Biomni (Huang et al., 2025).

Region 1967193–1999047 was found to be dominated by genes related to metabolic functions. The proteins within this region primarily suggest a systemic or intracellular role, potentially related to nutrient acquisition or host interaction in deeper tissues.

Region 366847–372320 contains toxins, such as hemolysin, cluster. This region is of particular interest because it is the shortest of all 3 regions and only contains 10 identified open reading frames. Most of its genes could be associated with drug resistance, such as a nitroreductase family protein, acyl-CoA thioesterase, and a metallothionein. However, the presence of an SSU ribosomal protein S20P raises questions about its classification as a classic pathogenicity island (PAI). Generally, essential genes like those for ribosomes are not typically mobile genetic elements and are not associated with virulence.

Region 1496613–1523855 exhibited characteristics consistent with a pathogenicity island. It contains mobile genetic elements and the presence of toxin-antitoxin systems. Furthermore, this region corresponded directly to a significant fluctuation identified during the initial GC skew analysis.

The Map of the Promoters reveals coexpression and gene organisation

The *F.nucleatum* ATCC 25586 genome contains 2,067 predicted protein-coding genes based on the National Institutes of Health (NIH) database. The PePPER analysis predicts that within this genome, there are 680 promoters, with promoter confidence scores ranging from 0.0 to 1.0, higher scores indicating stronger predicted promoter activity. All four promoters within PAI1 achieved the maximum prediction confidence score of 1.0, and 8 of 10 promoters in PAI2 scored ≥ 0.82, suggesting these regions may contain highly expressed elements. The genome-wide promoter-to-gene ratio is 0.33. When estimating 3 predicts PAI region using Biopython (Cock et al., 2009), PAI 2 and PAI 3 have promoter-to-gene ratios closely matching the genome average, with 0.31 and 0.33, respectively. However, PAI 1 has a notably higher ratio of 0.7. Focusing on PAI 1 and its surrounding genomic region, a hemolysin cluster is found located between 296–307 kb, which includes PAI1. Within PAI1 itself (coordinates: 366,847–372,320 bp), the gene at locus tag FN1885 encodes hemolysin III, a pore-forming cytolytic toxin with a potentially high pathogenic potential due to its direct membrane-damaging activity.

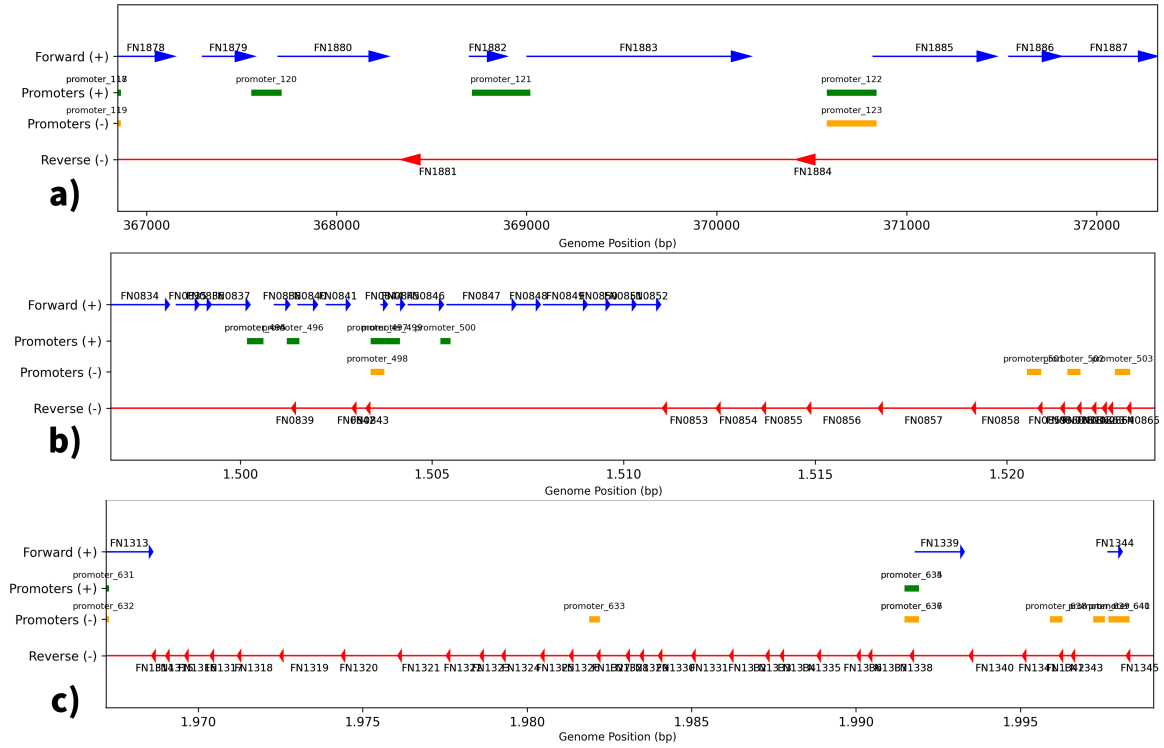

Figure 3: Promoter mapping of 3 predicts potential Pathogenicity Islands (PAIs), showing the distribution of genes on both strands and ORF. Genes are represented by arrows, with blue arrows indicating forward strand (+) genes and red arrows indicating reverse strand (−) genes. Promoters are shown as bars, with green bars representing predicted promoters on the forward strand and yellow bars representing predicted promoters on the reverse strand. a) PAI 1 (region 366847–372320) includes 7 predicted promoters (promoter_117 to promoter_123) and 10 genes (locus tags FN1878 to FN1887). b) PAI 2(region 1496613–1523855) with 10 predicted promoters (promoter_494 to promoter_503) and 32 genes (FN0834 to FN0865). c) PAI 3(region 1967193–1999047) contain 11 predicted promoters (promoter_631 to promoter_641) and 33 genes (FN1313 to FN1345).

After promoter mapping of predicted PAIs, several functional clustering appear to suggest potential coexpression and co-regulation patterns. In PAI 1, the hemolysin III (locus tag FN1885) and the adjacent unknown protein (locus tag FN1886) on the forward strand, with only one promoter upstream, suggest a potential linkage in their transcriptional regulation. Hemolysin III (FN1885), as a pore-forming toxin, provides computational evidence consistent with virulence potential by allowing host cell lysis and immune evasion. The downstream unknown protein (FN1886) could imply an accessory factor that will either enhance the function of hemolysin or represent some form of interaction with the host in an unknown and uncharacterized manner.

In PAI 2 there appears to be a biotin synthesis cluster on forward strand FN0849 (synthase) → FN0850 → FN0851 (bioC) → FN0852 (hydrolase) and the glycogen metabolism operon on the reverse strand(FN0853-FN0858, all glucose related), potentially implementing co-adaptation to metabolic processes possibly coordinating vitamin biosynthesis with energy storage/utilization under nutrient limited situations. The 2 predicted operons do not directly contribute virulence factors, but may facilitate colonization in host cells. Biotin synthesis genes (FN0849–FN0852) may aid the survival in nutrient-restricted niches where the host limits vitamin availability. The glycogen metabolism operon (FN0853–FN0858) allows energy storage, supporting persistence during fluctuating nutrient conditions. Overall, both operons may contribute to pathogenicity through nutritional adaptation, helping the bacterium withstand stresses imposed by its host.

In PAI 3, 2 potential coexpression clusters could be RNA polymerase sigma factors cluster on the reverse strand with FN1317 to FN1318 (rpoD) and the close link metabolic enzymes with FN1322 (metalloprotease) → FN1323 (thymidylate kinase)→ FN1324 (reductoisomerase). One could be regulatory

control of metabolic actives, while the other could function in tissue degradation (protease activity) and nucleotide/amino acid metabolism (thymidylate kinase, reductoisomerase), ensuring replication inside the host. Those 2 operons support PAI3's role as a metabolic regulator but not as a pathogenicity island.

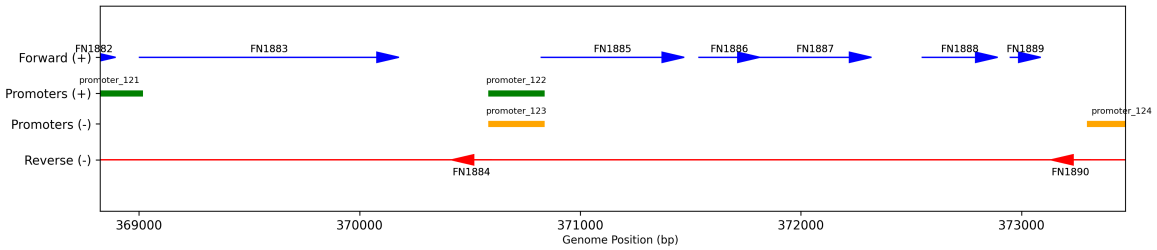

Figure 4: Genomic organization of the region spanning 368,823–373,471 bp surrounding locus tag FN1885. Blue arrows show coding sequences (ORFs) on the forward strand (+), while red arrows show on the reverse strand (−). Predicted promoters are depicted as colored bars. Green forward, yellow reverse. Promoter positions are plotted relative to their genomic coordinates and strand orientation.

Analyzing the genomic environment of the gene with locus tag FN1885 includes 9 genes and 4 putative promoters, of which FN1885 is in a forward-strand gene arrangement with 7 other genes in the same direction and has no large intergenic gaps. The reverse-strand genes (FN1884 and FN1889) are interspersed but not aligned in the same transcriptional direction, making them less likely to be part of the same operon. Within the forward strand gene, there are 2 upstream promoters, particularly promoter_122, located proximal to FN1885, indicating overlapping or sequential transcriptions and coregulation of this gene cluster. Based on the promoter positioning, FN1885 is most likely coexpressed with FN1886 to 1889, potentially in response to signals for coordinated biological function.

Codon Adaptation Index as a Proxy for Protein Abundance

High CAI values suggest strong codon optimization, likely for efficient translation. A large proportion of the top genes in the bacterial genome were ribosomal proteins (for example, L24P, S19P, L20P), which are typically highly expressed and exhibit high CAI values. Some non-ribosomal genes (e.g., flavodoxin, transcriptional regulators, hypothetical proteins) also show high CAI, which may indicate functional importance or horizontal gene transfer.

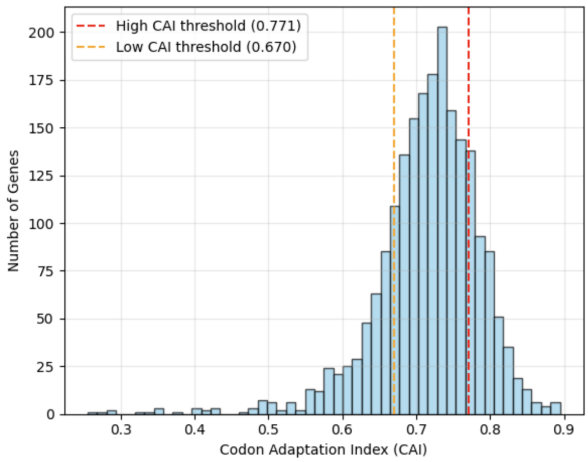

Figure 5: Distribution of Codon Adaptation Index (CAI) values of *F.nucleatum* subsp. *nucleatum* ATCC 25586 genome. using ribosomal protein sequences as reference, calculated based on the API Reference (Bagnoli and Liò, 1995; Lee, 2019). The mean CAI value is 0.716 (SD = 0.071) with values ranging from 0.256 to 0.895. The red line is the threshold for a high CAI (top 20%, CAI $\geq$ 0.771), while the yellow line is the threshold for a low CAI (bottom 20%, CAI $\leq$ 0.670). The distribution of genes with high CAIs is ribosomal proteins, translation factors (EF-G, EF-Ts, aminoacyl-tRNA synthetases), and metabolic or energy-related proteins (ATP-binding proteins, kinases, dehydrogenases). The distribution of genes with low CAIs is majoritly hypothetical proteins, with some other transport proteins (permeases and ABC transporters), metabolic or energy-related proteins (ATP synthase subunits), and transcriptional regulators.

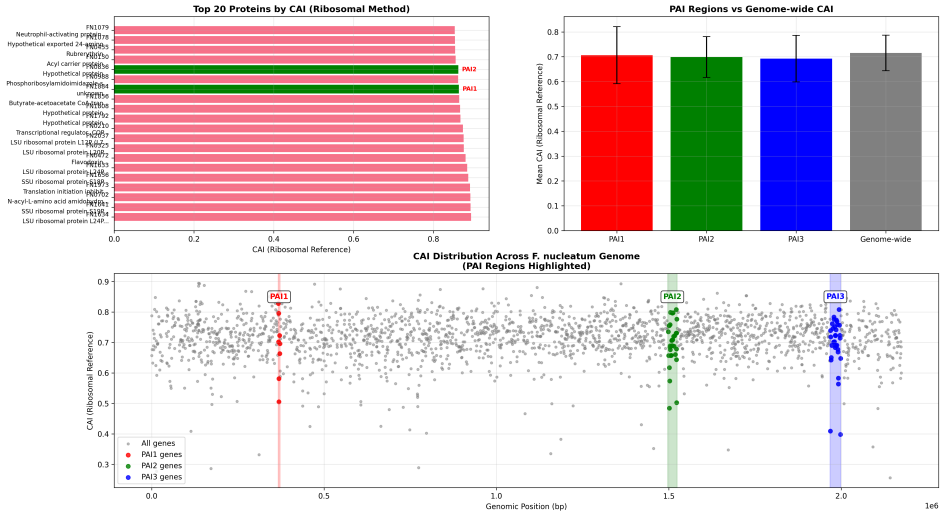

Figure 6: Distribution of the Codon Adaptation Index (CAI) along the genome of Fusobacterium nucleatum, with specific attention to potential Pathogenicity Islands (PAI1: 366,847–372,320; PAI2: 1,496,613–1,523,855; PAI3: 1,967,193–1,999,047). The analysis was performed using a ribosomal gene set as the reference for optimal codon usage.

PAI1 (region 366847–372320) contains genes with moderate to high CAI, including a ribosomal protein and a hemolysin, which may suggest functional integration into the host genome.

Table 1: Summary of genes in the predicted PAI1 region in *Fusobacterium nucleatum*, including predicted products, pathogenic potential, coexpression evidence, genomic context, and Codon Adaptation Index (CAI) values. High CAI values suggest codon optimization.

Locus Tag	Predicted Product	Pathogenic Potential	Coexpression Evidence	PAI Support	CAI
FN1885	Hemolysin III	HIGH – Direct cytolysis	Target gene	Primary virulence factor	0.722699
FN1884	Hypothetical protein	Unknown	Same strand, 406 bp upstream	Typical PAI composition	0.862872
FN1886	Hypothetical protein	Unknown	Same strand, 715 bp downstream	Typical PAI composition	0.663412
FN1881	Esterase	Medium – Membrane modification	Opposite strand, overlapping region	Metabolic modification	0.795734
FN1880	Oxygen-insensitive NAD(P)H nitroreductase	Low – Metabolic advantage	Same strand, 3.1 kb upstream	Anaerobic adaptation	0.702174
FN1879	Ribosomal protein S20P	Low – Housekeeping	Same strand, 3.5 kb upstream	Housekeeping (contradicts PAI)	0.827737
FN1882	Hypothetical protein	Unknown	Same strand, 2.1 kb upstream	Typical PAI composition	0.505487
FN1883	Transposase	Low – Mobile element	Same strand, 1.8 kb upstream	HGT evidence	0.581472
FN1887	Transposase	Low – Mobile element	Same strand, 1.0 kb downstream	HGT evidence	0.696104

CAI values ranged from 0.505 to 0.863, with most genes exceeding 0.65, indicative of moderate-to-high codon optimization relative to ribosomal gene reference sets. High CAI values in hypothetical proteins AAL93983.1 (locus tag: FN1884, CAI: 0.863) and AAL93978.1 (SSU ribosomal protein S20P, CAI: 0.828) suggest potential functional relevance, although one has uncertain annotation. At the same time, lower CAI scores in protein AAL93981.1 (locus tag: FN1882, CAI: 0.505) and AAL93982.1 (locus tag: FN1883, CAI: 0.581) may indicate either lower expression under standard conditions or acquisition from distinct genomic backgrounds via horizontal gene transfer. Functionally, the cluster contains a potentially highly pathogenic hemolysin gene and a likely coexpressed gene that could potentially facilitate the hemolysin pathway. Also, this region contains accessory hypothetical proteins and mobile genetic elements (transposases), consistent with horizontally acquired virulence modules. However, the inclusion of a ribosomal protein (housekeeping function) and metabolic enzymes (e.g., NAD(P)H nitroreductase) diverges from the traditional pathogenicity islands' composition.

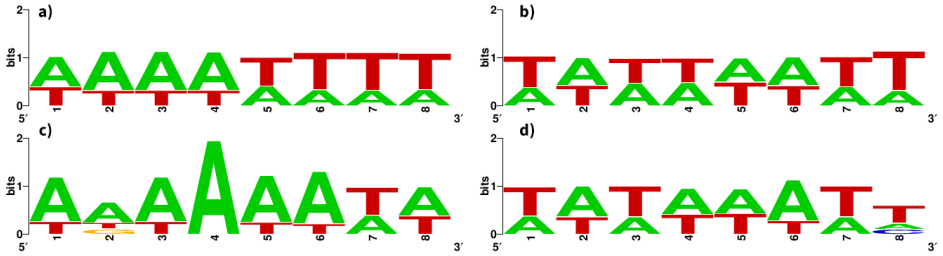

Figure 7: Promoter motif analysis of high and low CAI genes and PAI loci using WebLogo (Crooks et al., 2004; Schneider and Stephens, 1990). a) Sequence logo representation of enriched 8-mer motifs found in the promoters of high CAI genes (CAI > 0.771). Promoter motifs were found using the coding pipeline that calculates CAI, filtered sequences by minimum occurrence, sorted by frequency, and generated consensus motifs with WebLogo. b) Sequence logo representation of promoter motifs from low CAI genes (CAI< 0.670), generated using the same approach. c)Promoter motif analysis of the gene in PAI2. d)Promoter motif analysis of genes in PAI1.

Despite containing a high CAI gene, promoter motif analysis of the gene in the PAI1 region shows a pattern highly similar to low-CAI promoters, suggesting limited or intermittent transcriptional activity. Pathogenic islands often contain virulence genes that are actively transcribed when needed (Hacker and Kaper, 2000; Zhao et al., 2018); a region with weak or inconsistent promoter motifs may indicate that these genes are not strongly expressed. Combined with previous analyses, PAI1 may be more likely a remnant of DNA from horizontal transfer, and a coexpressed cluster of virulence genes rather than a functional PAI.

PAI2 (region 1496613–1523855) mostly contains hypothetical proteins and metabolic enzymes. Inspection of the region via blast and UniProt ID mapping found several transposases and toxin-antitoxin protein-coding genes. The high-CAI hypothetical proteins, which could be recently acquired genes that are being adapted for expression, and the presence of metabolic enzymes, suggest potential functional roles in niche adaptation. Promoter motif analysis reveals poly-A-tracts, which are associated with strong DNA curvature (Fiorini et al., 2011). This has been further supported by DNA curvature analysis, which indicated that 7 of 10 annotated promoters exhibit pronounced macroscopic curvature. On the other hand, only 5 of 32 protein-coding genes within the same region show comparable curvature (Gohlke, 2025). This aligns with the established role of DNA shape in transcriptional regulation, which

recognizes that intrinsic curvature tends to be concentrated in promoter regions than coding sequences in the same genome (Kozobay-Avraham, 2006) and is consistent with the identification of the promoters. DNA curvature enhances promoter–RNA polymerase interactions and stable open complex formation, which can associate with stronger transcription (Asayama and Ohyama, 2000). The presence of mobility genes, adaptive metabolic functionalities, and strong promoter motifs supports the interpretation that PAI2 represents a horizontally acquired genomic region with adaptive significance, and it is reasonable to propose that PAI2 may be a functional pathogenicity island, although additional experimental validation will be necessary to verify its role in virulence.

PAI3 (region 1967193–1999047) contains genes involved in translation and nucleotide metabolism, which are typically core functions. Core functional regions are generally not integrated from other genes.

Proteomics

Prediction of Protein interaction within PAI2

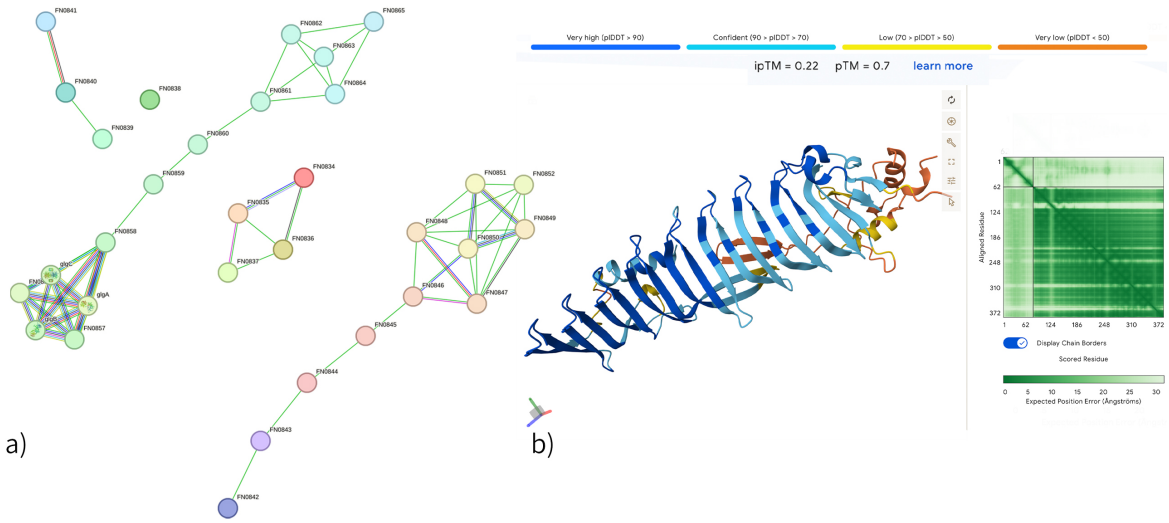

Figure 8: Predicting structural and interaction analysis for 32 proteins in region 1496613–1523855 (PAI2). a) STRING network analysis of predicted protein–protein interactions among proteins in PAI2 (Szklarczyk et al., 2024) b) AlphaFold-Multimer predicted 3D structure of FN0845 and FN0846 interaction with high pTM (0.70), suggesting confident structure prediction (Jumper et al., 2021).

The genomic region PAI2 (1496613–1523855) encodes 32 protein-coding genes, several of which are associated with mobility elements (e.g., tyrosine-type recombinase/integrase and transposases), consistent with acquisition via horizontal gene transfer. To further characterize the functional organization of PAI2, we STRING database analysis combined with BLAST-based functional annotation of all 32 encoded proteins, transmembrane domain prediction, followed by AlphaFold-Multimer structural validation of predicted interactions (Hallgren et al., 2022; Jumper et al., 2021; Krogh et al., 2001; Szklarczyk et al., 2024).

STRING network analysis organizes the 32 proteins encoded within PAI2 into 4 visually connected components and 1 isolated protein, as shown in Figure 8a (Szklarczyk et al., 2024). The proteins showing minimal connectivity to the main network are FN0838 (uncharacterized). This suggests they could not be functionally linked to any pathway, likely represent housekeeping genes captured within the island boundaries rather than a core PAI2 component.

Deep Learning tools, Biomni, analysis confidence scores derived from STRING interaction data using NetworkX reveal an average clustering coefficient of 0.560 and a network density of 0.138, indicating moderate connectivity with distinct functional modules (Huang et al., 2025). Community detection

identified 6 functional modules, with 2 inter-module connections suggesting pathway crosstalk (FN0845 and N0846; FN0858 and FN0859).

A core glycogen metabolism cluster comprising six proteins represents the largest and most densely connected module observed. This cluster includes glycogen synthase (*glgA*, FN0853), glucose-1-phosphate adenylyltransferase (*glgC*, FN0854; *glgC2*, FN0855), 1,4-α-glucan branching enzyme (*glgB*, FN0856), glycogen phosphorylase (*glgP*, FN0857), and 4-α-glucanotransferase (*malQ*, FN0858), with STRING confidence interaction scores ranging from 0.976 to 0.999. The tight clustering of these enzymes suggests coordinated expression and potential physical interaction, which may facilitate efficient substrate channeling during glycogen metabolism.

Another prominent cluster is associated with biotin biosynthesis, including 8-amino-7-oxononanoate synthase (*bioF*, FN0849), pimeloyl-ACP methyl esterase (*bioG*, FN0850), malonyl-ACPO-methyltra nsferase (*bioC*, FN0851), and an α/β hydrolase (FN0852). These proteins exhibit confidence scores of 0.914 to 0.988 and are supported by gene neighborhood conservation and co-expression evidence across multiple bacterial species.

A third cluster comprises proteins with potential virulence association. BLAST analysis indicated that FN0846 belongs to the YwqK family of type II toxin-antitoxin systems, frequently linked to pathogenicity islands. FN0846 interacts with a TPR-repeat-containing protein (FN0847) and DUF4125 domain-containing protein (FN0848), with a confidence score of 0.971, suggesting functional cooperation in stress response or host adaptation. Notably, interaction is observed between this virulence cluster and the biotin biosynthesis module, suggesting that *F. nucleatum*'s virulence may depend on metabolic pathways such as biotin production. While there is no direct evidence linking biotin (vitamin B7) biosynthesis to virulence in *F. nucleatum*, this metabolic-virulence interface warrants further investigation.

A cluster containing a RiboL-PSP-HEPN domain protein (FN0845), a BLAST-verified DUF2335 domain-containing protein (FN0842), and two uncharacterized proteins (FN0843, FN0844) was identified. HEPN (Higher Eukaryotes and Prokaryotes Nucleotide-binding) domains possess RNase activity and are commonly found in toxin-antitoxin systems. The genomic proximity of FN0845 to the YwqK antitoxin (FN0846) and their STRING-predicted interaction suggests these proteins may constitute a functional toxin-antitoxin module, with FN0845 potentially serving as the cognate toxin.

In addition to metabolic and virulence clusters, clusters of mobility-associated proteins were identified, consistent with the horizontally acquired nature of PAI2. Four mobility-associated proteins, a DUF6882 domain-containing exported protein (FN0834), a DUF6882 domain-containing protein (FN0835), a Dabb family stress-response protein (FN0836), and a tyrosine-type recombinase/integrase (FN0837), where integrase likely mediates chromosomal integration of PAI2. The transposase cluster (FN0840 and FN0841) may facilitate internal rearrangements or mobilization events.

Out of 32 analyzed genes, eight encode uncharacterized proteins. FN0843 and FN0844 show a potential link with the HEPN domain module; FN0838 cannot be functionally related to known pathways, and FN0839 is in proximity to the transposase cluster (FN0840, FN0841), possibly implicating mobility.

The remaining hypothetical proteins appear to interact with each other and are predicted to be membrane-associated. This cluster includes a predicted lipoprotein (FN0859), a TM2 domain-containing protein (FN0862), and four uncharacterized proteins (FN0860, FN0861, FN0863, FN0864). The presence of a lipoprotein suggests surface localization and potential host-pathogen interaction. STRING analysis indicated functional associations based on gene neighborhood conservation, although the specific functions are undetermined.

Structure Support

To further characterize these proteins, we performed transmembrane domain analysis using TMHMM2.0 and DeepTMHMM (Table 2) (Hallgren et al., 2022; Krogh et al., 2001).

Table 2: Comparison of TMHMM and DeepTMHMM predictions. Only proteins with TMHs are shown; displays number of predicted TMHs.

locus_tag	protein_id	TMHMM	DeepTMHMM	product
FN0839	AAL95035.1	1	0	unknown
FN0842	AAL95038.1	2	2	DUF2335 domain-containing protein
FN0844	AAL95040.1	2	2	unknown
FN0859	AAL95055.1	1	0	Lipoprotein
FN0862	AAL95058.1	2	2	TM2 domain-containing protein
FN0864	AAL95060.1	2	0	unknown
FN0865	AAL95061.1	0	12	Outer membrane protein

Minor discrepancies were observed for FN0839, FN0859, and FN0864 between the two prediction methods. A notable difference was observed for FN0865, where TMHMM2.0 predicted no transmembrane helices, while DeepTMHMM predicted 12 membrane spanning regions. Detailed analysis of the DeepTMHMM output indicated that FN0865 is predicted to adopt a beta-barrel topology with alternating beta-sheets and periplasmic/extracellular loops, characteristic of outer membrane proteins (OMPs) in Gram-negative bacteria. This beta-barrel architecture, rather than alpha-helical transmembrane domains, explains the discrepancy between prediction methods, as TMHMM2.0 focused on alpha helical TM topology (Hallgren et al., 2022). The OMP classification of FN0865 is consistent with its potential role in host-pathogen interactions and supports its inclusion as a virulence-associated factor.

To validate predicted protein-protein interactions and assess the structural plausibility of complex formation, we performed AlphaFold-Multimer analysis on protein pairs/complexes prioritized based on STRING evidence and functional relevance (Jumper et al., 2021). Interface predicted TM-score (ipTM) values above 0.5 indicate confident interactions, while predicted TM-score (pTM) values exceeding 0.7 reflect reliable overall folding.

The FN0845 and FN0846 pair (putative HEPN toxin and YwqK antitoxin) exhibited the highest overall structural confidence (pTM = 0.70). FN0846 formed a canonical beta-barrel fold characteristic of DNA-binding antitoxins, and FN0845 displayed a helical domain typical of HEPN-family toxins (Figure 8b). Although the interface confidence (ipTM = 0.22) was not ideal, suggesting transient or conditional interaction, the predicted architectures align with established toxin-antitoxin modules.

For the glycogen metabolism pair (FN0853 and FN0854), individual domains folded with moderate confidence (pTM = 0.55), but interface confidence was low (ipTM = 0.14). Despite a strong functional association (STRING score 0.999), this pattern supports their role as sequential enzymes in glycogen biosynthesis rather than as stable complex partners.

The predicted structure for the four-protein membrane-associated complex (FN0860, FN0861, FN0863, FN0864) had low overall confidence (pTM = 0.37, ipTM = 0.28), forming a helical bundle. The low confidence likely reflects the limitations of AlphaFold-Multimer in modeling membrane environments, as these proteins are predicted to be membrane-associated.

Among transmembrane protein pairs, FN0839 and FN0860 showed the highest interface confidence (ipTM = 0.32), with helical architectures suggesting potential membrane complex formation. However, overall low confidence scores for membrane protein pairs further underscore AlphaFold-Multimer's constraints when predicting interactions dependent on lipid bilayer context.

Co-expression relationships among PAI2-encoded proteins

To further support computational predictions of protein interactions, we employed Bayesian network structure learning using the bnlearn R package (Scutari, 2010). Expression-based evidence was obtained from GSE161360 (GEO), a publicly available RNA-seq data from Ponath et al. (2021), who characterized the transcriptional landscape of *F. nucleatum* across multiple subspecies and growth conditions. We extracted 18 samples corresponding to *F. nucleatum* subsp. nucleatum ATCC 25586, containing three growth phases (exponential, mid-log, and stationary) and two treatment conditions. The original study focused on identifying novel non-coding RNAs and transcriptional start sites; here, we used their expression data and analysis with bioinformatics tools to infer protein co-expression networks within the PAI2 region (Ponath et al., 2021). Although optimal Bayesian network learning

typically requires 30–50 samples, we are only focused on 32 genes rather than the entire genome (18 samples provide sufficient statistical power for meaningful inference).

Using the hill-climbing algorithm with BIC score and 500 bootstrap iterations, bnlearn identified 56 significant edges (bootstrap strength > 0.5, direction > 0.5). Comparison with STRING database predictions revealed 14 overlapping interactions supported by both expression-based and database-based evidence. The strongest overlapping interactions involve glycogen metabolism genes (FN0853 (glgA) and FN0854 (glgC), FN0856 (glgB) and FN0857 (glgP), FN0855 (glgC) and FN0856 (glgB)). This concordance is biologically expected, as these enzymes function coordinately in glycogen biosynthesis and degradation. Similarly, the biotin biosynthesis module showed strong co-expression((FN0849 (bioF) and FN0851 (bioC), FN0849 and FN0850, FN0850 and FN0852)).

For the hypothetical protein cluster (FN0860-FN0865), bnlearn detected several co-expression relationships not captured by STRING. There is interaction of FN0864 and FN0837 (strength = 0.774, direction = 0.651), FN0860 and FN0848 (strength = 0.620, direction = 0.619), and FN0859 and FN0848 (strength = 0.598, direction = 0.540). Moreover, within the PAI2 hypothetical protein seems to coexpress with each other even when not adjacent to each other, for instance, FN0861 exhibited strong co-expression with FN0844 (strength = 0.950), suggesting potential functional coordination between uncharacterized proteins in this region. Regarding the hypothesis that hypothetical proteins may interact at the membrane based on function and structure evidence, bnlearn data also provided partial support. FN0865 (outer membrane protein) showed co-expression with FN0842 (strength = 0.712), FN0849 (strength = 0.722), and FN0850 (strength = 0.706). FN0859 (lipoprotein) exhibited co-expression with FN0844 (strength = 0.638) and FN0848 (strength = 0.598). Additionally, FN0863 showed strong co-expression with FN0846 (exported protein; strength = 0.726).

The expression data is related limited on sample size (n=18), which may result in both false negatives (missed true interactions) and false positives (spurious edges). Expression-based evidence suggests the protein pair identified showed co-expression patterns, which may reflect shared regulatory control, membership in common pathways, or indirect functional relationships rather than direct protein-protein binding. The low overlap between bnlearn and STRING (only 14 pairs) is expected, as STRING aggregates literature and experimental evidence across organisms, while bnlearn captures condition-specific co-expression patterns unique to the experimental conditions examined.

PAI predicted interactome

The integration of STRING network analysis, BLAST functional annotation, transmembrane domain prediction, and AlphaFold-Multimer structural validation suggests PAI2 as a mosaic genomic island with characteristics of metabolic adaptation, virulence, and mobility.

All computational tools used in this section, including STRING, AlphaFold-Multimer, bnlearn, none supported direct physical interactions among these hypothetical proteins. AlphaFold-Multimer interface predictions showed generally low confidence (ipTM < 0.5), indicating that predicted interactions may be transient, require additional cofactors, or occur in membrane contexts not captured by the algorithm. Furthermore, the functional associations predicted by STRING may represent co-expression or genomic proximity rather than direct physical interactions. Finally, 8 of 32 proteins remain functionally uncharacterized despite predictions made by computational analysis. To verify that those proteins actually form complexes and exhibit pore-forming or toxin-related activity in vitro, biochemical and functional assays will be necessary.

Taken together, our analysis of PAI2 suggests its function as a pathogenicity island combining mobility-associated genes, virulence factors including a toxin-antitoxin system, metabolic adaptation genes, and membrane-associated hypothetical proteins. The AlphaFold-Multimer structural validation of the FN0845 and FN0846 toxin-antitoxin pair provides structural evidence consistent with the virulence-associated nature of this island. The transmembrane domain predictions and STRING-derived interaction networks predicted that some uncharacterized proteins may assemble into membrane-associated complexes. However, these insights are primarily based on computational predictions. Further experimental validation is needed to confirm biological significance.

Pathogenic cluster analysis and interesting pathogenic activity

Since region 1496613–1523855 is identified as a potential PAI, regions around it that also seem to deviate significantly from the normal GC skew pattern were analyzed. Adjacent to region 1496613–1523855,

a 150-kb region (1.7–1.85 Mb), which contains 136 annotated proteins, was analyzed for pathogenic potential. The same process of ID mapping, homology-based search (BLAST), and functional categorization was performed on this region. Of the 136 proteins present in this region, 37/136 (27.21%) are unannotated in the NCBI database. The incomplete characterization of this region underscores the potential presence of novel or strain-specific functions.

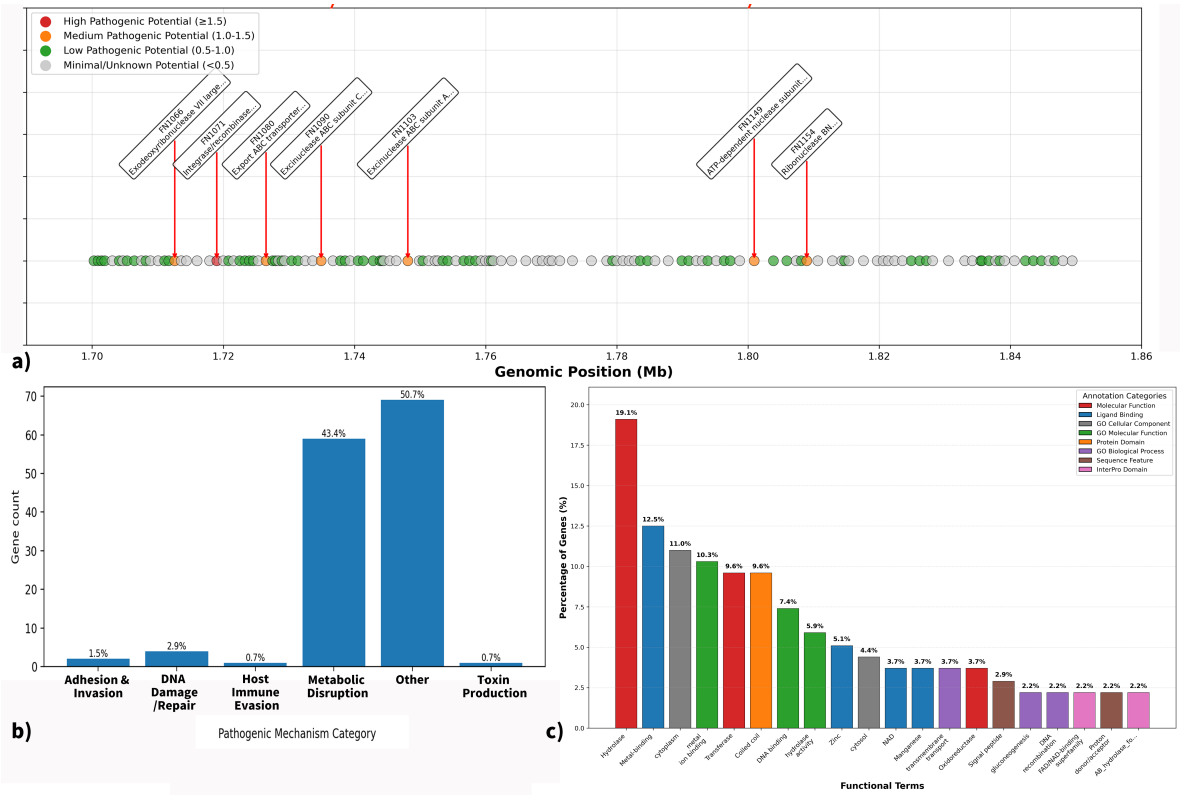

Figure 9: Functional and Pathogenic Annotation of the 1.7–1.85 Mb Genomic Region in *F.nucleatum* ATCC 25586. 1) Gene organization schematic for 1.70-1.85 Mb region. The genes of high and medium pathogenicity are labeled. Classification of potential pathogenicity was completed by Biomni's (Huang et al., 2025) annotation pipeline. b) Predicted pathogenicity mechanisms among genes in these genomic regions based on PubMed-assisted text mining. c) The functional annotation output of DAVID for the same gene set (Huang et al., 2008; Sherman et al., 2022). The total percentages are higher than 100%, as this is a feature of DAVID where gene-term associations overlap and are counted multiple times across similar or redundant terms.

Functional classification based on PubMed-assisted text mining showed that Genes for Metabolic Disruption had the highest abundance, while Genes for DNA Damage/Repair were second. This suggests that these regions may be playing a key role in the metabolic flexibility and / or genomic integrity of bacteria. DAVID functional annotation provided further support, where the most enriched categories were for Hydrolase activity (19.1%), Metal-ion binding (12.5%), and Ion Transport (11%). These functions are likely involved in host interaction, nutrient acquisition, and stress-facilitation, providing additional evidence consistent with possible pathogenic capacity. The biomni analysis needs to be interpreted cautiously, because the method of evaluating pathogenic potential is unknown, and the prediction of pathogenic protein products could potentially involve either virulence or normal cellular pathways.

Similarly, constraints are present for the PubMed-assisted text mining analysis. Most genes have roles in multiple pathways, and the same protein in different pathways may play entirely different roles. The abundance of hypothetical proteins limits functional interpretability, and automatically mined text introduces variability in the results based on the time of database indexing, revisions to entries, and the sensitivity of the keyword search. In addition, the percentage values in DAVID are

greater than 100%; this is a recognized artifact resulting from overlapping gene-term associations, for instance, where genes belong to more than one functional category.

The analysis suggests that this area of the genome may represent a contribution to the overall pathogenic potential, and the metal-ion pathway potentially related to iron acquisition should be given special attention. However, these are computational results that may be strain dependent. Experimental evidence, likely involving transcriptomics or functional assays, could help validate the involvement of this region in virulence.

The iron in "bacterial economy": hemolysin and iron ions transporters

Hemolysin genes are identified throughout the genome. This gene encodes a TlyA family protein, present in phylogenetically distant bacteria, being primarily found in distantly related bacteria such as *Mycobacterium*, *Helicobacter*, and *Brachyspira* species, rather than within the Fusobacteriota phylum (Arenas et al., 2011; Rahman et al., 2015). This TlyA protein shows pathogen-specific distribution patterns, where it is observed in pathogenic species like *M. tuberculosis* and *H. pylori* but is absent in non-pathogenic relatives, suggesting a strong selective pressure for its retention in pathogens (Keith et al., 2022; Sharma and Singh, 2022). Functionally, hemolysin operates as a pore-forming cytotoxin that lyses host erythrocytes, releasing iron and nutrients, allowing F. nucleatum to acquire nutrients in inflamed periodontal pockets (Ma et al., 2023; Samainukul et al., 2019).

Hemolysin and iron acquisition mechanism in F. nucleatum

The identification of hemolysin suggests that F. nucleatum (Fn) may specialize in a heme utilization pathway to acquire iron from the host, indicating its potential ability to colonize and thrive in blood-rich environments such as periodontal pockets and tumor microenvironments.

From genome annotation, key proteins and genes involved in the heme-based iron acquisition pathway are identified. This system relies on ABC transporters (HmuUV), which comprise Hemin Transport System Permease Protein HmuU (encoded by genes hmuU_1 and hmuU_2) and the Heme Import ATP-Binding Protein HmuV (encoded by gene hmuV). HmuV, as the ATP-binding component, provides the energy to power HmuU. HmuU then facilitates the physical transport of the heme molecule from the blood and tissues across the cell membrane to the bacterial cells. While these proteins are dedicated to heme uptake, other ABC transporters for various metals, such as an Inner Membrane ABC Transporter Permease Protein (ydcV), a Zinc Import ATP-Binding Protein (znuC), Manganese ABC transporter substrate-binding lipoprotein (psaA), Iron import ATP-binding/permease protein (irtA), and many others, were also identified, highlighting a broader strategy for metal acquisition (Delepelaire, 2019; Louvel et al., 2006; McGregor et al., 2023).

Following import, the heme molecule is enzymatically degraded to release iron, a process in which flavodoxin (encoded by fldA, ispg, and isiB; highly expressed, cai of 0.880656) acts as an electron-transfer protein, reducing the ferrous complex and facilitating catalysis (Boynton et al., 2009; McGregor et al., 2023).

The iron is processed so it can be used by the cell for various cellular processes. In addition to basic metabolism sustaining bacteria, these iron acquisition mechanisms may disrupt erythrocytes and potentially promote metastasis in colorectal cancer. The virulence protein, DNA protection during starvation protein (Dps), for instance, serves as a bacterial ferritin to protect DNA from oxidative stress while also acting as a potent pathogenic factor. Besides disrupting and lysing the host's erythrocytes to acquire iron, Dps also upregulates the chemokine CCL2/CCL7 to enhance the bacterium's survival in macrophages. The expression of CCL2/CCL7 also contributes to cancer progression by inducing epithelial-mesenchymal transition (EMT) and promoting colorectal cancer (CRC) metastasis in vivo (Wu et al., 2023).

Genomic evidence showed complete iron acquisition machinery, and *F. nucleatum* relies on iron-dependent enzymes for core metabolic processes. An additional metabolic model analysis is done to see if iron is essential for *F. nucleatum* survival and proliferation.

How scavenging iron helps metabolic growth

The annotated genome *F. nucleatum* ATCC 25586 was utilized with CarveMe (Machado et al., 2018) to reconstruct a genome-scale metabolic model, comprising 1,914 reactions and 1,339 metabolites, with 582 associated genes. Flux Balance Analysis (FBA) was applied to determine the optimal growth rate, which was predicted to be 67.42 mmol/gDW/hr under ideal nutrient-rich conditions. The flux through the biomass reaction represents the rate at which the bacterium produces those metabolites necessary for its growth (e.g., amino acids, lipids, cofactors, and proteins). The biomass flux returned by COBRA is the specific growth rate that can be used to calculate the specific biomass of the bacteria at a specific time using an exponential function. Therefore, biomass flux can be used as an indicator of the growth rate of the bacterium. Iron is required in a minimal medium for maintaining the growth rate as indicated by the CarveMe-generated metabolic model. The model identified uptake fluxes of 0.452702 mmol/gDW/h for ferrous iron (EX_fe2_e) and 0.524388 mmol/gDW/h for ferric iron (EX_fe3_e) as necessary to maintain the organism's growth rate in a minimal medium.

Although it contains iron requirements in the composition of the biomass and the model contains reactions involving iron metabolites, COBRApy analyses performed using the metabolic model generated by Carveme evaluate biomass production under all iron conditions, including complete absence of iron, limitation of iron uptake, and unlimited iron availability, give identical results (67.416545 mmol / gDW / hour). This indicates missing iron dependencies or unrealistic constraints in the model.

The failure of the stimulation could be explained by FBA's inherent reliance on functional optimization in the face of biological redundancy. The metabolic networks contain multiple iron-chelate compounds (enterobactin, salmochelin, heme) and alternative metabolic pathways that can perform equivalent functions, masking true iron dependencies. In this simulation, only the EX_fe2_e and EX_fe3_e are set to (0,0), while other iron exchange reactions allow unrealistic, unlimited iron uptake. The FBA algorithm, pursuing optimization of biomass flux output, can simply re-route metabolic flux through these redundant pathways. Consequently, the simulation predicts no impact on growth, leading to a classification of the gene as non-essential (Boone, 2025).

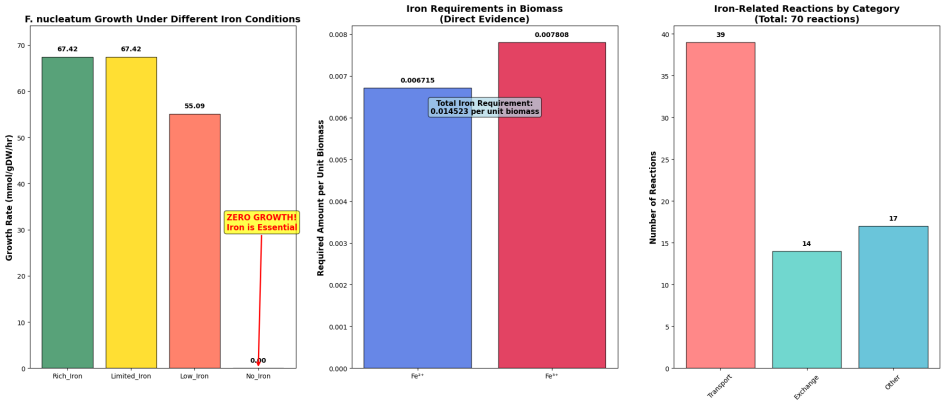

Figure 10: Multi-panel visualization that provides a comprehensive overview of iron's essentiality for *F. nucleatum*, FBA flux analyzed using cobrapy. The panels show that iron deprivation completely halts growth, while rich and limited iron conditions support high growth rates, emphasizing iron's metabolic importance. Additionally, the presence of diverse iron acquisition systems in the model supports iron critical role in survival and virulence.

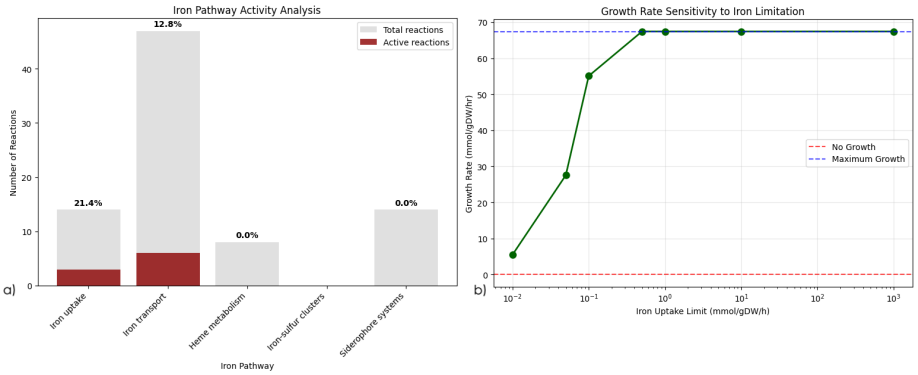

Figure 11: Detailed flux analysis, sensitivity curves, and pathway activity, using COBRApy. a) Bar graph compares the total and active reactions across five iron-related metabolic pathways: Iron Uptake, Iron Transport, Heme Metabolism, Iron-Sulfur Clusters, and Siderophore Systems. Grey bars represent the total number of reactions, while red bars indicate the subset of reactions that are active under the studied conditions. Percentages above each bar denote the proportion of active reactions, highlighting differential pathway utilization and potential regulatory prioritization. b) This plot illustrates the relationship between iron uptake limits and cellular growth rate. The x-axis shows iron uptake limits (mmol/gDW/hr) on a logarithmic scale, while the y-axis indicates growth rate (mmol/gDW/hr).

Identifying this potential limitation, 70 reactions and 40 metabolites related to iron across all three cellular compartments (cytoplasm, periplasm, and extracellular space) are used as the new constraints. The refined code indicates iron is required in biomass composition, for *F. nucleatum* cannot produce biomass without iron. FBA under different iron availability conditions is then performed. The model predicts that complete iron deprivation results in the arrest of growth. The metabolic consequences of iron limitation show that although iron is essential for bacterial growth, it only requires a limited amount of iron for maximum growth ($< 1\ mmol/gDW/h$). Under iron-limited conditions, the model shows a compensatory activation of multiple siderophore systems to maximize scavenging, although this still results in an 18.3% reduction in growth rate. In addition, the bacterium possesses complex iron acquisition mechanisms that involve direct iron transport (ferrous iron (Fe^{2+} transport, ferric iron (Fe^{3+}) transport, and ABC transporters), siderophore-mediated iron acquisition (enterobactin system, salmochelin system, and staphyloferrin systems), and heme iron uptake.

Furthermore, this bacterium appears to demonstrate a strong preference for ferric iron (Fe^{3+}) over ferrous iron (Fe^{2+}). An examination of iron transport systems in *F. nucleatum* using the metabolic model generated by Carvme from annotated Prokka through an AI-assisting tool, Biomni. The results suggested that 77.7% of total iron transport within the model is dedicated to Fe^{3+} uptake, with direct ferric iron uptake flux being 0.526, ferripyoverdine complex formation flux being 0.526, and Fe^{3+}-pyoverdine exchange flux being 0.526. The direct ferrous iron transport flux is only 0.453.

Ferrous iron (Fe^{2+}) is usually more soluble over a larger pH range than ferric iron (Fe^{3+}), which hydrolyzes and precipitates as iron(III) hydroxide at near-neutral and higher pH values. The preference of *F. nucleatum* for Fe^{3+} may represent a metabolic adaptation to obtain iron in acidic niches, typically seen in tumor microenvironments.

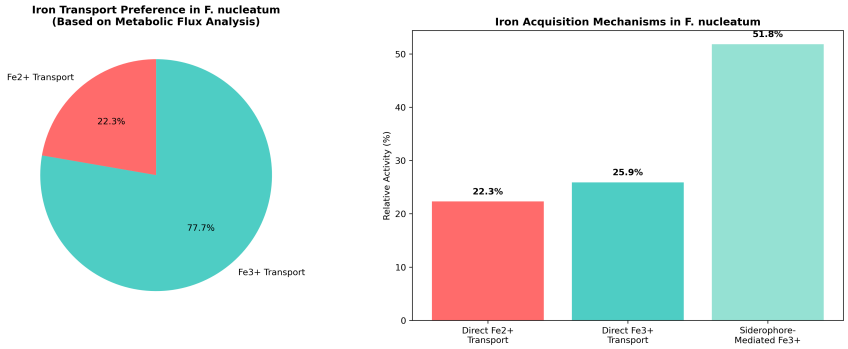

Figure 12: Examination of iron transport systems in *F. nucleatum* using the metabolic model generated by Carvme from annotated Prokka through an AI-assisting tool, Biomni.(Huang et al., 2025)

F. nucleatum and cancer

Previous studies show *F.nucleatum* is frequently found in cancer tissue as well as associated with promoting tumor progression, supporting colonization in the cells, and altering the immune response within the TME. High levels of *F. nucleatum* is associated with the number of tumor-promoting immune cells, including tumor-associated macrophages (TAMs), tumor-associated neutrophils (TANs), dendritic cells (DCs), and myeloid-derived suppressor cells (MDSCs), and a decrease in the function of T-cells and natural killer (NK) cells. *F. nucleatum* contains virulence factors, such as Fap2, FadA, RadD, and FomA, which enable it to interact with cancer cells, modulating key signaling pathways, altering cytokine secretion, and promoting cancer progression. This bacterium can also secrete outer membrane vesicles that facilitate metastasis by promoting autophagy in oral cancer (Rubinstein et al., 2013; Willenbockel et al., 2024).

Moreover, lysozyme inhibitor Fn1792, a recently identified virulence factor, inhibits the apoptosis of host phagocytes through CX3CR1 expression. CX3CR1 expression can recruit CX3CR1$^+$ PD-L1$^+$ phagocytes to tumor tissues to adapt their microenvironment, resulting in immune evasion (Chen et al., 2024). *F.nucleatum* has lysed erythrocytes to free hemoglobin for growth as an iron source. Iron accumulated in macrophages enhances Fn-induced chemokine production through the TLR4/NF-κB signaling pathway (Yamane et al., 2022). Therefore, iron availability may be predicted to affect the CX3CR1 induction by Fn1792, where further experimental studies would be required to confirm the possibility.

Recently, emerging evidence suggests it also contributes to mutagenesis (genomic instability) and DNA damage, particularly in the context of chronic inflammation. *F.nucleatum* contributes to mutagenesis indirectly through inflammatory responses and directly through metabolite production rather than directly inserting mutations, but these interactions accelerate the acquisition of mutations in tumor cells. One potential interpretation is that inflammation signals by *F.nucleatum* enhance the production of reactive oxygen species (ROS) and recruit inflammatory cytokines (Zhao et al., 2022).

Our genomic analysis identified a predicted hemolysin cluster (coordinates: 366,847–372,320 bp). Hemolysin III (FN1885) with CAI of 0.723 (moderate-high codon optimization) is a pore-forming cytotoxin with direct membrane-damaging activity. Promoter mapping indicated that FN1885 shares a promoter with the downstream hypothetical protein FN1886, suggesting coordinated expression. This genomic organization is consistent with a functional virulence module acquired through horizontal gene transfer, as evidenced by the presence of flanking transposases (FN1883, FN1887).

Hemolysis is generally associated with iron overload and excess hemoglobin in the bloodstream. Hemolysins lyse red blood cells, releasing iron, indirectly increasing iron concentration in blood (Millot et al., 2016). Free iron undergoes Fenton reactions, generating reactive oxygen species (ROS) such as hydroxyl radicals $Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + OH \cdot + OH^-$, causing oxidative damage to lipids, proteins, and DNA (Xuan et al., 2025; Zhao, 2019).

Hemolysis expression is context dependent

A critical observation is that hemolysin genes are typically not highly expressed under standard laboratory culture conditions. Transcriptomic analysis of *F. nucleatum* subsp. animalis during Caco-2 cell infection indicated that hemolysin III was among only 5 genes consistently upregulated during invasion (Cochrane et al., 2020). Computational essentiality analysis using NetGenes identified 202 essential genes in *F. nucleatum ATCC 25586* (Database, 2024). Several entries in the NetGenes essential gene set lacked specific functional annotation. To address this, we cross-referenced all essential genes with genome annotation from the NCBI GenBank database (NCBI accession: AE009951.2). 20 essential genes with predicted roles in iron homeostasis and transport is identified. Among these, FbpC (FN0376), a central ATP-binding protein for Fe^{3+}, several ABC transporters, and additional iron transport genes were included in the essential gene set. Notably, hemolysin genes were not among the essential genes. This is consistent with the understanding that virulence factors are dispensable for in vitro growth but critical for in vivo colonization and pathogenesis (Kitamoto et al., 2015; Los et al., 2013; Nicholas et al., 1999; Opal et al., 1990). The absence of hemolysin from the essential gene set, combined with transcriptomic evidence of host-induced expression, suggests hemolysin represents a conditionally expressed virulence factor activated specifically during host infection. This explains why hemolysin activity may be underestimated in laboratory studies while being highly relevant in clinical contexts such as tumor colonization.

Proposed mechanism: Hemolysin-Iron-ROS-Hippo pathway

Our flux balance analysis (FBA) using a genome-scale metabolic model indicated that iron is essential for *F. nucleatum*. Through genomic analysis of *F. nucleatum strain ATCC 25586*, direct cytotoxicity virulence factors are identified (shlB_3, shlB_2, shlB_1, tlyA). Iron acquisition by *F. nucleatum* through hemolysis is proposed to be responsible for *F. nucleatum* modulation of the hippo pathway, which may allow *F. nucleatum* to play a role in cancer progression and therapy resistance (Han et al., 2020; Millot et al., 2016; Wang et al., 2024).

Hemolysin enables *F. nucleatum* to acquire iron through a specific pathway. Hemolysin secretion leads to pore formation in erythrocyte membranes, releasing hemoglobin into the extracellular environment (Samainukul et al., 2019). Intracellular heme catabolism liberates iron, which participates in the Fenton reaction, generating reactive oxygen species (ROS) (Xuan et al., 2025; Zhao, 2019). ROS inhibit LATS1/2 activity, leading to YAP/TAZ dephosphorylation and nuclear entry. Activated YAP-TEAD signaling drives anti-apoptotic responses by upregulating BCL-2 and suppressing caspase-3–mediated GSDME cleavage, thereby reducing host cell pyroptosis(Amanda et al., 2024; Chen et al., 2017; Wang et al., 2024).

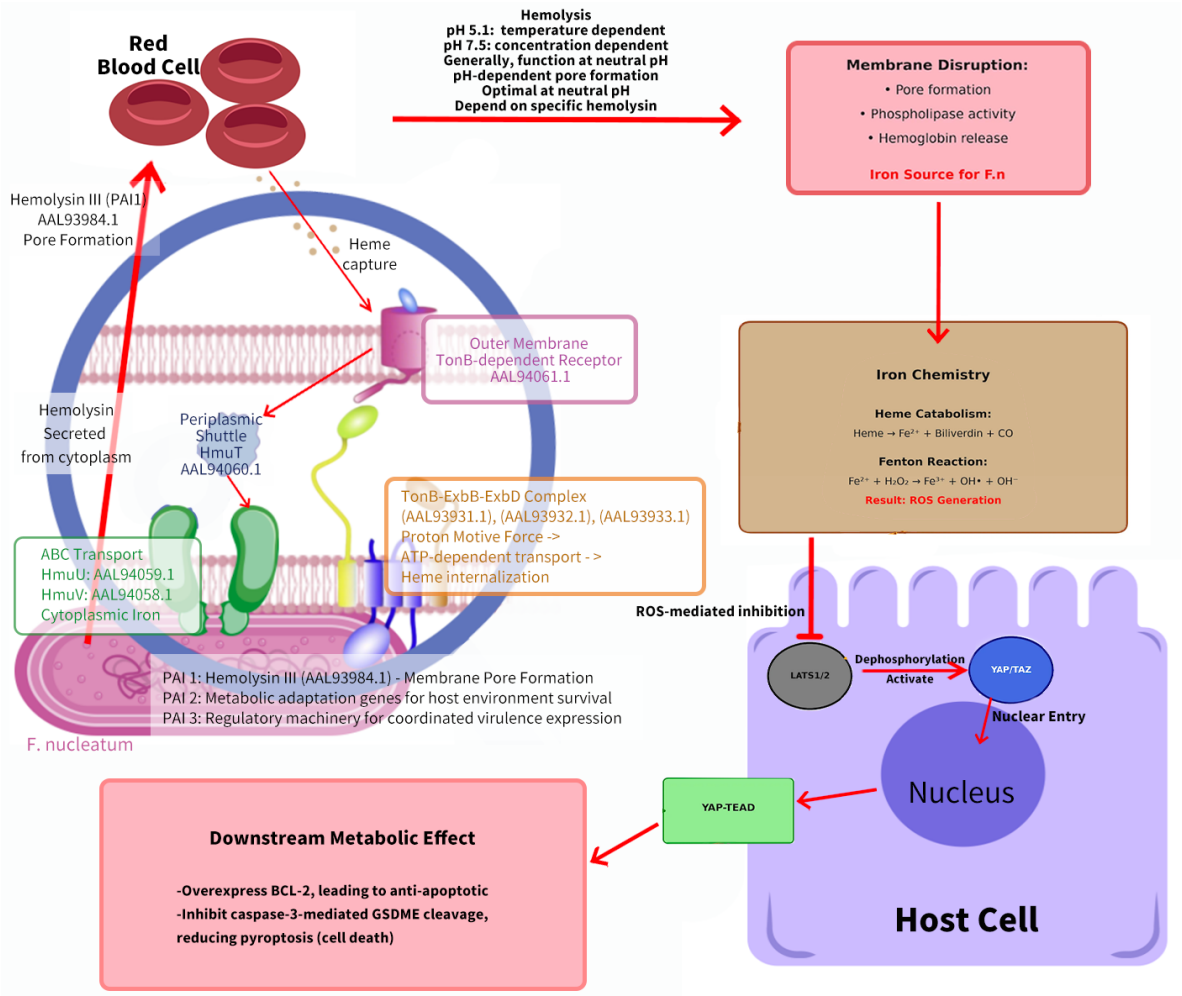

Figure 13: We propose that *F. nucleatum*'s iron acquisition through hemolysis has downstream effects on host cell signaling that promote cancer progression. The pathway produces reactive oxygen species (ROS), which could be associated with oxidative DNA damage (Albesa et al., 2004). Key proteins involved in this pathway, including TonB-dependent receptors, HmuTUV ABC transporters, and the TonB-ExbB-ExbD energy transduction system, with their respective protein IDs, are labeled (Braun et al., 2023; Richard et al., 2019; Silale and van den Berg, 2023).

Iron-induced oxidative stress inhibits LATS1/2 kinases, which dephosphorylate and nuclear translocate YAP and TAZ, the downstream effectors of the Hippo pathway. YAP/TAZ overactivation can promote pluripotency, tumor initiation, and chemotherapy resistance, which are common signs in cancer cells. They upregulate multidrug transporters (e.g., ABCB1, ABCC1, ABCG2) and anti-apoptotic proteins (e.g., Bcl-xL, cIAP1, Survivin), while suppressing p53 via MAPK/NF-κB signaling. Furthermore, YAP/TAZ activate TEAD-mediated transcription, and BCL-2 is a downstream target of the YAP-TEAD1 transcriptionally active complex (Amanda et al., 2024; Moroishi et al., 2015; Nguyen and Yi, 2019; Xuan et al., 2025). BCL-2 regulates apoptosis, and its overexpression is found in over 50% of oral squamous cell carcinoma (OSCC) tissues. The balance between pro-apoptotic (e.g., BAK, BAD) and anti-apoptotic (e.g., BCL-2, MCL-1) proteins determines tumor cell survival (Chen et al., 2017). *F. nucleatum*'s iron-mediated activation of YAP/TAZ may thus contribute to OSCC progression and resistance to anti-tumor therapies.

Further annotation of the Prokka-annotated genome was performed using eggNOG-mapper (see supplementary file `FN.annotations.xlsx`) (Huerta-Cepas et al., 2018). The result indicated the presence of the adhesin protein FadA (Fusobacterium adhesin A), but didn't direct the identification of Fap2 or RadD in their classic forms. However, related proteins with similar functions were identified.

Two FadA proteins were found in the dataset (EJMLJHKB_00032 and EJMLJHKB_00869), both annotated as "Adhesion protein FadA" with the FadA protein family domain (PFAMs: FadA); several other proteins with the FadA binding domain were also identified. FadA can activate the β-catenin and Wnt pathway via the binding to E-cadherin on the surface of CRC cells, causing relocation. Under stress conditions, FadA increase pathogenicity via conformational changes (Chen et al., 2022; Rubinstein et al., 2013).

While direct Fap2 was not found, several galactose-related proteins were found, including galactose kinase (EJMLJHKB_00595) and galactose-1-phosphate uridyl transferase (EJMLJHKB_00596), along with trimeric autotransporter adhesins (EJMLJHKB_01075 and EJMLJHKB_01330) which have structural homology with Fap2. Fibroblast activation protein 2 (Fap2), an outer membrane protein from F. nucleatum, is a galactose-sensitive hemagglutinin and adhesive protein that facilitate bacterial co-aggregation and host cell adhesion through binding to D-galactose-β (1-3)-N-acetyl-D-galactosamine (Gal-GalNAc) on host cell surfaces. Fap2 can bind to inhibitory receptors on immune cells, preventing immune recognition and facilitating immune evasion (Coppenhagen-Glazer et al., 2015; Ganesan et al., 2019). Although RadD is not directly present, multiple DNA repair and recombination proteins were identified, including RecG helicase (EJMLJHKB_00161), DNA topoisomerases (EJMLJHKB_00613, EJMLJHKB_00614), transcription-repair coupling factor (EJMLJHKB_00637), and Holliday junction resolvase RuvC (EJMLJHKB_00821). RadD-related proteins function in DNA repair and recombination pathways. The identified protein also has a role in both of these pathways.

Our analysis identified one hemolysin cluster and a potential pathogenicity island, which could contribute to the proposed mechanism. PAI1 (366,847–372,320) may provide effector (hemolysin), initiating iron acquisition and membrane disruption. PAI2 (1,496,613–1,523,855), with metabolic adaptation genes (biotin synthesis, glycogen metabolism), may support survival in nutrient-limited host environments.

Conserved Mechanism Across Pathogens

The hemolysin-ROS-genotoxicity pathway is not unique to *F. nucleatum*. Similar mechanisms have been documented in other cancer-associated bacteria, such as *E. coli* (Jamali et al., 2026; Nhu et al., 2019; Trouillon et al., 2023). *Uropathogenic Escherichia coli* (UPEC) originates from the gut, migrates to the urethra, is asscioated with bladder cancer. It secretes α-hemolysin (HlyA), a pore-forming toxin that disrupts mitochondrial function and promotes the generation of reactive oxygen species (ROS) in targeted host cells (Brito et al., 2018). ROS burst involves *E. coli* can lead to oxidative stress in blood, oxidative damage to macromolecules (proteins and lipids), oxidative DNA damage(Albesa et al., 2004; Baronetti et al., 2013). ROS can cause DNA damage during cell replication, and subsequently, mismatched or unpaired damaged DNA can cause the formation of mutations (Verma et al., 2020; Waris and Ahsan, 2006).In specialized cells like spermatozoa and renal epithelial cells, exposure to hemolytic *E. coli* strains leads to increased intracellular ROS, which subsequently causes DNA strand breaks and genomic instability (Tvrdá et al., 2022).

The hemolysin-ROS pathway in bacteria can function as a mutagenic mechanism, particularly by inducing oxidative stress that causes DNA damage and subsequently increases mutation rates. This is not only proven in *E. coli*, but also shown in *Staphylococcus aureus*, which secretes α-toxin (Hla) leading to ROS production and resulting in DNA damage and apoptosis (Deplanche et al., 2019; Ha and Edwards, 2021; Menzies and Kourteva, 2000). Given the conserved mechanisms among pore-forming toxins, it is plausible that *F. nucleatum* hemolysin acts similarly.

However, not all bacteria containing hemolytic activity are associated with cancer, such as *Streptococcus pyogenes*, *Vibrio cholerae*, *Bacillus cereus*. S. pyogenes produces two major hemolysins, streptolysin O (SLO) and streptolysin S (SLS), which activate inflammasomes, induce pro-inflammatory cytokine release, and cause tissue destruction during acute infections such as pharyngitis, necrotizing fasciitis, and toxic shock syndrome (Brouwer et al., 2023; Langshaw et al., 2023; Pinho-Ribeiro et al., 2018; Richter et al., 2021; Shannon et al., 2024). Similarly, V. cholerae disrupts cell membranes and triggers inflammatory responses by β-barrel pore-forming hemolysin/cytolysin (VCC) (Gandhi et al., 2025; Singh et al., 2025). B. cereus produces multiple hemolysins, including hemolysin BL (Hbl), non-hemolytic enterotoxin (Nhe), and cytotoxin K, which cause food poisoning manifesting as diarrheal or emetic syndromes (Jovanovic et al., 2021). However, all are acute infections with a transient nature, which provides no opportunity for the persistent bacterial-epithelial interaction required for oncogenic transformation (Mager, 2006; Parsonnet, 1995).

Bacteria epidemiologically linked to cancer have the ability to establish chronic, persistent colonization of host tissues. The long persistence allows cumulative DNA damage from sustained oxidative stress and chronic immune activation. The proposed hemolysin-ROS-Hippo pathway alone is insufficient for cancer promotion; it requires sustained exposure to be oncogenically relevant (Jain et al., 2021; Mager, 2006; Parsonnet, 1995).

Different from acute pathogens, *F. nucleatum* establishes persistent colonization in the oral cavity and can migrate to other sites via the hematogenous route, such as colorectal cancer (Abed et al., 2020). The persistence is facilitated by the adhesins of FadA to E-cadherin on epithelial cells, while Fap2 recognizes tumor-displayed Gal-GalNAc (Alon-Maimon et al., 2022).

Hemolysin activity adds to the cancer-promoting potential of *F. nucleatum*, but it alone is insufficient. The temporal and spatial context of bacterial-host interaction is important. While the proposed pathway of hemolysin induces ROS, acting as genotoxic, is biologically plausible and fits current biology, experimental validation is required to establish causality. The computational framework only provides testable predictions for future mechanistic studies.

Tumor Microenvironment Compatibility

Cancer cells generally create a more acidic microenvironment than normal tissues, which favors their progression (Tafech and Stéphanou, 2024). Hemolysin secretion and pore formation are pH-dependent, with optimal activity near neutral pH (Kasschau et al., 1995).

Based on experimental literature, *F. nucleatum* demonstrates optimal growth at pH 7.4, a relatively neutral pH. The bacterium exhibits significantly reduced metabolic activity, with glucose consumption decreasing to less than 50% and cell yield being halved at pH 6.8. Under moderate acidic conditions (pH 5.8), growth is substantially reduced, while under slightly alkaline conditions (pH 7.8), growth fails (Rogers et al., 1991). While Rogers's team (1991) found a narrow optimal range, other studies indicate a broader survival capacity. *F. nucleatum* subsp. polymorphum demonstrates the ability to form biofilms at alkaline pH 8.2, indicating subspecies-specific alkaline adaptation capabilities(Chew et al., 2012). However, *F. nucleatum* ATCC 25586 (subsp. nucleatum), the strain commonly found in oral cavities, probably exhibits tolerance to alkaline conditions, as *F. nucleatum* is found to survive in root canal systems at pH 9.0 after $Ca(OH)_2$ treatment (Zilm and Rogers, 2007). Furthermore, a key literature finding suggests *F. nucleatum* possesses extreme acid tolerance, due to its cell membrane containing erucic acid, which is regulated by the FnFabM protein, giving it the ability to resist the extreme acidic conditions of the stomach (pH 1.5) when translocated into the intestine (Li et al., 2025).

F. nucleatum ability to modulate the tumor microenvironment(TME) can also promote proliferation, metastasis, and chemoresistance. In CRC tissue, iron deposits within the macrophages were correlated with a poor prognosis for the *F. nucleatum*-positive patients (Dadgar-Zankbar et al., 2024). Considering the proposed mechanism of iron dependency and potential cancer activity, *F. nucleatum*'s activity within the TME may be with respect to iron availability. It is predicted that iron concentration evaluation in the TME may be a signal for *F. nucleatum* to induce a phenotypic switch, enabling the upregulation of its Pathogenicity Island (PAI) genes, allowing *F. nucleatum* to exploit the TME and facilitate cancer progression. The upregulation of these genes involved with *F. nucleatum*'s virulence will, in turn, increase its adhesive and invasive behavior to the host tissues, immune modulation, and ultimately may work with the TME by providing an environment conducive to the growth of tumors (Yamane et al., 2022).

While this study focuses on strain ATCC 25586, future analysis can be done on PAI and the rest of the unannotated genes across *F. nucleatum* subspecies, using phylogenetic analyses and positive selection, which may illuminate whether evolutionary pressures are shaping these virulence-associated functions and potentially characterize unknown proteins by evolutionary relatedness (Charlotte West, 2025).

This study has several limitations. All pathogenicity island predictions, protein functions, and interaction networks are computational and require experimental validation. The AlphaFold-Multimer interface with predictions showed low confidence (ipTM < 0.5) should be interpreted cautiously. Our co-expression analysis was limited by sample size (n=18). This study focused on strain ATCC 25586, and findings may not generalize to other *F. nucleatum* strains. The proposed Hemolysin-Iron-ROS-Hippo pathway remains a hypothesis requiring experimental testing.

Conclusion

This work show how integrating bioinformatics and machine learning approaches enables us to provide further details into the pathogenic mechanisms of bacteria. By combining genomic analysis, metabolic modeling, protein structure prediction, and transcriptomic data, we have constructed a coherent framework pinpointing potential pathogenicity islands of *F. nucleatum* and connect to a proposed mechanism of cancer promotion through iron-mediated oxidative stress.

Our multi-omics approach identified complete hemolysin and iron acquisition machinery within the pathogenicity islands, established iron as metabolically essential through flux balance analysis, and revealed host-induced upregulation of hemolysin expression. Contextualized within the broader literature on ROS-mediated Hippo pathway oncologically relevant and distinction between acute hemolytic pathogens and chronic colonizers, this finding proposes a biologically plausible model for *F. nucleatum*-associated carcinogenesis.

However, one should be conscious of the interpretive boundaries of this work. The result in this study is predictive and contextual rather than causal. In complex biological systems, functional redundancy and compensatory pathways often mean that no single component is strictly determinative of growth or virulence. The current results are state-level predictions rather than mechanistic claims. The study suggests configurations of genomic programs, metabolic constraints, and environmental conditions that are consistent with pathogenic behavior, rather than proving that any individual gene or island is causative on its own.

Furthermore, many of the predicted proteins and genomic regions remain speculative without direct experimental or transcriptomic validation. Rather than being a weakness, it is a clear guide for the next stage of investigation—identifying which predicted states are actually realized under specific host or environmental conditions, and how perturbations shift the system between them.

As the amount of molecular sequence data of pathogens in the public domain grows, so does the range of bioinformatics and machine learning methods that operate on them through functional, structural, and evolutionary considerations. The methodology presented here provides a template for systematic investigation of bacterial virulence factors and their potential roles in human disease. However, advancement of the field is hampered by the discrepancy between bioinformatics and deep learning for many of the analysis methods.

References

Jawad Abed, Naseem Maalouf, Abigail L. Manson, Ashlee M. Earl, Lishay Parhi, Johanna E. M. Emgård, Michael Klutstein, Shay Tayeb, Gideon Almogy, Karine A. Atlan, Stella Chaushu, Eran Israeli, Ofer Mandelboim, Wendy S. Garrett, and Gilad Bachrach. Colon cancer-associated fusobacterium nucleatum may originate from the oral cavity and reach colon tumors via the circulatory system. *Frontiers in Cellular and Infection Microbiology*, 10, August 2020. ISSN 2235-2988. doi: 10.3389/fcimb.2020.00400. URL http://dx.doi.org/10.3389/fcimb.2020.00400.

Inés Albesa, M.Cecilia Becerra, Paola C Battán, and Paulina L Páez. Oxidative stress involved in the antibacterial action of different antibiotics. *Biochemical and Biophysical Research Communications*, 317(2):605–609, April 2004. ISSN 0006-291X. doi: 10.1016/j.bbrc.2004.03.085. URL http://dx.doi.org/10.1016/j.bbrc.2004.03.085.

Tamar Alon-Maimon, Ofer Mandelboim, and Gilad Bachrach. Fusobacterium nucleatum and cancer. *Periodontology 2000*, 89(1):166–180, March 2022. ISSN 1600-0757. doi: 10.1111/prd.12426. URL http://dx.doi.org/10.1111/prd.12426.

Stephen F. Altschul, Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403–410, October 1990. ISSN 0022-2836. doi: 10.1016/s0022-2836(05)80360-2. URL http://dx.doi.org/10.1016/S0022-2836(05)80360-2.

Bella Amanda, Rangga Pragasta, Haris Cakrasana, Arifa Mustika, Zakiyatul Faizah, and Delvac Oceandy. The hippo signaling pathway, reactive oxygen species production, and oxidative stress: A two-way traffic regulation. *Cells*, 13(22):1868, November 2024. ISSN 2073-4409. doi: 10.3390/cells13221868. URL http://dx.doi.org/10.3390/cells13221868.

Nelson E. Arenas, Luz M. Salazar, Carlos Y. Soto, Carolina Vizcaíno, Manuel E. Patarroyo, Manuel A. Patarroyo, and Arley Gómez. Molecular modeling and in silico characterization of mycobacterium tuberculosis tlya: Possible misannotation of this tubercle bacilli-hemolysin. *BMC Structural Biology*, 11:16, 2011. doi: 10.1186/1472-6807-11-16.

Munehiko Asayama and Takashi Ohyama. Curved dna and prokaryotic promoters: A mechanism for activation of transcription. In *Madame Curie Bioscience Database [Internet]*, chapter "Curved DNA and Prokaryotic Promoters". Landes Bioscience, Austin (TX), 2000. URL https://www.ncbi.nlm.nih.gov/books/NBK6157/. NCBI Bookshelf ID: NBK6157.

Franco Bagnoli and Pietro Liò. Selection, mutations and codon usage in a bacterial model. *Journal of Theoretical Biology*, 173(3):271–281, 1995.

José Luis Baronetti, Natalia Angel Villegas, Virginia Aiassa, María Gabriela Paraje, and Inés Albesa. Hemolysin from escherichia coli induces oxidative stress in blood. *Toxicon*, 70:15–20, August 2013. ISSN 0041-0101. doi: 10.1016/j.toxicon.2013.03.014. URL http://dx.doi.org/10.1016/j.toxicon.2013.03.014.

Claire Bertelli, Matthew R Laird, Kelly P Williams, Britney Y Lau, Gemma Hoad, Geoffrey L Winsor, and Fiona SL Brinkman. Islandviewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Research*, 45(W1):W30–W35, May 2017. ISSN 1362-4962. doi: 10.1093/nar/gkx343. URL http://dx.doi.org/10.1093/nar/gkx343.

J. Boone. A topology-based machine learning model decisively outperforms flux balance analysis in predicting metabolic gene essentiality. *ArXiv.org*, 2025. URL https://arxiv.org/abs/2507.20406. Accessed via arXiv.

Tye O. Boynton, Lauren E. Daugherty, Tamara A. Dailey, and Harry A. Dailey. Identification of escherichia coli hemg as a novel, menadione-dependent flavodoxin with protoporphyrinogen oxidase activity. *Biochemistry*, 48(29):6705–6711, July 2009. ISSN 1520-4995. doi: 10.1021/bi900850y. URL http://dx.doi.org/10.1021/bi900850y.

Volkmar Braun, Anna C. Ratliff, Herve Celia, and Susan K. Buchanan. Energization of outer membrane transport by the exbb exbd molecular motor. *Journal of Bacteriology*, 205(6), June 2023. ISSN 1098-5530. doi: 10.1128/jb.00035-23. URL http://dx.doi.org/10.1128/jb.00035-23.

Cláudia Brito, Didier Cabanes, Francisco Sarmento Mesquita, and Sandra Sousa. Mechanisms protecting host cells against bacterial pore-forming toxins. *Cellular and Molecular Life Sciences*, 76(7):1319–1339, December 2018. ISSN 1420-9071. doi: 10.1007/s00018-018-2992-8. URL http://dx.doi.org/10.1007/s00018-018-2992-8.

Stephan Brouwer, Tania Rivera-Hernandez, Bodie F. Curren, Nichaela Harbison-Price, David M. P. De Oliveira, Magnus G. Jespersen, Mark R. Davies, and Mark J. Walker. Pathogenesis, epidemiology and control of group a streptococcus infection. *Nature Reviews Microbiology*, 21(7):431–447, March 2023. ISSN 1740-1534. doi: 10.1038/s41579-023-00865-7. URL http://dx.doi.org/10.1038/s41579-023-00865-7.

Conor Walker Charlotte West. Omegaai. GitHub repository, June 2025. urlhttps://github.com/charlotte-west/OmegaAI, accessed 2025-08-31.

Fangfang Chen, Songhe Guo, Yiqiu Li, Yongfan Lu, Le Liu, Shengxin Chen, Jun An, and Ge Zhang. Fusobacterium nucleatum-driven cx3cr1+ pd-l1+ phagocytes route to tumor tissues and reshape tumor microenvironment. *Gut Microbes*, 17(1), December 2024. ISSN 1949-0984. doi: 10.1080/19490976.2024.2442037. URL http://dx.doi.org/10.1080/19490976.2024.2442037.

Xiyan Chen, Weiting Gu, Qi Wang, Xucheng Fu, Ying Wang, Xin Xu, and Yong Wen. C-myc and bcl-2 mediate yap-regulated tumorigenesis in oscc. *Oncotarget*, 9(1):668–679, December 2017. ISSN 1949-2553. doi: 10.18632/oncotarget.23089. URL http://dx.doi.org/10.18632/oncotarget.23089.

Yuanxin Chen, Zhijie Huang, Zhengming Tang, Yisheng Huang, Mingshu Huang, Hongyu Liu, Dirk Ziebolz, Gerhard Schmalz, Bo Jia, and Jianjiang Zhao. More than just a periodontal pathogen –the research progress on fusobacterium nucleatum. *Frontiers in Cellular and Infection Microbiology*, 12, February 2022. ISSN 2235-2988. doi: 10.3389/fcimb.2022.815318. URL http://dx.doi.org/10.3389/fcimb.2022.815318.

Zhenzhen Chen and Leaf Huang. Fusobacterium nucleatum carcinogenesis and drug delivery interventions. *Advanced Drug Delivery Reviews*, 209:115319, June 2024. ISSN 0169-409X. doi: 10.1016/j.addr.2024.115319. URL http://dx.doi.org/10.1016/j.addr.2024.115319.

Jactty Chew, Peter S Zilm, Janet M Fuss, and Neville J Gully. A proteomic investigation of fusobacterium nucleatum alkaline-induced biofilms. *BMC Microbiology*, 12(1), September 2012. ISSN 1471-2180. doi: 10.1186/1471-2180-12-189. URL http://dx.doi.org/10.1186/1471-2180-12-189.

Kyla Cochrane, Avery V. Robinson, Robert A. Holt, and Emma Allen-Vercoe. A survey of fusobacterium nucleatum genes modulated by host cell infection. *Microbial Genomics*, 6(2), February 2020. ISSN 2057-5858. doi: 10.1099/mgen.0.000300. URL http://dx.doi.org/10.1099/mgen.0.000300.

Peter J. A. Cock, Tiago Antao, Jeffrey T. Chang, Brad A. Chapman, Cymon J. Cox, Andrew Dalke, Iddo Friedberg, Thomas Hamelryck, Frank Kauff, Bartek Wilczynski, and Michiel J. L. de Hoon. Biopython: freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11):1422–1423, March 2009. ISSN 1367-4803. doi: 10.1093/bioinformatics/btp163. URL http://dx.doi.org/10.1093/bioinformatics/btp163.

S. Coppenhagen-Glazer, A. Sol, J. Abed, R. Naor, X. Zhang, Y. W. Han, and G. Bachrach. Fap2 of fusobacterium nucleatum is a galactose-inhibitable adhesin involved in coaggregation, cell adhesion, and preterm birth. *Infection and Immunity*, 83(3):1104–1113, March 2015. ISSN 1098-5522. doi: 10.1128/iai.02838-14. URL http://dx.doi.org/10.1128/IAI.02838-14.

Gavin E. Crooks, Gary Hon, John-Marc Chandonia, and Steven E. Brenner. Weblogo: A sequence logo generator: Figure 1. *Genome Research*, 14(6):1188–1190, June 2004. ISSN 1088-9051. doi: 10.1101/gr.849004. URL http://dx.doi.org/10.1101/gr.849004.

Leila Dadgar-Zankbar, Zahra Elahi, Aref Shariati, Azad Khaledi, Shabnam Razavi, and Amin Khoshbayan. Exploring the role of fusobacterium nucleatum in colorectal cancer: implications for tumor proliferation and chemoresistance. *Cell Communication and Signaling*, 22(1), November 2024. ISSN 1478-811X. doi: 10.1186/s12964-024-01909-y. URL http://dx.doi.org/10.1186/s12964-024-01909-y.

NetGenes Database. Essential genes for fusobacterium nucleatum atcc 25586. https://rbc-dsai-iitm.github.io/NetGenes/organisms/Fusobacterium-nucleatum-ATCC25586.html, 2024. Accessed: 2024-06-06.

Anne de Jong, Hilco Pietersma, Martijn Cordes, Oscar P Kuipers, and Jan Kok. Pepper: a webserver for prediction of prokaryote promoter elements and regulons. *BMC Genomics*, 13(1), July 2012. ISSN 1471-2164. doi: 10.1186/1471-2164-13-299. URL http://dx.doi.org/10.1186/1471-2164-13-299.

Philippe Delepelaire. Bacterial abc transporters of iron containing compounds. *Research in Microbiology*, 170(8):345–357, November 2019. ISSN 0923-2508. doi: 10.1016/j.resmic.2019.10.008. URL http://dx.doi.org/10.1016/j.resmic.2019.10.008.

Martine Deplanche, Nassim Mouhali, Minh-Thu Nguyen, Chantal Cauty, Frédéric Ezan, Alan Diot, Lesly Raulin, Stephanie Dutertre, Sophie Langouet, Patrick Legembre, Frederic Taieb, Michael Otto, Frédéric Laurent, Friedrich Götz, Yves Le Loir, and Nadia Berkova. Staphylococcus aureus induces dna damage in host cell. *Scientific Reports*, 9(1), May 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-44213-3. URL http://dx.doi.org/10.1038/s41598-019-44213-3.

Ali Ebrahim, Joshua A Lerman, Bernhard O Palsson, and Daniel R Hyduke. Cobrapy: Constraints-based reconstruction and analysis for python. *BMC Systems Biology*, 7(1), August 2013. ISSN 1752-0509. doi: 10.1186/1752-0509-7-74. URL http://dx.doi.org/10.1186/1752-0509-7-74.

Adriana Fiorini, Quirino Alves De Lima Neto, and Maria Aparecida Fernandez. Sequence-directed dna curvature in replication origins segments, 2011. URL http://rgdoi.net/10.13140/2.1.3377.7604.

Shraddha Gandhi, Sindhoora Puravankara, Anish Kumar Mondal, Aakanksha Chauhan, Shashi Prakash Yadav, Kausik Chattopadhyay, and Arunika Mukhopadhaya. Vibrio cholerae cytolysin induces pro-inflammatory and death signals through novel tlr assembly. *PLOS Pathogens*, 21(4):e1013033, April 2025. ISSN 1553-7374. doi: 10.1371/journal.ppat.1013033. URL http://dx.doi.org/10.1371/journal.ppat.1013033.

Kumar Ganesan, Songhe Guo, Sundaz Fayyaz, Ge Zhang, and Baojun Xu. Targeting programmed fusobacterium nucleatum fap2 for colorectal cancer therapy. *Cancers*, 11(10):1592, October 2019. ISSN 2072-6694. doi: 10.3390/cancers11101592. URL http://dx.doi.org/10.3390/cancers11101592.

Christoph Gohlke. cgohlke/dnacurve: v2025.5.8, 2025. URL https://zenodo.org/doi/10.5281/zenodo.7135499.

Kam Pou Ha and Andrew M. Edwards. Dna repair in staphylococcus aureus. *Microbiology and Molecular Biology Reviews*, 85(4), December 2021. ISSN 1098-5557. doi: 10.1128/mmbr.00091-21. URL http://dx.doi.org/10.1128/MMBR.00091-21.

Jörg Hacker and James B. Kaper. Pathogenicity islands and the evolution of microbes. *Annual Review of Microbiology*, 54(1):641–679, October 2000. ISSN 1545-3251. doi: 10.1146/annurev.micro.54.1.641. URL http://dx.doi.org/10.1146/annurev.micro.54.1.641.

Jeppe Hallgren, Konstantinos D. Tsirigos, Mads Damgaard Pedersen, José Juan Almagro Armenteros, Paolo Marcatili, Henrik Nielsen, Anders Krogh, and Ole Winther. Deeptmhmm predicts alpha and beta transmembrane proteins using deep neural networks. April 2022. doi: 10.1101/2022.04.08.487609. URL http://dx.doi.org/10.1101/2022.04.08.487609.

Hu Han, Yan Li, Wan Qin, Lu Wang, Han Yin, Beibei Su, and Xianglin Yuan. Fusobacterium nucleatumfacilitates cetuximab resistance in colorectal cancer via the pi3k/akt and jak/stat3 pathways. December 2020. doi: 10.1101/2020.12.21.423755. URL http://dx.doi.org/10.1101/2020.12.21.423755.

Yiping W Han. Fusobacterium nucleatum: a commensal-turned pathogen. *Current Opinion in Microbiology*, 23:141–147, February 2015. ISSN 1369-5274. doi: 10.1016/j.mib.2014.11.013. URL http://dx.doi.org/10.1016/j.mib.2014.11.013.

Kah Yan How, Keang Peng Song, and Kok Gan Chan. Porphyromonas gingivalis: An overview of periodontopathic pathogen below the gum line. *Frontiers in Microbiology*, 7, February 2016. ISSN 1664-302X. doi: 10.3389/fmicb.2016.00053. URL http://dx.doi.org/10.3389/fmicb.2016.00053.

Da Wei Huang, Brad T Sherman, and Richard A Lempicki. Systematic and integrative analysis of large gene lists using david bioinformatics resources. *Nature Protocols*, 4(1):44–57, December 2008. ISSN 1750-2799. doi: 10.1038/nprot.2008.211. URL http://dx.doi.org/10.1038/nprot.2008.211.

Kexin Huang, Serena Zhang, Hanchen Wang, Yuanhao Qu, Yingzhou Lu, Yusuf Roohani, Ryan Li, Lin Qiu, Gavin Li, Junze Zhang, Di Yin, Shruti Marwaha, Jennefer N. Carter, Xin Zhou, Matthew Wheeler, Jonathan A. Bernstein, Mengdi Wang, Peng He, Jingtian Zhou, Michael Snyder, Le Cong, Aviv Regev, and Jure Leskovec. Biomni: A general-purpose biomedical ai agent. June 2025. doi: 10.1101/2025.05.30.656746. URL http://dx.doi.org/10.1101/2025.05.30.656746.

Jaime Huerta-Cepas, Damian Szklarczyk, Davide Heller, Ana Hernández-Plaza, Sofia K Forslund, Helen Cook, Daniel R Mende, Ivica Letunic, Thomas Rattei, Lars J Jensen, Christian von Mering, and Peer Bork. eggnog 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, 47(D1):D309–D314, November 2018. ISSN 1362-4962. doi: 10.1093/nar/gky1085. URL http://dx.doi.org/10.1093/nar/gky1085.

Utkarsh Jain, Kirti Saxena, and Nidhi Chauhan. Helicobacter pylori induced reactive oxygen species: A new and developing platform for detection. *Helicobacter*, 26(3), March 2021. ISSN 1523-5378. doi: 10.1111/hel.12796. URL http://dx.doi.org/10.1111/hel.12796.

Hossein Jamali, Tylor Pereira, and Charles M. Dozois. Environmental and regulatory control of rtx toxins in gram-negative pathogens. *Toxins*, 18(1):27, January 2026. ISSN 2072-6651. doi: 10.3390/toxins18010027. URL http://dx.doi.org/10.3390/toxins18010027.

Jelena Jovanovic, Vincent F. M. Ornelis, Annemieke Madder, and Andreja Rajkovic. Bacillus cereus food intoxication and toxicoinfection. *Comprehensive Reviews in Food Science and Food Safety*, 20 (4):3719–3761, June 2021. ISSN 1541-4337. doi: 10.1111/1541-4337.12785. URL http://dx.doi.org/10.1111/1541-4337.12785.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Zidek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021. doi: 10.1038/s41586-021-03819-2.

Margaret R. Kasschau, Mary P. Byam-Smith, Deborah S. Gentry, and Fay N. Watson. Influence of ph and temperature on hemolysis by adult schistosoma mansoni membranes. *Journal of Experimental Zoology*, 271(4):315–322, March 1995. ISSN 1097-010X. doi: 10.1002/jez.1402710409. URL http://dx.doi.org/10.1002/jez.1402710409.

Brandon A. Keith, John C.S. Harding, and Matthew E. Loewen. Mutational analysis of tlya from brachyspira hampsonii reveals two key residues conserved in pathogenic bacteria responsible for oligomerization and hemolytic activity. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1866(2):130045, February 2022. ISSN 0304-4165. doi: 10.1016/j.bbagen.2021.130045. URL http://dx.doi.org/10.1016/j.bbagen.2021.130045.

Sho Kitamoto, Hiroko Nagao-Kitamoto, Peter Kuffa, and Nobuhiko Kamada. Regulation of virulence: The rise and fall of gastrointestinal pathogens - journal of gastroenterology, Nov 2015. URL https://link.springer.com/article/10.1007/s00535-015-1141-5.

L. Kozobay-Avraham. Involvement of dna curvature in intergenic regions of prokaryotes. *Nucleic Acids Research*, 34(8):2316–2327, April 2006. ISSN 1362-4962. doi: 10.1093/nar/gkl230. URL http://dx.doi.org/10.1093/nar/gkl230.

Anders Krogh, Björn Larsson, Gunnar von Heijne, and Erik LL Sonnhammer. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *Journal of Molecular Biology*, 305(3):567–580, 2001. doi: 10.1006/jmbi.2000.4315.

Emma L. Langshaw, Simone Reynolds, Victoria Ozberk, Jessica Dooley, Ainslie Calcutt, Mehfuz Zaman, Mark J. Walker, Michael R. Batzloff, Mark R. Davies, Michael F. Good, and Manisha Pandey. Streptolysin o deficiency in streptococcus pyogenes m1t1 covr/s mutant strain attenuates virulence in in vitro and in vivo infection models. *mBio*, 14(1), February 2023. ISSN 2150-7511. doi: 10.1128/mbio.03488-22. URL http://dx.doi.org/10.1128/mbio.03488-22.

B. D. Lee. Python implementation of codon adaptation index. *Journal of Open Source Software*, 3(30): 905, 2019. doi: 10.21105/joss.00905. URL https://cai.readthedocs.io/en/latest/api.html.

Xiaocong Li, Shipeng Zhang, Huafang Sheng, Yan Zhen, Buling Wu, Zhuang Li, Dingqiang Chen, and Hongwei Zhou. Oral fusobacterium nucleatum resists the acidic ph of the stomach due to membrane erucic acid synthesized via enoyl-coa hydratase-related protein fnfabm. *Journal of Oral Microbiology*, 17(1), January 2025. ISSN 2000-2297. doi: 10.1080/20002297.2025.2453964. URL http://dx.doi.org/10.1080/20002297.2025.2453964.

Serene Lim, Wan-Hsin Hsueh, Ni-Hung Wu, Claire Hodges, Christopher Vuong, Kynzi Smith, Jenn-Ren Hsiao, Jeffrey S Chang, Jang-Yang Chang, Jenn-Wei Chen, and I-Hsiu Huang. Characterization of clinical fusobacterium nucleatum isolates from oral squamous cell carcinoma patients. *ArXiv.org*, January 2025. doi: 10.1101/2025.01.08.631950. URL http://dx.doi.org/10.1101/2025.01.08.631950.

Ferdinand C. O. Los, Tara M. Randis, Raffi V. Aroian, and Adam J. Ratner. Role of pore-forming toxins in bacterial infectious diseases. *Microbiology and Molecular Biology Reviews*, 77(2):173–207, June 2013. ISSN 1098-5557. doi: 10.1128/mmbr.00052-12. URL http://dx.doi.org/10.1128/MMBR.00052-12.

H. Louvel, S. Bommezzadri, N. Zidane, C. Boursaux-Eude, S. Creno, A. Magnier, Z. Rouy, C. Médigue, I. Saint Girons, C. Bouchier, and M. Picardeau. Comparative and functional genomic analyses of iron transport and regulation in leptospira spp. *Journal of Bacteriology*, 188(22):7893–7904, November 2006. ISSN 1098-5530. doi: 10.1128/jb.00711-06. URL http://dx.doi.org/10.1128/JB.00711-06.

Xiaomei Ma, Tianyong Sun, Jiannan Zhou, Mengfan Zhi, Song Shen, Yushang Wang, Xiufeng Gu, Zixuan Li, Haiting Gao, Pingping Wang, and Qiang Feng. Pangenomic study of fusobacterium nucleatum reveals the distribution of pathogenic genes and functional clusters at the subspecies and strain levels. *Microbiology Spectrum*, 11(3), June 2023. ISSN 2165-0497. doi: 10.1128/spectrum.05184-22. URL http://dx.doi.org/10.1128/spectrum.05184-22.

Daniel Machado, Sergej Andrejev, Melanie Tramontano, and Kiran Raosaheb Patil. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Research*, 46(15):7542–7553, June 2018. ISSN 1362-4962. doi: 10.1093/nar/gky537. URL http://dx.doi.org/10.1093/nar/gky537.

DL Mager. Bacteria and cancer: cause, coincidence or cure? a review. *Journal of Translational Medicine*, 4(1), March 2006. ISSN 1479-5876. doi: 10.1186/1479-5876-4-14. URL http://dx.doi.org/10.1186/1479-5876-4-14.

Alexandra K. McGregor and Kirsten R. Wolthers. Fusobacterium nucleatum: strategies for adapting to aerobic stress. *Journal of Bacteriology*, 207(7), July 2025. ISSN 1098-5530. doi: 10.1128/jb.00090-25. URL http://dx.doi.org/10.1128/jb.00090-25.

Alexandra K. McGregor, Anson C.K. Chan, Megan D. Schroeder, Long T.M. Do, Gurpreet Saini, Michael E.P. Murphy, and Kirsten R. Wolthers. A new member of the flavodoxin superfamily from fusobacterium nucleatum that functions in heme trafficking and reduction of anaerobilin. *Journal of Biological Chemistry*, 299(7):104902, July 2023. ISSN 0021-9258. doi: 10.1016/j.jbc.2023.104902. URL http://dx.doi.org/10.1016/j.jbc.2023.104902.

Barbara E Menzies and Iordanka Kourteva. Staphylococcus aureusÎ±-toxin induces apoptosis in endothelial cells. *FEMS Immunology amp; Medical Microbiology*, 29(1):39–45, September 2000. ISSN 1574-695X. doi: 10.1111/j.1574-695x.2000.tb01503.x. URL http://dx.doi.org/10.1111/j.1574-695X.2000.tb01503.x.

Sarah Millot, Constance Delaby, Boualem Moulouel, Thibaud Lefebvre, Nathalie Pilard, Nicolas Ducrot, Cécile Ged, Philippe Lettéron, Lucia de Franceschi, Jean Charles Deybach, Carole Beaumont, Laurent Gouya, Hubert De Verneuil, Saïd Lyoumi, Hervé Puy, and Zoubida Karim. Hemolytic anemia repressed hepcidin level without hepatocyte iron overload: lesson from günther disease model. *Haematologica*, 102(2):260–270, November 2016. ISSN 1592-8721. doi: 10.3324/haematol.2016.151621. URL http://dx.doi.org/10.3324/haematol.2016.151621.

Marwan Mansoor Ali Mohammed, Veronika Kuchařová Pettersen, Audun H. Nerland, Harald G. Wiker, and Vidar Bakken. Protein interactions between the oral bacteriafusobacterium nucleatumandporphyromonas gingivalisin biofilm and planktonic culture. November 2020. doi: 10.1101/2020.11.23.393850. URL http://dx.doi.org/10.1101/2020.11.23.393850.

Toshiro Moroishi, Hyun Woo Park, Baodong Qin, Qian Chen, Zhipeng Meng, Steven W. Plouffe, Koji Taniguchi, Fa-Xing Yu, Michael Karin, Duojia Pan, and Kun-Liang Guan. A yap/taz-induced feedback mechanism regulates hippo pathway homeostasis. *Genes amp; Development*, 29(12):1271–1284, June 2015. ISSN 1549-5477. doi: 10.1101/gad.262816.115. URL http://dx.doi.org/10.1101/gad.262816.115.

Chan D.K. Nguyen and Chunling Yi. Yap/taz signaling and resistance to cancer therapy. *Trends in Cancer*, 5(5):283–296, May 2019. ISSN 2405-8033. doi: 10.1016/j.trecan.2019.02.010. URL http://dx.doi.org/10.1016/j.trecan.2019.02.010.

Nguyen Thi Khanh Nhu, Minh-Duy Phan, Brian M. Forde, Ambika M. V. Murthy, Kate M. Peters, Christopher J. Day, Jessica Poole, Timothy J. Kidd, Rodney A. Welch, Michael P. Jennings, Glen C. Ulett, Matthew J. Sweet, Scott A. Beatson, and Mark A. Schembri. Complex multilevel control of hemolysin production by uropathogenic escherichia coli. *mBio*, 10(5), October 2019. ISSN 2150-7511. doi: 10.1128/mbio.02248-19. URL http://dx.doi.org/10.1128/mbio.02248-19.

R. O. Nicholas, Tong Li, D. McDevitt, Andrea Marra, S. Sucoloski, P. L. Demarsh, and D. R. Gentry. Isolation and characterization of a sigb deletion mutant of staphylococcus aureus. *Infection and Immunity*, 67(7):3667–3669, July 1999. ISSN 1098-5522. doi: 10.1128/iai.67.7.3667-3669.1999. URL http://dx.doi.org/10.1128/iai.67.7.3667-3669.1999.

S. M. Opal, A. S. Cross, P. Gemski, and L. W. Lyhte. Aerobactin and -hemolysin as virulence determinants in escherichia coli isolated from human blood, urine, and stool. *Journal of Infectious Diseases*, 161(4):794–796, April 1990. ISSN 1537-6613. doi: 10.1093/infdis/161.4.794. URL http://dx.doi.org/10.1093/infdis/161.4.794.

J Parsonnet. Bacterial infection as a cause of cancer. *Environmental Health Perspectives*, 103(suppl 8):263–268, November 1995. ISSN 1552-9924. doi: 10.1289/ehp.95103s8263. URL http://dx.doi.org/10.1289/ehp.95103s8263.

Akash Patel, Monica Patel, and Joseph Glowacki. Fusobacterium nucleatum: More than just an oral anaerobe. *Cureus*, July 2022. ISSN 2168-8184. doi: 10.7759/cureus.26989. URL http://dx.doi.org/10.7759/cureus.26989.

Felipe A. Pinho-Ribeiro, Buket Baddal, Rianne Haarsma, Maghnus O'Seaghdha, Nicole J. Yang, Kimbria J. Blake, Makayla Portley, Waldiceu A. Verri, James B. Dale, Michael R. Wessels, and Isaac M. Chiu. Blocking neuronal signaling to immune cells treats streptococcal invasive infection. *Cell*, 173(5):1083–1097.e22, May 2018. ISSN 0092-8674. doi: 10.1016/j.cell.2018.04.006. URL http://dx.doi.org/10.1016/j.cell.2018.04.006.

Falk Ponath, Caroline Tawk, Yan Zhu, Lars Barquist, Franziska Faber, and Jörg Vogel. Rna landscape of the emerging cancer-associated microbe fusobacterium nucleatum. *Nature Microbiology*, 6(8): 1007–1020, July 2021. ISSN 2058-5276. doi: 10.1038/s41564-021-00927-7. URL http://dx.doi.org/10.1038/s41564-021-00927-7.

Jessica Queen, Zam Cing, Hana Minsky, Asmita Nandi, Taylor Southward, Jacqueline Ferri, Madison McMann, Thevambiga Iyadorai, Jamuna Vadivelu, April Roslani, Mun Fai Loke, Jane Wanyiri, James R. White, Julia L. Drewes, and Cynthia L. Sears. Fusobacterium nucleatum is enriched in invasive biofilms in colorectal cancer. *npj Biofilms and Microbiomes*, 11(1), May 2025. ISSN 2055-5008. doi: 10.1038/s41522-025-00717-7. URL http://dx.doi.org/10.1038/s41522-025-00717-7.

Md. Aejazur Rahman, Parveen Sobia, Ved Prakash Dwivedi, Aakansha Bhawsar, Dhiraj Kumar Singh, Pawan Sharma, Prashini Moodley, Luc Van Kaer, William R Bishai, and Gobardhan Das. Mycobacterium tuberculosis tlya protein negatively regulates t helper (th) 1 and th17 differentiation and promotes tuberculosis pathogenesis. *Journal of Biological Chemistry*, 290(23):14407–14417, June 2015. ISSN 0021-9258. doi: 10.1074/jbc.m115.653600. URL http://dx.doi.org/10.1074/jbc.M115.653600.

Kaylie L. Richard, Brittni R. Kelley, and Jeremiah G. Johnson. Heme uptake and utilization by gram-negative bacterial pathogens. *Frontiers in Cellular and Infection Microbiology*, 9, March 2019. ISSN 2235-2988. doi: 10.3389/fcimb.2019.00081. URL http://dx.doi.org/10.3389/fcimb.2019.00081.

Johanna Richter, Mercedes M Monteleone, Amanda J Cork, Timothy C Barnett, Victor Nizet, Stephan Brouwer, Kate Schroder, and Mark J Walker. Streptolysins are the primary inflammasome activators in macrophages during streptococcus pyogenes infection. *Immunology amp; Cell Biology*, 99(10): 1040–1052, September 2021. ISSN 1440-1711. doi: 10.1111/imcb.12499. URL http://dx.doi.org/10.1111/imcb.12499.

A. H. Rogers, P. S. Zilm, N. J. Gully, A. L. Pfennig, and P. D. Marsh. Aspects of the growth and metabolism of fusobacterium nucleatum atcc 10953 in continuous culture. *Oral Microbiology and Immunology*, 6(4):250–255, August 1991. ISSN 1399-302X. doi: 10.1111/j.1399-302x.1991.tb00486.x. URL http://dx.doi.org/10.1111/j.1399-302x.1991.tb00486.x.

Mara Roxana Rubinstein, Xiaowei Wang, Wendy Liu, Yujun Hao, Guifang Cai, and Yiping W. Han. Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating e-cadherin/β-catenin signaling via its fada adhesin. *Cell Host amp; Microbe*, 14(2):195–206, August 2013. ISSN 1931-3128. doi: 10.1016/j.chom.2013.07.012. URL http://dx.doi.org/10.1016/j.chom.2013.07.012.

Nitchakan Samainukul, Aung Khine Linn, Mohammad Bagher Javadi, Somsri Sakdee, Chanan Angsuthanasombat, and Gerd Katzenmeier. Importance of the cys124cys128 intermolecular disulfide bonding for oligomeric assembly and hemolytic activity of the helicobacter pylori tlya hemolysin. *Biochemical and Biophysical Research Communications*, 514(2):365–371, June 2019. ISSN 0006-291X. doi: 10.1016/j.bbrc.2019.04.096. URL http://dx.doi.org/10.1016/j.bbrc.2019.04.096.

Thomas D. Schneider and R.Michael Stephens. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Research*, 18(20):6097–6100, 1990. ISSN 1362-4962. doi: 10.1093/nar/18.20.6097. URL http://dx.doi.org/10.1093/nar/18.20.6097.

Marco Scutari. Learning bayesian networks with thebnlearnrpackage. *Journal of Statistical Software*, 35(3), 2010. ISSN 1548-7660. doi: 10.18637/jss.v035.i03. URL http://dx.doi.org/10.18637/jss.v035.i03.

Torsten Seemann. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30(14):2068–2069, March 2014. ISSN 1367-4803. doi: 10.1093/bioinformatics/btu153. URL http://dx.doi.org/10.1093/bioinformatics/btu153.

Blake A. Shannon, Jacklyn R. Hurst, Ronald S. Flannagan, Heather C. Craig, Aanchal Rishi, Katherine J. Kasper, Stephen W. Tuffs, David E. Heinrichs, and John K. McCormick. Streptolysin s is required for streptococcus pyogenes nasopharyngeal and skin infection in hla-transgenic mice. *PLOS Pathogens*, 20(3):e1012072, March 2024. ISSN 1553-7374. doi: 10.1371/journal.ppat.1012072. URL http://dx.doi.org/10.1371/journal.ppat.1012072.

Mukul Sharma and Pushpendra Singh. Role of tlya in the biology of uncultivable mycobacteria. *Combinatorial Chemistry amp; High Throughput Screening*, 25(10):1587–1594, August 2022. ISSN 1386-2073. doi: 10.2174/1386207325666220111150923. URL http://dx.doi.org/10.2174/1386207325666220111150923.

Brad T Sherman, Ming Hao, Ju Qiu, Xiaoli Jiao, Michael W Baseler, H Clifford Lane, Tomozumi Imamichi, and Weizhong Chang. David: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Research*, 50(W1):W216–W221, March 2022. ISSN 1362-4962. doi: 10.1093/nar/gkac194. URL http://dx.doi.org/10.1093/nar/gkac194.

Augustinas Silale and Bert van den Berg. Tonb-dependent transport across the bacterial outer membrane. *Annual Review of Microbiology*, 77(1):67–88, September 2023. ISSN 1545-3251. doi: 10.1146/annurev-micro-032421-111116. URL http://dx.doi.org/10.1146/annurev-micro-032421-111116.

Mahendra Singh, Sindhoora Puravankara, Arunika Mukhopadhaya, and Kausik Chattopadhyay. Enigmatic roles of vibrio cholerae hemolysin/cytolysin in the bacterial pathogenesis and host-pathogen interactions. *Journal of Bacteriology*, 207(8), August 2025. ISSN 1098-5530. doi: 10.1128/jb.00248-25. URL http://dx.doi.org/10.1128/jb.00248-25.

Damian Szklarczyk, Katerina Nastou, Mikaela Koutrouli, Rebecca Kirsch, Farrokh Mehryary, Radja Hachilif, Dewei Hu, Matteo E Peluso, Qingyao Huang, Tao Fang, Nadezhda T Doncheva, Sampo Pyysalo, Peer Bork, Lars J Jensen, and Christian von Mering. The string database in 2025: protein networks with directionality of regulation. *Nucleic Acids Research*, 53(D1):D730–D737, November 2024. ISSN 1362-4962. doi: 10.1093/nar/gkae1113. URL http://dx.doi.org/10.1093/nar/gkae1113.

Alaa Tafech and Angélique Stéphanou. On the importance of acidity in cancer cells and therapy. *Biology*, 13(4):225, March 2024. ISSN 2079-7737. doi: 10.3390/biology13040225. URL http://dx.doi.org/10.3390/biology13040225.

Julian Trouillon, Ina Attrée, and Sylvie Elsen. The regulation of bacterial two-partner secretion systems. *Molecular Microbiology*, 120(2):159–177, June 2023. ISSN 1365-2958. doi: 10.1111/mmi.15112. URL http://dx.doi.org/10.1111/mmi.15112.

Eva Tvrdá, Filip Benko, and Michal Ďuračka. Oxidative stress as an underlying mechanism of bacteria-inflicted damage to male gametes. *Oxygen*, 2(4):547–569, November 2022. ISSN 2673-9801. doi: 10.3390/oxygen2040036. URL http://dx.doi.org/10.3390/oxygen2040036.

Sergi Vallvé. Gc_skew-plot: Plot a gc skew to look for the origin of replication of a bacterial genome. https://github.com/sgvallve/GC_skew-plot, 2023. Accessed: 2025-08-09.

Vivek Verma, Parveen Kumar, Surbhi Gupta, Sonal Yadav, Rakesh Singh Dhanda, Henrik Thorlacius, and Manisha Yadav. α-hemolysin of uropathogenic e. coli regulates nlrp3 inflammasome activation and mitochondrial dysfunction in thp-1 macrophages. *Scientific Reports*, 10(1), July 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-69501-1. URL http://dx.doi.org/10.1038/s41598-020-69501-1.

Ni Wang, Lu Zhang, Xiao-Xu Leng, Yi-Le Xie, Zi-Ran Kang, Li-Cong Zhao, Lin-Hong Song, Cheng-Bei Zhou, and Jing-Yuan Fang. Fusobacterium nucleatum induces chemoresistance in colorectal cancer by inhibiting pyroptosis via the hippo pathway. *Gut Microbes*, 16(1), March 2024. ISSN 1949-0984. doi: 10.1080/19490976.2024.2333790. URL http://dx.doi.org/10.1080/19490976.2024.2333790.

Gulam Waris and Haseeb Ahsan. Reactive oxygen species: role in the development of cancer and various chronic conditions. *Journal of Carcinogenesis*, 5(1):14, 2006. ISSN 1477-3163. doi: 10.1186/1477-3163-5-14. URL https://pubmed.ncbi.nlm.nih.gov/16689993/.

Hanna F Willenbockel, Birte Dowerg, and Thekla Cordes. Multifaceted metabolic role of infections in the tumor microenvironment. *Current Opinion in Biotechnology*, 89:103183, October 2024. ISSN 0958-1669. doi: 10.1016/j.copbio.2024.103183. URL http://dx.doi.org/10.1016/j.copbio.2024.103183.

Yixian Wu, Songhe Guo, Fangfang Chen, Yiqiu Li, Yuying Huang, Wanli Liu, and Ge Zhang. Fn-dps, a novel virulence factor of fusobacterium nucleatum, disrupts erythrocytes and promotes metastasis in colorectal cancer. *PLOS Pathogens*, 19(1):e1011096, January 2023. ISSN 1553-7374. doi: 10.1371/journal.ppat.1011096. URL http://dx.doi.org/10.1371/journal.ppat.1011096.

Wei Xuan, Dandan Song, Jianghua Hou, and Xiuping Meng. Regulation of hippo-yap1/taz pathway in metabolic dysfunction-associated steatotic liver disease. *Frontiers in Pharmacology*, 16, January 2025. ISSN 1663-9812. doi: 10.3389/fphar.2025.1505117. URL http://dx.doi.org/10.3389/fphar.2025.1505117.

Taishi Yamane, Yohei Kanamori, Hiroshi Sawayama, Hiromu Yano, Akihiro Nita, Yudai Ohta, Hironori Hinokuma, Ayato Maeda, Akiko Iwai, Takashi Matsumoto, Mayuko Shimoda, Mayumi Niimura, Shingo Usuki, Noriko Yasuda-Yoshihara, Masato Niwa, Yoshifumi Baba, Takatsugu Ishimoto, Yoshihiro Komohara, Tomohiro Sawa, Tasuku Hirayama, Hideo Baba, and Toshiro Moroishi. Iron accelerates fusobacterium nucleatum–induced ccl8 expression in macrophages and is associated with colorectal cancer progression. *JCI Insight*, 7(21), November 2022. ISSN 2379-3708. doi: 10.1172/jci.insight.156802. URL http://dx.doi.org/10.1172/jci.insight.156802.

Martha A. Zepeda-Rivera, Floyd E. Dewhirst, Susan Bullman, and Christopher D. Johnston. Addressing controversy in fusobacterium nomenclature: what exactly does " f. nucleatum " refer to? *Gut Microbes*, 17(1), June 2025. ISSN 1949-0984. doi: 10.1080/19490976.2025.2514797. URL http://dx.doi.org/10.1080/19490976.2025.2514797.

Tengda Zhao, Xueping Wang, Liwu Fu, and Ke Yang. Fusobacterium nucleatum: a new player in regulation of cancer development and therapeutic response. *Cancer Drug Resistance*, page 424–438, 2022. ISSN 2578-532X. doi: 10.20517/cdr.2021.144. URL http://dx.doi.org/10.20517/cdr.2021.144.

Xin Zhao, Ying Zhang, and Xinxiang Huang. Pathogenicity-island-encoded regulatory rnas regulate bacterial virulence and pathogenesis. *Microbial Pathogenesis*, 125:196–204, December 2018. ISSN 0882-4010. doi: 10.1016/j.micpath.2018.09.028. URL http://dx.doi.org/10.1016/j.micpath.2018.09.028.

Zhongwei Zhao. Iron and oxidizing species in oxidative stress and alzheimer's disease. *AGING MEDICINE*, 2(2):82–87, June 2019. ISSN 2475-0360. doi: 10.1002/agm2.12074. URL http://dx.doi.org/10.1002/agm2.12074.

Peter S. Zilm and Anthony H. Rogers. Co-adhesion and biofilm formation by fusobacterium nucleatum in response to growth ph. *Anaerobe*, 13(3–4):146–152, June 2007. ISSN 1075-9964. doi: 10.1016/j.anaerobe.2007.04.005. URL http://dx.doi.org/10.1016/j.anaerobe.2007.04.005.