

**Refining the genetic signature of bacterial genotoxin
colibactin and its presence in premalignant lesions of
colorectal cancer**

Dissertation

in fulfilment of the requirements for the degree of Dr. rer. nat.

of the Faculty of Mathematics and Natural Sciences

at Kiel University

submitted by

Shihan Wang

Kiel, 2025

First examiner: (Prof. Dr. Thomas Bosch)

Second examiner: (Prof. Dr. Marc Bramkamp)

Date of the oral examination: 24.11.2025

Declaration

- 1) Apart from the supervisor's guidance, the content and design of this thesis are the doctoral researcher's own work; all collaborations are explicitly indicated in the respective figure legends.
- 2) This thesis has not been submitted either partially or wholly for a doctoral degree to any other examining body. However, parts of this contents are intended to be submitted for publication in ordinary scientific journals.
- 3) The thesis has been prepared in compliance with the Rules of (a) Good Scientific Practice as provided by the German Research Foundation and (b) the rules of the Faculty of Mathematics and Natural Sciences, as outlined in their Information Sheet, no. 4d.
- 4) An academic degree has never been withdrawn for my person.

Kiel, 16.09.2025

A handwritten signature in black ink that reads "Shihan Wang". The script is cursive and fluid, with the first name "Shihan" and the last name "Wang" clearly distinguishable.

Signed by:

Table of Contents

Declaration	3
Abstract	9
1. Introduction.....	13
1.1 Cancer: A Global Health Challenge.....	13
1.2. Genetic and Environmental Contributions to Cancer.....	13
1.2.1 Hereditary factors.....	14
1.2.2 Environmental factors	15
1.2.3 Intrinsic replication errors	15
1.3 Transforming Genes and Core Cancer Pathways	16
1.3.1 Oncogenes.....	16
1.3.2 Tumor suppressor genes	16
1.3.3 DNA repair machinery.....	17
1.4 Microbes and Cancer: From <i>Helicobacter pylori</i> to the Gut Microbiome	18
1.4.1 The importance of microbes in GI tract	18
1.4.2 Bacterial carcinogenesis paradigm	18
1.4.3 Microbiota and CRC.....	19
1.5 <i>Escherichia coli</i> and Colibactin: A Bacterial Genotoxin in Colorectal Cancer...20	
1.5.1 Biosynthesis of colibactin	20
1.5.2 Genotoxic Impact of <i>pks+</i> <i>E. coli</i>	22
1.6 Mutational Signatures in Human Cancer	22
1.6.1 Concept of mutational signatures.....	22
1.6.2 SBS88- colibactin signature	23
1.7 The Human Intestine: Anatomy, Epithelial Architecture, and Stem Cells	24
1.7.1 Key signaling pathways.....	25
1.7.3 Barrier function.....	25

1.8 Colibactin in the Context of Inflammatory Bowel Disease, Colon Polyps and CRC	27
1.9 Recent challenges, research aims and objectives.....	29
1.9.1 Challenges in Direct Detection and Study of Colibactin	29
1.9.2 Aim of the Work.....	30
2. Materials and Methods	31
2.1. Materials.....	31
2.1.1. Human Tissue	31
2.1.2. Cell lines	32
2.1.3. Bacterial Strains	32
2.1.4. Chemicals and Reagents	33
2.1.5. Buffer, solutions and medium.....	35
2.1.6. Kits	39
2.1.7. S-oligos (Phosphorothioate Oligonucleotides)	39
2.1.8. Plasmid	41
2.1.9. Antibodies	41
2.1.10. Consumables	42
2.1.11. Lab instrumentation.....	43
2.1.12. Software.....	44
2.2. Methods	45
2.2.1 Bacteria cultures and Growth Conditions	45
2.2.2 Plasmids generation and extraction	45
2.2.3 Co-incubation of bacteria and plasmid DNA	46
2.2.4 Cofactors influence in colibactin production and activity	46
2.2.5 Plasmid purification and digestion.....	47
2.2.6 S-oligos generation and visualization by native polyacrylamide gel electrophoresis.....	47

2.2.7 S-oligos co-incubation with <i>pks</i> + <i>E. coli</i>	48
2.2.8 Quantification of colibactin-associated genotoxic effect by alkaline agarose gel electrophoresis	48
2.2.9 Nanopore sequencing	49
2.2.10 Nanopore sequencing to quantify colibactin affinity to different DNA hexanucleotides	49
2.2.11 Re-estimation of the true oncogenic potential of colibactin.	49
2.2.12 Statistical Analysis	50
2.2.13 WRN conditioned medium	50
2.2.14 R-spondin1 conditioned medium.....	51
2.2.15 Human colon crypts isolations from biopsies	51
2.2.16 Establishment of 2D culture	52
2.2.17 Passage of colon organoids culture	53
2.2.18 Freezing and thawing of organoids	54
2.2.19 Clonal selection of organoids	54
2.2.20 DNA isolation	55
2.2.21 DNA sequencing	55
2.2.22. Western Blot.....	55
2.2.23. RNA and DNA techniques.....	57
2.2.24. Immunofluorescence	60
3. Results.....	61
Part I: The analyses of sequence specificity of colibactin on isolated DNA.....	61
3.1 Generation of a plasmid for testing colibactin specificity	61
3.2 Sequence specificity of Colibactin in plasmid DNA	63
3.3 Kinetics of DNA inter-strand cross-linking in plasmid DNA.....	65
3.4 Specificity of colibactin from wild type vs. recombinant <i>E. coli</i>	67
3.5 DNA super-helicity influence colibactin specificity	68

3.6 Relaxation of DNA favors colibactin binding	70
3.7 Validation of colibactin targeting specificity on phosphorothioated modified oligonucleotides.....	72
3.8 Applying nanopore sequencing for detecting colibactin crosslinks	74
3.9 Detecting colibactin targeting sites in large DNA molecules.....	77
3.10 <i>pks+</i> <i>E. coli</i> H5194 respond to the presence of cofactors shaping colibactin production	78
3.11 Influence of colibactin activity by various cofactors.	81
3.12 Influence of Copper on colibactin production and activity	83
3.13 Oxygen regulates colibactin production on <i>E. coli</i> strains	84
3.14 Genomic similarity of <i>E. coli</i> DH10B, <i>E. coli</i> Nissle 1917 and <i>E. coli</i> M 1/5	87
Part II: The presence of colibactin damage motifs in colon cancer precursors	91
3.15 Establishment of 3D Human Colon Organoids	91
3.16 Improvement of 3D organoids culture	93
3.17 Establishment of organoid-derived monolayers.....	96
3.18 Validation of organoid-derived monolayers regarding barrier dysfunction.....	97
3.19 Whole genome sequencing of human colon organoids.....	100
3.20 Re-estimation of the true oncogenic potential of colibactin	102
4. Discussion	106
4.1 Defining the sequence specificity of colibactin–DNA interactions	106
4.1.1 Comparison of colibactin affinity: inter-strand crosslinks to different colibactin damage motif on plasmid-based assays	107
4.1.2 Phosphorothioate Oligonucleotides: A High-Fidelity Platform for CDM Validation	108
4.1.3 Nanopore Sequencing as a Reporter of Colibactin Binding	108
4.1.4 Converging evidence for AAAGTT as a novel motif	109
4.1.5 Causal link to tumor suppressor gene mutations	110
4.2 Environmental and host cofactors modulating colibactin activity	111

4.2.1 Cofactors regulating pks gene expression	111
4.2.2 Cofactors regulating colibactin activity	112
4.2.3 Dietary and microbial ecosystem factors.....	113
4.3 Analysis of colibactin-induced DNA damage motifs across distinct colorectal cancer precursors.....	114
4.3.1 Organoid platform as a model for early colorectal carcinogenesis	114
4.3.2 Inflammation-driven barrier dysfunction increases susceptibility to colibactin	115
4.3.3 Detection of SBS88 in organoids from inflamed tissues.....	116
4.3.4 Regenerative responses and potential therapeutic modulation.....	117
4.3.5 Mechanistic model: inflammation–dysbiosis–mutation axis	117
5. Conclusion and outlook.....	120
Publications and Manuscripts Arising from This Thesis	122
Published	122
Under review.....	122
In Preparation	122
6. List of References:	123
7. Acknowledgements	129

Abstract

Colibactin is a genotoxin encoded by *pks*-island-determined *Escherichia coli* strains of the human gut microbiome that causes colorectal cancer (CRC) through inter-strand crosslinks (ICLs) and subsequent double-strand breaks (DSBs). In this work, I investigated the sequence-specific properties of colibactin-induced DNA damage. Specifically, modified plasmids with DNA sequences from 15 putative colibactin target structures (CDMs) and phosphorothioate-protected oligonucleotides were used as substrates. This experiment revealed a preference for cross-linking in AT-rich sequences, particularly the sequence motifs AAATTT, AATTTT, and the newly identified, high-affinity motif AAAGTT. Linear DNA was significantly more susceptible than superhelical DNA, suggesting that superhelicity influences the cross-linking efficiency of colibactin. These results, obtained with naked DNA as the target, were largely confirmed by nanopore sequence analysis. According to this analysis, the spectrum of colibactin target structures for the formation of DNA strand cross-links essentially comprises the previously confirmed CDMs. However, the sequence AAAGTT was additionally identified as a further motif.

To investigate the occurrence of the colibactin signature in the course of inflammatory diseases of the human colon, three-dimensional colon organoids were grown from biopsy samples from patients with inflammatory bowel disease (IBD) and existing adenomatous polyps. This was done using growth media and protocols developed specifically for this purpose. The epithelial cells of the cultured colon organoids underwent whole genome sequencing (WGS). As part of a collaboration, bioinformatic detection of mutation signatures in the genome of these organoid cells was performed. This revealed an enrichment of the colibactin-associated signature SBS88 in the majority, but not all, samples. In particular, the SBS88 signature was identified in organoids from inflamed sections of the human intestine. In addition, the typical age-related signatures (SBS1/SBS5) were also detected. SBS88 mutations were also detected in particular in canonical CRC driver genes such as APC, TP53, KRAS, and AXIN2. This demonstrated a link between microbial genotoxicity and the development of early malignant mutations in the inflamed colon.

Taken together, the results obtained here expand our current knowledge of the sequence target spectrum of colibactin and provide important insights into the

inflammation-related susceptibility of the human colon epithelium to the genotoxic function of colibactin. The work underscores the importance of colibactin as a sequence-specific microbial carcinogen, expands the spectrum of new target structures of this genotoxin, and highlights its potentially damaging effect on the human genome. Beyond these mechanistic findings, the results provide a framework for microbiome-oriented prevention and treatment strategies for CRC.

Zusammenfassung

Colibactin ist ein von der pks-Insel bestimmter *Escherichia coli*-Stämme des menschlichen Darmmikrobioms kodiertes Genotoxin, das die Entstehung von Darmkrebs (CRC) durch DNA-Strangvernetzungen (ICLs) und nachfolgende Doppelstrangbrüche (DSBs) bewirkt. In dieser Arbeit untersuche ich die sequenzspezifischen Eigenschaften von Colibactin-induzierten DNA-Schäden. Hierzu wurden spezifisch modifizierte Plasmide mit DNA-Sequenzen von 15 mutmaßlichen Colibactin-Zielstrukturen (CDMs) sowie Phosphorthioat-geschützte Oligonukleotide als Substrat verwendet. Dabei stellte sich eine bevorzugte Vernetzung bei AT-reichen Sequenzen heraus, insbesondere den Sequenzmotiven AAATTT, AATTTT und dem neu identifizierten, hochaffinen Motiv AAAGTT. Lineare DNA war deutlich anfälliger als superhelikale DNA, was auf einen Einfluss der Superhelizität auf die Vernetzungseffizienz von Colibactin hindeutet. Diese mit nackter DNA als Zielstruktur erzielten Ergebnisse konnten durch Nanopore-Sequenzanalysen weitgehend bestätigt werden. Demnach umfasst das Spektrum der Colibactin-Zielstrukturen für die Bildung von DNA-Strang-Quervernetzungen im Wesentlichen die zuvor bestätigten CDMs. Zusätzlich wurde jedoch die Sequenz AAAGTT als weiteres Motiv identifiziert.

Um das Auftreten der Colibactin-Signatur im Verlauf von entzündlichen Erkrankungen des menschlichen Colons zu untersuchen, wurden aus Biopsieproben von Patienten mit entzündlichen Darmerkrankungen (IBD) und vorhandenen adenomatösen Polypen dreidimensionale Kolonorganoide gezüchtet. Dies erfolgte auf der Grundlage eigens hierfür entwickelter Wachstumsmedien und Protokolle. Die Epithelzellen der gezüchteten Kolon-Organoiden wurden einer Gesamtgenomsequenzierung (WGS) unterzogen. Im Rahmen einer Zusammenarbeit erfolgte die bioinformatische Detektierung von Mutationssignaturen im Genom dieser Organoidzellen. Sie ergab eine Anreicherung der Colibactin-assoziierten Signatur SBS88 in der Mehrzahl, jedoch nicht in allen Proben. Insbesondere zeigte sich die SBS88-Signatur in Organoiden von entzündeten Abschnitten des menschlichen Darms. Daneben konnten die üblichen altersbedingten Signaturen (SBS1/SBS5) nachgewiesen werden. SBS88-Mutationen wurden insbesondere auch in kanonischen CRC-Treibergenen wie APC, TP53, KRAS und AXIN2 nachgewiesen. Damit konnte eine Verbindung zwischen mikrobieller Genotoxizität und dem Entstehen früher maligner Mutationen im entzündeten Kolon aufgezeigt werden.

In ihrer Gesamtheit erweitern die hier erzielten Ergebnisse unsere bisher verfügbaren Kenntnisse zum Sequenzspektrum von Colibactin und liefern wichtige Einblicke in die entzündungsbedingte Empfänglichkeit des menschlichen Colonepithels hinsichtlich der genotoxischen Wirkung von Colibactin. Die Arbeit unterstreicht die Bedeutung von Colibactin als sequenzspezifisches mikrobielles Karzinogen, erweitert das Spektrum neue Zielstrukturen dieses Genotoxins und unterstreicht seine potenziell schädigende Wirkung auf das menschliche Genom. Über diese mechanistischen Erkenntnisse hinaus liefern die Ergebnisse einen Rahmen für Mikrobiom-orientierte Präventions- und Therapiestrategien bei CRC.

1. Introduction

1.1 Cancer: A Global Health Challenge

Cancer is one of the leading causes of mortality and morbidity worldwide, responsible for nearly 10 million deaths in 2020 and accounting for 16.8% of all global deaths¹. It ranks among the top two causes of premature death in most countries and remains a major public health burden, particularly in Asia. This region bears nearly half of all new cases and over half of cancer-related deaths². Projections suggest that, by 2030, the annual incidence will increase to 21.4 million cases and 13.2 million deaths, primarily due to aging populations, demographic growth, and lifestyle-related risk factors associated with socioeconomic development^{3,4}. This escalating global burden underscores the urgent need to better understand cancer's molecular origins to inform prevention and therapy.

Cancer is traditionally viewed as a multistep process in which normal cells progressively acquire mutations that transform them into dysplastic, invasive, and ultimately metastatic states^{5,6}. Initiation occurs when DNA damage accumulates from genetic, metabolic, or environmental carcinogens—including radiation, chemicals, and oncogenic pathogens—disrupting normal cellular homeostasis^{7,8}. These mutational events often activate oncogenes or inactivate tumor suppressor genes, enabling uncontrolled proliferation, resistance to apoptosis, and genomic instability, all of which drive malignant progression⁹. Therefore, understanding these mutational processes and the balance between inherited, intrinsic, and environmental factors is central to elucidating cancer initiation and developing effective interventions.

1.2. Genetic and Environmental Contributions to Cancer

The initiation and progression of cancer are fundamentally driven by the accumulation of genetic alterations that disrupt normal cellular regulation. These mutations arise from three principal sources: hereditary predisposition, environmental exposures, and intrinsic errors during DNA replication. Their relative contributions differ, yet together they provide the framework through which mutational processes shape cancer development (**Figure1**).

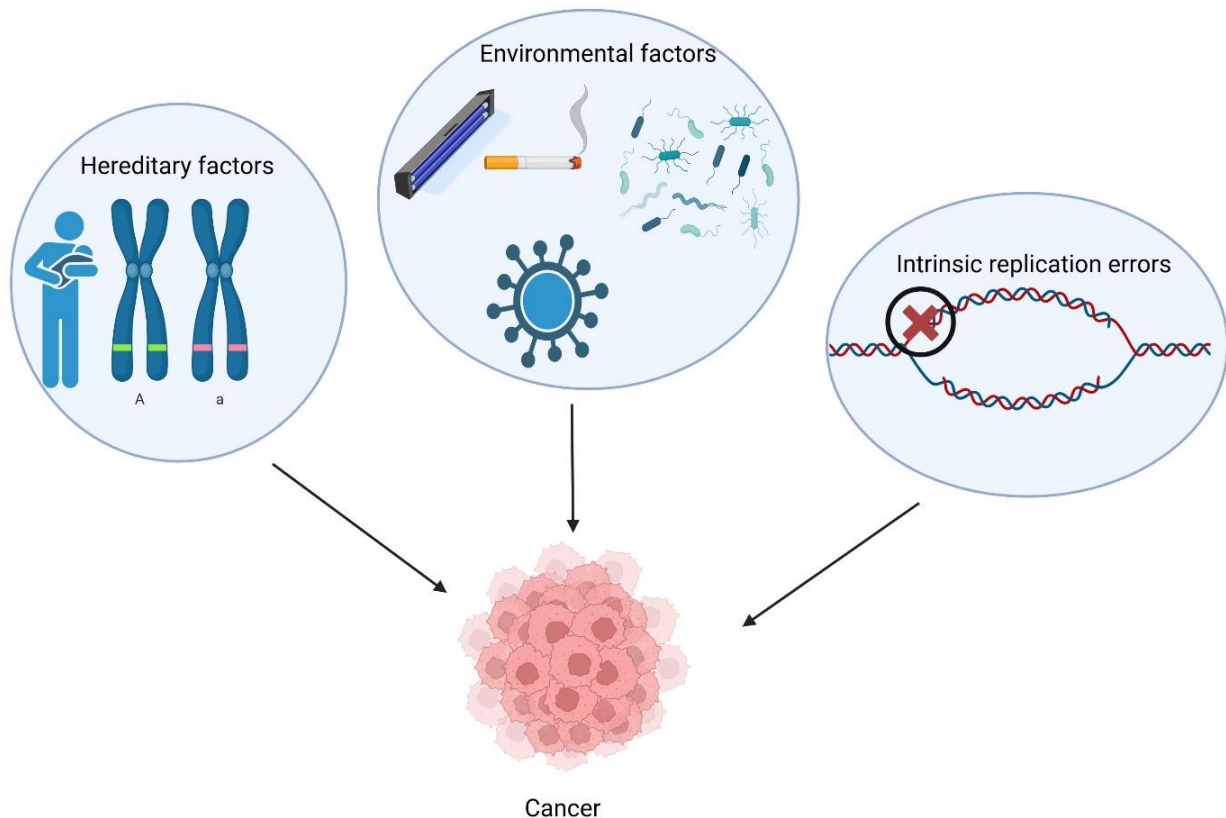


Figure 1- Major sources of mutational burden contributing to cancer development. Cancer arises through the interplay of three principal sources of genetic alterations: (i) hereditary factors, where germline mutations in cancer predisposition genes (e.g., mismatch repair genes in Lynch syndrome or APC in familial adenomatous polyposis) increase lifetime cancer risk; (ii) environmental factors, including ultraviolet (UV) radiation, tobacco smoke, viral infections, and microbial genotoxins (e.g., colibactin from *Escherichia coli*), which induce DNA damage and genomic instability; and (iii) intrinsic replication errors, arising from stochastic mistakes during DNA replication that escape repair mechanisms. Together, these processes introduce mutations that can disrupt oncogenes, tumor suppressors, or DNA repair pathways, ultimately driving malignant transformation. Created in <https://BioRender.com>.

1.2.1 Hereditary factors

Hereditary factors account for a relatively small, though clinically important, fraction of cancers, estimated at approximately 5–10% of cases ¹⁰. Germline mutations in key cancer susceptibility genes predispose individuals to malignancy, often at an earlier age. Examples include mutations in APC that underlie familial adenomatous polyposis (FAP) ¹¹, mismatch repair (MMR) gene defects that cause Lynch syndrome ¹², and pathogenic variants associated with breast and ovarian cancer ^{13,14}. While such syndromes are rare, their identification has provided invaluable insights into the molecular pathways that drive tumorigenesis and has enabled the development of preventive surveillance strategies for high-risk individuals.

1.2.2 Environmental factors

By contrast, environmental factors constitute the largest proportion of the cancer burden, estimated at ~30% of cases worldwide ¹⁵. These include chemical carcinogens (e.g., tobacco smoke, nitrosamines), physical agents (ultraviolet and ionizing radiation), dietary components, and chronic inflammatory states ¹⁶. Importantly, infection with oncogenic microbes is now recognized as a key environmental contributor ¹⁷. Viruses such as human papillomavirus (HPV) and hepatitis B and C viruses (HBV, HCV) are well-established causes of cervical and liver cancer, respectively ¹⁸. Likewise, bacterial infections contribute to carcinogenesis, with *Helicobacter pylori* providing the prototypical example by driving chronic gastritis, atrophy, and ultimately gastric carcinoma ¹⁹. These microbial examples illustrate how sustained environmental exposures can initiate mutagenic cascades and reshape tissue microenvironments to favor cancer development.

1.2.3 Intrinsic replication errors

Intrinsic mutational processes represent a third major source of genomic change. Errors that occur during DNA replication in proliferating stem and progenitor cells account for roughly 60–70% of mutations in human cancers ²⁰. Unlike germline predisposition or exogenous carcinogen exposure, these replication errors are stochastic and essentially unavoidable, representing a baseline mutational “background noise” in renewing tissues. Importantly, tissues characterized by rapid self-renewal, such as the intestinal epithelium, are particularly vulnerable to replication-driven mutagenesis, which in turn may explain their disproportionately high incidence of neoplastic transformation.

While hereditary, environmental, and intrinsic replication errors all contribute to the mutational burden in human tissues, these mutations alone do not inevitably lead to cancer. Their oncogenic potential depends on which genes are affected, particularly those that control cell growth, genome maintenance, and survival.

1.3 Transforming Genes and Core Cancer Pathways

The transformation of a normal cell into a malignant one is fundamentally driven by genetic alterations that disrupt core regulatory pathways. Genetic alterations contributing to oncogene activation and genetic disorders arise through various mechanisms, including chromosomal translocations (e.g., Bcr and Abl in chronic leukemia), point mutations (e.g., Ras in colorectal cancer), gene deletions (e.g., Erb-B in breast cancer), gene amplifications (e.g., N-myc in neuroblastoma), and insertional activation (e.g., C-myc in acute leukemia) ²¹. These alterations can be broadly classified into three categories: activation of oncogenes, loss of tumor suppressor function, and defects in DNA repair machinery.

1.3.1 Oncogenes

Oncogenes arise from mutations, chromosomal rearrangements, or gene amplifications in proto-oncogenes, resulting in gain-of-function alterations that drive malignant transformation. *Kirsten rat sarcoma* viral oncogene homologue (KRAS) is one of the best-characterized oncogenes and exhibits the highest mutation rate across human cancers, being strongly associated with highly lethal malignancies such as pancreatic ductal adenocarcinoma (PDAC), non-small cell lung cancer (NSCLC), and colorectal cancer (CRC), where mutations disrupt intrinsic GTPase activity and cause persistent activation of MAPK and PI3K pathways, leading to uncontrolled proliferation and survival ²². Similarly, the proto-oncogene MYC, which governs transcriptional programs for growth, metabolism, and proliferation, has been shown to causally induce tumorigenesis in both experimental mouse models and human cancers, with studies further revealing that its inactivation can trigger tumor regression, underscoring its dynamic role as an oncogenic driver ²³⁻²⁵.

1.3.2 Tumor suppressor genes

Tumor suppressor genes act as critical safeguards against malignant transformation by regulating cell-cycle checkpoints, DNA repair, and apoptosis. Among them, TP53 is often described as the “guardian of the genome,” as it orchestrates DNA damage responses and induces cell-cycle arrest or apoptosis when genomic integrity is

compromised ²⁶. Loss-of-function mutations in TP53 disable its protective functions, promote genomic instability, and confer therapy resistance, making it one of the most frequently altered genes across human cancers ²⁷. Similarly, APC is one of the most frequently mutated genes in colorectal cancer, with its inactivation representing a key and early event in tumorigenesis. Truncating mutations disrupt its regulatory role in the Wnt signaling pathway, leading to aberrant pathway activation and deregulation of key cellular processes that drive malignant transformation ²⁸. The RB1 gene was the first tumor suppressor identified, with its mutational inactivation established as the cause of the pediatric cancer retinoblastoma, and subsequent studies have shown that RB1 loss also contributes more broadly to deregulated cell cycle control and tumorigenesis across multiple cancer types ²⁹.

1.3.3 DNA repair machinery

DNA repair mechanisms are crucial for preserving genome stability, and defects in these pathways dramatically increase cancer risk. DNA mismatch repair (MMR) is a conserved pathway that corrects base–base mismatches and insertion/deletion loops generated during DNA replication and recombination, thereby preserving genomic stability ³⁰. In addition to error correction, MMR suppresses homologous recombination and participates in DNA damage signaling and checkpoint regulation. Mutations in core MMR genes such as MLH1, MSH2, and MSH6 cause microsatellite instability, a hallmark of Lynch syndrome and a subset of sporadic colorectal cancers ³¹. BRCA1 and BRCA2 are tumor suppressor genes most strongly associated with hereditary breast and ovarian cancers. Both play central roles in the DNA damage response, where they regulate cell cycle checkpoints and facilitate homologous recombination-mediated repair of double-strand breaks. Mutations in these genes impair genomic stability, thereby predisposing carriers to cancer development ³¹.

These transforming genes play a central role in colorectal cancer (CRC) development and progression. In the classical adenoma–carcinoma sequence, inactivation of the tumor suppressor APC represents the gatekeeping event that initiates adenoma formation by disrupting Wnt signaling ³². Subsequent activating mutations in oncogenes such as KRAS drive adenoma expansion, while later loss-of-function mutations in TP53 serve as key drivers of malignant transformation and invasion

^{22,27}.Collectively, these sequential alterations illustrate how oncogenes and tumor suppressor genes cooperate to shape the stepwise transformation from normal epithelium to invasive carcinoma, underscoring their central role in CRC pathogenesis.

1.4 Microbes and Cancer: From *Helicobacter pylori* to the Gut Microbiome

1.4.1 The importance of microbes in GI tract

Beyond genetic inheritance and intrinsic cellular processes, the tumor microenvironment is also shaped by microbial exposures. The human gastrointestinal tract harbors a diverse and complex community of approximately 100 trillion microorganisms—bacteria, fungi, viruses, and protists—collectively known as the gut microbiota ^{33,34}. This microbial ecosystem plays a crucial role in host physiology, contributing to nutrition, immune system development, and defense against pathogens ³⁵. For instance, in terms of nutritional benefits, certain bacterial species, such as *Bifidobacterium*, are involved in synthesizing critical compounds like vitamin K and B vitamins ³⁶. Moreover, the fermentation of dietary fibers by gut microbes produces short-chain fatty acids (SCFAs) ³⁷, a class of small fatty acids containing 2 to 5 carbons ^{38,39}. The predominant short-chain fatty acids (SCFAs) in the colon—acetate, propionate, and butyrate—support the growth of beneficial bacteria and stimulate regulatory T cells, reducing inflammatory mediators ⁴⁰. Once absorbed, SCFAs exhibit systemic effects, including immunomodulatory and anti-inflammatory properties, which are vital for maintaining the host's health. They also enhance colonic oxygen consumption by epithelial cells, contributing to improved immune regulation and gut barrier integrity ⁴¹. However, when this balance is disrupted—a condition known as dysbiosis—microbial alterations contribute to inflammation and disease ^{42,43}.

1.4.2 Bacterial carcinogenesis paradigm

Helicobacter pylori (*H. pylori*) is a gram-negative bacterium that colonizes the human stomach and infects nearly half of the world's population, representing the strongest risk factor for gastric cancer ⁴⁴⁻⁴⁶. Its persistence in the acidic gastric environment relies on urease, which hydrolyzes urea into ammonia to neutralize acid and enable

migration through the mucus layer ⁴⁷. Pathogenicity is largely driven by virulence factors, most notably the *cag* pathogenicity island, which encodes a type IV secretion system (T4SS) ⁴⁸. This injects CagA, ADP-heptose, and bacterial DNA into host epithelial cells, activating aberrant SHP2/ERK signaling, disrupting adhesion, and inducing the “hummingbird” phenotype, while also triggering innate immune responses via β 1 integrins, TLRs, and the cGAS–STING pathway ⁴⁹. In parallel, the secreted vacuolating cytotoxin A (VacA) forms membrane pores that cause vacuolation, mitochondrial damage, autophagy, and T-cell suppression ^{50,51}. Together, these mechanisms enable chronic colonization, drive inflammation, and promote epithelial transformation, establishing *H. pylori* as a paradigm of microbe-induced carcinogenesis.

1.4.3 Microbiota and CRC

Extending beyond *H. pylori*, the human gastrointestinal tract is colonized by trillions of microbes, collectively forming the gut microbiome. Under normal conditions, epithelial cells create an anaerobic environment in the gut by reducing oxygen availability through β -oxidation. However, during chronic inflammation, impaired β -oxidation increases oxygen levels and promotes the formation of nitrate (NO_3^-), leading to dysbiosis and the overgrowth of Proteobacteria, such as *Escherichia coli* (*E. coli*) and *Bacteroides fragilis*. Dysbiosis is marked by a loss of microbial diversity and an overgrowth of pathobionts, such as *Proteobacteria* and *Escherichia coli* (*E. coli*) inflammation, profoundly alters host–microbe interactions in the colon ⁵². Mucosal inflammation further reshapes the microbial community, reducing diversity and shifting species abundance, thereby creating an unstable ecosystem ⁵³. These disruptions impair epithelial barrier integrity and intestinal immune regulation, facilitating inappropriate immune responses ⁵⁴. At the same time, dysbiosis fosters a mutagenic environment through chronic inflammation, which elevates reactive oxygen and nitrogen species and enhances host exposure to microbial metabolites and toxins ⁵⁵. *E. coli* produces colibactin, a genotoxin that damages DNA and stimulates tumorigenesis (Thakur et al., 2025). BFT not only cleaves E-cadherin, a critical protein in cell adhesion and barrier integrity, but also recruits polymorphonuclear leukocytes (PMNs), further exacerbating inflammation and promoting colorectal cancer (CRC) progression (Quaglio et al., 2022).

Collectively, the interplay between microbiota, inflammation, and host immunity underscores the dual role of the gut microbiome in maintaining health and in driving the pathogenesis of inflammatory bowel disease (IBD) and colorectal cancer (CRC). Notably, among the microbial products implicated in these processes, the genotoxin colibactin produced by *pks*⁺ *E. coli* has emerged as a critical mediator linking dysbiosis, inflammation, and carcinogenesis.

1.5 *Escherichia coli* and Colibactin: A Bacterial Genotoxin in Colorectal Cancer

Colibactin, a bacterial genotoxin, induces DNA inter-strand crosslinks, leading to double-strand breaks, chromosomal abnormalities, and genomic instability—hallmarks of malignant transformation⁵⁶. *Escherichia coli* strains harboring the *pks* genomic island (*pks*⁺ *E. coli*), which encodes the biosynthetic machinery for colibactin production, are frequently detected as commensals within the intestinal microbiota. These strains have been linked to inflammatory bowel disease⁵⁷ and urinary tract infections and are disproportionately represented in colorectal cancer (CRC) tissues^{58,59}. Experimental studies using CRC mouse models further demonstrated that colibactin-producing *E. coli* accelerates tumor development, suggesting that these bacteria play a critical role in carcinogenesis⁵⁹.

1.5.1 Biosynthesis of colibactin

Colibactin is a genotoxin synthesized by the 54-kilobase *pks* pathogenicity island, encoded by 19 *clb* genes (*ClbA* to *ClbS*). This gene cluster, expressed in several members of the Enterobacteriaceae family—including *Citrobacter koseri*, *Klebsiella pneumoniae*, and *Enterobacter aerogenes*—confers the ability to produce colibactin⁶⁰. The *Clb* gene cluster responsible for colibactin biosynthesis, involving four non-ribosomal peptide synthetase (NRPS) proteins (*ClbE*, *H*, *J*, *N*), four polyketide synthase (PKS) enzymes (*ClbC*, *G*, *I*, *O*), two hybrid NRPS–PKS proteins (*ClbB*, *K*), and additional accessory, regulatory, and resistance proteins⁶¹. While most enzyme functions have been well-documented in previous reviews, the roles of *ClbO* (PKS) and *ClbL* (amidase) remain unclear^{62,63}.

The NRPS-PKS assembly synthesizes linear intermediates known as precolibactins, characterized by an N-myristoyl-D-Asn prodrug motif (**Figure 2**). Precolibactins are transported to the periplasm by the ClbM transporter^{64,65}, where the ClbP peptidase removes the prodrug motif, triggering spontaneous four-fold cyclization into the genotoxic colibactin structure. The resulting colibactin consists of two nearly symmetrical subunits, each containing an electrophilic cyclopropane warhead. These cyclopropane groups react with adenine residues on DNA to create inter-strand cross-links, a hallmark of colibactin's genotoxicity⁶⁶. ClbL incorporates two electrophilic cyclopropane warheads into colibactin, which becomes active upon ClbP-mediated activation, facilitating inter-strand cross-link formation and contributing to DNA damage and genomic instability^{65,67}.

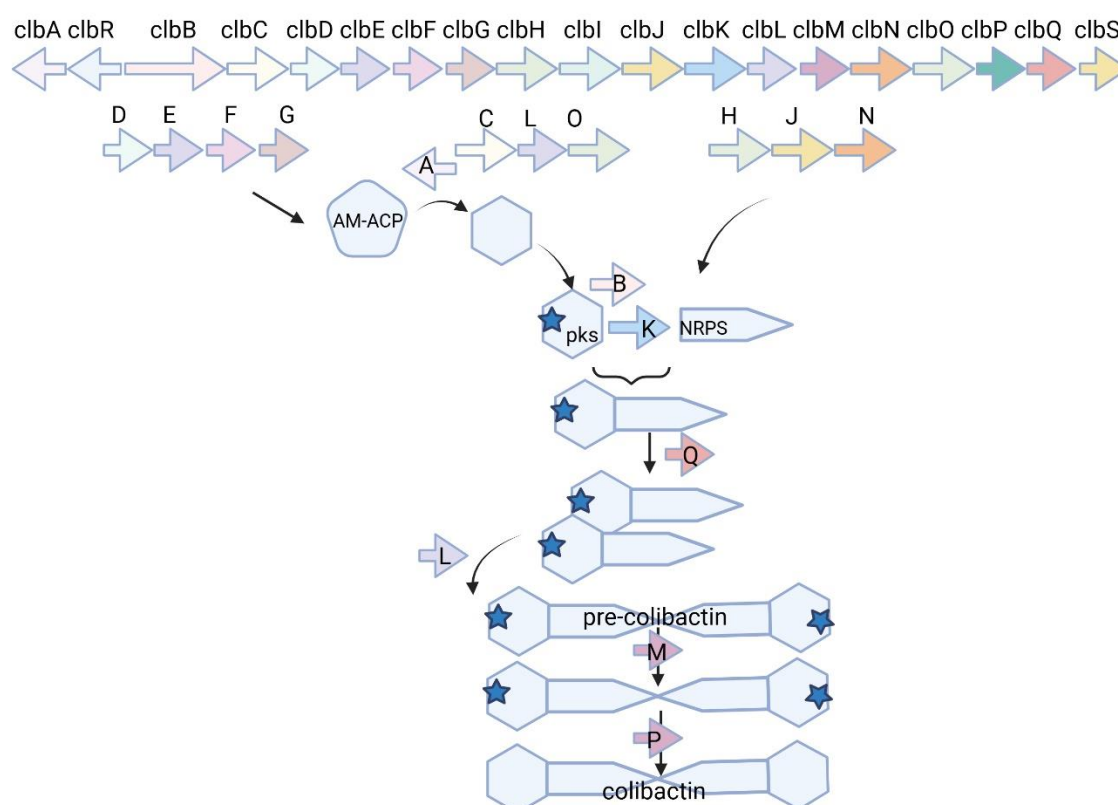


Figure 2- Biosynthesis of colibactin. The process is mediated by the *clb* gene cluster, comprising several enzymes responsible for distinct steps in colibactin synthesis. ClbA initiates the process with pantetheinyltransferase activity, while ClbD facilitates substrate conversion via 3-hydroxybutyryl-CoA dehydrogenase. ClbE acts as a freestanding thiolation domain (T), and ClbF functions as an acyl-CoA dehydrogenase. ClbG operates as a freestanding acyltransferase (AT), and ClbL, a peptidase, is critical in the later stages of the pathway. ClbM transports intermediates to the periplasm, where ClbP cleaves the prodrug motif to activate precolibactin. Additionally, ClbQ is involved in thioesterase editing, ClbR serves as a transcriptional

regulator, and ClbS confers resistance to colibactin toxicity, ensuring the host bacterium's survival during toxin production⁶⁶. Adapted from: ⁶⁵. Created in <https://BioRender.com>.

1.5.2 Genotoxic Impact of *pks+* *E. coli*

Colibactin-induced DNA damage activates a cascade of host signaling pathways. Infection of HeLa cells with *pks+* *E. coli* triggers the ATM/ATR-dependent DNA damage response (DDR), evidenced by phosphorylation of CHK2, RPA, and γ H2AX, and accumulation of CDC2 leading to G2-phase cell cycle arrest ⁶². Colibactin-induced DNA lesions also recruit FANCD2 to γ H2AX foci, implicating the Fanconi anemia pathway in the repair of inter-strand cross-links ⁶⁸. In addition, neutral comet provides direct visualization of double-strand breaks (DSBs) in host cells ⁶⁹. These findings collectively highlight the multifaceted genotoxic activities of colibactin, consistent with its alkylating and cross-linking properties.

In vivo studies reinforce these mechanistic insights. In gnotobiotic mice, *pks+* *E. coli* colonization alone does not impair epithelial integrity under normal homeostasis. However, following chemically induced mucosal injury, the bacteria gain direct access to the epithelium, resulting in epithelial damage and chronic colitis ⁷⁰. As a gram-negative facultative anaerobe residing in the gut, *pks+* *E. coli* not only promotes inflammation but also produces colibactin, a genotoxin with carcinogenic potential. Importantly, chronic intestinal inflammation compromises epithelial barrier function and fosters a mutagenic environment, thereby establishing a critical link between *pks+* *E. coli*, colitis-associated colorectal cancer (CRC), and the broader role of inflammation as a driver of tumor initiation and progression.

1.6 Mutational Signatures in Human Cancer

1.6.1 Concept of mutational signatures

The genomic landscape of cancer is shaped not only by the accumulation of random mutations but also by distinct mutational processes that leave characteristic imprints on the genome. These imprints, known as mutational signatures, are defined as reproducible patterns of base substitutions, small insertions or deletions (indels), and structural rearrangements that collectively reflect the underlying biological

mechanisms driving DNA damage and repair ^{71,72}. Advances in high-throughput sequencing have enabled systematic cataloguing of these signatures, culminating in the COSMIC (Catalogue of Somatic Mutations in Cancer) database, which currently describes more than 70 single base substitution (SBS) signatures across human malignancies.

Some mutational signatures arise from endogenous processes, such as SBS1 and SBS5, which reflect the gradual accumulation of mutations during aging through spontaneous 5-methylcytosine deamination and imperfect DNA repair ⁷³. Others are linked to environmental exposures or exogenous carcinogens, including tobacco smoke, ultraviolet radiation, and, more recently, microbial genotoxins. Thus, mutational signatures not only reveal mechanistic insights into tumorigenesis but also serve as molecular footprints connecting environmental or microbial exposures with their long-term genomic consequences in human cancer.

1.6.2 SBS88- colibactin signature

A major breakthrough in this field was the discovery of SBS88, a mutational signature uniquely associated with bacterial genotoxin colibactin, produced by *Escherichia coli* strains carrying the *pks* genomic island. First reported in colorectal cancers enriched with *pks*+ *E. coli* exposure ⁷⁴ and later validated by independent whole-genome sequencing studies in both clinical cohorts and experimental infection models ⁷⁵. SBS88 is characterized by T>A and T>C substitutions at AT-rich hexamer motifs such as AAATTT and AAAATT, which correspond to colibactin damage motifs (CDMs).

The recurrence of these motif-specific mutations across diverse experimental and clinical contexts highlights colibactin's mechanistic specificity, establishing it as a microbial carcinogen. SBS88 provides direct molecular evidence of microbial mutagenesis in human cancer and serves as a biomarker for studying host–microbiome interactions in colorectal carcinogenesis, with potential applications in personalized risk stratification and microbiome-targeted therapies.

1.7 The Human Intestine: Anatomy, Epithelial Architecture, and Stem Cells

The gastrointestinal (GI) tract represents the largest mucosal interface in the human body, where continuous exposure to dietary antigens and microorganisms requires an adjusted balance between nutrient uptake, immune regulation, and epithelial barrier protection ⁷⁶. Both the small intestine and the colon share a layered organization—mucosa, submucosa, muscularis propria, and adventitia—with the mucosa forming the dynamic luminal surface responsible for selective absorption and protection ⁷⁷. This epithelial layer undergoes rapid renewal, as Lgr5⁺ stem cells in the crypts continuously generate progenitors that differentiate into mature lineages, ensuring turnover of the epithelium every 3–5 days, thereby sustaining barrier integrity under mechanical and biochemical stress ⁷⁸. The small intestine is uniquely structured into villi and crypts, where villi contain absorptive enterocytes and secretory lineages, while crypts harbor stem cells protected by Paneth cell–derived antimicrobial peptides. In contrast, the colon lacks villi and is composed solely of crypts, where stem cells and goblet cells dominate, maintaining mucosal renewal and mucus production. Gradients of WNT, enriched at the crypt base, and BMP, increasing toward the lumen, establish spatial cues that regulate self-renewal and differentiation in both the small intestine and colon ⁷⁹ (Figure 3).

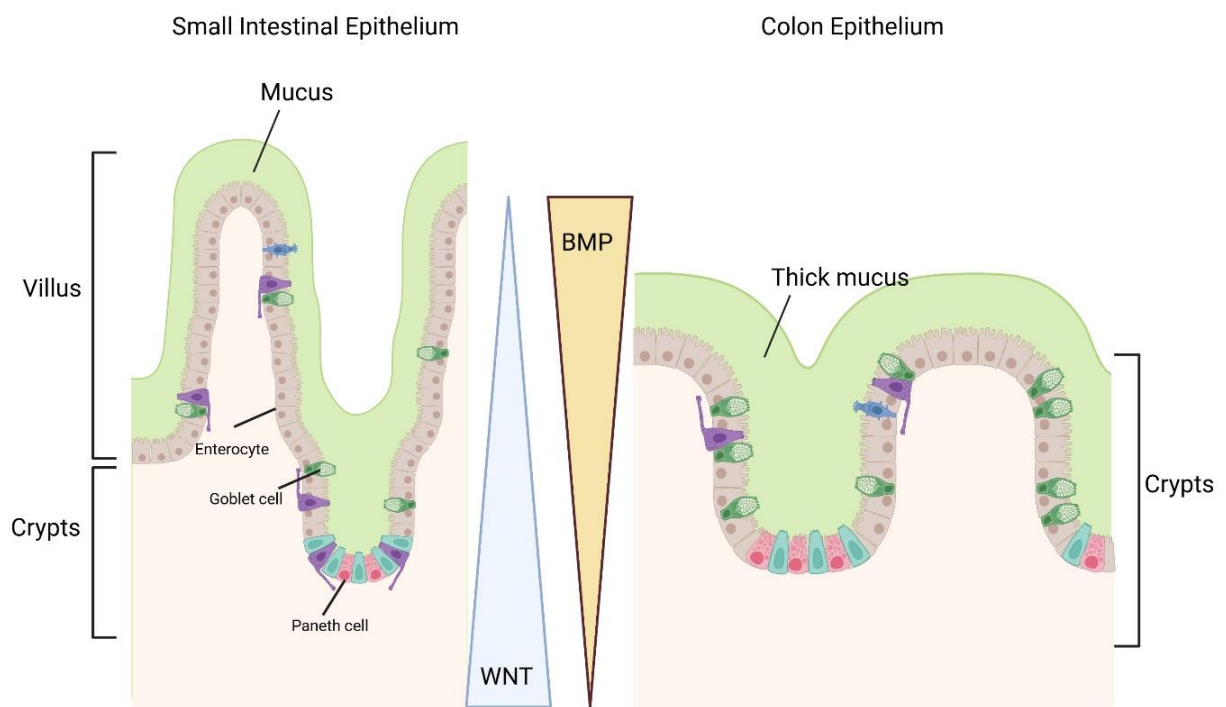


Figure 3- Structural and functional differences between the small intestinal and colonic epithelium.

The small intestinal epithelium is organized into villi and crypts, where absorptive enterocytes dominate the villus surface and are supported by secretory cell types, including goblet cells and Paneth cells, within the crypts. Paneth cells secrete antimicrobial peptides and contribute to stem cell niche maintenance, while goblet cells produce mucus that provides epithelial protection. In contrast, the colonic epithelium lacks villi and consists solely of crypts covered by a thick mucus layer, with abundant goblet cells and stem cells at the crypt base. Gradients of WNT (enriched at the crypt base) and BMP (increasing toward the lumen) orchestrate epithelial cell proliferation and differentiation, ensuring homeostatic renewal of both small intestinal and colonic epithelia. Created in <https://BioRender.com>.

1.7.1 Key signaling pathways

The intestinal epithelium forms a dynamic barrier that regulates interactions between the microbiota and the host immune system, thereby maintaining tissue homeostasis⁸⁰. Homeostatic renewal of the intestinal epithelium relies on a network of conserved signaling pathways that tightly control stem cell proliferation, lineage allocation, and growth constraints. WNT and EGF drive stem cell renewal and proliferation; NOTCH regulates absorptive versus secretory lineage allocation; while BMP and Hippo pathways impose growth constraints to preserve crypt–villus architecture⁸¹. Importantly, dysregulation of these pathways has direct implications for colorectal cancer (CRC): loss of APC hyperactivates WNT, driving uncontrolled proliferation; aberrant NOTCH signaling disrupts lineage balance; and impaired BMP or Hippo signaling removes essential growth brakes. Together, these alterations transform the stem cell compartment into a vulnerable site for malignant initiation, highlighting how the same pathways that govern epithelial renewal underlie the earliest steps of CRC development.

1.7.3 Barrier function

The intestinal barrier mediates selective exchange between the lumen and underlying tissues, a function maintained by the mucus layer, epithelial cells, and a network of intercellular junctions collectively termed the apical junctional complex (AJC)⁸². Tight junctions, composed of transmembrane proteins such as claudins, occludin, and tricellulin together with adaptor proteins like zonula occludens (ZO), form the paracellular gate that regulates ion and solute transport while preserving epithelial polarity and proliferation⁸³. Directly beneath them, adherens junctions—anchored by E-cadherin interactions with β -catenin and α -catenin—reinforce cell–cell adhesion and

integrate key signaling pathways such as Wnt⁸⁴. Desmosomes, through desmoglein and desmocollin, provide additional mechanical stability by linking to the cytoskeleton, while gap junctions composed of connexin-based connexons enable intercellular communication via diffusion of ions and small molecules⁸⁵. Together, these junctional systems coordinate barrier integrity, signaling, and epithelial renewal, functions that become compromised during chronic inflammation and are central to the pathogenesis of IBD and colorectal cancer.

When barrier integrity is compromised, epithelial permeability increases, permitting luminal bacteria and antigens to translocate into the mucosa (**Figure4**). This disruption triggers immune activation, epithelial stress, and chronic inflammation⁸⁶. Persistent barrier dysfunction therefore promotes chronic inflammation, establishing a pro-tumorigenic microenvironment central to the development of inflammatory bowel disease (IBD) and colorectal cancer^{76,87}.

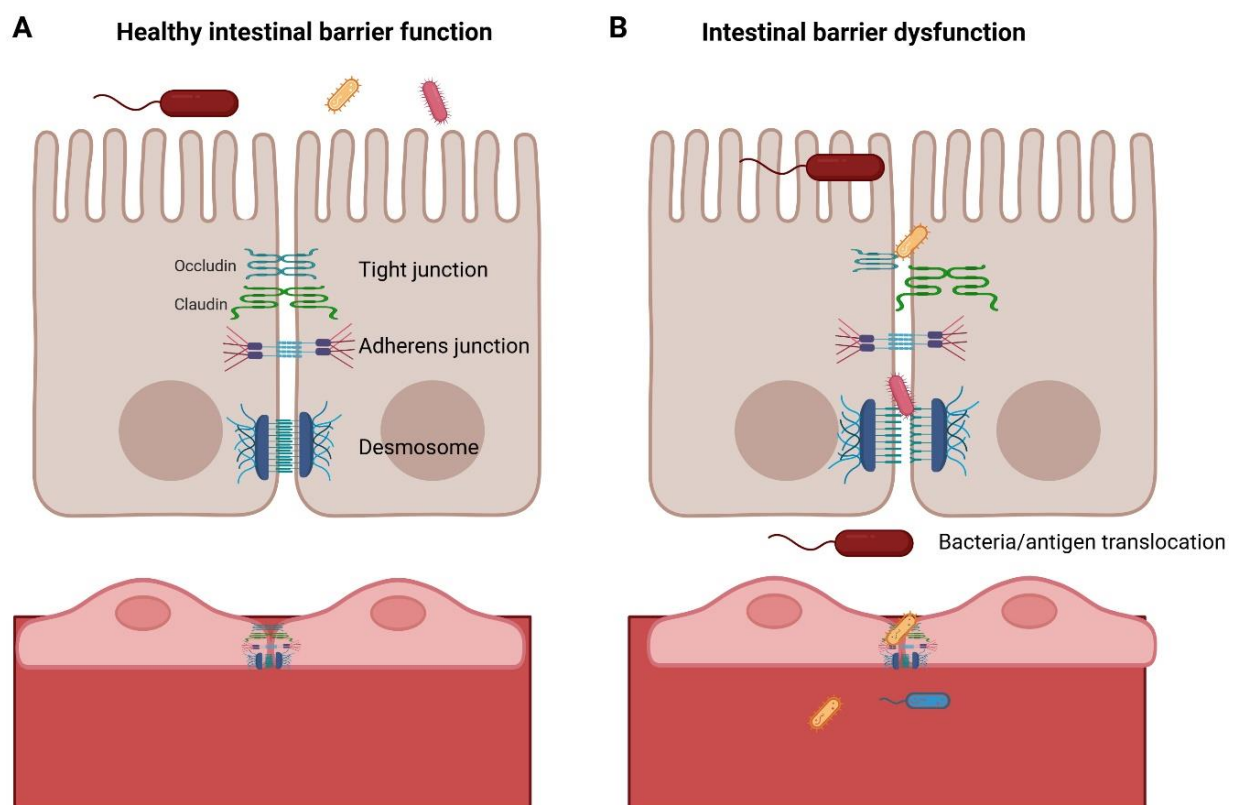


Figure 4- Intestinal barrier integrity and its disruption. Schematic illustration of healthy versus dysfunctional intestinal barrier integrity, highlighting how disruption of epithelial junctions facilitates bacterial and antigen translocation. (A) In healthy intestinal epithelium, the barrier is maintained by a coordinated network of junctional complexes, including tight junctions (occludin, claudins), adherens junctions (E-cadherin–catenin complex), and

desmosomes, which together regulate selective permeability, maintain epithelial polarity, and prevent microbial penetration. (B) During barrier dysfunction, tight junction disassembly and weakened adherens and desmosomal connections increase epithelial permeability, allowing bacteria and luminal antigens to translocate across the mucosa. This translocation promotes immune activation, chronic inflammation, and exposure of host tissues to microbial toxins, thereby contributing to inflammatory bowel disease (IBD) pathogenesis and colorectal cancer initiation. Created in <https://BioRender.com>.

1.8 Colibactin in the Context of Inflammatory Bowel Disease, Colon Polyps and CRC

Inflammatory bowel disease (IBD) is a chronic and relapsing disorder of the gastrointestinal tract, traditionally classified into ulcerative colitis (UC) and Crohn's disease (CD) ³⁶. Although the precise triggers of disease onset and relapse remain unclear ⁸⁸, both host genetics and microbial factors contribute to pathogenesis. Numerous IBD susceptibility genes are associated with host immune responses to gut bacteria ⁸⁹. Dysbiosis, characterized by a decrease in microbial diversity or a combination of the loss of beneficial bacteria and an increase in pathogenic bacteria, has been consistently implicated in disease development ⁹⁰. The most notable alterations in the microbiota of individuals with inflammatory bowel disease (IBD) include reduced diversity in bacterial species, particularly a decrease in the abundance of Bacteroidetes and Firmicutes ³⁵, coupled with an increase in the abundance of Proteobacteria ⁹¹. Among these, *Escherichia coli*, particularly strains harboring the *pks* island that encodes the genotoxin colibactin, are frequently enriched in the mucosa of IBD patients, especially those with UC ⁹²⁻⁹⁵.

Elevated mucosa-associated *pks*+ *E. coli* has also been reported in patients with familial adenomatous polyposis (FAP), an inherited cancer syndrome caused by germline mutations in the APC gene on chromosome 5q21 ^{93,96}. FAP is characterized by the development of hundreds to thousands of synchronous adenomatous polyps in the colon, which inevitably progress to CRC without intervention ⁹⁷. The incidence and frequency of polyp formation varies widely among families carrying the same APC gene mutation ⁹⁸, corroborating the notion that other factors contribute to the onset of the disease, as do components of the microbiome ⁹⁹.

Indeed, studies have shown that FAP mucosa is frequently colonized by bacterial biofilms, predominantly *pks*+ *E. coli* and enterotoxigenic *Bacteroides fragilis* (ETBF),

which significantly alter the mucosal microenvironment and promote tumorigenesis¹⁰⁰⁻¹⁰³ Strikingly, the prevalence of *pks+* *E. coli* and ETBF is markedly higher in FAP patients (68% and 60%, respectively) compared to healthy controls (22% and 30%), paralleling findings in sporadic CRC (Dejea et al., 2018). Mouse models further confirm that co-colonization with *pks+* *E. coli* and ETBF accelerates colonic tumorigenesis¹⁰³

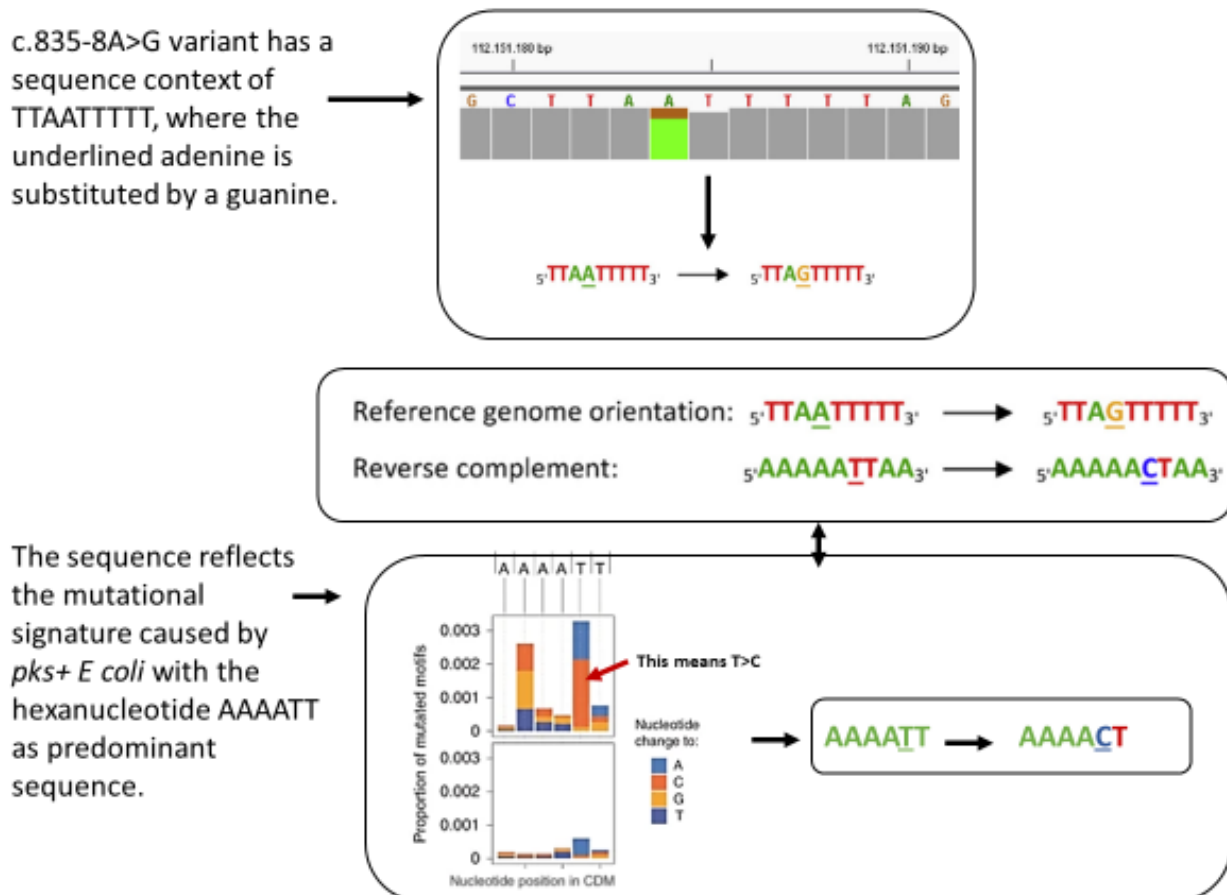


Figure 5- Linkage between colibactin and colonic polyps. This figure illustrates the connection between colibactin-induced DNA damage and colonic polyps. The variant c.835-8A>G is located within a sequence context of TTAATTTT, where the underlined adenine is substituted by a guanine. The sequence orientation is depicted in both the reference genome and its reverse complement, demonstrating its resemblance to the mutational signature caused by *pks+* *E. coli*, particularly the AAAATT hexanucleotide sequence, which predominates in colibactin-associated DNA damage. The lower panel shows mutational changes, such as T>C substitutions, reinforcing the characteristic patterns observed in colibactin-induced genotoxicity. This figure highlights how the mutational signature of colibactin aligns with observed genetic changes in colonic polyps, linking microbial activity to carcinogenesis. Adapted from^{74,104}.

Recent genomic studies have reinforced this connection by linking colibactin mutational signatures to colonic polyps. A recurrent *APC* splice variant (c.835-8A>G), frequently observed in unexplained polyposis patients, occurs within AT-rich motifs resembling colibactin damage motifs (CDMs), with mutation patterns consistent with SBS88—the signature of colibactin-induced genotoxicity (**Figure 5**)^{74,104} This suggests that colibactin-producing bacteria may contribute to both inherited and sporadic adenoma formation, providing a microbial explanation for some de novo cases of FAP, which account for 10–30% of patients variant⁹⁷.

Taken together, these findings highlight a critical interplay between genetic predisposition, microbial colonization, and epithelial mutagenesis. While FAP is classically regarded as a hereditary disease driven by *APC* mutations, accumulating evidence suggests that colibactin-producing bacteria act as potent co-drivers of early neoplastic lesions. Thus, elucidating the role of *pks*+ *E. coli* in adenoma formation is essential for understanding CRC initiation in both hereditary (FAP-associated) and sporadic contexts, and forms a central focus of this study.

1.9 Recent challenges, research aims and objectives

1.9.1 Challenges in Direct Detection and Study of Colibactin

Direct detection of colibactin remains a significant challenge due to its inherent chemical instability and transient nature, which make its isolation and characterization exceedingly difficult. In vivo studies are further complicated by the complexity of the human gut microbiome, where diverse microbial communities interact with each other and the host, making it difficult to attribute pathogenic effects specifically to *pks*+ *E. coli*. Additionally, host immune responses and other microbial metabolites often obscure the direct contributions of colibactin to pathogenesis. Animal models, while valuable, are limited to fully replicate human colonic physiology, the intricate host-microbe interactions, and the prolonged latency period associated with colorectal cancer development. Consequently, most research relies on indirect approaches, such as analyzing the mutational signatures associated with colibactin-induced DNA damage or using genetically modified bacterial strains that either produce or lack colibactin.

These limitations highlight the importance of refining preferential DNA-targeting sites and to develop reliable biomarkers and to better understand its role in colorectal cancer initiation, including its contributions to inflammatory bowel disease (IBD) and colon polyp progression.

1.9.2 Aim of the Work

The objectives of this study are to elucidate the mechanistic contribution of colibactin-producing bacteria to colorectal cancer initiation, with a particular focus on precancerous conditions such as IBD and adenomatous polyps.

First, the project aims to refine the sequence specificity of colibactin on naked DNA by characterizing its preferential targeting motifs, assessing the impact of chromatin context and environmental influences on DNA damage, and directly detecting colibactin-induced inter-strand crosslinks using Nanopore sequencing. This approach will provide deeper insight into the molecular determinants of colibactin's DNA-targeting mechanism.

Based on this characterization, the second aim is to investigate the mutational impact of colibactin in patient-derived organoid models of precancerous lesions. Organoids established from IBD and polyp tissues will undergo whole-genome sequencing to identify colibactin-associated signatures (e.g., SBS88) alongside other mutational events. Comparative analysis with matched normal mucosa, inflamed mucosa and blood samples will help to elucidate the contribution of colibactin-producing bacteria to the development of premalignant and malignant lesions in the human colon.

The insights gained from this work will establish a mechanistic framework for how colibactin contributes to the earliest stages of colorectal carcinogenesis, with particular emphasis on its mutational imprint in precancerous tissues. These findings are expected to guide the development of protective and therapeutic strategies that specifically target microbial genotoxin-driven cancer risk, thereby providing new opportunities for prevention and early intervention in colibactin-associated colorectal cancer.

2. Materials and Methods

2.1. Materials

2.1.1. Human Tissue

Human colon tissue biopsies were obtained from IBD and polyp patients undergoing routine endoscopy, under the ethics (AZ B231/98-1/13) approval by UKSH. All human materials used in this study were anonymized (Table 1 and Table 2).

Table 1: Human Tissue (IBD)

All following colon tissue were obtained from anonymized IBD patients.

Biopsy number	Age	Sex	Inflammation level	Disease
io-1	26	XY	0	UC
io-1	26	XY	2	UC
io-2	26	XY	1	CD
io-2	26	XY	0	CD
io-3	29	XX	0	UC
io-3	29	XX	2	UC
io-4	51	XY	0	UC
io-4	51	XY	2	UC
io-5	36	XX	2	UC
io-5	36	XX	0	UC
io-6	59	XX	2	UC
io-6	59	XX	0	UC
io-7	27	XY	2	UC
io-7	27	XY	0	UC
io-8	55	XX	0	UC
io-8	55	XX	2	UC
io-9	22	XX	0	UC
io-9	22	XX	3	UC
io-10	28	XY	0	UC
io-10	28	XY	3	UC
io-11	23	XY	0	UC
io-11	23	XY	2	UC

io-12	28	XX	0	UC
io-12	28	XX	3	UC

Table 2: Human Tissue (colon polyp)

All following colon tissue were obtained from anonymized colon polyp patients.

Biopsy number	Age	Sex	Disease
Nr.5 polyp	57	XX	Polyp
Nr.5 hc (healthy control)	57	XX	Polyp

2.1.2. Cell lines

Table 3: Cell lines

Strain	Description	Medium	Source
L-WRN producing cell line	Mouse fibroblasts; cells secrete biologically active Wnt3a protein, Rspodin3 and Noggin; selection marker G418	DMEM + 10% FCS, 2mM L-Glutamine, 1mM Na-Pyruvate, 0.5 mg/mL geneticin and 0.5 mg/mL hygromycin B	Kindly provided by Clevers Lab, Utrecht, Netherlands, RRID: CVCL_DA06
Cultrex HA-R-Spondin1-Fc 293T Cells	Human 293T cell line stably transfected to express murine Rspo1 with an N-terminal HA epitope tag and fused to a C-terminal murine IgG2a Fc fragment	DMEM + 10% FCS	Kindly provided by AG Aden, IKMB (Insitute of Clinical Molecular Biology, Catalog number: 3710-001-01
Noggin-producing cell line	HEK Noggin	DMEM + 10% FCS Puromycin 25 µg/mL	Kindly provided by K. Aden, IKMB (Insitute of Clinical Molecular Biology) ^{105,106}

2.1.3. Bacterial Strains

Table 4: Bacterial stains

Strain		Resistance	
E. coli M1/5	WT fecal isolate, with mutation of rpsL, which confers streptomycin	Streptomycin	U. Dobrindt (University of Münster)

	resistance. Colibactin positive		
E. coli H5194	DH10B pBACpks_Cam	Chloramphenicol	T.F. Meyer (Max Planck Institute for Infection Biology)
E. coli H5195	DH10B pBAC_Vector_Cam	Chloramphenicol	T.F. Meyer (Max Planck Institute for Infection Biology)
E. coli Nissle 1917			T.F. Meyer (Max Planck Institute for Infection Biology)

2.1.4. Chemicals and Reagents

Table 5: Chemicals and reagents

Reagent	Supplier	Catalog number
1×DPBS without Ca/Mg	Dulbecco's Balanced Salt Solution	14,190,250
Matrigel Matrix	Corning	11543550
Advanced DMEM/F12	Thermo Fisher Scientific	12634028
GlutaMax (100×)	Thermo Fisher Scientific	31765068
HEPES (1 M)	Capricorn Scientific	HEP-B
Fetal bovine serum (FBS), heat inactivated	Sigma	F7524
TrypLE	Gibco	11538856
Cryo-SFM	Sigma-Aldrich	C-29912
Penicillin/Streptomycin (100×)	Gibco	Cat. No. PS-B
Gentamycin Hygromycin B Geneticin (G418)	Sigma Carl Roth	CP12.2
Bovine Serum Albumin (BSA), Fraction V	Biomol	1,400,250
DMEM	Capricorn Scientific	DMEM-HPSTA
Sodium Pyruvate	Capricorn Scientific	NPY-B
Glutamine	Thermo Fisher Scientific	25030024
0.05% Trypsin/EDTA	Gibco	25300062
Zeocin	Fisher Scientific GmbH	R25001
B27 (NCS21 Neuronal Supplement), 50×	Capricorn Scientific	C21-H
N2 Supplement 100×	Capricorn Scientific	N2-K
Nicotinamide	Sigma	N0636

2 Materials and Methods

2.1 Materials

N-acetyl-L-cysteine	Sigma Aldrich	A9165
Human EGF, recombinant, PeproTech	Gibco	AF-100-15-1MG
A83-01 (TGF β inhibitor)	Biomol	Cay9001799
Human FGF, recombinant, PeproTech	Gibco	100-18B-1MG
Human Noggin, Recombinant, PeproTech	Gibco	100-10C-500UG
Human Gastrin I amide	Biomol	Cay31535-1
Y-27632 (ROCK inhibitor)	Biomol	Cay10005583
Tetracycline	Roth	HP63.1
Prostaglandin E2	Tocris	2296
CHIR99021	BioMol	Cay13122-5
RSPO1 conditioned medium	In-house preparation	
WRN conditioned medium	In-house preparation	
di-Sodium hydrogen phosphate (Na ₂ HPO ₄)	Merck	
Potassium dihydrogen phosphate (KH ₂ PO ₄)	Merck	1.04873
Sodium chloride (NaCl)	Roth	0601.2
Potassium chloride (KCl)	Roth	HN02.3
D(+)-Saccharose (Sucrose)	Roth	9097.1
D-Sorbitol	Roth	6213.1
DTT (dithiothreitol)	Thermo Scientific	R0861
Ethylenediamine tetraacetic acid disodium salt dihydrate (EDTA)	Roth	8043.2
D(+)-Glucose water-free	Sigma Aldrich	G-0750
L-Glutamine	Merck	1.00289.000
Coccarboxylase (Thiamine pyrophosphate)	Sigma Aldrich	T8637-5G
Iron (III)-nitrate-Nonahydrate	Sigma-Aldrich	157740
Thiamine hydrochloride (Vitamin B1)	Sigma Aldrich	95160
Nicotine amide adenine dinucleotide (NAD) free acid	Merck	10127965001
Hydrochloric acid (HCl)	Roth	P074.1
Chloramphenicol	Roth	56-75-7
Glycerol	Geyer Th. GmbH & Co. KG	A2926
Tris base	PanReac AppliChem ITW Reagents	A1086,1000
NaOH	Carl Roth	9356.1
30% Acrylamide/Bis Solution (ROTIPHORESE) 37.5:1	Carl Roth	3029.1

Ammonium persulphate (APS)	Sigma Aldrich	248614
TEMED	Carl Roth	2367.1

2.1.5. Buffer, solutions and medium

Table 6: Composition of ADF++ and ADF+++ medium

Component	ADF++	ADF+++
Advanced DMED/F12	+	+
GlutaMax (100×)	1×	1×
HEPES (1 M)	10 mM	10 mM
P/s	-	5%

Table 7: Human colon organoids medium

Component	Final concentration	Supplier	Catalog number
Wnt-3A conditioned medium	50%	In-house production	–
R-spondin-1 conditioned medium	25%	In-house production	–
Penicillin/Streptomycin (10,000 U/ml)	100/100 U/ml	Gibco	15140-122
N2 supplement (100X)	1X	Gibco	17502048
B27 supplement, minus vitamin A (50X)	1X	Gibco	12587001
N-acetylcysteine	1.25mM	Sigma-Aldrich	A9165
Human epidermal growth factor (hEGF)	50ng/ml	Peptotech	AF-100-15
Human noggin	150ng/ml	Peptotech	120-10C
Human gastrin	10nM	Sigma-Aldrich	G9145
A-83-01 (TGF- β inhibitor)	1 μ M	Cayman	Cay9001799-10
Nicotinamide	10mM	Sigma-Aldrich	N0636
CHIR99021	3 μ M	BioMol	GSK3 inhibitor, Cat. No. Cay13122-5
Prostaglandin E2 (PGE2)	10 μ M	Tocris Biosciences	2296
Y-27632 dihydrochloride (ROCK inhibitor)	9 μ M	Cayman	10005583
SB202190 (p38 inhibitor)	10 μ M	Sigma-Aldrich Chemie GmbH	S7067-5MG

Table 8: Human colon polyp medium

Organoid BMP Medium	Final concentration.	Supplier	Catalog number
---------------------	----------------------	----------	----------------

Adv. DMEM/F12			
B27/NCS21	2%	Gibco	12587001
Nicotine amide (NIC)	10 mM	Sigma-Aldrich	N0636
N2	1%	Gibco	17502048
N-acetyl cysteine	1.25 mM	Sigma-Aldrich	A9165
Human EGF	100 ng/ml	Peptrotech	AF-100-15
A83-01 (TGF- β inhibitor)	0.5 μ M		
PGE2	10 μ M	Tocris Biosciences	2296
Gastrin	10 nM	Sigma-Aldrich	G9145
RHOK I (Rock) Inhibitor	10 μ M	Cayman	10005583
GlutaMax	1x	Gibco™	35050061
HEPES	10 mM	Gibco™	15630080
SB202190 (p38 inhibitor)	10 μ M	Sigma-Aldrich Chemie GmbH	S7067-5MG
P/S	1x	Gibco	15140-122
BMP4 (120-05ET)	10 ng/ml	PeptoTech GmbH	120-05ET

Table 9: 5 x M9-Salt

Na ₂ HPO ₄ x2H ₂ O	64g
KH ₂ PO ₄	15g
NaCl	2,5g
NH ₄ Cl	5g
Dd water	Adjust in 1L

Table 10: 1L 1X M9 Medium

1 Liter 1xM9_CAP	
5 x M9 Salze	200ml
1M MgSO ₄	2ml
20% Glucose	20ml
1M CaCl ₂	0.1ml
AS solution	100ml

Table 11: 100ml AS solution for colibactin production of *pks+* *E. coli*

2 Materials and Methods

2.1 Materials

	AS	Percentage / %	5g/1 liter
1	3,6%Arg	3,60	0,17
2	21,1%Glu	21,10	1,02
3	2,7% His	2,70	0,13
4	5,6% Ile	5,60	0,27
5	8,4%Leu	8,40	0,40
6	7,5%Lys	7,50	0,36
7	4,6% Phe	4,60	0,22
8	9,9%Pro	9,90	0,48
9	4,2%Thr	4,20	0,20
10	1,1%Trp	1,10	0,05
11	6,1%Tyr	6,10	0,29
12	5%Val	5,00	0,24
13	4%Asn	4,00	0,19
14	4%Ala	4,00	0,19
15	4%Met	4,00	0,19
16	4%Gly	4,00	0,19
17	4%Cys	4,00	0,19
18	4% Ser	4,00	0,19
	In total	103,80	5,00

Table 12: General buffer and solutions

Name	Components	Conc.
LB medium	NaCl Tryptone Yeast extract dd H ₂ O, pH 7.4	5g/l 10g/l 5g/l
LB plates	To LB medium add agar Boil and cool down to 45-50°C. Pour into petri dishes.	1.5% (w/v)
MacConkey agar plates	MacConkey agar powder dd H ₂ O	50g 1000ml

2 Materials and Methods

2.1 Materials

	Boil and cool down to 45-50°C. Pour into petri dishes.	
EDTA	Disodium ethylenediaminetetraacetate (pH 8.0)	0.5M
5X Chelating solution	Na ₂ HPO ₄ KH ₂ PO ₄ NaCl KCl Sucrose D-Sorbitol H ₂ O	28mM 40mM 480mM 8mM 220mM 274mM
1X Chelating buffer	5X Chelating solution EDTA DL-Dithiothreitol (DTT) H ₂ O	1X 10mM 0.5mM
Alkaline Running buffer	NaOH EDTA	500mM 50mM
Neutralizing solution	Tris HCl	1M Adjust PH
5X TBE buffer	Tris base Boric acid EDTA	54g 27.5g 0.5M
loading buffer	6X loading buffer Running buffer	100µl 100µl
2% aklaine gel	agarose NaCl 50mM EDTA water	2% 50mM 1mM Adjust volume
6× Laemmli	Tris-HCl pH 6.8 SDS Glycerol Bromophenol blue	0.5 M 0.35 M 30%(v/v) 0.175mM

2× Laemmli buffer	6× Laemmli	0.9ml
	β-mercaptoethanol	0.1ml
	MilliQ H ₂ O	2ml
10× running buffer	Tris	0.25M
	Glycine	1.92M
	SDS	1%
1× running buffer	10× running buffer	100ml
	water	900ml

2.1.6. Kits

Table 13: Kits

Name	Application	Supplier
QIAprep Spin Miniprep Kit (50)	Plasmid DNA extraction kit	QIAGEN
Click-iT™ EDU Cell Proliferation Kit	proliferation assays fluorescence microscopy	Thermofisher
Quick-DNA Miniprep	Genomic DNA extraction	Zymo Research
QIAprep Spin Miniprep Kit (250)	Plasmid prep kit	QIAGEN
LunaScript® RT SuperMix Kit (E3010)	cDNA synthesis	NEB
Direct-zol RNA Microprep Kit	R2062 Direct-zol RNA Microprep	Zymo research

2.1.7. S-oligos (Phosphorothioate Oligonucleotides)

Table 14: Primer. All primers were purchased from Sigma and diluted to 10 μM. FW: forward; RV: reverse.

Oligo Name	Sequence (5' - 3')
FW: AAATTT	G*T*GCTGCCGGCAGGCGCGAGCCGAAATTTGCCGTGCTGCCCCTGGCGGAGCTG*C*A
FW: AATATT	G*T*GCTGCCGGCAGGCGCGAGCCGAATATTGCCGTGCTGCCCCTGGCGGAGCTG*C*A
RV: AAATTT	T*G*CAGCTCCGCCAGGGGCAGCACGGCAAATTTTCGGCTCGCGCCTGCCGGCAGC*A*C
RV: AATATT	T*G*CAGCTCCGCCAGGGGCAGCACGGCAATATTCGGCTCGCGCCTGCCGGCAGC*A*C
FW: AAATTG	G*T*GCTGCCGGCAGGCGCGAGCCCAAATTTGCCGTGCTGCCCCTGGCGGAGCTG*C*A
RV: AAATTG	T*G*CAGCTCCGCCAGGGGCAGCACGGCCAATTTGGGCTCGCGCCTGCCGGCAGC*A*C

FW: AAAATG	G*TVGCTGCCGGCAGGCGCGAGCCCAAATGGCCGTGCTGCCCCTGGCGGAGCTG*C*A
RV: AAAATG	T*G*CAGCTCCGCCAGGGGCAGCACGGCCATTTTGGGCTCGCGCCTGCCGGCAGC*A*C
FW: AATTGG	G*T*GCTGCCGGCAGGCGCGAGCCGAAGGTTGCCGTGCTGCCCCTGGCGGAGCTG*C*A
RV: AATTGG	T*G*CAGCTCCGCCAGGGGCAGCACGGCAACCTTCGGCTCGCGCCTGCCGGCAGC*A*C
FW: AAACCT	G*T*GCTGCCGGCAGGCGCGAGCCGAAACTTGCCGTGCTGCCCCTGGCGGAGCTG*C*A
RV: AAACCT	T*G*CAGCTCCGCCAGGGGCAGCACGGCAAGTTTCGGCTCGCGCCTGCCGGCAGC*A*C
FW: AGGCCT	G*T*GCTGCCGGCAGGCGCGAGCCGAGGCCTGCCGTGCTGCCCCTGGCGGAGCTG*C*A
RV: AGGCCT	T*G*CAGCTCCGCCAGGGGCAGCACGGCAGGCCTCGGCTCGCGCCTGCCGGCAGC*A*C
FW: AAGCCT	G*T*GCTGCCGGCAGGCGCGAGCCGAAGCCTGCCGTGCTGCCCCTGGCGGAGCTG*C*A
RV: AAGCCT	T*G*CAGCTCCGCCAGGGGCAGCACGGCAGGCCTCGGCTCGCGCCTGCCGGCAGC*A*C
FW: GAGCCT	G*T*GCTGCCGGCAGGCGCGAGCCGAGCCTGCCGTGCTGCCCCTGGCGGAGCTG*C*A
RV: GAGCCT	T*G*CAGCTCCGCCAGGGGCAGCACGGCAGGCCTCCGGCTCGCGCCTGCCGGCAGC*A*C
FW: TTAAAA	G*T*GCTGCCGGCAGGCGCGTGCCGTTTAAAGCCGGGCTTCCCCTGGCGGAGCTG*C*A
RV: TTAAAA	T*G*CAGCTCCGCCAGGGGAAGCCCGGCTTTAAACGGCACGCGCCTGCCGGCAGC*A*C
FW: TTTTTT	G*T*GCTGCCGGCAGGCGCGTGCCGTTTTTTGCCGTGCTGCCCCTGGCGGAGCTG*C*A
RV: TTTTTT	T*G*CAGCTCCGCCAGGGGCAGCACGGCAAAAAACGGCACGCGCCTGCCGGCAGC*A*C
FW: AAAAAAT	G*T*GCTGCCGGCAGGCGCGAGCCGAAAAATGCCGTGCTGCCCCTGGCGGAGCTG*C*A
RV: AAAAAAT	T*G*CAGCTCCGCCAGGGGCAGCACGGCATTTTTCGGCTCGCGCCTGCCGGCAGC*A*C
FW: AATTTA	G*T*GCTGCCGGCAGGCGCGAGCCGAATTTAGCCGTGCTGCCCCTGGCGGAGCTG*C*A
RV: AATTTA	T*G*CAGCTCCGCCAGGGGCAGCACGGCTAAATTCGGCTCGCGCCTGCCGGCAGC*A*C
FW: TAAATT	G*T*GCTGCCGGCAGGCGCGAGCCGTAAATTGCCGTGCTGCCCCTGGCGGAGCTG*C*A
RV: TAAATT	T*G*CAGCTCCGCCAGGGGCAGCACGGCAATTTACGGCTCGCGCCTGCCGGCAGC*A*C
FW: TAAAAA	G*T*GCTGCCGGCAGGCGCGAGCCGTAAAAAGCCGTGCTGCCCCTGGCGGAGCTG*C*A
RV: TAAAAA	T*G*CAGCTCCGCCAGGGGCAGCACGGCTTTTTACGGCTCGCGCCTGCCGGCAGC*A*C
FW: AAAATC	G*T*GCTGCCGGCAGGCGCGAGCCGAAAAATCGCCGTGCTGCCCCTGGCGGAGCTG*C*A
RV: AAAATC	T*G*CAGCTCCGCCAGGGGCAGCACGGCGATTTTCGGCTCGCGCCTGCCGGCAGC*A*C
FW: GAAATT	G*T*GCTGCCGGCAGGCGCGAGCCGAAATTGCCGTGCTGCCCCTGGCGGAGCTG*C*A
RV: GAAATT	T*G*CAGCTCCGCCAGGGGCAGCACGGCAATTTCCGGCTCGCGCCTGCCGGCAGC*A*C
FW: AATTTT	G*T*GCTGCCGGCAGGCGCGAGCCGAATTTCCGGGTGCTGCCCCTGGCGGAGCTG*C*A
RV: AATTTT	T*G*CAGCTCCGCCAGGGGCAGCACGGCGAAATTCGGCTCGCGCCTGCCGGCAGC*A*C
FW: GAATTA	G*T*GCTGCCGGCAGGCGCGAGCCGGAATTAGCCGTGCTGCCCCTGGCGGAGCTG*C*A

RV: GAATTA	T*G*CAGCTCCGCCAGGGGCAGCACGGCTAATTCCGGCTCGCGCTGCCGGCAGC*A*C
------------	--

2.1.8. Plasmid

Table 15: Plasmids

Strains	Description	CDM inserted
H713	pfdA8	
H5196	pMW1280_Kan	pfdA8+50bpOligo_mitMotif
H5197	pMW1281_Kan	pfdA8+50bpOligo_ohneMotif
E44	Dh2/6	
H5210	pMW1282_Kan	pfdA8+50bpOligo_mitAAAATT_2x_MM
H5211	pMW1283_Kan	pfdA8+50bpOligo_AAATTT_2x_MM
H5212	pMW1284_Kan	pfdA8+50bpOligo_AATATT_2x_MM
H5213	pMW1285_Kan	pfdA8+50bpOligo_AAATTT_1x_OM
H5214	pMW1286_Kan	pfdA8+50bpOligo_AAAATT_1x_OM
H5215	pMW1287_Kan	pfdA8+50bpOligo_AATATT_1x_OM
H5216	pMW1288_Kan	pfdA8+50bpOligo_AAATT_1x_OM
H5217	pMW1289_Kan	pfdA8+50bpOligo_AAAAT_1x_OM
H5218	pMW1290_Kan	pfdA8+50bpOligo_AAGGTT_1x_OM
H5219	pMW1291_Kan	pfdA8+50bpOligo_AAAGTT_1x_OM
H5220	pMW1292_Kan	pfdA8+50bpOligo_AAACCT_1x_OM
H5221	pMW1293_Kan	pfdA8+50bpOligo_AGGCCT_1x_OM
H5222	pMW1294_Kan	pfdA8+50bpOligo_AAGCCT_1x_OM
H5223	pMW1295_Kan	pfdA8+50bpOligo_GAGCCT_1x_OM
H5224	pMW1296_Kan	pfdA8+50bpOligo_TTTAAA_1x_OM
H5225	pMW1297_Kan	pfdA8+50bpOligo_TTTTTT_1x_OM

2.1.9. Antibodies

Table 16: Primary antibodies

Target	Host	Reactivity	Application (Dilution)	Supplier, Cat#
γ H2A.X	Rabbit	Human	WB (1:1000)	Cell signaling, 2577
β -actin	Mouse	Human	WB (1:1000)	Sigma, A5441

Claudin-1	Rabbit	Human	IF (1:200) WB (1:1000)	Cell signaling, 13255
Occludin	Rabbit	Human	IF (1:200) WB (1:1000)	Cell signaling, 91131
ZO-1	Rabbit	Human	IF (1:200) WB (1:1000)	Cell signaling, 13663

Table17: Secondary antibodies and antibody conjugates

Host	Reactivity	Labeling/conjugation	Application (Dilution)	Supplier, Cat#
Horse	Mouse	HRP	WB (1:10,000)	Cell signaling, 7076
Goat	Rabbit	HRP	WB (1:10,000)	Cell signaling, 7074
Donkey	Rabbit	Alexa Fluor 488	IF (1:200)	Jackson ImmunoResearch, 711-545-152
	Phalloidin	Alexa Fluor 647	IF (1:200)	Abcam, ab176759

2.1.10. Consumables

Table 18: Consumables

Name	Application	Supplier
0.5-2.0 ml tubes	General	Sarstedt
12-/24-/96-well plate	Primary cell culture	TPP
12 mm coverslips	IF	Marienfeld
Falcon tubes (15/50 ml)	General	Sarstedt
Filter tips (10/20/200/1000 µl)	Crypts isolation	Sarstedt
Petri dish	LB Agar plates	Sarstedt
Serological pipette (5/10/25 ml)	General	Sarstedt
Pipette tips (20/200/1000 µl)	Pipetting	Sarstedt
Storage Cryogenic Tubes	Organoids freezing	Thermo Scientific
Pasteur pipette	Liquid handling	VWR
Glass spreader	Bacteria culture	Carl Roth
Microscope slides	IF	Marienfeld
MilliCell® Cell culture inserts (0.4 µm / 12 mm)	Primary cell culture	Merck Millipore (PIHP01250)

PCR tube	DNA digestion	Carl Roth
Metal inoculation loop	Bacteria culture	VWR
Cuvettes	OD measurement	Sarstedt
T75/T150 culture flasks	WRN/RSP01 conditioned medium culture	TPP
Filter attachment bottle	General	LABSOLUTE

2.1.11. Lab instrumentation

Table 19: Lab instrumentation

Name	Application	Supplier
MaxQ 4000 Benchtop Orbital Shakers	Bacterial liquid culture	Thermo Scientific
SORVALL ST 16R Centrifuge	Primary cell culture	Thermo Scientific
VACUSAFE	Aspiration system	Integra
SAFE 2020	Primary cell culture laminar flow box	Thermo Scientific
Hera Safe Cell 150	Primary cell incubator	Thermo Scientific
Heraeus Hera Safe	Microbiology laminar flow box	Thermo Scientific
SUB Aqua Pro water bath	Pre-warming of samples, medium	Grant
Spectrophotometer OD600	OD measurement	IMPLEN
Micro Star 21R centrifuge	DNA isolation, et al.	VWR
AllegraTM 6R Centrifuge	E. coli infection	Beckman Coulter
ThermoMixer C	Heating protein samples	Eppendorf
Mini-PROTEAN Tetra System	SDS-PAGE running chamber	Bio-Rad
Mini-PROTEAN Tetra Handcast System	SDS-PAGE	Bio-Rad
PowerPacTM Basic	Power supply	Bio-Rad
WT12 Shaker	Washing gels	Biometra
Roller mixer SRT6	Sample agitation	Stuart
Accumet AE150 pH meter	pH measurement and adjustment	Fisher Scientific
Balances OEM4202i	Weighing	neoLab Line
MR Hei-Standard	Magnet roller	Heidolph
ChemiDoc Imaging System	Gel developing system	Bio-Rad
Infinite M Plex	DNA concentration measurement	TECAN
T100 Thermal Cycler	DNA digestion	Bio-Rad
RVL-100-M revolve microscope	Immunofluorescence microscopy	Discover ECHO

IX50-S8F inverted microscope	Light microscopy	Olympus
Colony counter SC6+	Colony counter	Stuart
Nalgene Mr. Frosty	Freezing of primary cell stocks	Thermo Scientific
Owl EASYCAST B2-BP	Gel electrophoresis	Bio-Rad
Avati JXN-30	Gentrifuge	BECKMAN COULTER
Concentrator plus	Concentrate DNA samples	Eppendorf

2.1.12. Software

Table 20: Software and databases

Name	Application	Supplier
GraphPad Prism 10.1.0	Data processing	GraphPad
Illustrator 2025	Image assembly	Adobe
Affinity designer 2.6.2	Image assembly	Serif
Fiji Version	Image processing	ImageJ
NCBI Primer-BLAST	Primer design tool	https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome
PowerPoint 2019	Presentation	Microsoft
Word 2019	Daily documentation	Microsoft
ECOH	Image processing	ECOH
SciPy v1.11.4	Data processing	SciPy
Seaborn v0.13.0	Data processing	Seaborn
Matplotlib v3.8.2	Data processing	matplotlib
Logomaker v0.8	Data processing	logomaker
Ggplot2 3.5.0	Data processing	Ggplot2
Maftools v.16.0	Data processing	Maftools
R v4.3.1	Data processing	R
HOMER v5.1	Data processing	HOMER

2.2. Methods

2.2.1 Bacteria cultures and Growth Conditions

The bacterial strains used in the study include two recombinant strains, *E. coli* H5194, *E. coli* H5195 (Nougayrède et al., 2006), and two wild type strains, *E. coli* M1/5 (Wallenstein et al., 2020) and *E. coli* Nissle 1917 (Massip et al., 2019). *E. coli* H5194 and H5195 are resistance to chloramphenicol and *E. coli* H5194 contain pks⁺ island. The wild type *E. coli* M 1/5 is a B2 pks⁺ human fecal isolate that have a rpsLK42R mutation, which renders the strain streptomycin-resistant (Wallenstein et al., 2020). *E. coli* Nissle 1917 strain has been used as a probiotic and does not contain antibiotic resistance.

All strains were grown at 37°C for 18 h with shaking at 250 rpm in Luria-Bertani Miller broth (LB) (Becton, Dickinson and Company - BD, Franklin Lakes, NJ, USA) under aerobic conditions. Chloramphenicol (20 µg/mL) or streptomycin (20 µg/mL) was added to the medium when required. The cultures were stored at -80°C in a -86°C deep freezer (Forma Scientific) in the presence of 50% (vol/vol) glycerol. For standard procedures, each strain was unstocked in LB agar plates with antibiotics, and propagated by sub-culturing a single colony in LB broth with antibiotics before each experiment. Unless otherwise specified, fresh bacteria suspensions were prepared by harvesting at 10,000× g for 3 min and resuspending 1 mL each bacteria culture on sterile M9 medium with Casamino Acids (Wallenstein et al., 2020) without antibiotics to subsequently subculture 5% (vol/vol) into 20 ml M9_Cap broth. The sub-cultured bacteria were incubated for 2 hours at 250 rpm and 37°C under aerobic conditions until an optical density (OD600) of 0.6 to 0.8 was reached.

2.2.2 Plasmids generation and extraction

Plasmids containing the different CDMs were generated by inserting 55 bp fragments containing only one potential colibactin targeting site (5'-AAATTT-3', 5'-AAAATT-3', 5'-AATATT-3', 5'-AAATT-3', 5'-AAAAT-3', 5'-AAGGTT-3', 5'-AAAGTT-3', 5'-AAACTT-3', 5'-AGGCCT-3', 5'-AAGCCT-3', 5'-GAGCCT-3', 5'-TTTAAA-3', 5'-TTTTT-3') within the restriction's sites of BamHI on the pMW1274.A plasmid. *E. coli* E44 was used as host for recombinant work and all cloning was performed according to standard procedures.

Therefore, plasmid digestion with BamHI generates a 302 bp fragment containing the inserted CDM and a bigger 1485bp fragment. To measure the DNA sequence specificity of colibactin, only the inter-strand crosslink generated in the small fragment (shifted up to 604bp in denaturing gels) was taken into consideration. A similar rationale was used to design the

plasmid used as internal control, whose digestion with BamHI resulted in a 450 bp small fragment migrating at 900bp size when crosslinked.

Plasmid extraction was done by culturing the *E. coli* strains containing the plasmids with the 13 different colibactin targeting sites in 50 mL LB broth (BD) with 50 µg/mL Kanamycin in an Erlenmeyer glass flask (250 mL) for 18 h at 37°C. Cells were collected by centrifugation (10,000 x g, 10 min, 4°C) and total plasmid DNA was isolated using the QIAfilter Plasmid Purification Kit (Qiagen, Venlo, Netherlands) following the instructions of the manufacturer. DNA concentrations were determined on TECAN Infinite® 200 PRO using NanoQuant Plate™ (Männedorf, Switzerland) and adjusted to 100 ng/µl with nuclease free water (Invitrogen, Waltham, Massachusetts, USA).

2.2.3 Co-incubation of bacteria and plasmid DNA

An aliquot of 150 µl of the corresponding bacteria suspension was transferred to a 96 well round bottom plate, and left to stand for 1 h and 30 minutes at 37°C. Subsequently, 3 µl of 50 mM EDTA and 1 µg plasmid DNA were added to the wells without mixing and the plate was incubated for additionally 10, 20, 40 and/or 80 min. In the assays where internal control was used 0.5 µg test plasmid DNA and 0.5 µg internal control plasmid were added. After the bacteria-DNA co-incubation time, the plasmid DNA was recovered from the suspension by centrifugation at 7500 x g for 1 min

2.2.4 Cofactors influence in colibactin production and activity

To determine the influence of five cofactors, previously reported in the literature, on colibactin production, *E. coli* H5195 was grown on M9_Cap media in the presence of FeSO₄, Iron II, FeCl₃, Iron III, Spermidine (Carl Roth, Karlsruhe, Germany), Inulin from chicory and Fructooligosaccharides, FOS (Sigma-Aldrich, Munich, Germany). Additionally, CuCl₂ was included to the experiment and to determine the influence of such cofactors not only in the genotoxin production but in its activity, the cofactors were added instead to the reaction together with the plasmid. Three different concentrations of FeSO₄ (Iron II) and FeCl₃ (Iron III), 25 µM, 50 µM and 100 µM were used for the colibactin- production and colibactin activity test. The assays with Spermidine, included four concentrations, 5 µg/ml, 10 µg/ml, 20 µg/ml and 40 µg/ml and for Inulin and Fructooligosaccharides 5 mg/ml, 10 mg/ml, 20 mg/ml and 40 mg/ml. To test the effects of copper on colibactin production 5 µM, 10 µM, 20 µM, 40 µM, 50 µM, 60 µM were used while to study the influence on colibactin activity 1.25 µM, 2.5 µM, 5 µM, 10 µM

and 20 μ M were set. Additionally, to discard any toxic effect of the cofactors on the bacteria growth, all the treatments were plated on LB agar after adjusting the optical density.

2.2.5 Plasmid purification and digestion

Plasmid DNA was purified using QIAquick spin PCR Purification Kit (Qiagen) according to the manufacturer instructions. Lastly, the plasmid was digested by mixing 1.2 μ l 10X FastDigest buffer without dye (Thermo Scientific, Waltham, MA, USA), 1 μ l FastDigest BamHI enzyme (Thermo Scientific) and 400 ng plasmid DNA ($\approx$ 10 μ l) per reaction and incubating at 37°C in T100TM Thermal Cycler for 1 h.

2.2.6 S-oligos generation and visualization by native polyacrylamide gel electrophoresis

S-oligos were generated 55bp contains potential colibactin targeting site (5'-AAATTT-3', 5'-AAAATT-3', 5'-AATATT-3', 5'-AAATT-3', 5'-AAAAT-3', 5'-AAGGTT-3', 5'-AAAGTT-3', 5'-AAACTT-3', 5'-AGGCCT-3', 5'-AAGCCT-3', 5'-GAGCCT-3', 5'-TTTAAA-3', 5'-TTTTT-3'). Apart for the targeting site, all the oligos contains the same sequence, e.g. CDM-AAAATT,

FW:5'G*T*GCTGCCGGCAGGCGCGAGCCGAAAATTGCCGTGCTGCCCCTGGCGGAGCT
G*C*A 3',

RV:5'T*G*CAGCTCCGCCAGGGGCAGCACGGCAATTTTCGGCTCGCGCCTGCCGGCAGC
*A*C 3'

S-oligos were annealed with the following PCR program: 37°C for 30 min, 95°C for 5 min and then ramp down to 25°C at (5°C/min). After the annealing, the duplex oligonucleotides are ready to be used or can be stored. Successful annealing of S-oligos was checked by native polyacrylamide 10%) gel electrophoresis. To prepare a 15 ml native polyacrylamide gel, the following components were mixed: 1.5 ml of 10x TBE buffer, 5 ml of 30% acrylamide/bis solution (19:1 ratio), and 8.5 ml of deionized water, resulting in a total volume of 15 ml. All components were thoroughly mixed in a beaker before proceeding with the gel casting.

To complete the preparation of the native polyacrylamide gel, 15 μ l of TEMED was added to the mixture, followed by gentle swirling of the beaker to ensure thorough mixing. Immediately after, 150 μ l of freshly prepared 10% (w/v) ammonium persulfate (APS) was incorporated into the solution, and the mixture was again thoroughly mixed. The acrylamide solution was then poured between the gel plates, and a comb was carefully inserted to create wells for sample

loading. The gel was left to polymerize for 30 minutes to ensure a fully solidified matrix, ready for electrophoresis. After polymerization, the gel is mounted onto the holder, placed in a water bath at 60-70°C, and filled with pre-heated 1x TBE running buffer, ensuring the buffer covers the wells. The gel is pre-run at 60 V for 30 minutes while maintaining the temperature. Samples (20 µl each) and a 50 bp DNA ladder are loaded into the wells, and the gel is run at 60 V for 60 minutes. After the run, the gel is washed twice with MilliQ water, stained with SYBR™ Gold solution, and visualized using the ChemiDoc Imaging System under UV light.

2.2.7 S-oligos co-incubation with *pks*+ *E. coli*

An aliquot of 400 µl of the corresponding bacteria suspension was transferred to a 96 well round bottom plate, and left to stand for 1 h and 30 minutes at 37°C. Subsequently, 3 µl of 50 mM EDTA and 0.5 µg plasmid DNA were added to the wells without mixing and the plate was incubated for additionally 120 min. After the bacteria-DNA co-incubation time, the plasmid DNA was recovered from the suspension by centrifugation at 7500 x g for 1 min and the supernatant was purified using QIAquick spin PCR Purification Kit (Qiagen) according to the manufacturer instructions

2.2.8 Quantification of colibactin-associated genotoxic effect by alkaline agarose gel electrophoresis

Colibactin DNA inter-strand cross-link formation was visualized by alkaline agarose gel electrophoresis. On that account, the products of the enzymatic digestion were mixed with 4 µl alkaline running buffer (50 mM NaOH and 1 mM EDTA) and 4 µl SafeRed™ Loading Dye (ABP Bioscience, Rockville, MD, USA) and loaded into 1% or 2% agarose gels. The gels were made by melting agarose into an Erlenmeyer flask and adding 1 M NaCl and 50 mM EDTA to a final concentration of 50 mM NaCl and 1 mM EDTA. Before loading the samples, the gels were immersed in alkaline running buffer for 30 min and ran at 5V/cm. Afterwards, gels were neutralized by 15 min immersion in distilled water and 15 minutes in 0.5 M Tris-HCl pH 7.5 and stained with SYBR™ Gold (Invitrogen) for 30 min. Lastly, the gels were documented in ChemiDoc Imaging (BIORAD, Hercules, CA, USA) and the ratio of cross-linked/non cross linked small DNA fragments (604 bp) was quantified using the ImageJ software.

2.2.9 Nanopore sequencing

DNA samples were provided to collaborators in University of Helsinki, Finland. After proceeding the samples, we got raw data and do bioinformatic analysis.

2.2.10 Nanopore sequencing to quantify colibactin affinity to different DNA hexanucleotides

Raw reads were processed with Porechop (<https://github.com/rrwick/Porechop>) to remove adapters and split DNA concatemers. Then, clean were aligned to plasmid references (Addgene plasmid #130662 and pMW1274.A) using minimap2 (<https://github.com/lh3/minimap2>). Using Rsamtools (v2.16.0), the resulting alignments were filtered to select the best alignment coming reads shorter than the plasmid, classified as signal_positive by the sequencer, with a mapping quality greater than 20. Then, a random sample of up to 200,000 alignments from both strands were selected to generate pileups at every plasmid position using samtools mpileup (Danecek et al., 2021). Custom python scripts were used to compute the difference in proportion of low quality basecalls (PropQ<10). Peaks were called using Scipy.signal (v1.11.4) find_peaks function. Visualizations were created using seaborn (v0.13.0) and matplotlib (v3.8.2). Sequence logos were created using logomaker (v0.8, <https://github.com/jbkinney/logomaker>).

2.2.11 Re-estimation of the true oncogenic potential of colibactin.

HOMER (<http://homer.ucsd.edu/homer/>) was used to find the position of the top ranked CDMs in the human genome (hg38 UCSC assembly ID) and annotated with canonical transcripts from Gencode v46 based on overlaps. The rsID of pathogenic mutations was downloaded from dbSNP (v155) and mapped to hg38 coordinates using UCSC's Table Browser on *dbSnp155ClinVar* table. Then, mutations located within the position of the top ranked CDMs were extracted and plotted using ggplot2(3.5.0) and maftools (v.16.0) in R (v4.3.1).

The evaluation of mutational rates in CDMs was performed as described previously (Dziubańska-Kusibab et al., 2020). Briefly, somatic variants from TCGA-COAD whole exome sequece dataset were obtained from Genomic Data Commons portal. Samples were classifies based in total SNVs (quartiles 1-4 and otulier) and POLE mutated. Then we computed the normalized mutation frequency as the number of mutations at each hexanucleotide (Canonical with A/T=6, new and random with A/T=5) were divided by the number of occurrences of each

hexanucleotide in the human exome. Differences between the normalized mutation rates at the motifs of interest compared to all other motifs or those with similar AT content were tested using Mann-Whitney U-tests, whose P values were adjusted using the Benjamini-Hochberg method.

For the quantification of clb DSB in the new CDMs, data from (Dziubańska-Kusibab et al., 2020) was used. Briefly, sBLISS is used in order to detect DSBs in Caco-2 cells co-incubated with pks+ *E. coli* strains. The proportion of DSBs at each position of the hexanucleotide was averaged across the 4 independent experiments and used to group DSB profiles in 4 clusters using the K-means algorithm. Profiles from the same cluster members AAAGTT, AAAATC and AAAATT were drawn using R.

2.2.12 Statistical Analysis

The results are presented as mean $\pm$ SEM and statistical significance was determined from the analysis of the results of three independent experiments using the GraphPad Prism 9 software (San Diego, CA, USA). Statistical analyses were performed using GraphPad Prism software 6.0 (GraphPad Software, La Jolla, Calif). Groups were compared by one-way ANOVA, and differences were adjudicated using the post hoc analysis recommended by PRISM. Statistical outliers were determined via the ROUT test ($p < 0.05$) and removed. $p < 0.05$ was considered to be statistically significant unless otherwise indicated.

2.2.13 WRN conditioned medium

The L-WRN (ATCC CRL-3276) cell line stably producing WRN conditional media (including Wnt3a, R-spondin3 and Noggin) respectively, were kindly provided by Prof. Hans Clevers (Hubrecht Institute, Utrecht, Netherlands). Cell lines were thawed and seeded in 75cm³ cell culture flasks with DMEM, high glucose with GlutaMAX Supplement and pyruvate and 10% Heat inactivated FBS. Incubate the flask at 37°C and change the medium after 24 h to L-WRN growth medium supplemented with 0.5 mg/mL geneticin and 0.5 mg/mL hygromycin B. At confluency, cells were rinsed with 10ml PBS and then detached by incubating with trypsin for ~5 minutes and seeded in required number of 175cm³ cell culture flasks. Pre-warmed conditioning medium (without geneticin, hygromycin B and P/S) were added to each T175 flask. After second passaging, the cells were seeded in the T150 flasks at 37°C for 4 days or until the cells become over-confluent.

The medium was discarded and the cell layer were rinsed once with 10 mL of conditioning

medium and discard the rinse. 25 mL of fresh pre-warmed conditioning medium were added and incubated for 24 h at 37°C. The conditioned medium was harvested every 24 hours for 5 days. The conditioned medium was centrifuged at $2,000 \times g$, 4°C for 5 min and the supernatant were filtered through Steritop bottle top filters (0.22µm) into sterile glass bottles. The conditioned medium was aliquoted and stored in -20°C.

Quality control was performed using the HEK 293 SuperTopFlash (ATCC CRL-3249). Quality control was performed by Gudrun Kliem at the Robert-Koch Institute using established protocols¹⁰⁷.

2.2.14 R-spondin1 conditioned medium

293T HA Rspo1-Fc cell line produce R-spondin1 stably. Cells were thawed and seeded in 75cm³ cell culture flasks with DMEM and 10% heat inactivated FBS. Incubate cells in cell culture incubator at 37°C and 5% CO₂ and obtained the cells until a confluency of 75% is reached. Briefly rinse the cell layer with pre-warmed PBS and detached by 3.5 mL Trypsin-EDTA (0.05 %). Incubate at 37°C + 5% CO₂ for 5 mins and observe cells under an inverted microscope until cell layer is dispersed.

Stop trypsin reaction by adding 15 mL conditioned growth medium. Homogenize the cell suspension by 10-times pipetting up and down. Spread cells in a ratio from 1:10 into 2 new flasks and add fresh culture medium to achieve a total volume of 25 ml each. Zeocin is added to the medium at a final concentration of 125µg/ml for selection of Rspo1 producing cells. Incubate at 37°C 5%CO₂ until 70% confluency. Repeat the passage step twice and add cell growth medium without zeocin at passage 3 after defrost. Start harvesting conditioned media after 4 days centrifuge at 300g for 5 minutes and collect. Add 25ml fresh medium to the cells. After 3 days, harvest media, centrifuge at 300g for 5 minutes and collect. Mix the entire collected medium, filter sterile, aliquot and freeze in -20 freezer. Quality control was performed as for WRN.

2.2.15 Human colon crypts isolations from biopsies

Human colon biopsies were obtained from colonoscopies performed at UKSH, with ethical approval granted under permission number AZ B231/98-1/13. The biopsies, approximately 2 mm x 2 mm in size, were preserved in Advanced DMEM medium (Gibco™ #12634028) at 4°C, and crypt isolation was initiated within 2–3 hours post-surgery. To begin, biopsies were

transferred to a 50 ml Falcon tube precoated with 0.1% BSA (A7030-100G) and rinsed three times with 10 ml cold PBS.

The biopsies were then incubated on ice in 10 ml of cold 1X PBS containing 0.05 mM DTT, Anti Fungin (1:1000), 50 µg/ml Gentamycin, 1X P/S (1:100), 50 µg/ml Tetracycline (1:1000), and 10 mM EDTA, with gentle shaking on a roller mixer for 45–60 minutes. After incubation, the supernatant was discarded to minimize contamination, and the biopsies were transferred into a precoated 15 ml Falcon tube. Approximately 3–5 ml of cold 1X PBS with EDTA was added, and the contents were pipetted up and down about 10 times. The biopsies were allowed to settle, and the supernatant was collected into another precoated 50 ml Falcon tube. This process was repeated three times.

The 15 ml tube with the remaining biopsies was filled with 8 ml of 1X PBS, agitated thoroughly for 2 minutes, and the 8 ml supernatant was collected in a 50 ml Falcon tube and centrifuged at 300 g for 5 minutes at 4°C. The resulting pellet was washed with 5 ml of Advanced DMEM/F-12, followed by another centrifugation at 300 g for 5 minutes at 4°C. The crypts were then counted using a hemocytometer and resuspended for seeding at 40 µl per Matrigel drop in a 24-well plate. The Matrigel was allowed to polymerize at 37°C for 30 minutes, after which each well received 500 µl of medium and was incubated at 37°C. For the first three days, fungin and gentamycin were added to the medium, which was refreshed every three days.

2.2.16 Establishment of 2D culture

Before seeding the cells, the filters are carefully placed into the wells of a 24-well plate using forceps, and 100 µl of collagen dilution is added to each filter. Collagen must be kept on ice during preparation, as it polymerizes at room temperature. The collagen-coating solution is calculated per filter, with adjustments based on the total number of filters. For a collagen stock concentration of 5 mg/ml (5 µg/µl) and a target coating of 10 µg/cm², the calculation per filter (0.6 cm²) requires 6 µg of collagen. This corresponds to 6 µl of collagen stock per filter, diluted with 94 µl of sterile water to a final volume of 100 µl per filter. The prepared dilution should be kept on ice until use.

The filters are incubated with the coating solution for 30–60 minutes at 37°C (not exceeding this time). Following incubation, the filters are washed once with pre-warmed PBS. A second PBS wash is performed, and the filters are left submerged in PBS at room temperature until cell seeding.

The organoids were harvested from 3D culture by mechanical shearing and counted using a Neubauer chamber to ensure single-cell visibility. For each filter, 20,000 cell pieces were

prepared in a seeding volume of 200 μ l. The cells were centrifuged at 300 $\times$ g for 5 minutes at 4°C, and the resulting pellet was resuspended in ALI medium supplemented with 2.5 μ l/ml ROCK inhibitor. A total of 200 μ l of the cell suspension was seeded onto each prepared filter, with 400–600 μ l of medium added to the surrounding well, followed by incubation at 37°C and 5% CO₂ for 3 days. After this period, the 2D culture was established. Surrounding medium and top medium were refreshed, and the use of high-concentration Y-27632 dihydrochloride was discontinued. Media in the wells were replaced every 3–4 days. Contamination checks using a small sample of media under a microscope were routinely performed. Filters were ready for harvesting or further experiments after 2 weeks of culture, with passage required every 2–3 weeks to maintain the system.

2.2.17 Passage of colon organoids culture

For organoid processing, 15 ml tubes and pipette tips were first pre-coated with 0.1% BSA in PBS. Tubes were prepared by adding the BSA solution and decanting it, while pipette tips were coated by pipetting the solution up and down. Once pre-coated, tubes were placed on ice, and 6–10 ml of ice-cold ADF++ was added depending on the number of wells, with up to 12 wells per 15 ml tube. Using a 1000 μ l pipette tip, Matrigel domes were scraped from each well, pooling up to 12 wells into a single 15 ml tube.

The tubes were then centrifuged at 500 $\times$ g for 5 minutes at 4°C, and the supernatant and the top Matrigel layer were gently removed, leaving only the cell pellet. If the pellet was not clearly visible, only the supernatant was discarded to avoid losing organoids. The pellet was then resuspended in 500–1000 μ l of TrypLE with 10 μ M Y-27632 dihydrochloride and incubated at 37°C for 7–10 minutes.

To shear the organoids, the suspension was pipetted up and down 5–15 times using pre-coated 1 ml tips, depending on organoid size, taking care to avoid contamination by not recapping the pipette tip. Cells were then washed by adding ADF++ to a total volume of 12 ml, followed by a second centrifugation at 500 $\times$ g for 5 minutes at 4°C. The supernatant and Matrigel layer were removed, leaving the cell pellet, with an optional small amount of supernatant retained for direct Matrigel addition. The pellet was resuspended as needed in ADF++.

Next, the cell pellet was mixed with Matrigel, and 40–50 μ l of the cell-Matrigel suspension was dispensed into each well of a 24-well plate, which was incubated at 37°C for at least 30 minutes to solidify the Matrigel. Finally, 500 μ l of 3D medium (prepared as per the 3D medium table)

was gently added to each well, ensuring the Matrigel dome remained intact. The medium was refreshed every 2–3 days.

2.2.18 Freezing and thawing of organoids

To preserve organoids, begin by mechanically disrupting organoids and Matrigel in 24 well plate. Transfer this mixture to a 15 ml Falcon tube precoated with 0.1 % BSA. The 15 ml falcon tube were filled up to 13ml with cold Advanced DMED F12. Centrifuge at 300 g for 5 minutes at 4°C, then resuspend the resulting pellet in Cryo-SFM freezing medium. 500µl per 3 Matrigel drops of organoids. Place the suspension into a cryovial, position the vial in a freezing container filled with isopropanol, and store overnight at -80°C. Transfer the cryovial to liquid nitrogen after 24 hours for extended storage.

To thaw frozen organoids, place the cryovial in a 37°C water bath until small ice inside. Immediately transfer the contents to a 15 ml Falcon tube with 5 ml of advanced DMEM/F-12, then centrifuge at 300 g for 5 minutes at 4°C. Resuspend the pellet in 2 ml of advanced DMEM/F-12, and repeat centrifugation at 300 g for 5 minutes at 4°C. Finally, resuspend the organoids pellet in the desired volume of Matrigel and seed 45 µl of the cell-Matrigel suspension into each well of a 24-well plate. The Matrigel drop was incubated at 37°C for 30 minutes. Each well was then supplemented with 500µl of medium and incubated at 37°C.

2.2.19 Clonal selection of organoids

The organoids were slightly passaged, serial diluted and seeded in 96-well plate.

For organoid seeding, the pellet from one well of organoids is resuspended in 150 µl of Matrigel, maintaining a 1:3 splitting ratio. Two 50 µl drops of the cell suspension (Matrigel and cells) are distributed into a single well of a 24-well plate and incubated at 37°C for at least 30 minutes to allow the Matrigel to solidify. After solidification, 500 µl of 3D medium is gently added. Based on the number of visible cell clusters or single cells, 10 µl of the 50 µl cell suspension is diluted with 10–50 µl of Advanced DMEM F12, mixed with 50 µl of Matrigel, and 10 µl of this mixture is seeded into a well of a 96-well plate. The cell count is then examined under a microscope. If 10–20 single cells or multiple clusters are present in each Matrigel drop, an additional four wells of the 96-well plate are seeded with 10 µl of the cell/Matrigel mixture. If fewer cells are observed, the dilution step is repeated until the desired cell density per drop is achieved. The plate is then incubated at 37°C for 15 minutes to solidify the Matrigel, followed by the gentle

addition of 100 µl of 3D medium. Finally, 2–4 organoids are selected for further analysis or experimentation.

2.2.20 DNA isolation

The Quick-DNA Miniprep Kit was used to extract DNA. To extract DNA from organoids, organoids from 5-6 wells were collected into a 15 ml Falcon tube, followed by the addition of 500 µl cell recovery solution (Corning BV: 11543560) per three wells. The suspension was incubated in ice for 45-60 minutes to let the organoids settle down and then centrifuged at 500 × g for 5 minutes, after which the supernatant was removed and 500 µl of Genomic Lysis Buffer were added directly to the cell pellet. The pellet was resuspended by vortexing for 4–6 seconds and allowed to stand at room temperature for 5–10 minutes. The mixture was then transferred to a Zymo-spin IICR column in a collection tube and centrifuged at 10,000 × g for 1 minute, after which the collection tube containing the flow-through was discarded. The column was placed in a fresh collection tube, and 200 µl of DNA Pre-Wash Buffer was added, followed by centrifugation at 10,000 × g for 1 minute. Subsequently, 500 µl of g-DNA Wash Buffer was added to the column and centrifuged under the same conditions. The column was then transferred to a clean microcentrifuge tube, and ≥50 µl of DNA Elution Buffer or nuclease-free water was added. After incubating for 2–5 minutes at room temperature, the DNA was eluted by centrifuging at maximum speed for 30 seconds.

2.2.21 DNA sequencing

Whole genome sequencing (WGS) was performed at Novogene, UK using target coverage of 30X.

All tertiary analysis was done using R Core Team (2018), including circlize package ¹⁰⁸. Bioinformatics analysis was performed by A. Quevedo-Olmos.

2.2.22. Western Blot

2.2.22.1. Whole cell lysate preparation

The infection cultures were washed once with 1 ml DPBS. The cultures were harvested in 2 × Laemmli buffer containing β-mercaptoethanol and heated for 5 min at 95°C. The samples were stored at -20°C.

2.2.22.2. SDS-PAGE

Sodium-dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) was performed to separate proteins by their molecular weight. SDS gel consists of a resolution gel to separate proteins and a stacking gel to concentrate the samples. Resolution gel was prepared according to target protein size. Normally, resolution gel is 10%-13.5%, stacking gel is 5%. The protocol of gel preparation is shown in the table below. PageRuler™ was used as a protein marker. 20 µl protein sample was loaded to defined well. The electrophoresis was run with a constant voltage of 80 V for 15 min and later of 120 V for 90 min.

Table 21: SDS gel preparation

Component	5% stacking gel	10% resolution gel	13.5% resolution gel
MilliQ water	3.4 ml	4 ml	2.8 ml
30% Acrylamide mix	1 ml	3.3 ml	4.5 ml
1.5 M Tris-HCl pH 8.8	-	2.5 ml	2.5 ml
0.5 M Tris-HCl pH 6.8	0.5 ml	-	-
10% APS	50 µl	100 µl	100 µl
10% SDS	50 µl	100 µl	100 µl
TEMED	5 µl	10 µl	10 µl

Proteins were separated on the gel and then were transferred to Nitrocellulose (NC) Membrane by electrophoresis. For protein transfer, the wet blot method was used. NC membrane and Whatman paper were cut into 6 cm × 9 cm size and they were soaked in 1 × Transfer buffer. The transfer cassette was assembled on the transparent side in the following order: Black sponge→3 Whatman papers →NC membrane→Gel→3 Whatman papers→Black sponge. The cassette was closed and placed in the Trans-Blot chamber, filled with transfer buffer. The transfer was run at 4°C with a constant voltage of 100 V for 3 h. After transfer, the membrane was stained in Ponceau S solution to check the quality of protein transfer on membrane. The membrane was washed with TBST to wash off Ponceau and blocked with 1× Roti Block at RT for 1 h. Membrane was incubated in primary antibody (diluted in 5% BSA/TBST) at 4°C overnight. On the second day, the membrane was washed 3 times with TBST. Next, the membrane was incubated with secondary antibody at RT for 1 h. After incubation, the membrane was washed 3 times with TBST again, and the target protein on the membrane was developed by ChemiDoc Imaging System. For low expression protein, Select ECL was used on the membrane. For most target protein, Prime ECL was used to develop the signal. ECL

was mixed freshly in a 1:1 ratio and the membrane was incubated with the mixture for 1-2 min in the dark.

2.2.23. RNA and DNA techniques

2.2.23.1. RNA isolation

RNA isolation was performed with RNeasy Mini Kit and RNase-Free DNase Set. Colonic primary cells were washed with 1 ml DPBS. 350 μ l RLT buffer + 10% β -mercaptoethanol (freshly) was added to each well. The cell lysate was pipetted around 5 times to homogenize it and transferred to a 1.5-ml reaction tube. The lysate could be stored at -80°C at this step and it was thawed at RT for later RNA isolation. 350 μ l of 70% ethanol was added to each tube and mixed immediately by pipetting. The cell lysate was transferred to a RNeasy mini spin column placed in a 2 ml collection tube and centrifuged at $\geq 8000 \times g$ for 15 s. The spin column was washed with 350 μ l RW1 buffer, followed by centrifugation ($\geq 8000 \times g$, 15 s). DNase I vial provided by the RNase-Free DNase Set was dissolved in 550 μ l RNase-free water and aliquoted into small stocks. DNase I stock solution was diluted in RDD buffer (DNase I: RDD=1:7). 80 μ l DNase I/RDD mix was dropped onto the column membrane and incubated at RT for 15 min. After the digestion, the column was washed with 350 μ l RW1 buffer, followed by centrifugation ($\geq 8000 \times g$, 15 s). The column was washed with 500 μ l RPE buffer twice, followed by centrifugation ($\geq 8000 \times g$, 15 s). The RNeasy spin column was placed in a new 2 ml collection tube and performed an extra spin at full speed for 1 min to dry the membrane. The column was transferred to a new 1.5 ml reaction tube (supplied) and 22 μ l of RNase-free water was dropped to the membrane. After 2-min incubation at RT, the column was spun for 1 min at $\geq 8000 \times g$ to elute the RNA. RNA measurement was determined by using a TECAN reader at absorbance 260 nm. A 260/280 ratio value of ≤ 2.0 for RNA indicates contamination of the sample with proteins and phenols, and a 260/230 ratio ≥ 2.0 indicates contamination with carbohydrates, salts, or organic solvents. RNA samples were stored at -80°C and otherwise always handled on ice.

2.2.23.2. RNA reverse transcription (cDNA synthesis)

Complementary DNA (cDNA) synthesis was performed with LunaScript® RT SuperMix Kit. The reverse transcription was performed following the manufacturer's protocol. All steps were performed under nuclease-free conditions. The reaction was prepared as described below in PCR tubes:

Table 1: cDNA synthesis reaction preparation

COMPONENTS	20 µl REACTION	FINAL AMOUNT
LunaScript RT SuperMix (5X)	4 µl	1X
RNA Sample	x µl	500ng-1 µg
Nuclease-free Water	16-x µl	

The reaction was incubated in T100 Thermal Cycler with the following steps:

Table 2: Reverse Transcription PCR setting

CYCLE STEP	TEMPERATURE	TIME	CYCLES
Primer Annealing	25°C	5 minutes	1
cDNA Synthesis	55°C	20 minutes	
Heat Inactivation	95°C	1 minute	
Preservation	4°C	∞	

After reverse transcription, the cDNA sample was diluted with nuclease-free water in 1:10-1:5 ratio (according to RNA sample amount) and stored at -20°C for further use.

2.2.23.3. Quantitative reverse transcription polymerase chain reaction (RT-qPCR)

Quantitative reverse transcription polymerase (RT-qPCR) is used to detect and quantify the relative mRNA expression of a target gene. mRNA is first transcribed into cDNA. The cDNA is then used as the template for the quantitative PCR (qPCR) or real-time PCR reaction which is based on the measurement of a fluorescence signal during the elongation phase of the PCR. Primers designed specifically for the gene of interest bind to the template DNA and elongate the cDNA strand. The fluorescence dye SYBR Green binds to the big groove of double-stranded DNA which is present in the elongation phase of the PCR. The fluorescence signal is detected during the elongation phase. The intensity of the fluorescence signal is proportional to the amount of double-stranded DNA. After the PCR is completed, a melting curve analysis is performed to ensure the purity of the PCR product and the quality of the primers and the product. Luna Universal qPCR Master Mix was used for qPCR. All steps were done under nuclease-free conditions. One reaction was performed in 20 µl reaction volume, the composition of a reaction was list in Table 24. Master Mix and cDNA sample were loaded in a 96-well-PCR plate. The plate was covered with the provided firm and centrifuged for 30 sec. Each reaction was measured in duplicate. The setting of StepOne Plus Real-Time System was as the table (Table 25) below:

Table 3: Composition of one qPCR reaction

Component	20 µl reaction	Final concentration
Luna Universal qPCR Master Mix	10 µl	1 ×
Forward primer (10 µM)	0.5 µl	0.25 µM
Reverse primer (10 µM)	0.5 µl	0.25 µM
Template DNA	2 µl	10 ng
Nuclease-free Water	7 µl	

Table 4: Program setting of the qPCR cyclers

Cycle step	Temperature	Time	Cycles
Initial Denaturation	95°C	60 sec	1
Denaturation	95°C	15 sec	45
Extension	60°C	30 sec (+plate read)	
Melt curve	95°C	15 sec	1
	60°C	60 sec	
	95°C	15 sec	

The results were automatically determined using the StepOne Software v2.3. Melt curve analysis was performed to confirm gene specific amplification by a single peak. The data analysis was done using the $\Delta\Delta C_t$ method which gives a relative quantification of a gene of interest by normalization on an endogenous control. In this thesis, the endogenous control was RNA18SN5, uninfected control (NI) was the reference sample. The primary efficiency was calculated. For this, all the samples were pooled together and diluted with water to 1:10 and 1:100. A scatter plot was made by Excel between the average C_t values (y) and the log values (x). A trendline was determined and the slope of the formula was used in the formula below to calculate primer efficiency:

$$\text{Efficiency (\%)} = (10^{\frac{-1}{\text{slope}}} - 1) \times 100$$

The efficiency of primers, endogenous control and reference sample were set in the “Relative quantitation setting” channel. The software calculated the relative expression of target genes automatically. The figures were made according to the RQ value.

2.2.23.4. RNA Sequencing (Novogene)

RNA isolation was performed as described in 2.2.5.1. Gastric primary cells (23a, 42a and 54a) were infected with *H. pylori* P12 WT and Δcgt for 24 h. After RNA isolation, 200ng of every sample was reverse transcribed for pre-test. Then the samples were transported to the company in dry ice.

2.2.24. Immunofluorescence

Immunofluorescence (IF) is a method to visualize proteins by labeling the target protein in two steps with a specific primary antibody and a fluorescence dye coupled secondary antibody.

2.2.24.1. Staining

Cells were seeded on collagen I coated coverslips or filters and cultured for several days until confluent. Cells were washed with DPBS twice and fixed with 4% PFA/PBS at 37°C for 10 min. After fixation, cells were washed with DPBS 3 times and permeabilized with permeabilization buffer (0.1 M glycine, 0.2% Triton X-100, 0.1% Tween 20/DPBS) at RT for 60 min. Later the cells were blocked with blocking buffer (5% BSA, 0.2% Triton X-100, 0.1% Tween 20/DPBS) at RT for 60 min. After blocking, cells were washed with DPBS once and incubated with primary antibody which was diluted in blocking buffer at 4°C overnight. On the second day, cells were washed 3 times with DPBS and incubated with secondary antibody/Phalloidin which were diluted in 5%BSA/DPBS at RT for 60 min. From this step, the plates should always protect from light. Cells were washed 3 times with DPBS to get out of secondary antibody and followed by DAPI (1:1000 diluted in DPBS) staining at RT for 5 min. Cells were washed 5 times with DPBS and once with water in the end. Coverslips were mounted and sealed properly. Samples were stored in dark at 4°C. The method that I used followed the published paper¹⁰⁹.

2.2.24.2. Immunofluorescence Microscopy and image analysis

After IF staining, images were recorded at the immunofluorescence microscope with a 10x, 20x or 60x oil objective using the RVL-100-M revolve microscope (Discover ECHO) and the corresponding ECHO software. Afterward, images were further processed and analyzed using Fiji ImageJ (open source). This software was also used for processing Z-stack with plugin (Bio-Formats).

3. Results

Part I: The analyses of sequence specificity of colibactin on isolated DNA

3.1 Generation of a plasmid for testing colibactin specificity

On average, Colibactin's canonical target sequences match the AAWWTT motif, where 2 adenines in opposite strands separated by 2 other nucleotides (W: A or T) are bound by the genotoxin, creating an inter-strand crosslink (ICL). Although simulations suggest that these sequences create the optimal docking site for colibactin, an *in vitro* validation of colibactin binding to defined sequences is still missing. Therefore, by introducing small variations into the canonical CDM, we designed fifteen CDMs to test, where the 2 adenines are placed at different distances and/or with different central nucleotides, with the hope to identify the positions that determine colibactin binding (**Figure 6**). We used the efficiency of DNA crosslinks, measured using denaturing gel electrophoresis, as a proxy of colibactin affinity. Because the size of the CDM (6nt) is below the resolution limit of the gels, we surrounded the candidate CDMs with GC rich sequences by cloning them into plasmid pfdA8 (**Figure 6A**), so that after BamHI digestion, a 302bp fragment that contains a single copy of CDM to test is released. Successful crosslinking by colibactin creates a band at 604bp (**Figure 6B(e)**), after normalizing its size by the amount of not crosslinked DNA (302bp band), we obtained the crosslink efficiency (ratio 604bp/302bp).

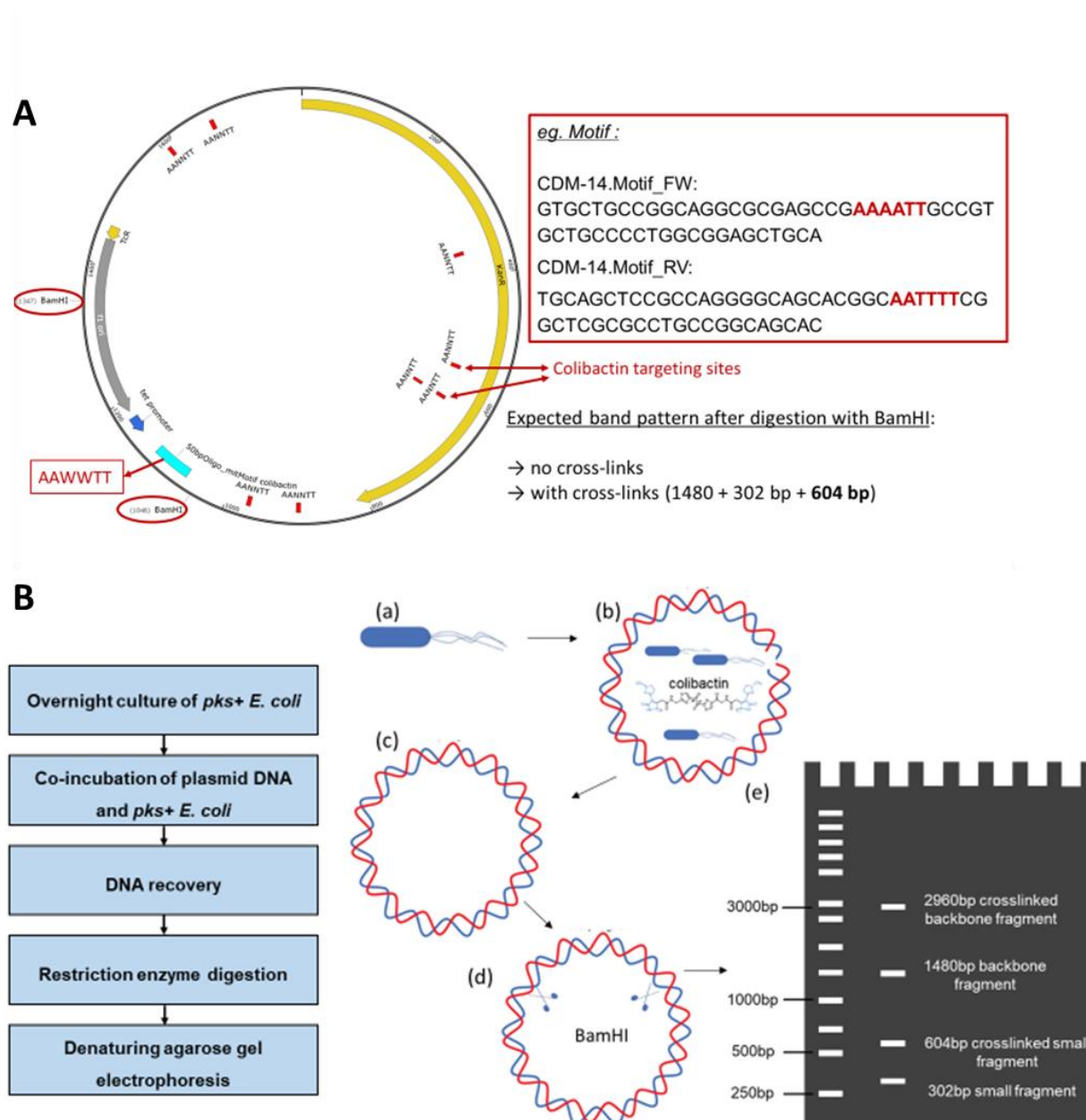


Figure 6- Design of the Plasmid. This figure presents the plasmid design used to study colibactin-induced DNA crosslinking (A) and the corresponding experimental workflow for detection and analysis (B). The plasmid generation were done by M. Sørensen.

(A) This figure illustrates the design of a plasmid specifically engineered for colibactin targeting sites research. The plasmid can be digested with the restriction enzyme BamHI into two fragments: a small fragment (302 bp) containing a single colibactin binding site introduced by a 50 bp oligonucleotide, and a larger backbone fragment (1480 bp) containing additional potential colibactin targeting sites. Upon co-incubation with *pks+* *E. coli* and subsequent restriction digestion, four distinct products are expected: the small fragment (302 bp), the crosslinked small fragment (604 bp), the backbone fragment (1480 bp), and the crosslinked backbone fragment (2960 bp). This design allows the detection and quantification of colibactin-induced inter-strand crosslinks in both the small fragment and the backbone.

(B) Experimental workflow for detecting DNA inter-strand crosslinks. The process involves overnight culture of *pks*⁺ *E. coli*, co-incubation with plasmid DNA, DNA recovery, BamHI restriction enzyme digestion, and separation of crosslinked vs. un-crosslinked products via denaturing agarose gel electrophoresis. Created in <https://BioRender.com>.

3.2 Sequence specificity of Colibactin in plasmid DNA

Although colibactin displayed a marked preference for the canonical motifs AAAATT and AAATTT, it also generated crosslinks at AAAGTT, despite the introduction of a guanine in the central position that disrupts the canonical CDM consensus (**Figure 7C**). Quantification confirmed significantly higher ICL levels for AAATTT ($p < 0.01$), AAAATT ($p < 0.01$), 2×AAAATT ($p < 0.005$), and AAAGTT ($p < 0.05$) compared with the control sequence GCCGTT (CDM–) (Figure 7D). No significant enrichment was observed for the other tested motifs, including the canonical-like AATATT ($p > 0.05$), indicating that not all AT-rich contexts support efficient crosslinking. Importantly, no ICLs were detected when either 2×AAAATT or GCCGTT were co-incubated with a *pks*– *E. coli* strain, confirming that the observed crosslinking activity is specifically attributable to colibactin.

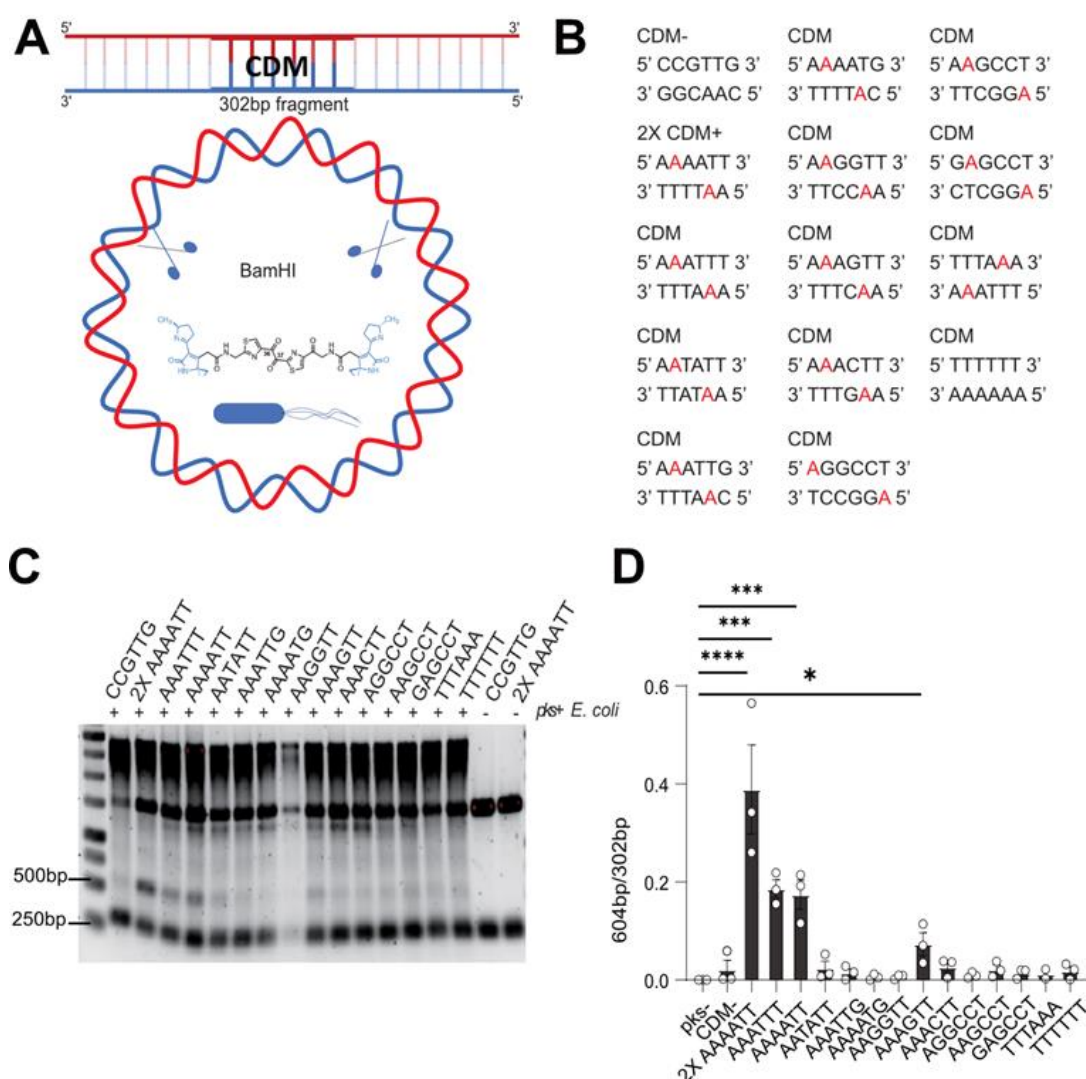


Figure 7- Comparison of colibactin affinity to different plasmid DNA sequences. This figure illustrates the experimental design and results used to assess the sequence specificity of colibactin-induced DNA inter-strand crosslinking across a panel of candidate damage motifs (CDMs) inserted into a plasmid model system. Figure 7C was generated in collaboration with Dr. K. Arellano-Ayala.

(A) The graphic representation of the plasmid was designed to be digested by BamHI into two distinct fragments: a smaller 302 bp fragment and a larger 1485 bp fragment. A 55 bp sequence, containing a single potential colibactin targeting site, was positioned within the smaller BamHI fragment. In contrast, the larger 1485 bp fragment, corresponding to the plasmid backbone, contained multiple potential targeting sites.

(B) The 15 different putative colibactin damage motifs within the small fragment after digestion with BamHI restriction enzyme. The bases in red correspond to the sites where colibactin might induce the inter-strand cross-link formation, according to the chemical constraints of colibactin binding (2 A's on opposite strands, positions 2 and 5).

(C) 1% alkaline gel with the different plasmid-CDMs after 80 minutes co-incubation with pks+ *E. coli* H5194 and digestion with BamHI. The 300 bp fragment corresponds to the digested plasmid without inter-strand cross-link while the 600 bp corresponds to the cross-linked fragment on the indicated CDM.

(D) Quantification of the cross-links by CDM as a ratio between 604bp and 302 bp. Statistical differences were compared to *pks*- *E. coli* treatment by one-way ANOVA and Dunnett's multiple comparisons test where, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$ and **** $p < 0.0001$.

3.3 Kinetics of DNA inter-strand cross-linking in plasmid DNA

To confirm the time dependency of inter-strand cross-link (ICL) formation induced by colibactin, the established *in vitro* assay was performed at multiple co-incubation time points (10, 20, 40, and 80 minutes). Cross-links were visualized on alkaline agarose gels, where the intensity of the 604 bp fragment increased progressively with longer exposure to *pks*+ *E. coli* H5194. For the highest-affinity motifs, AAATTT and AAAATT, ICLs were evident as early as 10 minutes, whereas most motifs only showed detectable crosslinks after 80 minutes. CDM-AAAGTT displayed intermediate kinetics, with crosslinks visible at 40 minutes (**Figure 8A, E**).

Quantification of ICLs confirmed a time-dependent increase across motifs (**Figure 8D**). Significant differences ($p < 0.05$) were observed for CDM-AAATTT, CDM-AATTTT, and CDM-AAAGTT across time points, with 80 minutes showing the highest levels ($p < 0.005$ compared with 10–40 minutes). These results indicate that canonical AT-rich motifs (AAATTT, AATTTT) are rapidly targeted by colibactin, while near-canonical sequences such as AAAGTT require longer exposure for efficient crosslinking.

By contrast, motifs including AAATTG, AAGTTT, AAGCCT, AGGCTC, TTAAAA, and TTTTTT showed consistently low crosslinking, with only weak ICL signals detectable even after 80 minutes. This suggests a clear hierarchy of motif susceptibility, with AAATTT and AAAATT as primary targets, AAAGTT as a secondary target, and other tested motifs largely resistant under the assay conditions.

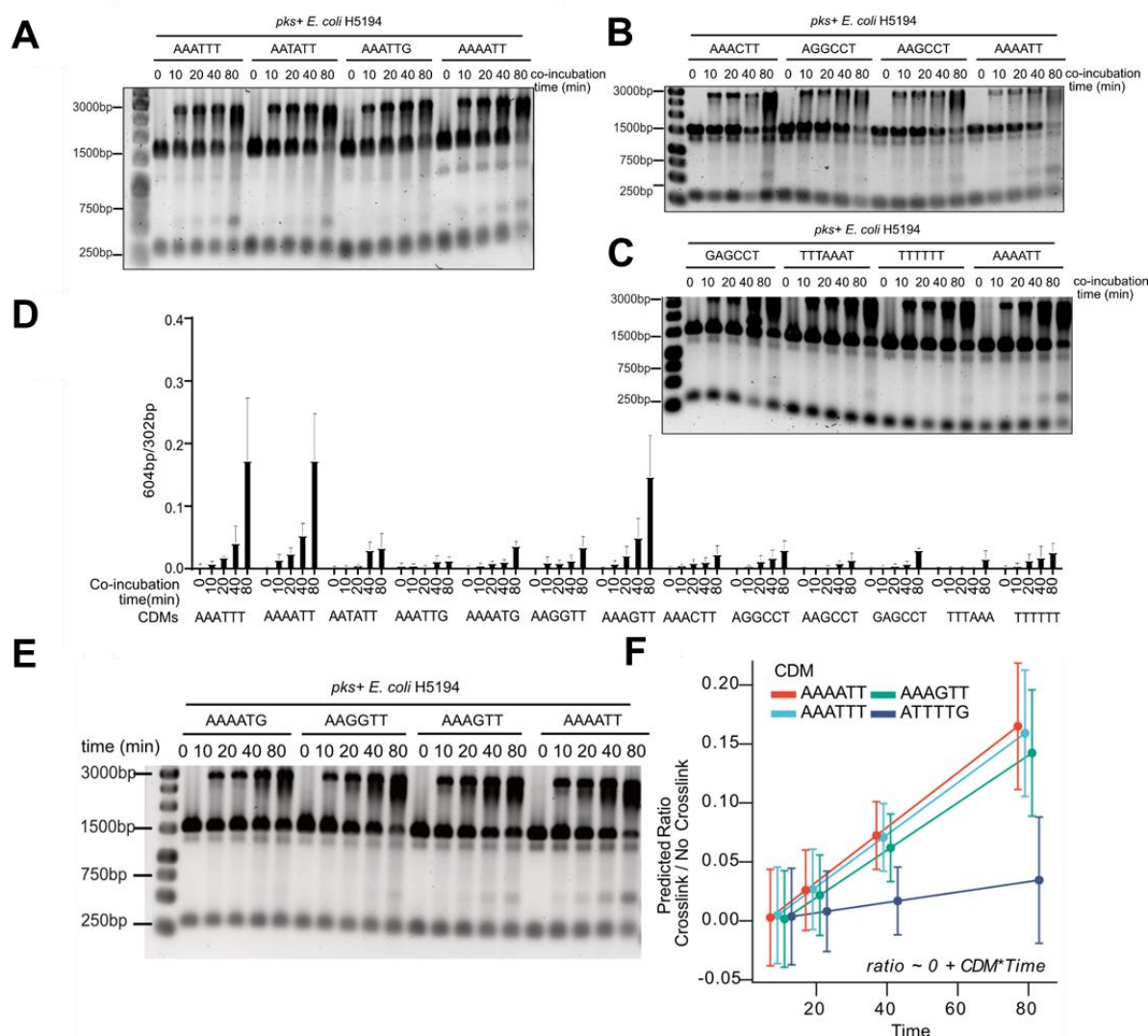


Figure 8- DNA inter-strand cross-link formation by colibactin is time dependent. This figure demonstrates the kinetics of colibactin-induced DNA inter-strand crosslink formation across multiple colibactin damage motifs (CDMs) using a time-course co-incubation with *pks+* *E. coli* H5194. Figure 8A-E were generated in collaboration with K. Arellano-Ayala and figure 8F was made by A. Quevedo-Olmos.

(A) CDM - AAATTT, CDM - AATATT, CDM - AAATTG, CDM - AAAATT, (B) CDM - AAACCT, CDM - AGGCCT, CDM - AAGCCT, CDM - AAAATT and (C) CDM - GAGCCT, CDM - TTAAAT, CDM - TTTTTT, CDM - AAAATT time dependent DNA inter-strand cross-link formation after co-incubation with *pks+* *E. coli* H5194. 1% denaturing agarose gel.

(D) Quantifications of the cross-links by CDM as a ratio between 604bp and 302 bp at 0 10 20 40 80 min.

(E) CDM - AAAATG, CDM - AAAATG, CDM - AAGGT, CDM - AAAGT 1% alkaline gel after 0, 10, 20, 40- and 80-minutes co-incubation with *pks+* *E. coli* H5194 and digestion with BamHI. DNA cross-link induced by colibactin is observed on the gel as a 604 bp fragment and it increase as a function of the time.

(F) Predicted ratio 604bp/302 with 95% Confidence Intervals using linear regression with slopes learned from the quantification of 3 independent experiments of the gel shown in E. AAAAT, AAAT and AAAGT show a

significant increase of colibactin-dependent ICLs after 80 minutes compared with the non-colibactin binding ATTTTG.

3.4 Specificity of colibactin from wild type vs. recombinant *E. coli*

Since *pks*⁺ *E. coli* H5194 and *pks*⁻ *E. coli* H5195 are recombinant strains, we validated our findings using two wild-type colibactin-producing strains: *pks*⁺ *E. coli* Nissle 1917 (a probiotic isolate) and *pks*⁺ *E. coli* M1/5 (a clinical fecal isolate). CDM- (GCCGTT), CDM+/AAAATT, CDM-AAATTT, and CDM-AATTTT were co-incubated with both strains across four time points, and the resulting ICL profiles were consistent with those obtained using the recombinant strain (**Figure 9**). These results confirm that the preferential crosslinking activity at AAAATT, AAATTT, and AATTTT is reproducible in wild-type *pks*⁺ *E. coli*, reinforcing the robustness and biological relevance of our assay.

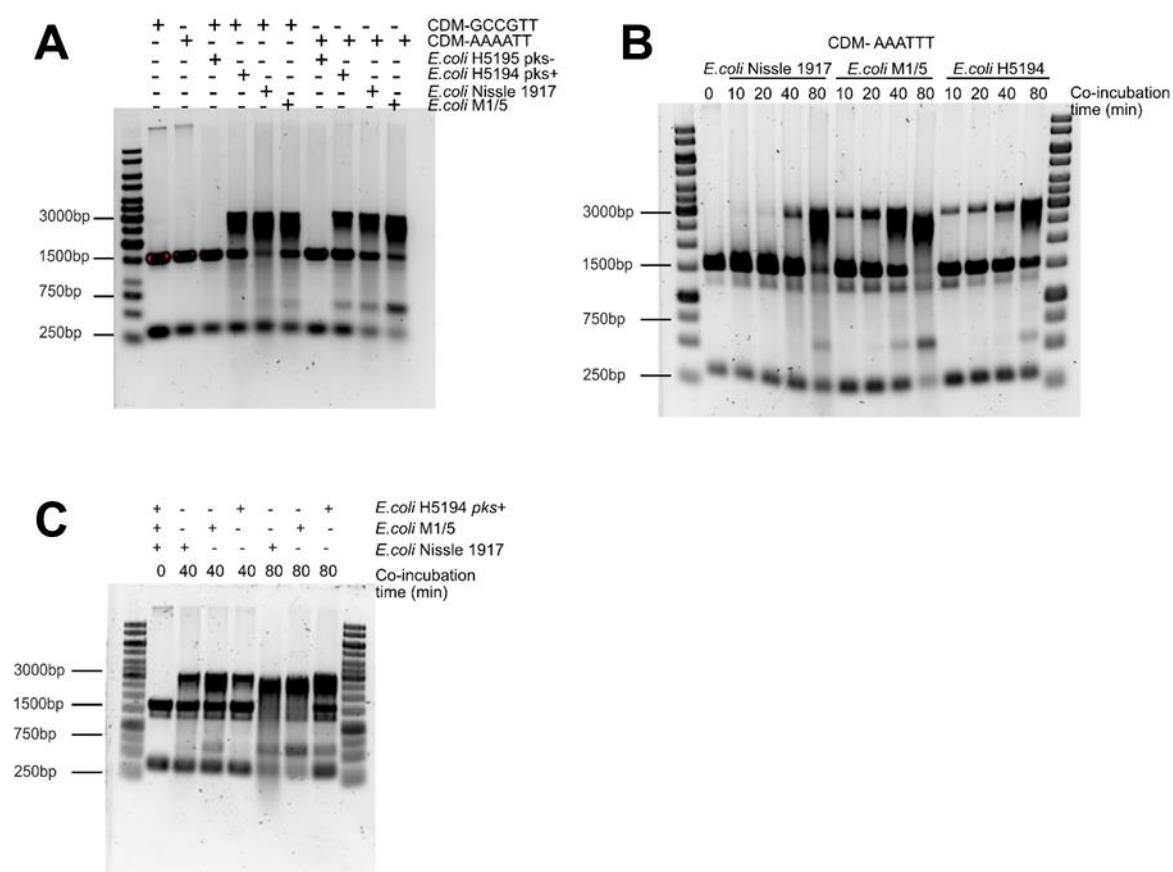


Figure 9- Further confirmation by 2 WT colibactin producing strains. This figure validates the sequence-specific induction of DNA inter-strand cross-links (ICLs) by colibactin across multiple *pks*⁺ *E. coli* strains, using a denaturing agarose gel-based assay. Figure 9 was created in collaboration with Dr. Karina Arellano Ayala.

(A) DNA inter-strand cross-link, observed as a 608 bp fragment, on CDM negative, CDM positive which contain 2 AAAATT sequences,

(B-C) CDM - AAATTT and CDM - AAAATT by *pks+* *E. coli* H5195 (recombinant strain), *pks-* *E. coli* H5194 (recombinant strain), *pks+* *E. coli* Nissle 1917 (wild type strain) and *pks+* *E. coli* M 1/5 (wild type strain). 1% denaturing agarose gel.

3.5 DNA super-helicity influence colibactin specificity

Colibactin affinity for CDM-AAAATT and CDM-AAATTT confirmed previously reported results identifying these sequences as canonical colibactin damage motifs (CDMs). However, in the present work, CDM-AAAGTT also exhibited measurable affinity, despite not conforming to the canonical AAWWTT motif. To validate this observation, we generated an internal control plasmid (CDM_IC) derived from CDM-AAAATT with an additional 167 bp fragment inserted at the BamHI cleavage site. Digestion of CDM_IC with BamHI yielded a 469 bp fragment, and following co-incubation with *pks+* *E. coli* H5194, an inter-strand cross-linked fragment of 938 bp was detected (**Figure 10E**).

Co-incubation experiments using CDM_IC alongside CDM-AAATTT, CDM-AATTTT, and CDM-AAAGTT at 10, 20, 40, and 80 minutes demonstrated consistent detection of the CDM_IC ICL fragment, while time-dependent accumulation of ICLs was evident for all test motifs. Although crosslinks were visible for CDM-AAAATT and CDM-AAAGTT, their signal intensities (604 bp fragments) were consistently lower than those of the internal control (**Figure 10B–C**). Quantitative analysis using the ratio of test CDMs relative to CDM_IC (604/302 bp over 938/469 bp) revealed no significant differences either between time points within each CDM or between CDM-AAAATT and CDM-AAAGTT ($p > 0.05$; **Figure 10B–D**).

Together, these findings highlight the sequence specificity of colibactin–DNA interactions and show how minor changes in sequence context can alter crosslinking efficiency. Notably, AAAGTT, which deviates from the canonical AAWWTT motif, displayed one of the strongest affinities in our assays, whereas AATATT, despite being reported as a canonical motif, showed comparatively weak activity.

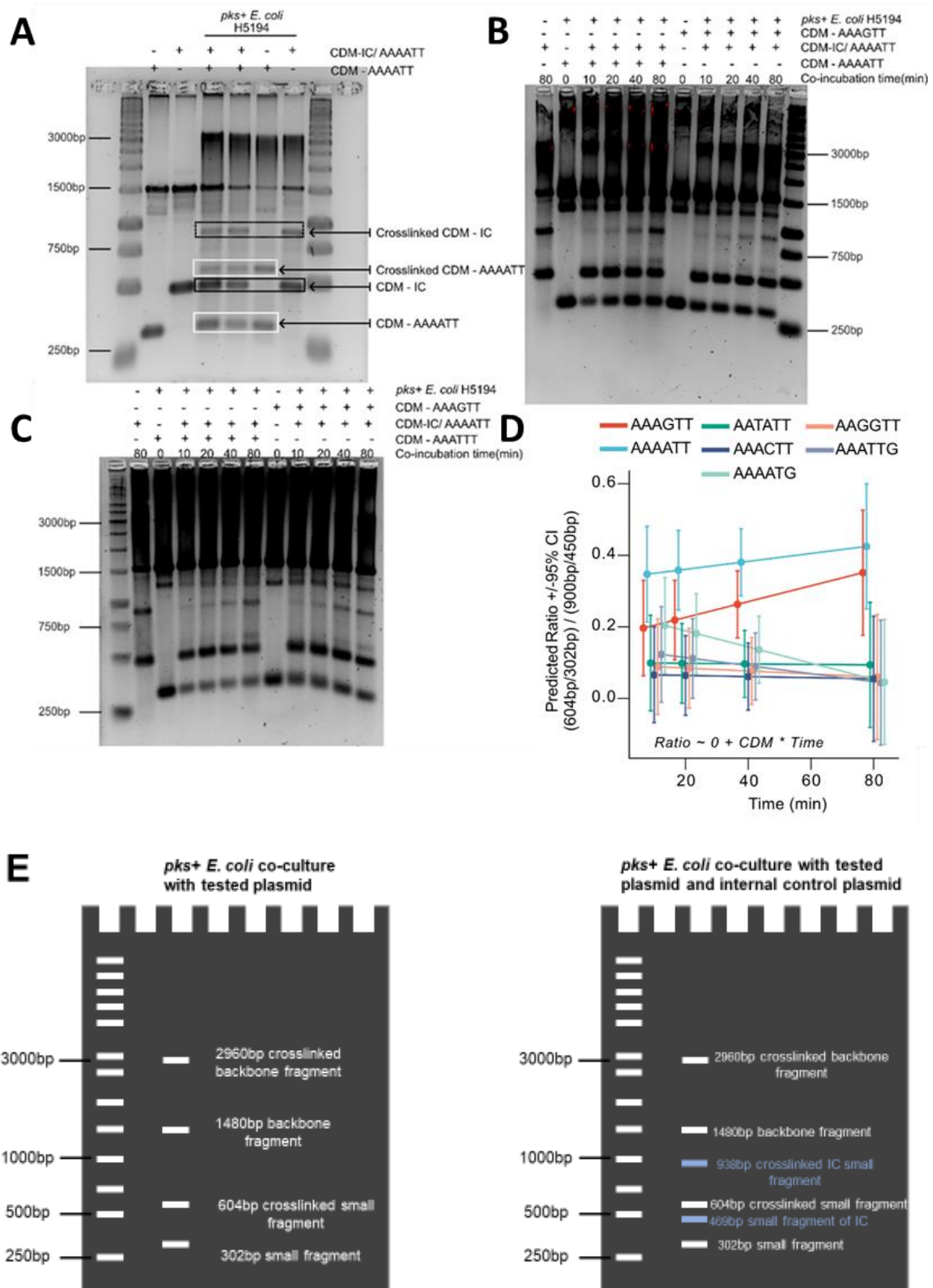


Figure 10- Usage of an experimental internal control to confirm crosslinking potential of AAAGTT as a novel CDM. This figure demonstrates the application of a plasmid-based internal control (CDM_IC) to normalize and

validate colibactin-induced DNA inter-strand crosslinks (ICLs), confirming AAAGTT as a functional CDM. Figure 10D was created by A. Quevedo-Olmos.

(A) Internal control test on 2% denaturing agarose gel. CDM_IC small fragment after digestion (BamHI) observed as 469 bp (white box) and after co-incubation with *pks+* *E. coli* H5194 seen as 938 bp (white dotted box). CDM14 small fragment after digestion indicated as 302 bp (black box), and after co-incubation with *pks+* *E. coli* H5194 detected as 604 bp (black dotted box).

(B-C) Colibactin sequence specificity kinetics on CDM13-AAATTT, CDM14-AAAATT and CDM19-AAAGTT in co-incubation with CDM_IC and *pks+* *E. coli* H5194. All the plasmids were digested after co-incubation with *pks+* *E. coli* H5194 and the cross-links detected on 2% denaturing agarose gel. Figure 10A-C were generated in collaboration with Dr. Karina Arellano Ayala.

(D) CDM-AAAATT and CDM-AAAGTT inter-strand crosslink quantification (600/300) in comparison to the internal control CDM (900/450) after 10, 20, 40, and 80 minutes co-incubation with *pks+* *E. coli* H5194

(E) This image shows the outcomes of DNA crosslinking activity in *pks+* *E. coli* co-culture experiments with and without an internal control (IC) plasmid.

3.6 Relaxation of DNA favors colibactin binding

We observed notable differences in colibactin target sequences between our *in vitro* assay and previous *in vivo* results (Dziubańska-Kusibab et al., 2020). We speculated that some could arise as a matter of DNA superhelicity differences. Circular plasmid DNA is always negatively supercoiled after purification (Patrick Higgins & Vologodskii, 2015) while in eukaryotic DNA the superhelicity status varies across different positions of the genome (Baranello et al., 2012). In supercoiled DNA, particularly when negatively supercoiled, torsional strain builds up, causing the DNA to twist further upon itself (Noy et al., 2016). This results in a more compact structure, with both grooves narrowing, particularly the minor groove. The narrowing of the minor groove in supercoiled DNA alters its geometry and increases the electrostatic potential in specific regions, which can significantly influence the binding efficiency of proteins and small molecules, especially those that target AT-rich sequences in the minor groove, such as colibactin (Antonino Lauria et al., 2007).

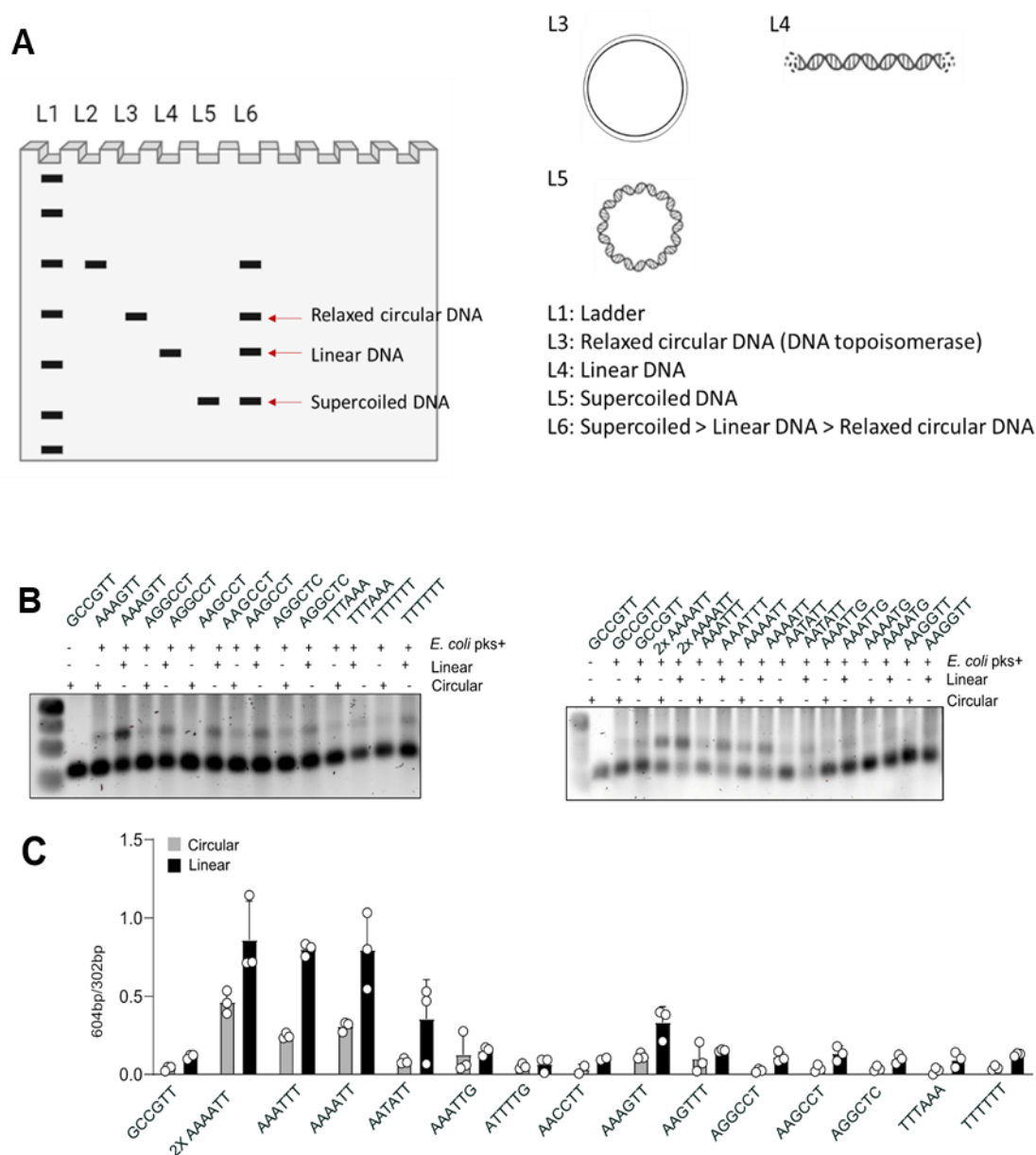


Figure 11- Influence of DNA super helicity on colibactin specificity. This figure investigates how DNA topology—specifically supercoiling versus linearization—affects colibactin's sequence specificity and inter-strand crosslinking efficiency. Figure 11B-C were created in collaboration with Dr. K. Arellano-Ayala.

(A) Schematic overview of plasmid DNA topological states, ranked from fastest to slowest migration on agarose gels: supercoiled denatured DNA > supercoiled > linear > relaxed circular > nicked open-circular. These forms reflect varying degrees of torsional strain and structural accessibility, which may influence colibactin binding. (The figure was created in <https://BioRender.com>.)

(B) Inter-strand crosslink detection on circular and linearized plasmids with EcoRI. Plasmids containing CDM19-CDM25. Both plasmids, circular and EcoRI linearized, were first co-incubated with *pks+* *E. coli* H5194 for 80 minutes and subsequently linearized with BamHI. The crosslinks were detected on 1% denaturing agarose gel.

(C) Inter-strand crosslink quantification in circular and linearized plasmids with EcoRI after 80 minutes co-incubation with *pks+* *E. coli* H5194. Inter-strand crosslink quantification (ratio of band size 604bp/302bp) in circular (gray) or linearized (black) plasmids.

To directly assess the influence of DNA topology on colibactin binding, plasmids containing AAATTT, AATTTT, AAAGTT, and other CDMs were linearized with EcoRI to relax superhelicity and then co-incubated with *pks+* *E. coli* H5194 for 80 minutes. Parallel experiments were performed with their supercoiled counterparts. Following incubation, both forms were digested with BamHI, and ICL formation was quantified by denaturing gel electrophoresis. Across all tested CDMs, linearized DNA exhibited a stronger colibactin response, reflected by an increased intensity of the ICL band at 604 bp compared to supercoiled plasmids (**Figure 11B**). Quantitative analysis confirmed a consistent increase in crosslinking efficiency in relaxed DNA for each motif (**Figure 11C**). Importantly, relaxation enhanced overall binding efficiency but did not change colibactin's preference for AT-rich motifs, as AAAATT and AAATTT remained the strongest targets, followed by AAAGTT.

In summary, these findings demonstrate that DNA topology modulates colibactin's activity: relaxation of super-helical stress favors inter-strand cross-linking without altering sequence specificity. This suggests that local DNA topology, such as relaxed or nucleosome-free regions in eukaryotic chromatin, could represent preferential binding sites for colibactin *in vivo*.

3.7 Validation of colibactin targeting specificity on phosphorothioated modified oligonucleotides

While linearized plasmids demonstrated higher colibactin binding affinities than supercoiled substrates, their susceptibility to degradation by bacterial nucleases and the presence of additional CDMs in the plasmid backbone introduced experimental variability. To address these limitations, we developed an assay based on phosphorothioate-modified oligonucleotides (S-oligos). In these constructs, a sulfur atom replaces a non-bridging oxygen in the phosphate backbone, markedly increasing nuclease resistance. In this study, phosphorothioate linkages were incorporated at both the 5' and 3' termini, ensuring stability during bacterial co-incubation.

S-oligos share the relaxed topology of linearized DNA but offer two additional advantages: (i) they are resistant to exonucleolytic degradation, and (ii) they exclude background CDMs located outside the target sequence, providing a clean substrate for motif-specific binding analysis. Quantification of ICL formation using S-oligos revealed results consistent with plasmid-based assays. AAAATT exhibited the highest colibactin affinity, followed by AATTTT. Interestingly, AAAGTT—although not matching the canonical AAWWTT motif—displayed comparable binding activity to AATATT, which aligns with its recurrent detection in plasmid-based assays (**Figure 12A–B**).

Taken together, these results validate the use of S-oligos as a high-fidelity platform for characterizing colibactin–DNA interactions. The data confirm three canonical AAWTT motifs (AAAATT, AATTTT, AATATT) and one noncanonical motif (AAAGTT) as colibactin targeting motifs *in vitro*. Importantly, the consistent detection of AAAGTT across multiple independent platforms (plasmids and S-oligos) suggests that colibactin may target a broader sequence repertoire than previously assumed, potentially expanding the genomic contexts at risk for colibactin-induced lesions.

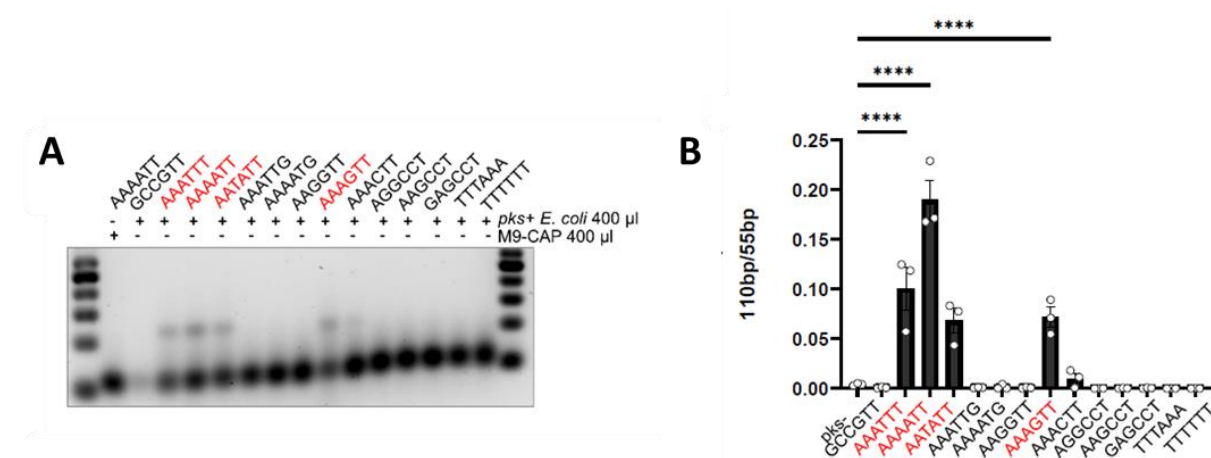


Figure 12- Confirmation of new CDM 19 AAAGTT on S-oligos. This figure demonstrates sequence-specific DNA inter-strand crosslink formation by colibactin in S-oligos containing 15 different candidate damage motifs, including the novel motif AAAGTT.

(A) Inter-strand crosslink detection on S-oligos containing 15 different CDMs. All S-oligos were co-incubated with *pks+* *E. coli* H5194 for 120 minutes. The crosslinks were detected on 3% denaturing agarose gel. C. Inter-strand crosslink quantification in S-oligos after 120 minutes co-incubation with *pks+* *E. coli* H5194.

(B) Quantification of the cross-links by CDM as a ratio between 110bp and 55 bp. Statistical differences were compared to *pks- E. coli* treatment by one-way ANOVA and Dunnett's multiple comparisons test where, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$ and **** $p < 0.0001$.

3.8 Applying nanopore sequencing for detecting colibactin crosslinks

Next, we wondered whether it could be possible to establish a high-throughput version of the previous assays. Both alkylated DNA bases (Rand et al., 2017; Wang et al., 2021) and covalent inter-strand crosslinks (Zhang et al., 2015) can be detected using nanopore sequencing. As colibactin causes both types of modifications in the DNA, we wonder whether nanopore could be also used to detect such colibactin-dependent modifications. Mechanistically, a covalent ICL prevents the separation of the DNA strands required during sequencing. As a result, the pore is blocked and the newly generated sequence read is terminated prematurely (Zhang et al., 2015). However, only a nucleotide from one strand is alkylated, this modified nucleotide, with a new molecular structure, will alter the ionic current in a yet unknown pattern which cannot be accurately deciphered by the basecaller. As a consequence, a reduction in the quality of the current called nucleotide is introduced at positions where colibactin binds DNA (Wang et al., 2021). In conclusion, we hypothesized that a reduction of quality in nanopore basecalls can be used as a reporter of colibactin binding to DNA, which can be used to quantify colibactin affinity to different DNA sequences in a high-throughput manner.

To test this idea, we incubated a 9.8kb plasmid (Addgene #130662) with the same *pks*⁺ or *pks*⁻ *E. coli* strains (recombinant strains H5194 and H5195). Furthermore, we also explored the role of superhelicity by keeping plasmids in linear (NdeI digested) or circular conformations during co-incubation with bacteria. After 40 minutes of incubation, where unspecific ICLs were not yet observed in our gel assay, plasmid DNA was purified and sequenced using a Nanopore sequencer. After preprocessing aligning raw sequences to the plasmid reference, we computed the proportion of low quality basecalls, defined as those with a Phred quality score less than 10 ($Q < 10$), for each position in the plasmid (See Methods). Finally, we normalized the proportion $Q < 10$ by taking the difference with the values obtained in the *pks*⁻ and no-colibactin samples (*pks*⁺ minus *pks*⁻ and *pks*⁻ minus no-colibactin), therefore representing the difference in proportion of low quality basecalls with respect to the control condition, which we will refer hereafter as $\Delta\text{Prop}Q < 10$. The resulting $\Delta\text{Prop}Q < 10$ values along the whole plasmid are depicted in Figure 8A. We observed that some positions contain

a higher $\Delta\text{PropQ} < 10$ values, which we called peaks. If those peaks are caused by colibactin, then they must occur where colibactin binding sequences exist. For that reason, when we compared the location of peaks and the positions of canonical Colibactin Damage Motif variations (CDM, 5-AAWWTT-3', Dziubańska-Kusibab et al., 2020), we observed a clear colocalization between both, especially for the known CDMs also detected in gels (AAAATT), suggesting that the signal is colibactin dependent (**Figure 13A**). A closer look at the region surrounding the highest peak reinforced this idea (**Figure 13B**). When GC rich sequences exist around the peak, no increase in the $\Delta\text{PropQ} < 10$ scores are observed in them. But, at the positions where CDM like motifs exist, the quality of basecalls drops abruptly. In order to confirm the colibactin specificity of the peaks (marked with “x” in **Figure 13A**), we took the $\pm 10\text{nt}$ surrounding their highest point and draw sequence logos (**Figure 13C**). In the pks+ samples, the obtained logos around the peak summit are reminiscent of the canonical CDM, where Adenines are followed by Thymines in 5'>3' orientation, whereas in pks- samples, either no peaks are called (minus strand) or the only 4 detected lack any consensus.

Based on these observations, we conclude that nanopore sequencing is a feasible method to detect colibactin binding to DNA.

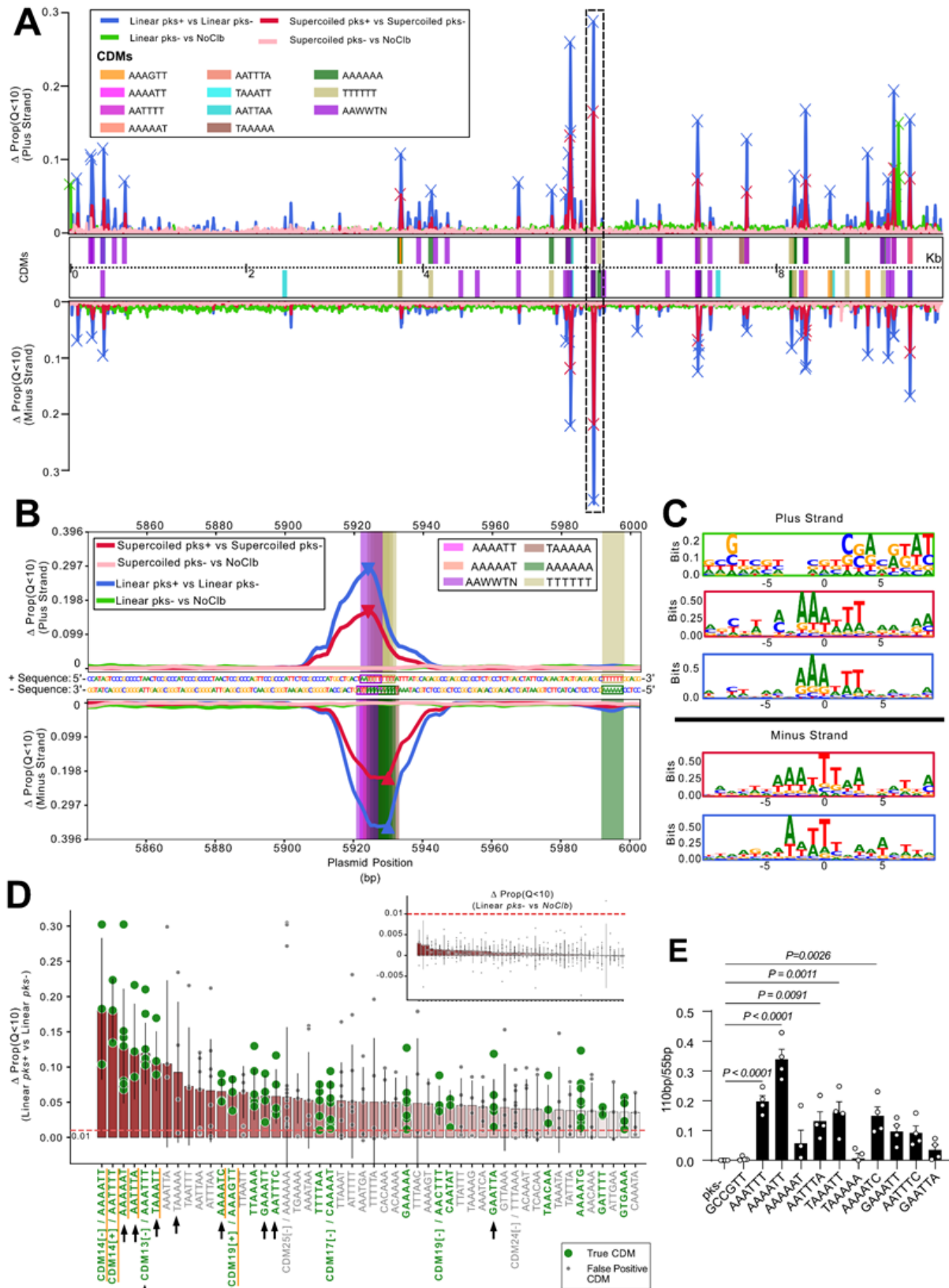


Figure 13- Detection of colibactin binding to DNA using nanopore seq. This figure illustrates the application of long-read nanopore sequencing to identify colibactin-induced DNA basecall disruptions at sequence-specific motifs, further validated by oligonucleotide-based crosslinking assays. Figure13A-D were created by A. Quevedo-Olmos.

(A) Difference in proportion of low quality basecalls (Prop Q<10) profiles in both plus (top) and minus (bottom) strands of the sequenced plasmid. Peaks of quality drops marked by "X" occur at positions that contain colibactin damage motifs (CDMs) depicted in the middle panel. Negative $\Delta\text{PropQ}<10$ values are collapsed to Zero. Only Replicate 1 shown.

(B) Magnification of the highest peak outlined in A). Higher Prop Q<10 values are found around the only AT rich sequence in the region. When TTTTTT/AAAAAA are not in the neighborhood of other AT rich stretches, no basecall quality drops are observed. Middle panel shows the plasmid sequence for both strands. Only Replicate 1 shown

(C) Logos of sequences found at +/-10nt of peaks depicted with "X" in panel A. The canonical AANTT motif is conserved at peak summits. Only Replicate 1 shown.

(D) CDMs ranked by median difference in proportion of low quality basecalls (Prop Q<10) in linear *pks+* vs *pks-* condition. False positive CDMs are those with at least 1 position below 0.01 or those where median minus standard deviation is less than Zero. CDMs tested using short oligos are labelled with their corresponding ID number and strand from which the sequence is read. Inset shows the same values in for the linear *pks-* vs No-colibactin condition, where none of the median values is above 0.01.

(E) 8 selected CDMs from nanopore sequencing data were validated by ODNs. Quantification of the cross-links by CDM as a ratio between 110bp and 55 bp. Statistical differences were compared to *pks-* *E. coli* treatment by one-way ANOVA and Dunnett's multiple comparisons test where, * $p<0.05$, ** $p<0.01$, *** $p<0.005$ and **** $p<0.0001$.

3.9 Detecting colibactin targeting sites in large DNA molecules

Going back to our initial purpose of establishing a high-throughput method to quantify colibactin affinity to DNA sequences, we aimed to confirm the results obtained in the gel-based assay. To do so, we decided to rank the affinity of colibactin for different hexamers (6nt CDMs) using the median $\Delta\text{PropQ}<10$ value for those CDMs with 3 or more occurrences in the plasmid. However, due to the width of the $\Delta\text{PropQ}<10$ signals (**Figure 13B**), it is possible that sequences not bound by colibactin are present within a region of basecall quality drop. In this situation, the $\Delta\text{PropQ}<10$ assigned to that CDM occurrence will be high and could led us to wrongly attribute colibactin binding to that sequence. On the other hand, for CDMs with true colibactin binding, the basecall quality changes should be similar independently of the context those CDMs are found. An example of this scenario can be observed in the **Figure 13B**. Several occurrences of TTTTTT on the plus strand lie under the peak created, probably, by colibactin bound

to the nearby sequence AATTTT. However, when this same hexamer is found within a GC rich context several nucleotides downstream, there are no increases in the $\Delta\text{PropQ}<10$ signal. In addition, when extreme different $\Delta\text{PropQ}<10$ values are given to the same motif, the dispersion estimates increase. With this idea in mind, after ranking hexamers by median $\Delta\text{PropQ}<10$, we marked them as false positive if 1) at any position they have a $\Delta\text{PropQ}<10$ value less than 0.01, or if 2) their median $\Delta\text{PropQ}<10$ is less than one standard deviation from zero. The 50 CDMs with the highest median $\Delta\text{PropQ}<10$ value (**Figure 13D**) detected using nanopore confirm the gel results. Both assays identify AAAATT, AATTTT and AAAGTT among the CDMs with the highest colibactin binding potential. But this new assay also reveals other hexamers not tested so far: namely the AT only AAAAAT, AATTTA, TAAATT and TTAAAA; and the G/C containing AAAATC, GAAATT. When the same approach is performed in the *pks*- samples, none of the median values exceed 0.005 (**Figure 13D**), again confirming the colibactin specificity of the method. The binding of colibactin to the new sequences was experimentally confirmed on the *in vitro* assay (**Figure 13E**).

Altogether, the evidence presented here shows how both high and low throughput assays can be used in combination to define putative colibactin target sequences.

3.10 *pks*+ *E. coli* H5194 respond to the presence of cofactors shaping colibactin production

Previous works have already identified several molecules that regulate genes in the *pks* island. Those has been investigated by luciferase assays where the activity of the enzyme indicates the upregulation or downregulation of specific genes in the island. Then we investigated the influence of previously reported cofactors and copper on colibactin production by the *pks*+ recombinant strain *E. coli* H5194 on the CDM-AAAATT plasmid, in which colibactin had showed higher specificity according to the former results. Colibactin activity was visualized in alkaline gels where the crosslink in the CDM-AATTTT was observed as a 602 bp fragment (**Figure 14G**). Therefore, *E. coli* H5194 was cultured on M9_Cap media in the presence of FeSO_4 and FeCl_3 (iron II and III) at 25, 50 and 100 μM , CuCl_2 at 50, 100 and 150 μM , the polyamide, spermidine at 10, 20, 40 $\mu\text{g/mL}$ and the fermentable fibers, inulin and fructooligosaccharides, FOS at 10, 20, 40 mg/mL at exponential phase. The cultures

were let stand for 90 minutes to subsequently add the plasmid containing the CDM14/AATTTT to further incubate for 80 minutes more. The quantification of the cross-links in the agarose gels indicated that iron II and III reduced significantly ($p < 0.05$), the cross-links in a concentration dependent manner compared to the cross-links observed in the control media without iron (**Figure 14A-B**). The highest reduction was to 100 μM of iron and the lowest to 25 μM ; being the effect of iron III stronger compared to iron II ($p < 0.005$). Copper similarly inhibited colibactin production by reducing the crosslinks in CDM-AAAATT, showing a significant inhibition ($p < 0.005$) from the lowest concentration tested, 50 μM (**Figure 14C**). On the other hand, the polyamide spermidine, and the fermentable fibers inulin and FOS, showed an increasement on the DNA crosslinked by colibactin, reaching the highest area at the highest concentration analyzed, 40 $\mu\text{g/mL}$ for spermidine and 40 mg/mL for the fermentable fibers; however, such difference was not significant ($p > 0.05$) compared to the values detected on the control media without the cofactors. The aminoacidic serine has been also reported to influence colibactin production in the luciferase assay, nevertheless our results do not show any difference between the control media M9_Cap which contain 0.19 g/L serine and the M9_Cap without serine, data not shown (**Figure 14H**). *pks+* *E. coli* H5194 was also quantified by plate counting all the cultures with the different treatments on LB agar to discard the influence of the cofactors on bacteria growth and numbers. The results indicated that there is not significantly difference in the CFU/mL ($p > 0.05$) within the different growth media (**Figure 14I**).

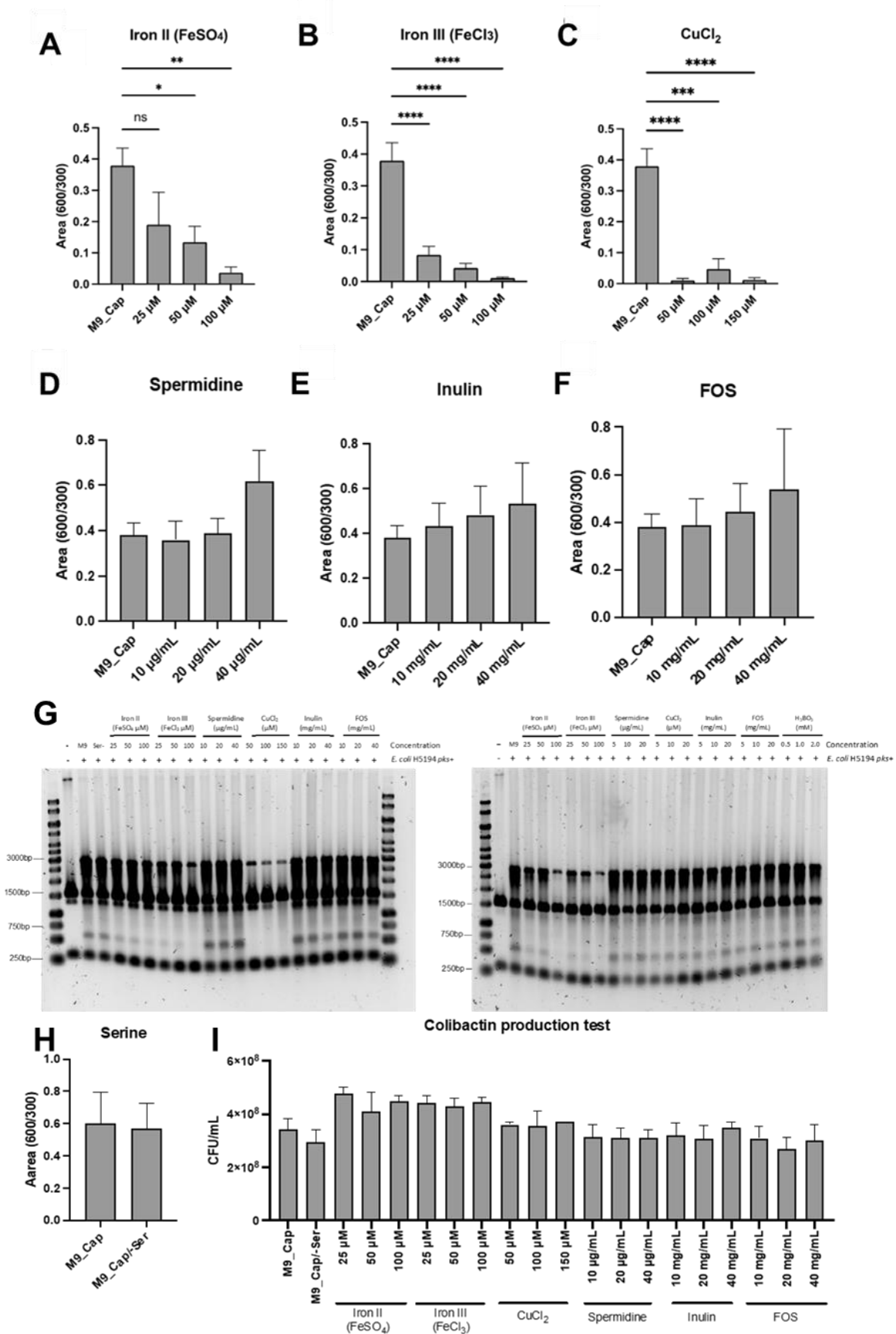


Figure 14- Colibactin production is influenced by the presence of cofactors in the media. This figure explores how metal ions, polyamines, and fermentable fibers impact colibactin-induced DNA inter-strand crosslink (ICL) formation by *pks+* *E. coli* H5194, using CDM-containing plasmids and a 1% alkaline or denaturing agarose gel readout. This figure was made in collaboration with Dr. K. Arellano-Ayala.

(A-F) Quantification of cross-links produced by colibactin on CDM14-AATTTT during 80 minutes after growing *E. coli* H5194 *pks+* to exponential phase on M9_Cap supplemented with iron II, FeSO₄ (25, 50 and 100 μ M), iron III, FeCl₃ (25, 50 and 100 μ M), CuCl₂ (50, 100 and 150 μ M), polyamide, spermidine (10, 20, 40 μ g/mL) and fermentable fibers, inulin and fructooligosaccharides, FOS (10, 20, 40 mg/mL). M9_Cap without supplements was used as control. The values correspond to the ratio between the cross-links and the non-cross-linked DNA, observed on 1% agarose alkaline gel as a 604bp fragment and 302bp fragments respectively.

(G) CDM-AAAATT inter-strand cross link visualization (600 bp) on 1% denaturing gel after 80 minutes co-incubation with *pks+* *E. coli* H5194 under the presence of different cofactors on the colibactin production test.

(H) CDM-AAAATT cross-link's quantification after 80 minutes co-incubation with *pks+* *E. coli* H5194 on M9_Cap media without serine.

(I) *E. coli* H5194 viability on CFU/ml at logarithmic phase on M9_Cap media supplemented with different cofactors concentrations during the colibactin production test.

Statistical differences were compared to the cross-links produced on M9_Cap without cofactors by one-way ANOVA and Dunnett's multiple comparisons test where, * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.005$.

3.11 Influence of colibactin activity by various cofactors.

To further investigate how iron, copper, spermidine, inulin and FOS interact with the genotoxin colibactin and the plasmid DNA. *pks+* *E. coli* H5194 was grown in M9_Cap without additions and successively the CDM-AAAATT plasmid DNA and the different cofactors were added at the same time to the reaction. The DNA, cofactors and *pks+* *E. coli* H5194 were co-incubated for 80 min and the effects of the different combinations were determined on an alkaline gel where it was observed that only Iron II (FeSO₄) and copper (CuCl₂) interfered in the reaction (**Figure 15G**). Whereas, iron III (FeCl₃), that has hindered colibactin production (**Figure 15A**), spermidine, inulin and FOS presented not significantly difference compared to the control, $p > 0.05$ (**Figure 15B, D, E and F**). The quantification of the CDM-AAAATT crosslinks in the treatment with iron II shown a significant reduction at 50 μ M ($p < 0.05$) and 100 μ M ($p < 0.01$) but not at 25 μ M ($p > 0.05$) with no differences between 50 and 100 μ M. In the treatment with copper, there was a significantly concentration dependent reduction on the crosslinks area ($p < 0.005$), from 5 to 10 and 10 to 20 μ M compared to the control, having the strongest inhibition 20 μ M (**Figure 15C**). The non-crosslinked CDM14/AATTTT fragment was

visualized as a 302bp on the gel, indicating the absence of DNA degradation as a consequence of the ions (**Figure 15G**). In addition, to discard any adverse effects of the different cofactors in the bacteria viability and differences on the bacteria numbers on the assay, all the bacteria cultures were also serial diluted and plate counted on LB agar. The results indicated that there was not significantly difference ($p>0.05$) in the *E. coli* H5194 CFU/mL within the different treatments and the M9_Cap without cofactors used as a control (**Figure 15H**).

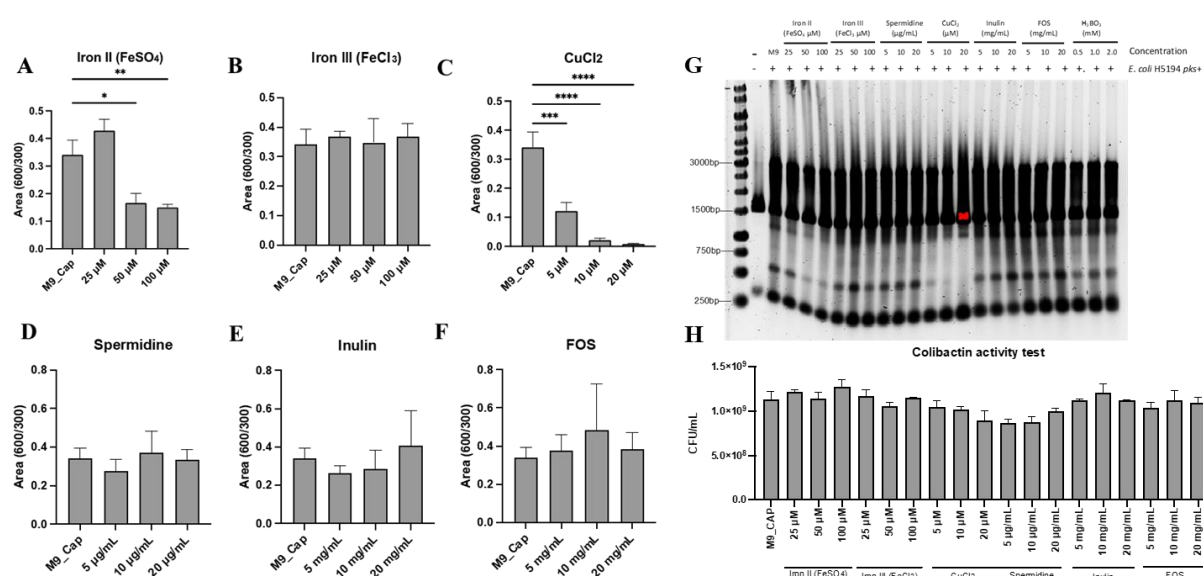


Figure 15- Colibactin-CDM-AATTTT interaction is altered by the addition of cofactors to the reaction. This figure investigates how the direct addition of cofactors to the colibactin-DNA interaction system affects the efficiency of DNA inter-strand crosslink (ICL) formation on the CDM14-AATTTT motif. This figure was made in collaboration with Dr. K. Arellano-Ayala.

(A-F) Quantification of cross-links produced by colibactin after addition of iron II, FeSO₄ (25, 50 and 100 μ M), iron III, FeCl₃ (25, 50 and 100 μ M), CuCl₂ (50, 100 and 150 μ M), polyamide, spermidine (10, 20, 40 μ g/mL) and fermentable fibers, inulinid and fructooligosaccharides, FOS (10, 20, 40 mg/mL) to the reaction simultaneously with the CDM-AATTTT plasmid. H5194 pks+ was grown to stationary phase in M9_Cap without supplements and 1 μ g plasmid DNA and the different cofactors were added at the same time to the reaction and co-incubated for 80 min. M9_Cap was used as control. The values correspond to the ratio between DNA cross-links (604bp) and non-crosslinked DNA (302 bp).

(G) CDM14/AATTTT inter-strand cross link visualization (604 bp) on 1% denaturing gel after 80 minutes co-incubation with pks+ *E. coli* H5194 under the presence of different cofactors on the colibactin activity test.

(H) H5194 viability on CFU/mL at stationary phase on M9_Cap media supplemented with different cofactors concentrations in the colibactin activity test. Statistical differences were compared to the cross-links produced on M9_Cap by one-way ANOVA and Dunnett's multiple comparisons test where, * $p<0.05$, ** $p<0.01$ and *** $p<0.005$.

3.12 Influence of Copper on colibactin production and activity

Since we had observed that copper interfere in colibactin production and activity. It was investigated the inhibitory concentrations of copper following the same approached as described earlier. In the production test, 5 different copper concentrations including 5, 10, 20, 40 and 60 μM were added to the M9_Cap media. While for the activity test the concentrations tested were lower comprising 1.25, 2.5, 5, 10 and 20 μM . As it was mentioned, in the production test copper was added from the beginning in the media to culture *pks+* H5194 whereas in the activity test copper was added to the reaction simultaneously with the CDM-AAAATT plasmid DNA. To let colibactin react with the DNA for 80 minutes. The visualization of the CDM-AAAATT cross-links on the denaturing gel showed a reduction on the intensity of the 604 bp band (**Figure 16E**) while its quantification confirmed a significantly difference ($p < 0.01$) between the areas of the band observed in the control M9_Cap media without copper and the media supplemented with copper (**Figure 16A and C**). In the production test, the inhibition was observed from the lowest concentration tested, 5 μM , which was gradually increasing reaching the highest inhibition at 60 μM . Similar tendency was observed in the activity test, where the highest concentration tested, 20 μM , reached the highest DNA inter-strand-cross link inhibition, yet it was not significantly different to the other concentration included in the assay (1.25, 2.5, 5 and 10 μM) which also showed significantly differences compared to control (**Figure 16C**). Furthermore, the presence of copper on the media did not affect the H5194 viability since there were not significantly differences ($p > 0.05$) in the bacteria numbers expressed as CFU/mL between the control M9_Cap media without copper and the M9_Cap media where different copper concentrations were added.

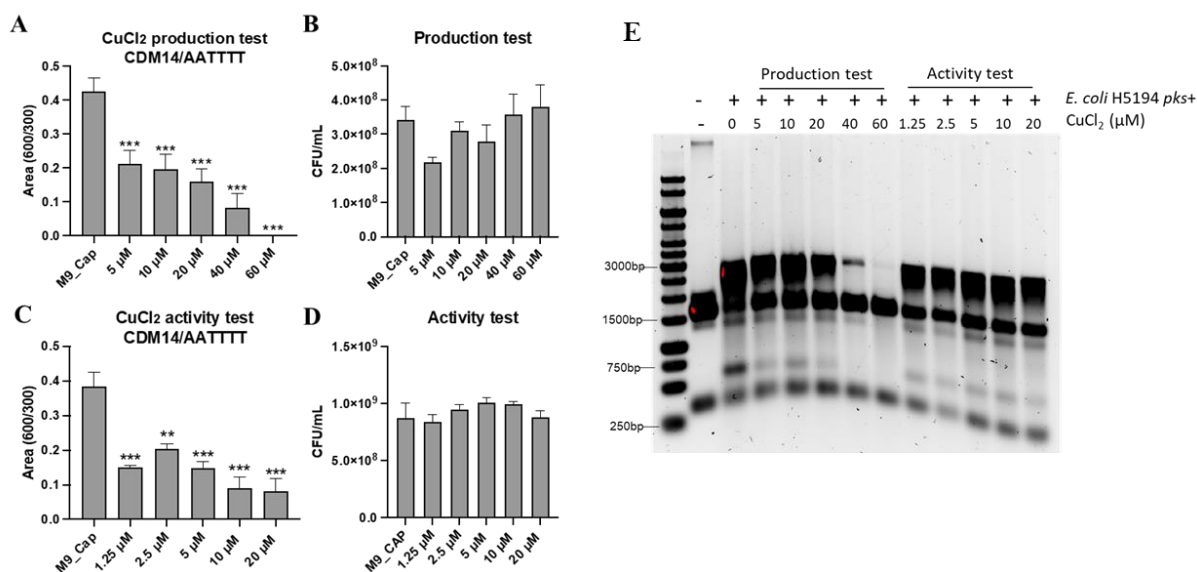


Figure 16- Copper reduces colibactin production by H5194 *pks+* *E. coli* and interfere between colibactin-CDM-AAAATT interaction. This figure examines the dual role of copper (CuCl₂) in suppressing colibactin biosynthesis and inhibiting its DNA crosslinking activity at the CDM-AAAATT motif. This figure was made collaborate with Dr. K. Arellano-Ayala.

(A) Cross-links quantification after growing *pks+* *E. coli* H5194 to exponential phase in M9_Cap with CuCl₂ (5, 10, 20, 40 and 60 μM) and 80 minutes co-incubation with CDM-AAAATT.

(B) H5194 viability at exponential phase in M9_Cap with CuCl₂.

(C) Cross-links quantification after CDM-AAAATT - CuCl₂ addition to *pks+* *E. coli* H5194 culture on stationary phase.

(D) H5194 viability at stationary phase after 80 minutes CDM14/AATTTT- CuCl₂ co-incubation. M9_Cap without copper was used as control.

(E) CDM-AAAATT inter-strand cross link visualization (600 bp) on 1% denaturing gel after 80 minutes co-incubation with *pks+* *E. coli* H5194 under the presence different of copper. For the colibactin production test H5194 was grown to logarithmic phase on the presence of copper while the activity test was done by adding the CDM-AAAATT plasmid simultaneously with the copper to the bacteria culture. Statistical differences were compared to the cross-links produced on M9_Cap by one-way ANOVA and Dunnett's multiple comparisons test where, *p<0.05, **p<0.01 and ***p<0.005.

3.13 Oxygen regulates colibactin production on *E. coli* strains

Since it has been observed that oxygen might play a role in the regulation of the *pks* island and considering the colon is an environment mostly anaerobic besides the association of *pks+* strains with colorectal cancer. It was imperative to investigate the role of oxygen on colibactin regulation and DNA cross-link formation on the CDM-

AAAATT. First, in order to determine the influence of oxygen on colibactin production by different strains, two recombinant strains, *pks*⁺ H5194 and *pks*⁻ H5195, and two wild type strains, *pks*⁺ Nissle 1917, a well-known probiotic strain, and *pks*⁺ M1/5, a fecal isolated were cultured on M9_Cap media under aerobic and anaerobic conditions. The bacteria cultures were co-incubated with the CDM-AAAATT plasmid for 20min and 40mins and the ICLs were visualized on denaturing gel as described. Surprisingly the recombinant strain *pks*⁺ H5194 did not induce ICLs under anaerobic conditions merely in aerobic conditions at 20 and 40 minutes (**Figure 17F**). Conversely, the wild type strains, Nissle 1917 and M1/5, induced ICLs on CDM-AAAATT in the presence and absence of oxygen after 20 minutes showed a higher intensity on the 604 bp crosslinked DNA under anaerobic conditions. The *pks*⁻ H5195 used as a control did not react in any of the conditions. Interestingly, in the treatments with the wild type strains after 40 minutes incubation under both conditions the DNA started to degrade and it was difficult to detect the 604 bp band on the gels (**Figure 17F**). Therefore, the subsequent tests were done by co-incubating the DNA with the strains for 10 and 20 minutes. Such phenomena might be due to the presence of DNases produced on the media by the wild type strains but not by the recombinant. To confirm whether the recombinant *pks*⁺ H5194 require oxygen to produce colibactin and to determine the earliest time point in which Nissle 1917 and M1/5 start to produce colibactin, the same experiment was performed but now incubating the DNA with the bacteria for 10 and 20 minutes in the absence of oxygen. The results indicated that Nissle 1917 and M1/5 start to produce colibactin after 10 minutes, detected as ICLs in the CDM-AAAATT (604 bp fragment) while the recombinant strain did not (**Figure 17A**). The increasement was time dependent, yet it was not significantly different ($p>0.05$) between 10 and 20 minutes (**Figure 17B**). The quantification and analysis of the different strains CFU/mL did not reveal any significantly difference ($p>0.05$) between them, discarding the influence of the oxygen on the bacteria growth (**Figure 17E**).

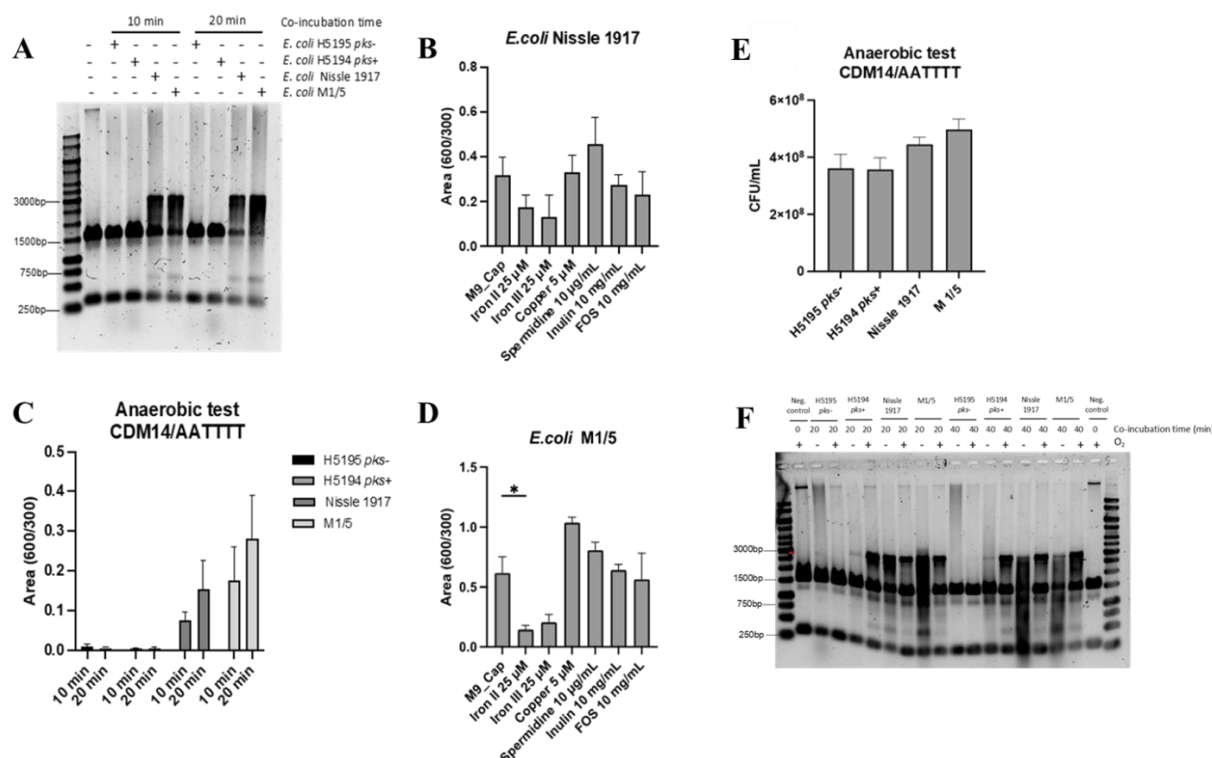


Figure 17- Oxygen regulates colibactin and inter-strand DNA cross-link formation on naked DNA. This figure explores how aerobic versus anaerobic conditions influence the capacity of recombinant and wild-type *pks*⁺ *E. coli* strains to induce colibactin-mediated DNA inter-strand crosslinks (ICLs) on plasmids containing the AATTTT motif (CDM14). This figure was created in collaboration with Dr. K. Arellano-Ayala.

(A) CDM14/AATTTT cross-link visualization (600 bp) on 1% denaturing gel after 10- and 20-minutes co-incubation with *pks*⁺ H5194, *pks*⁻ H5195, recombinant strains, Nissle 1917 and M1/5 wild type strains, cultured on M9_Cap under anaerobic conditions.

(B) CDM14/AATTTT cross-links' quantification after 10- and 20-minutes co-incubation with recombinant and wild type strains.

(C-D) CDM14/AATTTT cross-links quantification after 20 minutes co-incubation with Nissle 1917 and M1/5 on the presence of iron II, FeSO₄ (25 μ M), iron III, FeCl₃ (25 μ M), CuCl₂ (5 μ M), polyamide, spermidine (10 μ g/mL) and fermentable fibers, inulin and fructooligosaccharides, FOS (10 mg/mL). The wild type strains were grown to exponential phase on M9_Cap supplemented with the cofactors. M9_Cap was used as control and statistical differences were compared to the cross-links produced on M9_Cap by one-way ANOVA and Dunnett's multiple comparisons test where, **p*<0.05, ***p*<0.01 and ****p*<0.005.

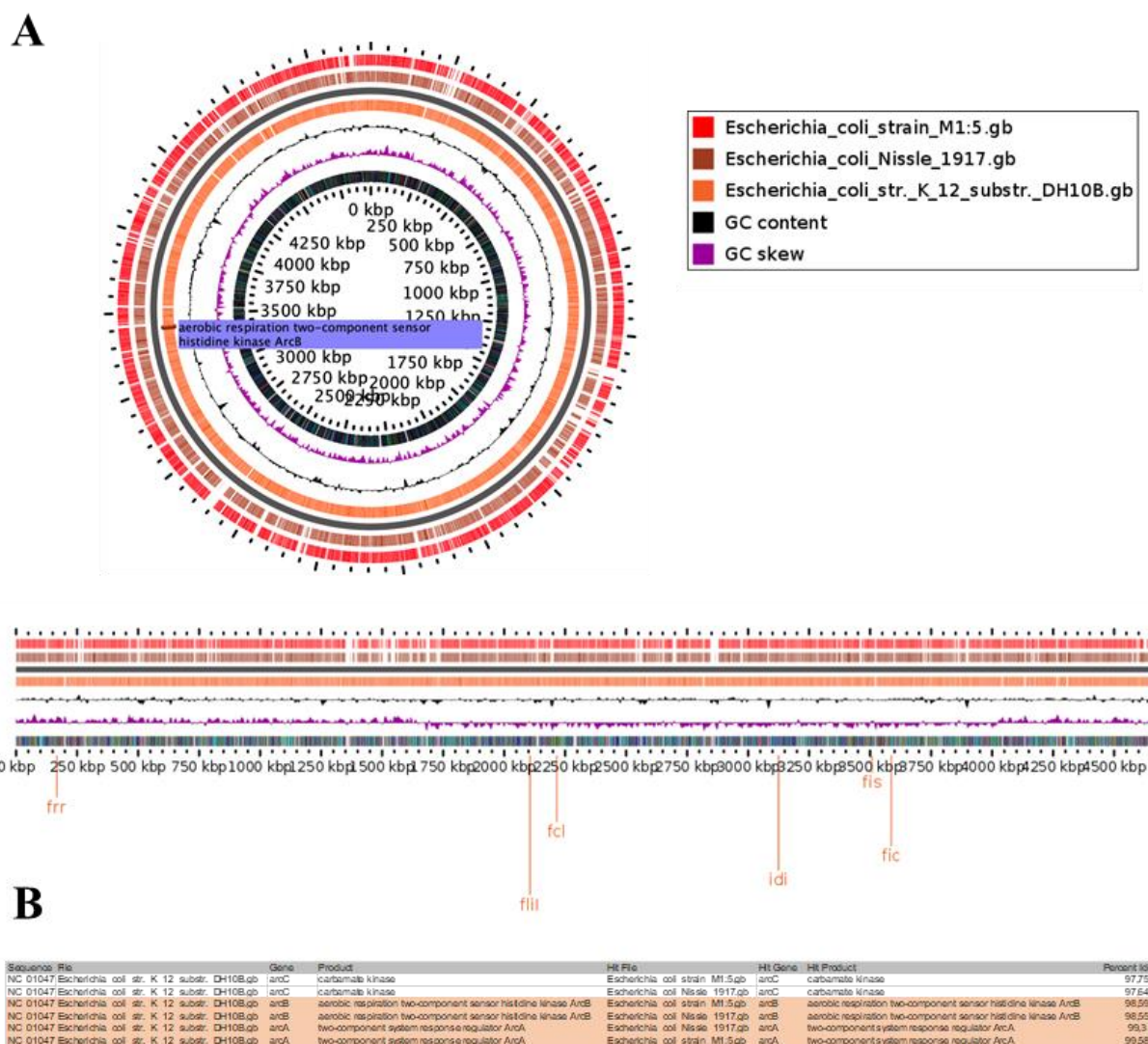
(E) *pks*⁻ H5195, *pks*⁺ H5194, *pks*⁺ Nissle and *pks*⁺ M 1/5 viability on CFU/mL at logarithmic phase on M9_Cap media under anaerobic conditions.

(F) CDM-AAAATT intra strand cross link visualization, 600 bp, on 1% alkaline gel after 20- and 40-minutes co-incubation with *pks*⁻ H5195, *pks*⁺ H5194, *pks*⁺ Nissle and *pks*⁺ M 1/5 under aerobic and anaerobic conditions.

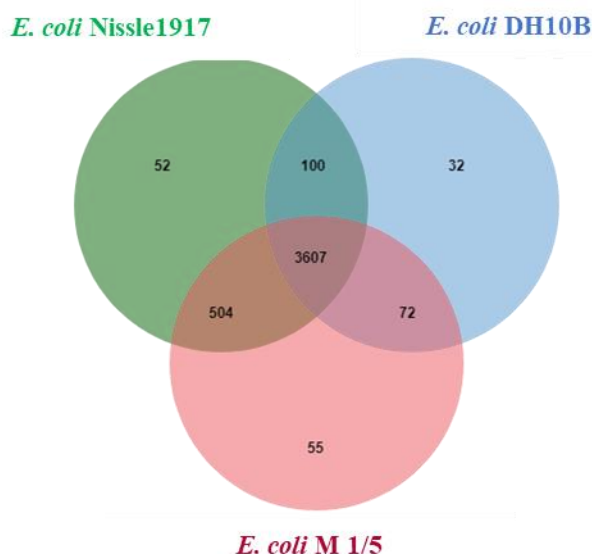
3.14 Genomic similarity of *E. coli* DH10B, *E. coli* Nissle 1917 and *E. coli* M 1/5

In addition, to investigate the mechanism underlying the downregulation of colibactin on the recombinant *pks+* H5194 strain. Whole genome sequences of H5194 strain that was generated from DH10B, the M1/5 and Nissle 1917 were retrieved from the NCBI and compared using the web-program Gview server (**Figure 18A-B**). The Aerobic respiration two-component sensor histidine kinase *ArcB* gene and two-component system response regulator *ArcA* gene that have been previously associated to colibactin regulation under anaerobic conditions were identified on the three strains. The BLAST hits for the *ArcB* gene indicated a 98.5% identity within DH10B, the M1/5 and Nissle 1917 while the *ArcB* has a 99.3 and 99.58% of identity between DH10B and Nissle 1917 and DH10B and M1/5 respectively (**Figure 18B**). Furthermore, the protein sequences of the three strains were downloaded from the NCBI and used to generate a Venn diagram to identify differences within the strains using the web-program orthoVenn2. The results denoted that DH10B, Nissle 1917 and M 1/5 share 3607 genes. On the other hand, 32 genes are specific for DH10B, 52 genes for Nissle 1917 and 55 genes for M 1/5 (**Figure 18C**). The protein sequence comparison for the aerobic respiration control protein *ArcA* of DH10B, Nissle 1917 and M 1/5 also indicated a 100% identity (**Figure 18D**). Lastly, to study the influence of the cofactors on colibactin production under anaerobic conditions by Nissle 1917 and M 1/5, the strains were cultured on media containing, 25 μ M FeSO₄ (iron II) and FeCl₃ (iron III), 5 μ M CuCl₂, 10 μ g/mL spermidine and 10 mg/mL inulin and FOS. The CDM14/AATTTT plasmid was co-incubated with the bacteria cultures for 20 minutes and the ICLs visualized following the same approach. The analysis of the bands in the gel pointed out similar results as the one observed with the *pks+* H5194 recombinant strain for iron, under aerobic conditions where it inhibited the ICLs. However, that was not confirmed for copper in both strains (**Figure 18G**). The quantification of the bands showed that such reduction was only significant ($p < 0.005$) for iron II in co-incubation with M 1/5 compared to the M9_Cap control media (**Figure 18B and D**). Similar CFU/mL number of Nissle 1917 and M 1/5 under anaerobic conditions and in the presence of the different

cofactors also discarded any difference associated to the bacteria viability (**Figure 18E-F**).



C



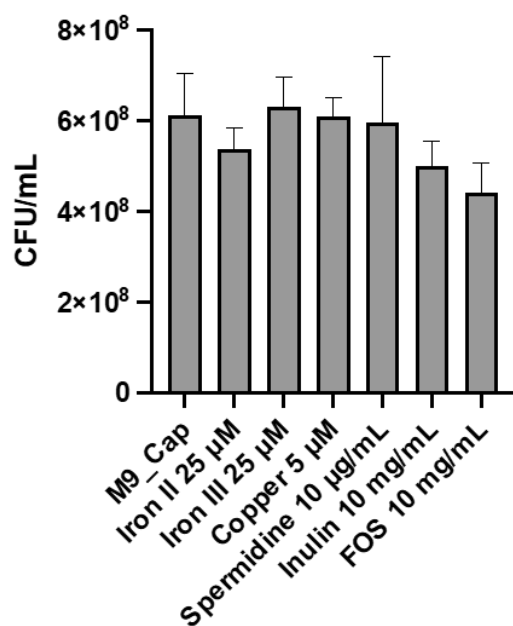
D

Aerobic respiration control protein ArcA

Escherichia_coli_Nissle_1917_pro	MQTPHILIVEDELVTRNTLKSIFEAEGYDVFEATDGAEMHQILSEYDINLVIMDINLP GK
Escherichia_coli_str NC_010473.1	MQTPHILIVEDELVTRNTLKSIFEAEGYDVFEATDGAEMHQILSEYDINLVIMDINLP GK
Escherichia_coli_strain_M15_prot	MQTPHILIVEDELVTRNTLKSIFEAEGYDVFEATDGAEMHQILSEYDINLVIMDINLP GK
Escherichia_coli_Nissle_1917_pro	NGLLARELREQANVALMFLTGRDNEVDKILGLEIGADDYITKPFNPRELTIRARNLLSR
Escherichia_coli_str NC_010473.1	NGLLARELREQANVALMFLTGRDNEVDKILGLEIGADDYITKPFNPRELTIRARNLLSR
Escherichia_coli_strain_M15_prot	NGLLARELREQANVALMFLTGRDNEVDKILGLEIGADDYITKPFNPRELTIRARNLLSR
Escherichia_coli_Nissle_1917_pro	TMNLGTVSEERRSVESYKFNWELDINSRSLIGPDGEQYKLP RSEFRAMLHFCENPGKIQ
Escherichia_coli_str NC_010473.1	TMNLGTVSEERRSVESYKFNWELDINSRSLIGPDGEQYKLP RSEFRAMLHFCENPGKIQ
Escherichia_coli_strain_M15_prot	TMNLGTVSEERRSVESYKFNWELDINSRSLIGPDGEQYKLP RSEFRAMLHFCENPGKIQ
Escherichia_coli_Nissle_1917_pro	SRAELKKMTGRELKP HORTVDVTIRIRKHFESTPDTPEIIATIHGEGYRFGDLED
Escherichia_coli_str NC_010473.1	SRAELKKMTGRELKP HORTVDVTIRIRKHFESTPDTPEIIATIHGEGYRFGDLED
Escherichia_coli_strain_M15_prot	SRAELKKMTGRELKP HORTVDVTIRIRKHFESTPDTPEIIATIHGEGYRFGDLED

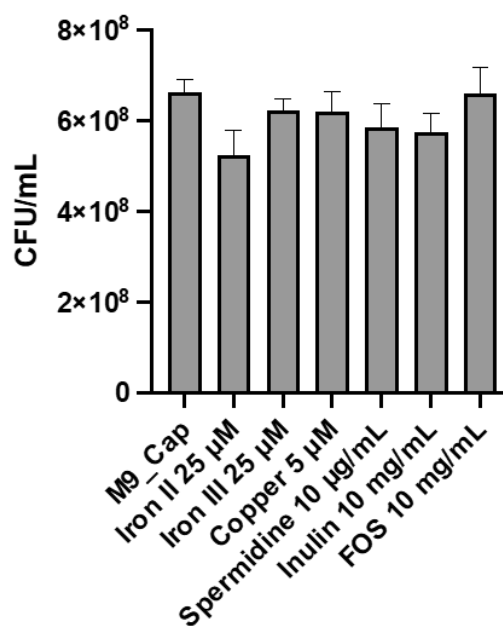
E

Anaerobic test
E. coli Nissle 1917



F

Anaerobic test
E. coli M 1/5



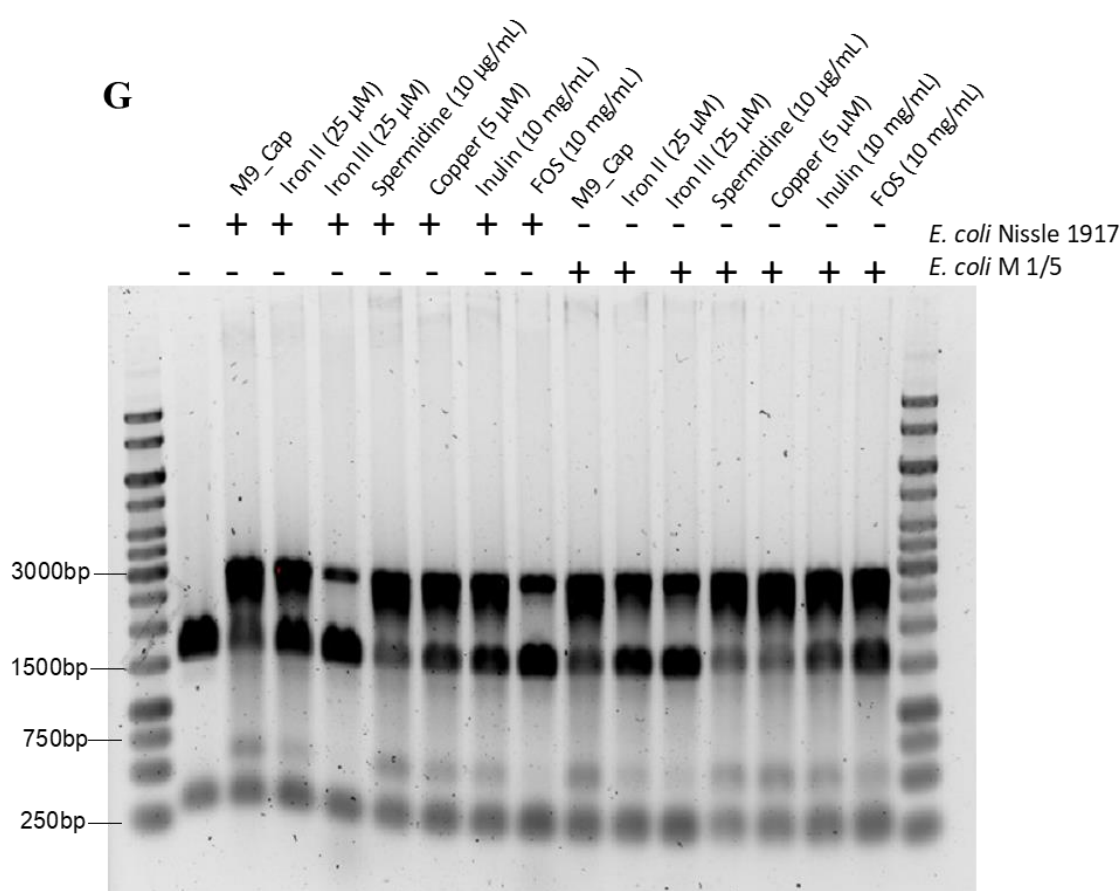


Figure 18- Circular and linear representation of the whole-genome comparison of DH10B, Nissle 1917 and M 1/5. This figure integrates whole-genome alignments with functional assessments to investigate strain-specific differences in gene content, regulatory proteins, and colibactin-mediated DNA damage under varying cofactor conditions. Figure 18 was created in collaboration with Dr. K. Arellano-Ayala.

(A) The outer and inner lines represent the scale in kbp. Circle 1 (red): all genes of M 1/5; circle 2 (brown): all genes of Nissle 1917; circle 4 (orange): all genes of DH10B. Circle 5 (black line): guanine and cytosine content. Circle 7 (multicolored): clusters of orthologous groups. Figure generated by web-program Gview server.

(B) Table of BLAST hits for Aerobic respiration two-component sensor histidine kinase ArcB and two-component system response regulator ArcA.

(C) Venn diagram for the comparison of DH10B, Nissle 1917 and M 1/5. Numbers inside each region indicate the number of genes contained. The intersection of the 3 circles contains the orthologous genes for the 3 genomes with 3607 genes. The outer part of the circle represents the specific genes for each strain, corresponding to 32 genes for DH10B, 52 genes for Nissle 1917 and 55 genes for M 1/5.

(D) Protein sequence comparison for the aerobic respiration control protein ArcA of DH10B, Nissle 1917 and M 1/5.

(E) *pks*⁺ Nissle and *pks*⁺ M 1/5 (F) viability on CFU/mL at logarithmic phase on M9_Cap media supplemented with different cofactor concentrations under anaerobic conditions.

(G) CDM14/AATTTT inter-strand cross link visualization, 600 bp, on 1% alkaline gel after 20 minutes co-incubation with *pks*⁺ Nissle and *pks*⁺ M 1/5 under anaerobic conditions on the presence of different cofactor concentrations. The bacteria were cultured on M9_Cap media with the cofactors at exponential phase.

Part II: The presence of colibactin damage motifs in colon cancer precursors

Having established the sequence specificity of colibactin-induced DNA damage in vitro and validated the utility of Nanopore sequencing for detecting inter-strand crosslinks, we next sought to investigate the role of colibactin in colorectal cancer (CRC) precursors, including inflammatory bowel disease (IBD) and colonic polyps. Therefore, we developed a patient-derived 3D organoid model to systematically analyze colibactin-associated mutational signatures in a physiologically relevant system

3.15 Establishment of 3D Human Colon Organoids

The first aim of this project was to extend the analyses of sequence specificity of colibactin on naked DNA, we validated the canonical colibactin sequences via gel electrophoresis. We also detected new colibactin binding sites and confirmed Nanopore sequencing could be novel tool to detect ICLs. Therefore, we would like to investigate the role of the genotoxin colibactin in initiating colorectal cancer (CRC). To achieve this, we established organoids derived from precancerous conditions, such as inflammatory bowel disease (IBD) and the presence of polyps, and subjected these tissues to whole-genome sequencing. Three-dimensional (3D) organoid models were selected as the optimal platform for this study due to their ability to recapitulate the diverse cell types of the primary epithelium, making them particularly valuable for infectious disease research.

A 3D human colon organoid model was developed, based on the method initially reported by Sato et al. (Sato, Vries et al., 2009). They demonstrated that epidermal growth factor (EGF), essential for stem cell proliferation; Noggin, a BMP antagonist; and R-spondin, a Wnt agonist, are critical for organoid growth. Notably, Wnt signaling is dispensable for the growth of small intestinal organoids, as Paneth cells within the culture endogenously produce Wnt ligands. However, the colon epithelium lacks Paneth cells, making colon organoid cultures reliant on external Wnt supplementation and highly sensitive to its concentration.

This challenge was addressed by using conditioned media including WRN, R-spondin and Noggin¹⁰⁷. This approach ensured the successful establishment of a reliable and

reproducible 3D human colon organoid model for studying colibactin's role in CRC initiation.

Crypts were isolated from fresh patient biopsies and cultured into spherical organoids (**Figure 19A**). Under optimal conditions, human organoids can be maintained in culture without limitation. Immunofluorescence staining confirmed the presence of stem cells within the organoids (**Figure 19C**), demonstrating their successful preservation and representation in the 3D culture system.

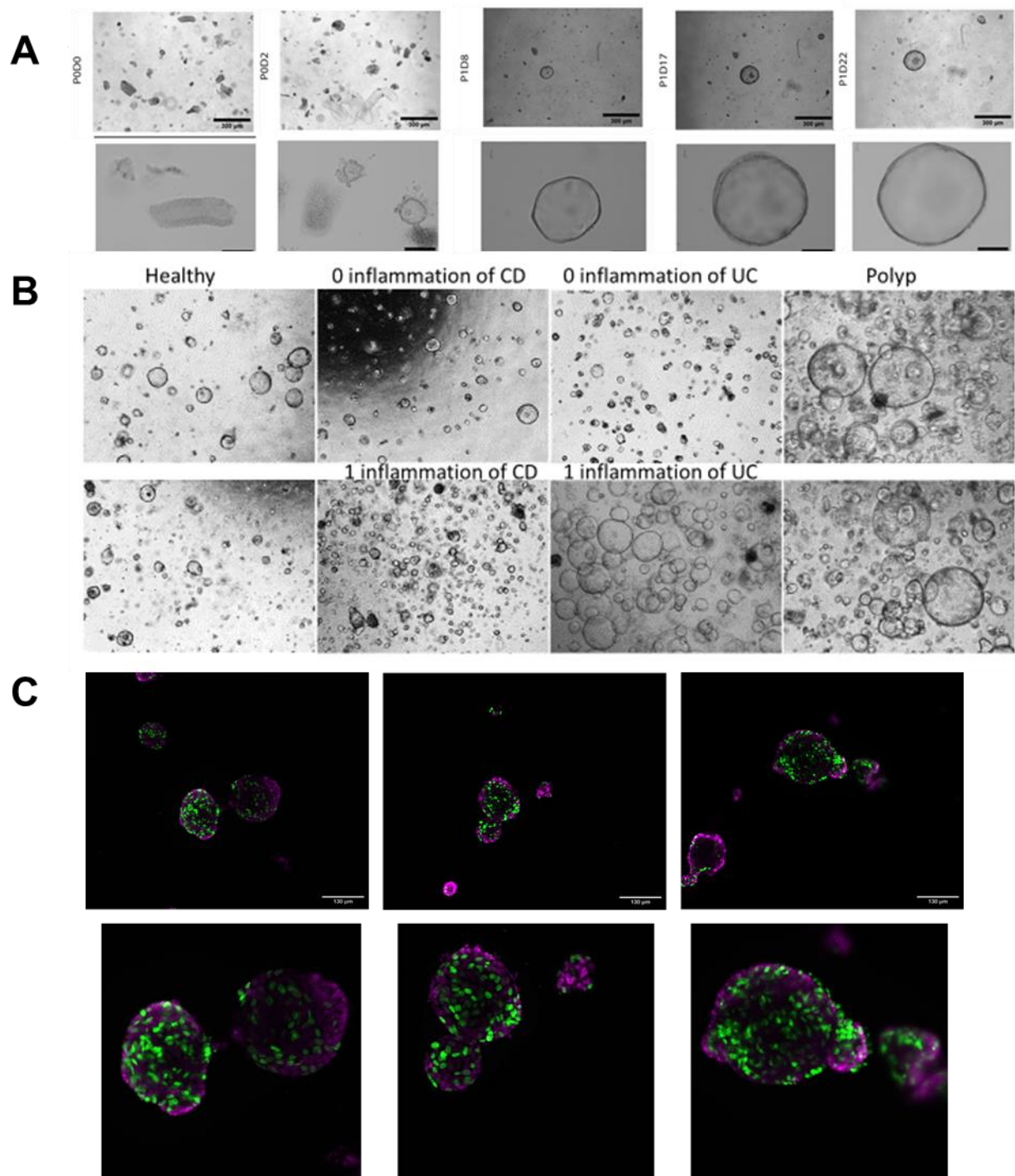


Figure 19- Establishment and Characterization of Human Colon Organoid Models. This figure demonstrates the successful isolation, long-term culture, and stem cell characterization of patient-derived 3D human colon organoids. Figure19A were created by Dr. D. Holthaus.

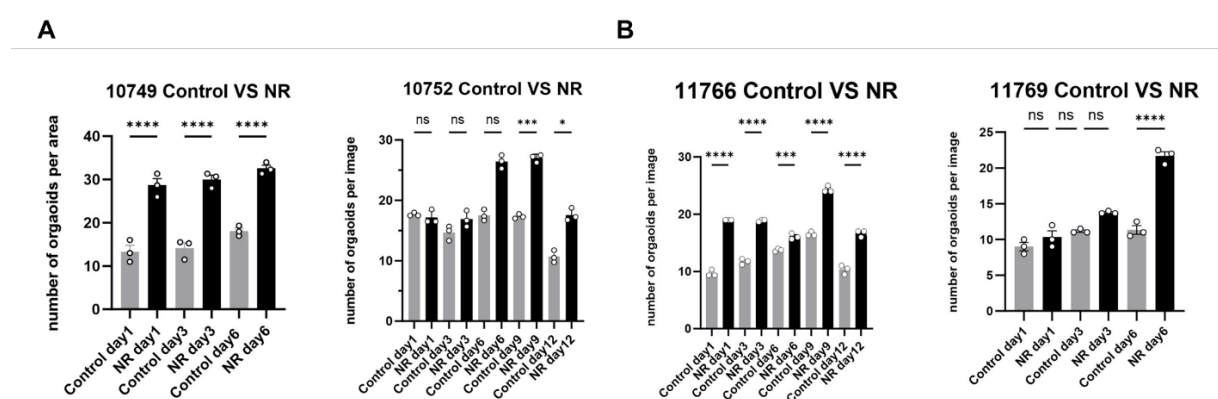
(A) Isolated epithelial crypts from human colon biopsies developed into self-organizing 3D organoids when cultured in defined growth medium. Organoids were maintained and expanded for up to four weeks post-isolation (passage 0, P0). Scale bar: 300 μ m.

(B) A panel of organoid lines was established from diverse clinical sources, including healthy colonic mucosa and precancerous tissues such as inflammatory bowel disease (IBD) mucosa and colonic polyps, enabling comparative studies of epithelial behavior under different disease contexts.

(C) EdU staining revealed proliferative stem cell populations within the organoid structure (green), while DAPI staining (purple) marked nuclear DNA, confirming the cellular integrity and spatial organization of the epithelial architecture. Scale bars: 130 μ m and 330 μ m.

3.16 Improvement of 3D organoids culture

Since NAD⁺ repletion can directly improve mitochondrial function and cellular fitness, it might be assumed that the insufficient provision of NAD⁺ via QPRT might impact stem cell function and wound repair in intestinal inflammation. Therefore, research on NAD⁺ rescue could be done.



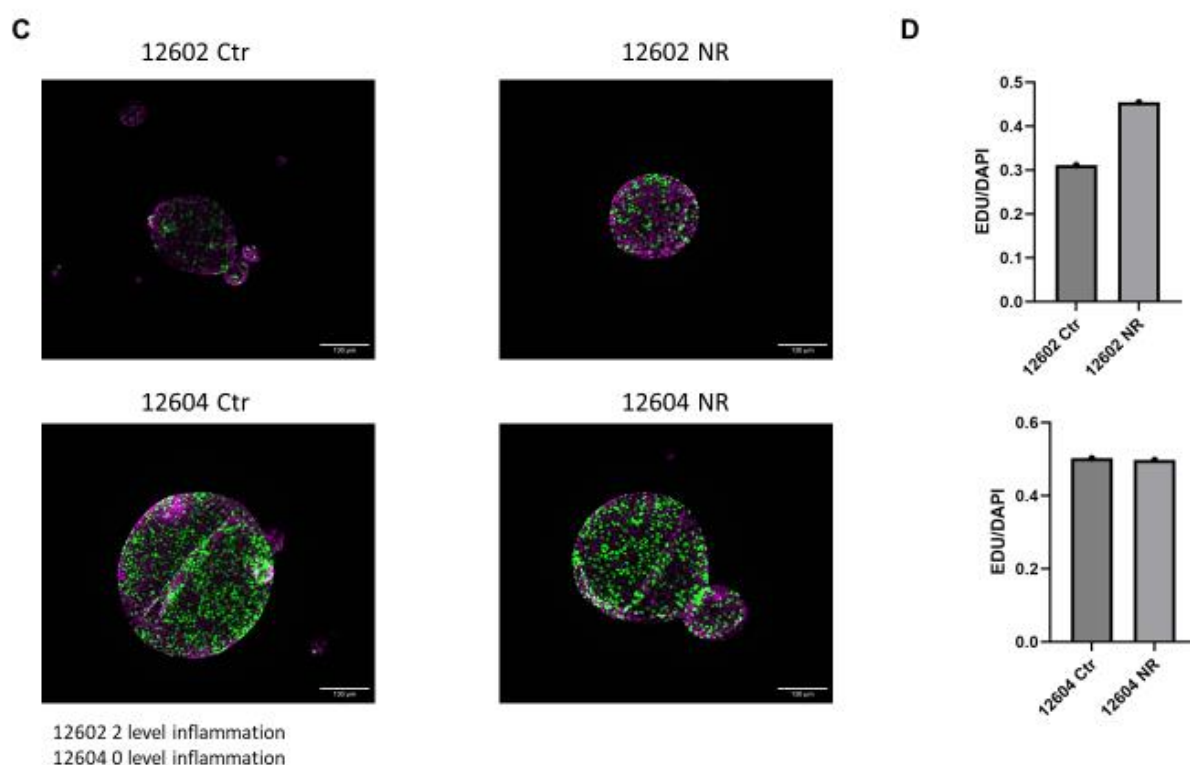


Figure 20- Nicotinamide riboside enhances the growth of patient-derived colonic organoids over time. This figure demonstrates that treatment with nicotinamide riboside (NR) promotes proliferation and expansion of colonic organoids derived from patients with varying levels of inflammation. Fig20C were created jointly with H. Xiang.

(A–B) Quantification of organoid formation from two matched patient-derived organoid line pairs:

- 11766 (inflammation level 0) and 11769 (inflammation level 2) from the same ulcerative colitis (UC) patient
- 11749 (inflammation level 0) and 11752 (inflammation level 2) from another UC patient

Bar graphs show the number of organoids per image or area at days 1, 3, 6, 9, and 12 following passage and treatment with either control or NR. NR treatment significantly enhanced organoid growth over time, especially at later time points and in organoids derived from inflamed tissues.

(C) Representative confocal images of 3D organoids from additional patient lines 12602 and 12604 treated with NR or control. EdU (green) marks proliferating cells; DAPI (purple) stains nuclei. NR-treated organoids show visibly increased EdU incorporation, suggesting enhanced proliferation. Scale bar = 100 μ m.

(D) Bar graphs quantify EdU incorporation as the EdU/DAPI ratio, indicating the proportion of proliferative cells in each condition. NR-treated organoids showed significantly increased EdU incorporation compared to controls across both patient lines.

Data are presented as mean $\pm$ SEM from three biological replicates per condition. Statistical analysis was performed using two-way ANOVA followed by Tukey's multiple comparisons test. Significance is denoted as follows: $p < 0.05$, $p < 0.01$, $p < 0.001$, $p < 0.0001$; ns = not significant.

Nicotinamide riboside (NR) treatment significantly increased organoid formation in ulcerative colitis patient-derived colonic lines. In patient line 11766, the number of organoids progressively increased from day 3 through day 12 following NR treatment, particularly after two passages, with a marked enhancement evident from day 6 onward ($p < 0.0001$; **Figure 20A**). Similarly, in line 11769, NR induced a robust increase in organoid formation as early as day 3 ($p < 0.05$), which became even more pronounced by day 6 ($p < 0.0001$). Comparable effects were observed in lines 11752 and 11749, reflecting a consistent pro-proliferative response across distinct patient backgrounds and levels of inflammation. Notably, while organoid expansion plateaued under control conditions, NR-treated cultures continued to grow, indicating a sustained regenerative effect. This response was especially evident in lines 11769 and 11752 (**Figure 20A and B**), both derived from tissue with moderate (level 2) inflammation, suggesting a greater sensitivity to NR under inflammatory conditions. These findings implied that NR may not only promote epithelial regeneration but also enhance organoid viability and expansion efficiency in 3D culture systems derived from inflamed colonic tissue.

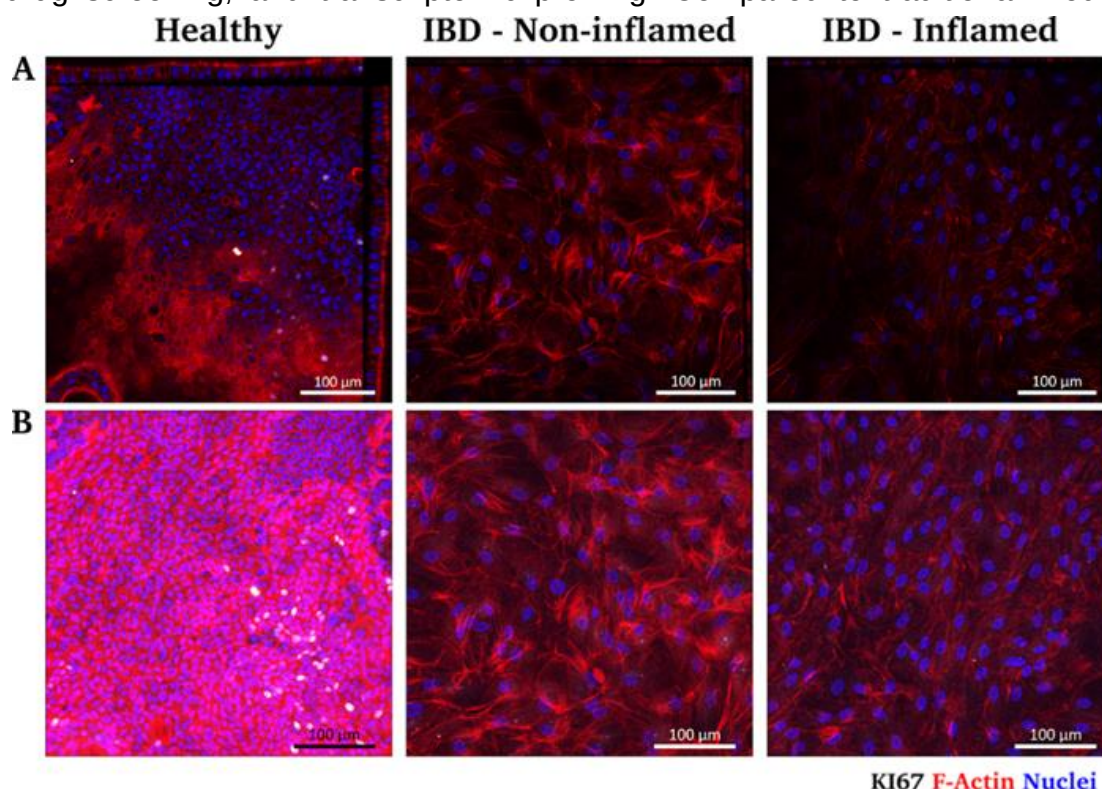
To evaluate the effect of Nicotinamide riboside (NR) on epithelial cell proliferation, EdU incorporation assays were performed on 3D colonic organoids derived from the same ulcerative colitis patient, representing distinct inflammation levels: line 12602 (level 2 inflammation) and line 12604 (non-inflamed, level 0). In 12602, NR treatment markedly increased the number of EdU-positive cells, as shown by enhanced EdU staining and a significant elevation in the EdU/DAPI ratio compared to the untreated control (**Figure 20C and D**). This finding suggested that NR effectively promotes epithelial proliferation in inflamed organoids. By contrast, organoids from the non-inflamed line 12604 exhibited a high baseline level of proliferation, and NR treatment did not lead to a further increase in EdU incorporation, which indicated that the proliferative response to NR may be inflammation-dependent, with greater sensitivity observed in organoids derived from inflamed tissue.

Collectively, the data demonstrated that Nicotinamide riboside (NR) enhances epithelial regenerative capacity by promoting proliferation and increasing organoid-forming efficiency, particularly in tissues with reduced basal proliferation due to inflammation. Notably, this effect was most evident after passaging, indicating that NR may restore or maintain stemness in inflamed epithelial tissues. These findings

highlight NR's potential as a targeted therapeutic strategy for rescuing stem/progenitor cell function under inflammatory conditions, offering a promising avenue for personalized regenerative therapies in chronic intestinal disorders such as ulcerative colitis.

3.17 Establishment of organoid-derived monolayers

Three-dimensional intestinal organoid cultures are *in vitro* models that have shown significant potential for modeling human intestinal physiology and disease. Traditional intestinal organoid systems, while faithfully mimicking the *in vivo* architecture, are difficult to access from the apical or luminal side of the epithelium. Therefore, it is not optimal to use 3D organoids when apical access is required, i.a. for drug screening and microbial infection purposes. We would like to establish organoid-derived 2D monolayer cultures enable the direct access to both the apical and basolateral compartments that are divided by a functional epithelial barrier (**Figure 22A**). Individual organoid lines and organoid-derived monolayers showed patient, disease, inflammation-specific characteristics (**Figure 21A and B**) that make them suitable for personalized medicine applications such barrier function assays, microbial co-culture, drug screening, and transcriptomic profiling. Compared to traditional models, this



approach remains patient-specific epithelial characteristics and allows side-by-side comparisons of disease- and non-disease-associated tissues under controlled conditions.

Figure 21- Immunofluorescence of organoid-derived monolayers. This figure provides a comparison of immunofluorescence staining of healthy and IBD-affected organoids, categorized as "Healthy," "Healthy-IBD," and "Inflamed-IBD." The ODMs show clear differences in height and intensity of F-Actin staining, with IBD samples being flatter and less compartmentalized than the healthy control. Inflamed regions also show less F-Actin than non-inflamed regions of the same patient. Images are presented as orthogonal (A) and maximum intensity projections (B) of z-stacks. (Projection: A 3Dprojection to demonstrate the overall spatial arrangement of the organoid) Healthy organoids show dense nuclei and organized F-Actin with scattered Ki67-positive cells, indicating controlled proliferation. "Healthy-IBD" organoids display reduced Ki67 staining, suggesting decreased proliferation, along with altered actin organization. "Inflamed-IBD" organoids reveal disorganized actin, fewer proliferative cells (lower Ki67), and disrupted cellular architecture, consistent with inflammation-related tissue damage. Scale bars are 100µm. IF staining and pictures taken in cooperation with D. Holthaus (Department of Department of Gynecology and Obstetrics). Figure 21 was created jointly with Dr. D. Holthaus.

3.18 Validation of organoid-derived monolayers regarding barrier dysfunction

In order to investigate the effect of inflammation on epithelial barrier function, we compared monolayers derived from colonic tissues with inflammation level 0 (non-inflamed) and level 3 (highly inflamed). Western blot analysis revealed a marked reduction in ZO-1 expression in inflamed samples compared to non-inflamed counterparts (**Figure 22C and D**), suggesting impaired tight junction integrity. Functional barrier assessment using TEER showed that while both non-inflamed and inflamed monolayers developed increasing resistance over time, the inflamed group consistently exhibited lower TEER values throughout the 12-day period (**Figure 22B**), indicating compromised epithelial sealing capacity. The expression of tight junction proteins varied across patient lines; in some cases (e.g., patient 13297/13301), inflammation was associated with reduced ZO-1 and increased occludin levels, suggesting patient-specific alterations in epithelial barrier components (**Figure 22D and E**). Immunofluorescence staining detected more expression of Ki67 proliferation marker in non-inflamed monolayers compared to inflamed monolayers (**Figure 23A and B**). These results collectively demonstrated that inflammation disrupts epithelial barrier maturation and tight junction organization in organoid-derived monolayers.

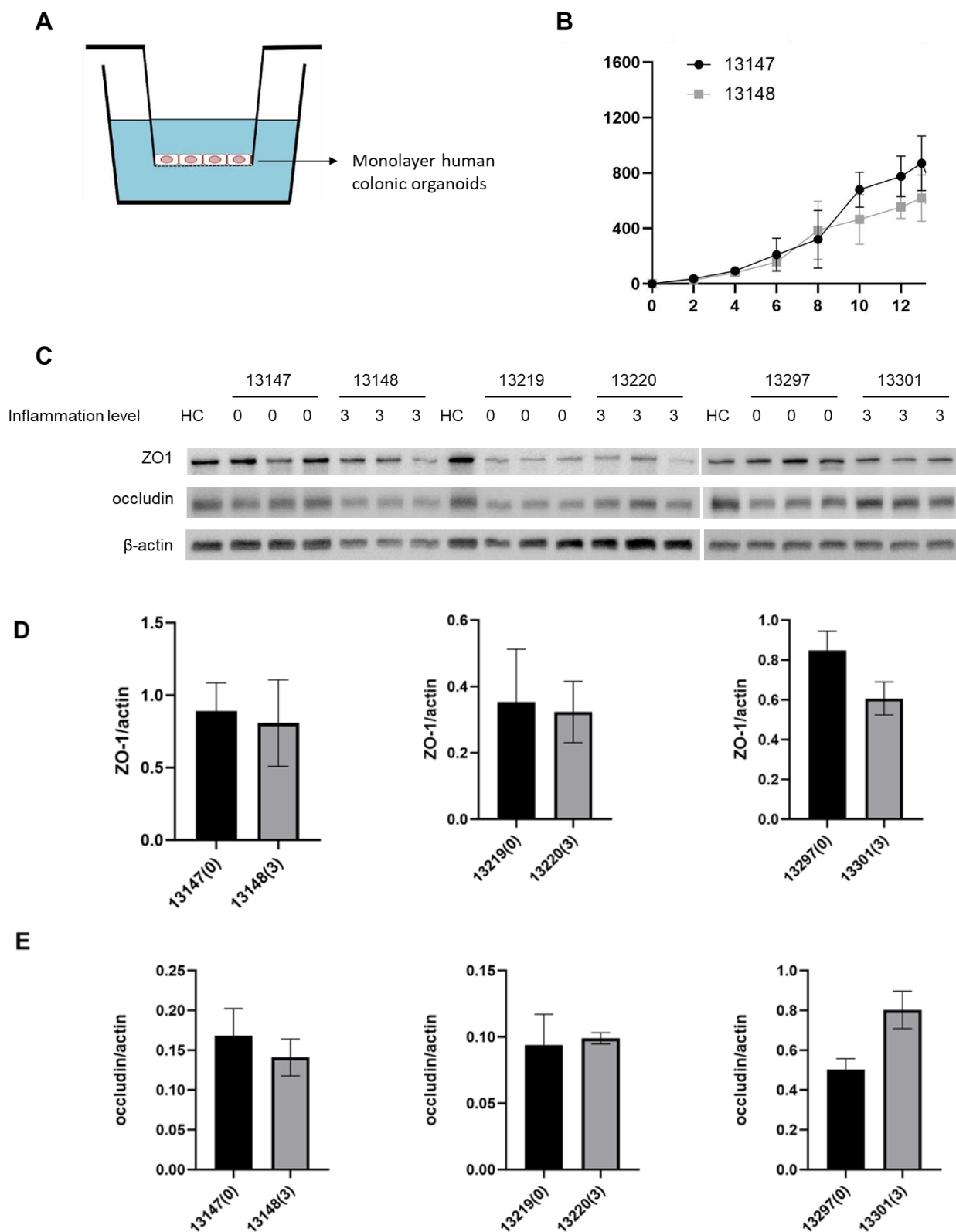


Figure 22- Inflammation impairs intestinal epithelial barrier function and tight junction protein expression.

Figure 22C was generated jointly with X. Zhang.

(A) Western blot analysis of Occludin in monolayers derived from tissues with inflammation levels 0 and 3. β -actin served as a loading control. Occludin expression was reduced in highly inflamed samples (level 3).

(B) Transepithelial electrical resistance (TEER) measurements over 12 days in organoid-derived monolayers from non-inflamed (black line) and inflamed (gray line) tissues. Barrier integrity increased progressively in both groups, but monolayers from inflamed tissue exhibited significantly lower TEER values across timepoints.

(C) Immunofluorescence staining of monolayers at day 12. Occludin (green), nuclei (DAPI, blue), and F-actin (phalloidin, red) were visualized. Monolayers from inflamed tissue (right) displayed irregular occludin localization and disorganized cytoskeletal architecture compared to non-inflamed controls (left). Scale bar = 50 μ m.

(D-E) Bar graphs represent Western blot quantification of ZO-1 (D) and occludin (E) normalized to β -actin in organoid-derived monolayers from three individual patients. Each pair of bars compares protein expression in monolayers derived from tissues with inflammation level 0 (black bars) and level 3 (gray bars). Data are presented as mean $\pm$ SEM.

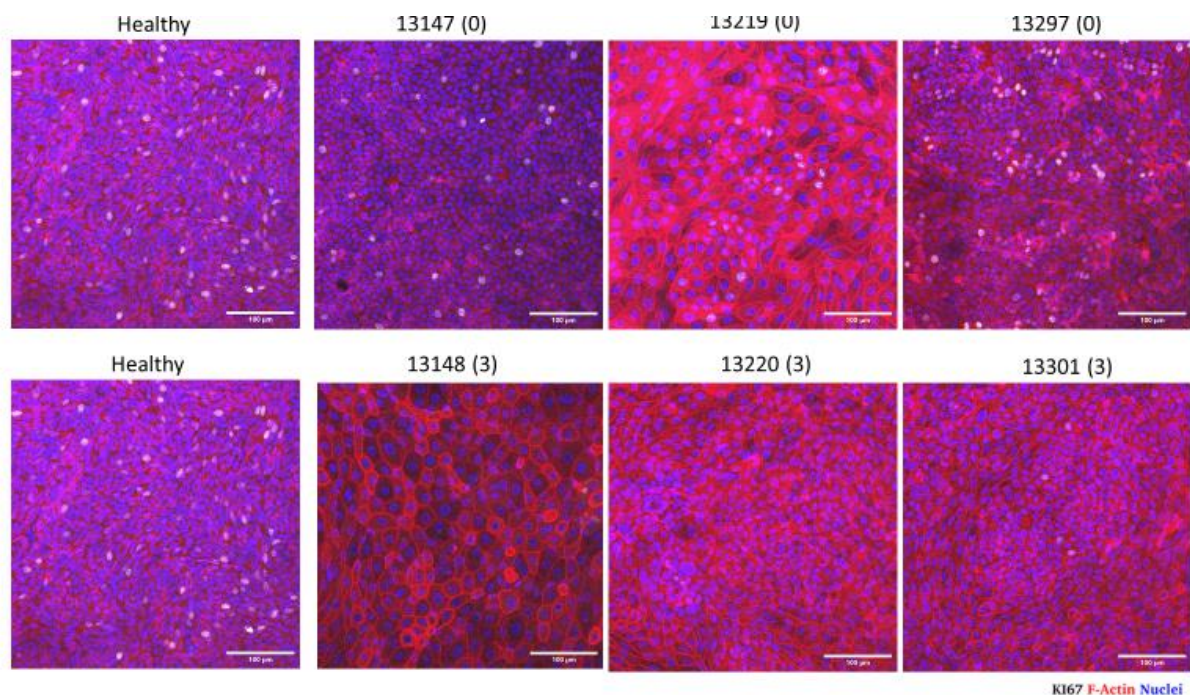


Figure 23- Immunofluorescence staining of colon epithelial monolayers. (A) A maximum intensity projection of a z-stack is a 2D image that shows the brightest pixels from each layer of a 3D image stack, making it easier to visualize complex structures in a single image. The monolayers derived from healthy individuals and IBD patients with inflammation scores 0 (top row) and 3 (bottom row). Ki67-positive nuclei (white) indicate proliferating cells; phalloidin (F-actin) staining is shown in red, and nuclei (DAPI) in blue. Healthy and score 0 samples exhibit more Ki67 expression, while inflammation score 3 samples show decreased nuclear Ki67 signal. Scale bar = 100 μ m. Figure 23 was generated jointly with Dr. D. Holthaus.

3.19 Whole genome sequencing of human colon organoids

Mutational signatures, mathematically defined as distinct patterns of somatic mutations within the genome ¹¹⁰, ¹¹¹, provide critical insights into the underlying mutational processes, whether external or internal, and contribute significantly to understanding the etiology of various diseases, including cancer. The concept was first introduced in 2012 through an analysis of substitution mutations in 21 whole-genome-sequenced (WGS) breast cancers. This landmark study revealed consistent patterns of mutagenesis across tumors, establishing the foundation for identifying and characterizing distinct mutational processes ¹¹².

To investigate the underlying mutational processes contributing to the somatic mutation landscape in IBD derived paired organoids, mutational signature deconvolution was performed using established COSMIC single base substitution (SBS) signatures. The analysis included 12 paired comparisons of IBD organoids including noninflamed and inflamed organoids from same patient and 1 pair of polyp organoids as control. Results are visualized in two complementary panels, providing both qualitative and quantitative insights into the counts of mutations (**Figure 24A**) and the proportion of mutations attributed to specific mutational signatures (**Figure 24B**).

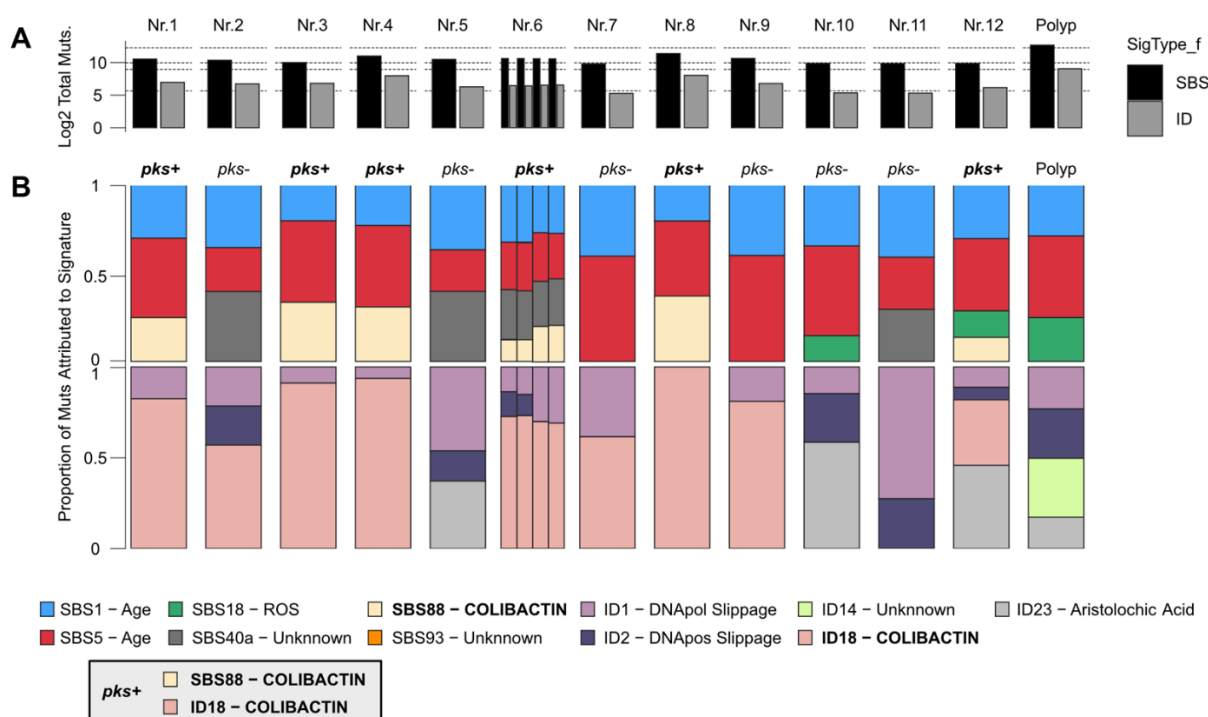


Figure 24- Mutational signature analysis across four paired samples. This figure illustrates the distinct

mutational signatures identified in organoid-derived samples, highlighting the enrichment of colibactin-associated patterns (SBS88, ID18) specifically in *pkS*⁺ conditions. Figure 24 were created by A. Quevedo-Olmos.

(A) Total mutation counts (log₂ scale) separated by single base substitutions (SBS, black) and small insertions/deletions (ID, gray) across 12 patient-derived organoid lines and one polyp sample.

(B) Stacked bar plots showing the proportion of mutations attributed to specific mutational signatures in each sample. Signatures include age-related (SBS1, SBS5), reactive oxygen species–associated (SBS18), colibactin-associated (SBS88, ID18), DNA polymerase slippage (ID1, ID2), and other known or unknown etiologies (e.g., SBS40a, SBS93, ID14, ID23). Notably, SBS88 and ID18—colibactin-associated signatures—were enriched in several *pkS*⁺ samples, indicating the mutagenic imprint of colibactin in inflamed colon epithelium. The polyp sample exhibited a distinct mixture of age- and inflammation-associated signatures.

Whole-genome sequencing (WGS) was performed on patient-derived colonic organoid lines established from both inflamed and non-inflamed tissues to assess the mutational consequences of exposure to *pkS*⁺ *Escherichia coli*. The total number of mutations varied considerably across samples, as illustrated in **Figure 24A**, with a general trend of higher mutational burdens observed in *pkS*⁺ organoids compared to *pkS*[−] controls. Mutations were categorized into single base substitutions (SBS) and small insertions/deletions (ID), both of which showed differential accumulation.

Mutational signature deconvolution (**Figure 24B**) revealed that well-known background signatures, including SBS1 and SBS5 (aging-associated) and SBS18 (reactive oxygen species), were present across all samples regardless of bacterial exposure. However, two mutational signatures strongly associated with colibactin—SBS88 and ID18—were observed. The enrichment of SBS88 was particularly evident in inflamed-derived lines, supporting the hypothesis that epithelial barrier dysfunction during inflammation increases exposure to microbial genotoxins.

These findings collectively support a model in which colibactin exposure leaves a reproducible genomic imprint in patient-derived colon organoids, linking microbial genotoxicity with inflammation-associated mutagenesis. The identification of SBS88 in this context strengthens its role as a biomarker of colibactin activity *in vivo* and demonstrates the utility of organoid-based WGS for studying host–microbe interactions in colorectal cancer initiation.

3.20 Re-estimation of the true oncogenic potential of colibactin

The estimation of colibactin mutagenesis is based on sequence analysis, mainly by comparing the sequence context of the mutations with the canonical CDM AAWWTT. However, a set of validated colibactin binding sequences was missing. Besides, our results showed that colibactin binds sequences outside the established damage motif. For that reason, I used the validated CDMs to evaluate the role of colibactin in CRC mutagenesis.

I started by focusing in the 8 CDMs with the highest $\Delta\text{PropQ} < 10$ values (Underlined in Figure 33D), which contain 2 sequences not matching the canonical CDM (AAAGTT, AAAATC). Their distribution across chromosomes in the human genome is uniform and correlated with chromosome length (**Figure 25A**). As tumorigenesis is driven by mutations in genes (Martínez-Jiménez et al., 2020), I focused on those that lie within transcribed regions (**Figure 25B**). Between 2 to 4 million CDMs are located within transcripts, but, as not all genes are cancer related genes nor the affected positions within the gene are equally important for malignant transformation, I had a closer look at mutations with clinical significance in CRC genes (ClinVar database on dbSNP v.155 CITE). For each CDM, approximately 100 clinically relevant variants affecting CRC genes were found (**Figure 25C**). When I looked at the genes affected by those variants (**Figure 25D**), I observed that some of them contain mutations located in the context of all top 8 CDMs, like APC, the most frequently mutated gene in CRC. Others, in contrast, contained mutations in just one of the top ranked motifs. For example, there was an insertion located within the AAAGTT motif of AXIN2 (rs1056103847), whose product acts as an inhibitor of WNT signaling, which was associated with a CRC predisposition syndrome (OMIM #608615). This insertion created a frameshift that leads to RNA nonsense mediated decay and loss of Axin-2 protein, therefore preventing WNT inhibition (**Figure 25E**). Several deleterious mutations in those new motifs were present in other genes like APC, the most frequently mutated gene in CRC, and also in the mTOR inhibitor and tumor suppressor TSC2 (**Figure 25E**). In summary, this analysis showed that more CRC related mutations can be attributed to colibactin after defining its target sequences.

Which is the relevance of the new CDMs in tumor samples? Are mutation rates in CRC samples higher in those colibactin targets than in other sequences with no evidence of colibactin binding? To answer this question, I computed the normalized mutation frequencies at different hexamer sequences using CRC data from TCGA as described previously (Dziubańska-Kusibab et al., 2020). We defined 3 groups based on the CDM sequence: 1) the canonical CMD matching sequences AAAATT/AATTT 2) The new CDMs AAAGTT/AAAATC and 3) random hexamers with 1 C or G allowed. The normalized mutation frequencies were significantly higher for the canonical and new CDMs and not for the random CDMs, confirming the new sequences as colibactin targets in vivo (Figure 21F). Moreover, when looking at the distribution of double strand breaks generated by colibactin in human cells in vivo (Dziubańska-Kusibab et al., 2020), both AAAGTT and AAAATC gave rise to similar location of DSB as the canonical motif AAAATT, with preferred break positions after the second adenine and before the thymine in position 5 (**Figure 25G**).

Overall, the present analyses demonstrate that colibactin is able not only to bind AT only sequences, but also G and C nucleotides at certain positions still are targeted by colibactin. In all cases, colibactin binding leads to increased mutation frequencies at such positions, that could promote tumorigenesis when affecting tumor driver genes.

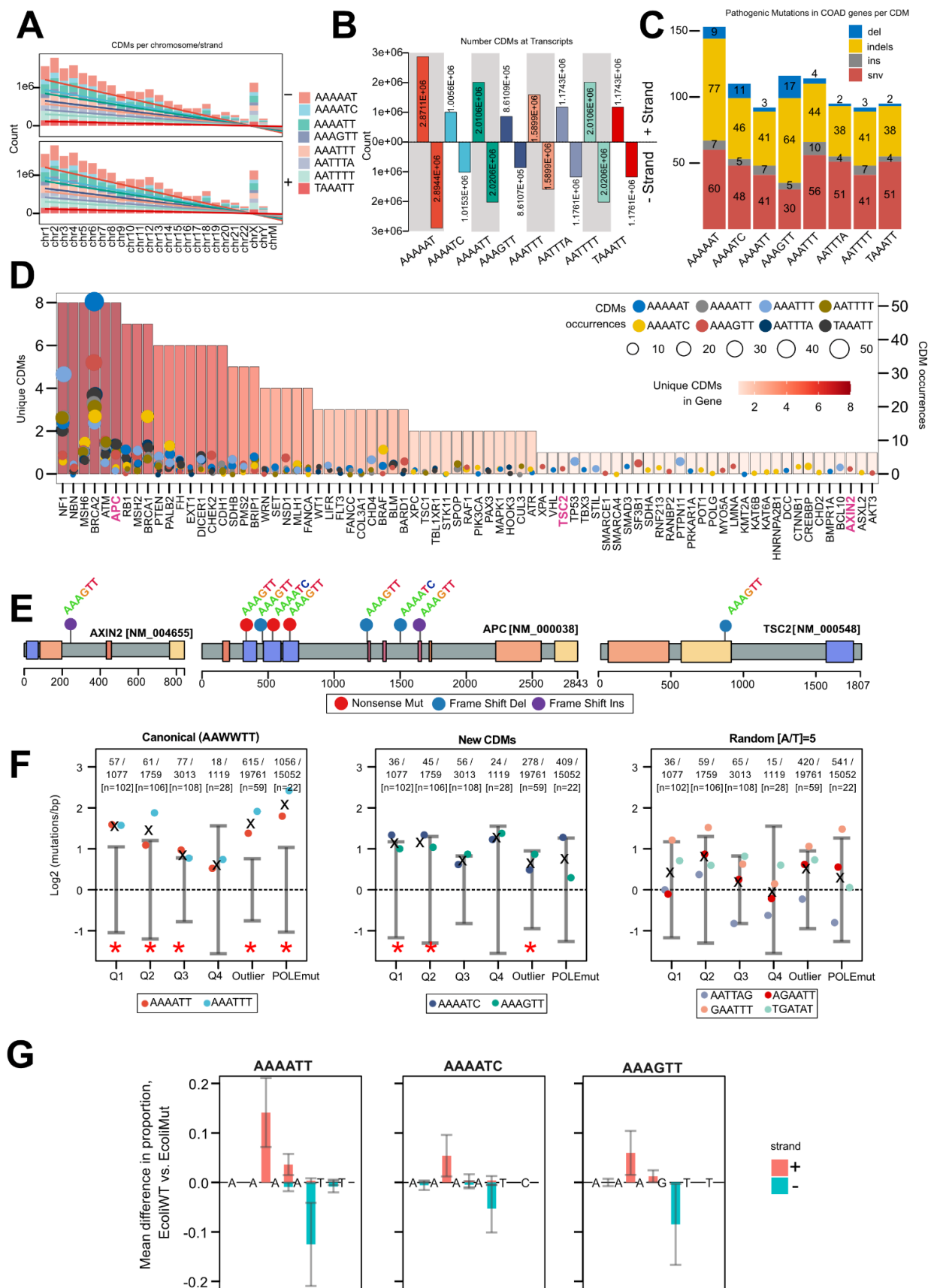


Figure 25- Re-estimation of colibactin carcinogenic potential in colorectal cancer This figure comprehensively illustrates the genomic enrichment, mutational impact, and DNA breakage potential of canonical and novel colibactin damage motifs (CDMs), highlighting their relevance in colorectal cancer pathogenesis. Figure 25 was created by A. Quevedo-Olmos.

(A) Occurrences of the top ranked CDMs underlined in orange in Figure 3D in the human chromosomes. Lines show the linear regression slopes of chromosome lengths.

(B) Occurrences of top ranked CDMs within canonical gene transcripts split by strand.

(C) Number of mutations with clinical relevance located within CDMs at colorectal cancer related genes.

(D) Colon cancer related genes containing any of the top CDMs (left y-axes) within their canonical transcript and the number of CDM occurrences (right y-axes).

(E) Lollipop plots of the 3 genes highlighted in pink from panel D that depict deleterious mutations at the C/G containing CDMs AAAGTT or AAAATC.

(F) Normalized mutational frequencies in colorectal cancer whole exome sequences from TCGA at Canonical (AAATT/AAATTT), New CDMS (AAAGTT/AAAATC) and random hexanucleotides with AT content = 5. Asterisks - significant difference (two-sided Mann-Whitney-U test, $p < 0.05$ and $FDR < 20\%$) between CDMs and: 1- all WWWWW motifs for canonical CDMs, 2- all other AT=5 hexamers for new and random CDMs. Error bars = $0 \pm 2MAD$ intervals of mutation rates (mutations/base-pair) of WWWWW sequences, excluding CDMs, after subtracting their mean. Dots - mutation rates for tested hexanucleotides after mean subtraction. Crosses - means of the CDMs. Q1-Q4 cohorts defined by the quantiles of total SNV numbers. Outliers - samples with SNV numbers significantly larger than the 95% CI in each cancer entity. POLE - polymerase epsilon mutated cases.

(G) Mean differences in proportions of double strand break ends detected using BLISS in *pks*⁺ vs *pks*⁻ infected cells for the canonical, AAAATT, and the G/C containing CDMs ($n = 4$ independent experiments). All 3 motifs belong to the same cluster identified using K-means ($k=4$) based on the mean difference values across the hexamer, characterized by increased proportion of breaks after the 2nd and before the 5th nucleotide.

4. Discussion

Colorectal cancer (CRC) is one of the leading causes of cancer-related mortality worldwide, arising from the interplay of genetic predisposition and environmental influences ¹¹³. Inflammatory bowel diseases (IBD), such as ulcerative colitis and Crohn's disease, are well established risk factors, where barrier dysfunction and persistent inflammation create a microenvironment favorable for carcinogenesis. Among microbial contributors, colibactin-producing bacteria e.g. *pks+* *Escherichia coli* have gained attention. Colibactin, a polyketide–nonribosomal peptide genotoxin, induces DNA inter-strand crosslinks and generate a distinct mutational signature (SBS88) in host cells, linking microbial activity to cancer initiation ^{74,93,114}.

Here I investigated the role of colibactin in CRC initiation through a multi-layered approach, using in vitro DNA assays, nanopore sequencing, and patient-derived colon organoids with whole-genome sequencing. At the molecular level, biochemical assays defined colibactin damage motifs (CDMs), confirming canonical sites (AAAATT, AAATTT) and identifying novel motifs such as AAAGTT. At the environmental level, cofactors including iron, copper, oxygen, and dietary inputs were shown to regulate colibactin production or activity. At the tissue level, organoid models revealed that colibactin-associated mutational signatures (SBS88) arise specifically in inflamed epithelium, where compromised barrier function and impaired DNA repair promote lesion fixation. Together, these findings contribute to a better understanding on how colibactin contributes to early carcinogenesis and highlight opportunities for prevention by targeting microbial, environmental, and host pathways.

4.1 Defining the sequence specificity of colibactin–DNA interactions

A central aim of this work was to define how colibactin interacts with DNA at the molecular level. Because its genotoxicity is thought to depend on sequence-specific inter-strand crosslinking, identifying and validating precise colibactin damage motifs (CDMs) was essential for linking biochemical activity to mutational signatures found in human tissues. To address this, I employed complementary platforms—plasmid-based assays, phosphorothioate-modified oligonucleotides, and nanopore sequencing—that together enabled a systematic evaluation of colibactin's binding specificity. These

approaches established the sequence motifs most vulnerable to colibactin, clarified the role of DNA topology in modulating crosslinking efficiency, and provided a mechanistic framework for understanding how microbial exposures shape cancer-relevant mutational landscapes.

4.1.1 Comparison of colibactin affinity: inter-strand crosslinks to different colibactin damage motif on plasmid-based assays

Plasmid-based assays provided the first platform to assess colibactin's DNA sequence specificity and crosslinking activity. Similar conclusions were drawn in earlier studies that used plasmid topology as a proxy for monitoring colibactin alkylation and ICL formation^{115,116}. Using isolated plasmid DNA, I found that colibactin induced inter-strand crosslinks (ICLs) preferentially at AAAATT and AAATTT motifs, while other proposed targets such as AATATT were less susceptible. These results are consistent with earlier reports that colibactin favors AT-rich motifs, particularly the AAWWTT consensus sequence⁷⁴. They highlight variability in motif activity, suggesting that local DNA context influences recognition and efficiency.

The plasmid system also revealed how DNA topology shapes colibactin activity. Relaxed linear DNA supported more efficient crosslinking than supercoiled DNA, indicating that torsional strain can affect binding or reactivity at colibactin motifs. While plasmids are artificial substrates compared with chromatinized human DNA, they capture certain structural parameters, such as supercoiling, but lack the dynamic relaxation processes present in eukaryotic chromatin that may influence how colibactin engages with the genome.

By confirming reproducible sequence preferences in a simplified system, plasmid-based assays established that colibactin is not a random DNA-damaging agent but a genotoxin with defined sequence selectivity. This insight provided the conceptual foundation for subsequent motif validation in more physiologically relevant systems and strengthens the interpretation of colibactin-associated mutational signatures observed in human tissues.

4.1.2 Phosphorothioate Oligonucleotides: A High-Fidelity Platform for CDM Validation

To refine the analysis of colibactin sequence specificity, I employed phosphorothioate-modified oligonucleotides (S-oligos). These synthetic substrates overcome limitations of plasmid-based assays, including susceptibility to nuclease degradation and interference from multiple potential binding sites within the plasmid backbone. By replacing non-bridging oxygen atoms in the phosphate backbone with sulfur, S-oligos achieve enhanced nuclease resistance while maintaining a linear topology more similar to eukaryotic DNA ¹¹⁷.

Using this system, I confirmed strong crosslinking at canonical motifs such as AAAATT and AAATTT, consistent with the AAWWTT consensus⁷⁴. Importantly, the approach also validated additional motifs, including AAAGTT and AATATT, broadening the spectrum of sequences susceptible to colibactin damage. These results indicate that colibactin recognition is not restricted to a single consensus but extends across a family of structurally related AT-rich hexamers.

Compared with plasmid assays, S-oligos yielded clearer, more reproducible readouts, establishing them as a high-fidelity platform for dissecting genotoxin–DNA interactions. Beyond technical improvement, this work strengthens the conclusion that colibactin damage is reproducible and sequence-specific, reinforcing the interpretation of SBS88 and related mutational signatures as genomic imprints of microbial exposure in colorectal carcinogenesis.

4.1.3 Nanopore Sequencing as a Reporter of Colibactin Binding

To complement plasmid- and oligonucleotide-based assays, I explored nanopore sequencing as a platform for detecting colibactin–DNA interactions. This technology measures changes in ionic current as DNA passes through a pore, enabling the detection of both base alkylation and inter-strand crosslinks at single-molecule resolution ¹¹⁸⁻¹²⁰. Our analyses showed reproducible read interruptions and quality dropouts at known motifs such as AAAATT, while GC-rich regions without canonical motifs remained largely unaffected. Sequence logo analysis around disrupted regions consistently recovered AT-rich patterns resembling the AAWWTT consensus ⁷⁴, confirming sequence-specific binding of colibactin.

Compared with plasmid and S-oligo systems, nanopore sequencing provided higher throughput and directly linked signal disruptions to sequence context, eliminating the need for gel-based readouts. Its scalability also opens the door to studying colibactin activity in more complex and clinically relevant DNA substrates.

By bridging biochemical damage to mutational patterns, nanopore sequencing provides a scalable and context-sensitive method for mapping genotoxin activity. Beyond validating colibactin's sequence selectivity, this approach offers a framework for applying similar analyses to patient-derived samples, where mutational signatures can be directly linked to microbial exposures. In the future, integrating nanopore-based damage profiling with clinical sequencing could help identify high-risk individuals, guide preventive interventions, and refine our understanding of how microbial genotoxins contribute to cancer evolution.

4.1.4 Converging evidence for AAAGTT as a novel motif

Across plasmid, S-oligo, and nanopore assays, I consistently detected crosslinking activity at AAAGTT, establishing it as a novel colibactin damage motif. This expands the repertoire of AT-rich sequences susceptible to colibactin and shows that recognition is more versatile than the canonical AAWWTT consensus⁷⁴. Subtle sequence variations, such as those in AAAGTT, can still serve as efficient binding sites, underscoring flexibility in colibactin–DNA interactions.

Notably, not all motifs validated *in vitro* overlap with those enriched in *in vivo* datasets. This discrepancy maybe explained by differences in DNA topology. Plasmid DNA is maintained in a negatively supercoiled state, whereas eukaryotic DNA is packaged into nucleosomes and dynamically relaxed by topoisomerases¹²¹. Such structural distinctions influence DNA accessibility and crosslinking efficiency, which may explain why certain motifs emerge more prominently in human tissues.

Our experiments further indicated that DNA relaxation enhances crosslink formation. This aligns with observations from Xue et al. (2020), who reported that prolonged colibactin exposure progressively converted supercoiled plasmids into relaxed forms, correlating with increased inter-strand crosslink accumulation. Thus, while supercoiling can initially limit accessibility, relaxation over time appears to favor colibactin's activity.

By confirming AAAGTT as a functional motif and linking its activity to DNA structure, this work refines the mechanistic basis of colibactin–DNA recognition. More broadly, it suggests that the vulnerability of certain genomic regions in human cells reflects both sequence composition and chromatin accessibility, providing insight into why colibactin-induced mutational signatures preferentially localize to AT-rich, structurally exposed sites.

4.1.5 Causal link to tumor suppressor gene mutations

The identification of sequence-specific colibactin damage motifs provides a mechanistic bridge between bacterial genotoxin exposure and colorectal cancer development. Whole-genome sequencing studies of organoids and patient tissues have demonstrated that colibactin-associated mutational signatures, such as SBS88, accumulate in AT-rich genomic regions and can overlap with functionally important loci, including colorectal cancer driver genes ^{74,122}. Our motif analyses suggest that this overlap arises because colibactin preferentially targets hexamers that are present within genes critical for intestinal homeostasis and tumor initiation.

A key advance of this work was the observation that recognized colibactin targeting sequences—including canonical AAWT motifs and the novel AAAGTT motif—were not randomly distributed but enriched in canonical CRC driver genes such as APC, TP53, KRAS, and AXIN2. These genes serve as gatekeepers of epithelial integrity, and their disruption defines critical milestones in the adenoma–carcinoma sequence. For instance, truncating mutations in APC initiate adenoma formation through deregulation of Wnt signaling, while TP53 loss promotes malignant progression ^{32,27}. Our finding that colibactin-induced substitutions map to these loci provides direct mutational evidence that microbial exposures can drive oncogenic transformation, thereby filling a long-standing gap in linking microbial genotoxicity to functional cancer pathways.

This mechanistic connection highlights that colibactin does not merely contribute to background mutational load but actively targets genes essential to carcinogenesis, suggesting that colibactin may serve as an upstream driver of malignant transformation in the context of chronic inflammation. More broadly, these results underscore the importance of identifying colibactin motifs within cancer genomes to pinpoint mutational hotspots most relevant to tumor initiation. From a translational perspective,

they provide a rationale for therapeutic strategies aimed at eliminating *pks*+ *E. coli* or inhibiting colibactin activity as a means of reducing early mutational events in high-risk patients.

Collectively, plasmid, S-oligo, and nanopore assays establish colibactin as a sequence-selective genotoxin that preferentially targets AT-rich motifs such as AAAATT, AAATTT, and the newly identified AAAGTT. These results show that sequence specificity is shaped not only by motif composition but also by DNA topology and accessibility. Importantly, integration with whole-genome sequencing data reveals that colibactin-associated signatures (SBS88) overlap with colorectal cancer driver genes, including APC and TP53, providing a mechanistic link between microbial mutagenesis and early tumor initiation. Looking ahead, systematic mapping of colibactin motifs across the human genome could help identify mutational hotspots most vulnerable to microbial exposure, informing prevention strategies and personalized cancer risk assessment.

4.2 Environmental and host cofactors modulating colibactin activity

Colibactin's mutagenic impact is not determined by DNA sequence alone but is strongly shaped by environmental, nutritional, and microbial cofactors. These variables influence either the expression of the *pks* gene cluster or the chemical reactivity of colibactin, thereby modulating the genotoxic burden in the intestinal epithelium. Understanding these modulators is critical for interpreting why colibactin signatures arise in some individuals but not in all carriers of *pks*+ *E. coli*.

4.2.1 Cofactors regulating *pks* gene expression

Previous studies demonstrated that ferric iron (Fe^{3+}) suppresses colibactin synthesis through both ferric uptake regulator (Fur)-dependent and independent pathways^{123,124}. Our results extend this concept by showing that ferrous iron (Fe^{2+}) can also modulate crosslinking efficiency, suggesting that multiple oxidation states of iron participate in fine-tuning colibactin activity. This observation may have direct clinical relevance, as

iron supplementation is widely used in IBD patients, where altered iron metabolism could inadvertently increase or suppress bacterial genotoxicity.

Oxygen tension is another determinant of *pks* expression. Bossuet-Greif et al. (2018) reported that anaerobic conditions enhance colibactin synthesis via the ArcA regulator, consistent with the hypoxic physiology of the gut lumen. In our assays, however, oxygen effects varied across strains: anaerobiosis suppressed activity in a recombinant *E. coli* but enhanced production in clinical isolates and in the probiotic strain Nissle 1917. These discrepancies point to the influence of strain-specific regulatory networks and genomic context, which may explain divergent findings across laboratories.

Taken together, these results highlight the dynamic interplay between microbial gene regulation and host environment. They suggest that iron metabolism and oxygen gradients in the inflamed colon not only shape bacterial growth but also influence the mutational risk associated with *pks*⁺ *E. coli*. Understanding these interactions helps answer the broader question of this thesis—how microbial genotoxins contribute to colorectal cancer—and points toward practical interventions, such as adjusting iron supplementation strategies or targeting regulatory pathways that control *pks* expression.

4.2.2 Cofactors regulating colibactin activity

In addition to transcriptional regulation, several host- and environment-derived factors directly influence the activity of colibactin once it is produced. These modulators determine the extent of DNA damage and, ultimately, the mutational footprint left in host tissues.

Early biochemical studies suggested that copper ions could potentiate DNA cleavage by colibactin ⁶⁹, but our results revealed the opposite: even low concentrations of copper inhibited crosslink formation. This discrepancy highlights the strong context-dependence of colibactin activity and suggests that local metal ion concentrations in the intestinal lumen may act as natural brakes on genotoxicity. Such findings emphasize the importance of validating in vitro chemistry within physiologically relevant systems. This finding contrasts with earlier work suggesting that colibactin–Cu²⁺ complexes can promote double-strand breaks ⁶⁹, underscoring the context-

dependence of metal interactions. Similarly, ferrous iron (Fe^{2+}) reduced ICL formation, suggesting that redox state and metal availability may act as natural brakes on genotoxicity.

Polyamines, particularly spermidine, are another class of modulators. Chagneau et al. (2019) reported that spermidine promoted colibactin-dependent DNA damage, possibly by stabilizing DNA conformations that favor binding¹²⁵. In our assays, however, spermidine did not markedly enhance crosslinking efficiency. Differences in bacterial strains, culture conditions, or detection methods may explain this divergence, again pointing to the complexity of host–microbe metabolic interactions.

These examples illustrate how cofactors can amplify or suppress colibactin activity without altering *pks* gene expression. From a clinical perspective, they underscore that the genotoxic potential of *pks*+ *E. coli* depends not only on bacterial genetics but also on the chemical milieu of the gut. Understanding this layer of regulation is essential for clarifying the relative significance of colibactin damage in colorectal cancer. Future work should aim to define the physiological ranges of these cofactors in IBD and CRC patients, which may reveal opportunities to therapeutically attenuate colibactin activity in situ.

4.2.3 Dietary and microbial ecosystem factors

Beyond direct gene regulation and biochemical modulation, the broader intestinal environment—shaped by diet and microbial ecology—also plays a decisive role in colibactin activity. Dietary components influence microbial metabolism and gut barrier properties, which in turn affect how *pks*+ *E. coli* interacts with host tissues. Previous work showed that inulin and fructooligosaccharides (FOS) stimulate *pks* expression and increase tumorigenesis in mouse models^{126,127}. However, under controlled in vitro conditions, our experiments did not detect significant effects of inulin or FOS on colibactin production. These contrasting results suggest that diet likely influences colibactin indirectly by reshaping microbial ecology or host physiology, rather than through direct stimulation of the *pks* expression.

Fiber deprivation has the opposite but equally important effect. Thakur et al. (2025) demonstrated that low-fiber diets alter gut redox balance by elevating nitrate (NO_3^-) levels, which in turn reshapes microbial composition to favor *pks*+ *E. coli*¹²⁸. This

dietary shift also thins the mucus barrier, creating conditions where bacteria gain direct access to the epithelium. In such an environment, colibactin-producing strains are both enriched and positioned closer to their epithelial targets, thereby intensifying the risk of DNA damage.

These findings highlight how diet and microbiota act as upstream regulators of microbial mutagenesis. Fiber availability not only affects colibactin expression indirectly through redox and ecological shifts but also influences barrier thickness, which controls bacterial access to host stem cells. From a translational perspective, maintaining adequate fiber intake and promoting mucus-preserving microbial communities may be a practical strategy to limit colibactin activity and reduce colorectal cancer risk in high-risk populations.

4.3 Analysis of colibactin-induced DNA damage motifs across distinct colorectal cancer precursors

Having established the environmental and molecular factors that modulate colibactin activity, the next step was to examine its consequences in human tissues, focusing on how colibactin-induced DNA damage contributes to mutational signatures and early colorectal carcinogenesis.

The ultimate goal of this work was to understand the conditions under which colibactin damage leaves a stable mutational imprint in human tissues. By integrating patient-derived colon organoids, inflammation models, and whole-genome sequencing (WGS), we provide evidence that colibactin signatures emerge preferentially in inflamed epithelium. These findings not only strengthen the causal link between microbial exposure and CRC risk but also address a broader question in cancer biology: how environmental and microbial factors become permanently recorded in the genome as mutational signatures.

4.3.1 Organoid platform as a model for early colorectal carcinogenesis

A key strength of this study was the use of patient-derived colon organoids to model early carcinogenesis. Compared with immortalized cell lines, organoids preserve the

genetic and epigenetic background of the donor, while reproducing essential epithelial functions such as stem cell renewal, polarity, and barrier formation^{129,130 131}. This makes them uniquely suited for investigating how colibactin acts in clinically relevant tissues. By generating cultures from inflamed and non-inflamed biopsies, I were able to capture a continuum of precancerous states, something that has been missing in earlier studies.

The adaptation of 3D cultures into 2D monolayers further extended their utility. This allowed quantitative monitoring of barrier function via transepithelial electrical resistance (TEER) and junctional protein assays, bridging observational pathology with mechanistic testing. Crucially, this approach revealed how inflammation-induced barrier dysfunction increases susceptibility to colibactin mutagenesis, providing a functional explanation for why colibactin signatures are preferentially observed in inflamed mucosa. Similar results have been reported in laser micro-dissected IBD tissues¹³², but organoid-based models provide causal evidence rather than correlation.

Taken together, organoids serve not only as a technical innovation but also as a conceptual bridge between descriptive clinical observations and mechanistic dissection. They demonstrate how chronic inflammation and microbial exposure interact at the epithelial interface, thereby clarifying the earliest steps by which colibactin contributes to CRC initiation.

4.3.2 Inflammation-driven barrier dysfunction increases susceptibility to colibactin

One of the central insights from this work is that chronic inflammation leaves a lasting imprint on epithelial barrier function, independent of immune infiltrates. Organoid-derived monolayers from inflamed tissues showed persistently reduced transepithelial electrical resistance (TEER), disorganized tight junction protein localization, and cytoskeletal disruption. These features mirror clinical pathology in IBD but, importantly, our model demonstrates that such defects persist even outside the inflammatory milieu, suggesting long-term “epigenetic scarring” of epithelial cells.

This compromised barrier state has direct implications for colibactin mutagenesis. By weakening epithelial integrity, inflammation allows *pks*⁺ *E. coli* and their genotoxin to reach the basal stem cell niche—the population most relevant for sustaining clonal

mutations. This mechanistic link helps explain why colibactin-associated mutational signatures (SBS88) are preferentially detected in inflamed mucosa ¹³². While earlier clinical studies documented the presence of SBS88 in IBD patients, they could not resolve whether barrier dysfunction was a prerequisite. Our 2D models provide causal evidence that barrier breakdown creates the necessary conditions for microbial mutagens to leave a durable genomic imprint.

From a broader perspective, this finding underscores how inflammation contributes to carcinogenesis not only through immune-mediated oxidative stress but also by structurally altering the epithelial interface. This dual effect—weakening the barrier while exposing stem cells to microbial toxins—clarifies why chronic inflammation represents a particularly high-risk context for CRC initiation.

4.3.3 Detection of SBS88 in organoids from inflamed tissues

Whole-genome sequencing of patient-derived organoids provided direct experimental evidence that colibactin mutagenesis occurs in human epithelial cells under inflammatory conditions. In approximately half of the inflamed lines, I detected SBS88 and ID18, the canonical colibactin-associated signature. Its reproducibility across patients highlights SBS88 as a robust biomarker of microbial genotoxicity, while its persistence during long-term culture suggests that once established, colibactin lesions are stably embedded in the genome.

Importantly, SBS88 did not occur in isolation. It frequently co-appeared with aging-associated signatures such as SBS1 and SBS5, supporting a “double-hit” model in which microbial mutagens act on top of endogenous processes to accelerate carcinogenesis. This context helps explain why colibactin exposure alone is not sufficient for transformation, but when combined with chronic inflammation or age-related damage, it can tip the balance toward malignant progression.

These results extend earlier infection-based organoid studies ¹¹⁴ and crypt analyses in healthy colon tissue ⁷⁴. By confirming SBS88 in patient-derived systems, this work moves beyond experimental models and provides clinical evidence that colibactin leaves a durable imprint in the human colon. It strengthens the link between bacterial exposure and early colorectal carcinogenesis, and highlights the value of mutational signatures as genomic records of microbial interactions with human epithelium.

4.3.4 Regenerative responses and potential therapeutic modulation

In addition to defining mutational consequences, this work examined whether regenerative capacity could be restored in the inflamed epithelium. Supplementation with nicotinamide riboside (NR), a precursor of NAD⁺, improved proliferation and survival in ulcerative colitis–derived organoids with impaired renewal. This supports evidence that chronic inflammation depletes NAD⁺ pools, weakening DNA repair and regenerative programs ¹³³. The effect was context-specific, as non-inflamed organoids showed little benefit, suggesting that NR selectively rescues metabolically stressed epithelia.

Although NR did not prevent the accumulation of mutational signatures such as SBS88, its ability to restore epithelial resilience highlights a complementary therapeutic strategy that focuses on repair rather than damage prevention. Strengthening regeneration may reduce long-term risk by promoting mucosal healing and limiting clonal expansion of damaged cells. These findings illustrate the translational value of patient-derived organoids for testing interventions and provide proof-of-principle that metabolic support could complement microbial or immune-directed therapies. More broadly, they emphasize that colorectal cancer initiation reflects not only the accumulation of colibactin-induced mutations but also the failure of epithelial repair, pointing to regeneration as a critical axis for intervention. Future work should test whether NAD⁺-boosting strategies like NR can be integrated with established IBD therapies or microbiome-targeted interventions, opening a path toward combinatorial approaches that both limit damage and enhance repair.

4.3.5 Mechanistic model: inflammation–dysbiosis–mutation axis

The organoid and sequencing data presented in this work converge on a unifying model in which chronic inflammation, microbial dysbiosis, and impaired DNA repair cooperate to initiate colorectal carcinogenesis. In inflamed organoids, persistent barrier dysfunction enabled *pks*⁺ *E. coli* to reach the stem cell compartment, where colibactin induced DNA inter-strand crosslinks and double-strand breaks. These lesions left a durable mutational imprint in the form of SBS88, detected in the majority of inflamed lines and confirmed as stably embedded in the genome. At the same time,

inflammation impaired repair capacity through oxidative stress and replication stress, amplifying the fixation of colibactin-induced mutations ¹³⁴. These findings position the inflamed colon as a high-risk environment where barrier failure, microbial invasion, and deficiencies in DNA repair synergistically contribute to the initiation of carcinogenesis.

The “inflammation–dysbiosis–mutation” axis that emerges from this work highlights how host and microbial factors converge to drive early colorectal carcinogenesis. Chronic inflammation weakens epithelial barrier function, allowing microbial translocation and activating innate immune pathways such as TLR4 and NOD2. This amplifies NF- κ B–driven inflammation, further compromising barrier integrity and sustaining a permissive environment for microbial colonization ¹²⁸. Within this context, expansion of *pks*+ *E. coli* and erosion of mucus protection expose the epithelium to colibactin for prolonged periods. Our organoid models capture this interplay, showing that SBS88 arises specifically in inflamed lines, consistent with clinical observations that *pks*+ *E. coli* are more prevalent in IBD patients than in healthy controls.

Importantly, colibactin-induced mutations did not occur randomly but were enriched at AT-rich motifs embedded in tumor suppressor genes such as *APC* and *TP53*, directly linking microbial exposure to key oncogenic pathways. The stability of these mutations in long-term organoid cultures suggests that they arise early and persist, potentially establishing precancerous fields that predispose to malignant transformation. In parallel, NAD⁺ depletion in inflamed epithelia may reduce PARP activity and compromise DNA repair, further coupling metabolic stress to mutagenesis. By integrating biochemical assays, patient-derived organoids, and genomic readouts, this thesis demonstrates that colibactin-driven carcinogenesis is not the outcome of a single insult, but the product of converging processes: inflammation, microbial imbalance, and impaired repair. This synthesis explains why mutational signatures such as SBS88 are enriched in inflamed mucosa and underscores the inflamed colon as a uniquely permissive environment for microbial genotoxin-driven tumor initiation.

By capturing this triangle experimentally, our work extends observational findings in patient tissues and provides causal evidence of how microbial exposures accelerate tumor initiation in a genetically vulnerable epithelium (**Figure 26**).

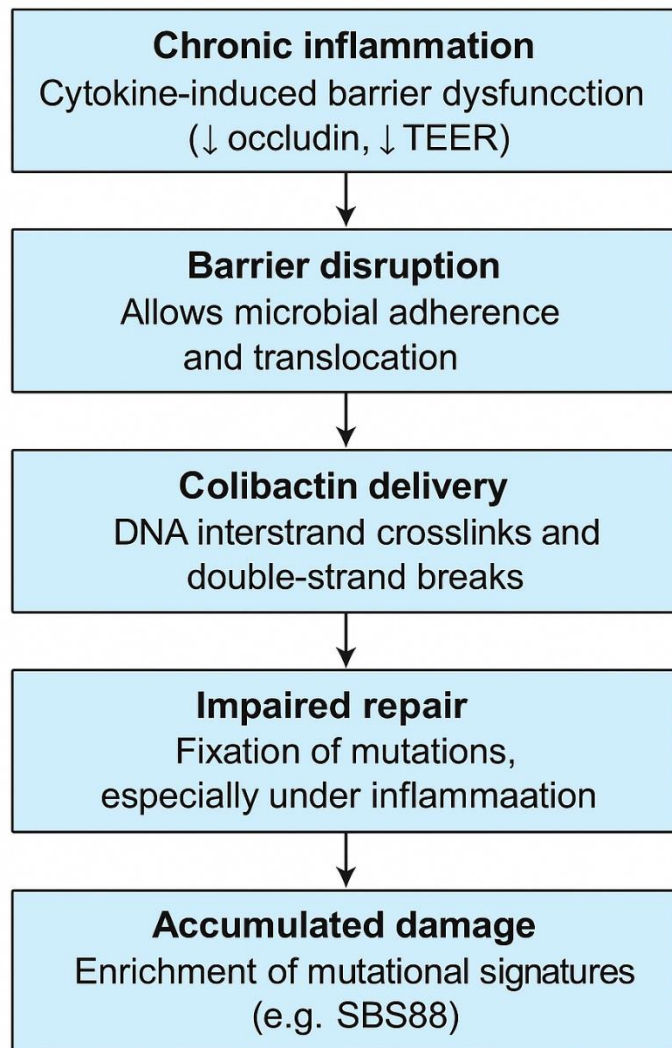


Figure 26- Inflammation-driven model of colibactin-associated mutagenesis in the intestinal epithelium.

Chronic inflammation leads to cytokine-induced barrier dysfunction, characterized by reduced tight junction proteins (e.g., occludin) and decreased transepithelial electrical resistance (TEER). This barrier disruption facilitates microbial adherence and translocation, enabling the delivery of colibactin, a genotoxin produced by *pks+*. Colibactin induces DNA inter-strand crosslinks and double-strand breaks. Under inflammatory conditions, impaired DNA repair mechanisms increase the likelihood of mutation fixation. Over time, this results in the accumulation of DNA damage and the enrichment of mutational signatures, such as SBS88, which are associated with colibactin exposure

5. Conclusion and outlook

This work contributes to a better understanding of the role of the colibactin in the initiation of colorectal cancer, by refining its molecular specificity and by the use of organoid models to reveal its action in the course of human pathogenesis. Colibactin was shown to target a family of AT-rich motifs, including canonical sequences such as AAATTT and AAAATT as well as novel variants like AAAGTT, with nanopore sequencing linking this specificity to mutational patterns. These motifs are enriched in cancer-relevant loci such as APC and TP53, providing a mechanistic link between microbial exposure and early oncogenesis. Beyond DNA sequence, colibactin activity was modulated by host and environmental factors: iron, oxygen, and copper influenced *pks* expression or activity, while diet and microbiota composition, including fiber deprivation and mucus erosion¹²⁸, shaped epithelial access. Together, these findings show that colibactin-driven mutagenesis arises from the interplay between microbial determinants, host physiology, and diet, illuminating why not all carriers of *pks*+ *E. coli* develop cancer.

A major achievement of this work was the detection of SBS88, the canonical colibactin mutational signature, in patient-derived inflamed organoids. SBS88 was observed in approximately half of the lines, persisted during long-term culture, and frequently co-occurred with aging-related signatures such as SBS1 and SBS5. This demonstrates that colibactin leaves a stable genomic record of microbial exposure and supports a “double-hit” model in which microbial mutagens accelerate tumor initiation against a background of endogenous processes. Importantly, SBS88 was enriched in colorectal cancer driver genes, linking bacterial genotoxicity to functional cancer pathways and providing direct evidence that microbial exposures can shape the earliest steps of oncogenesis.

Looking forward, these insights have considerable translational implications. The reproducibility of SBS88 across patients highlights its potential as a durable biomarker of microbial exposure, suitable for risk stratification and early detection in inflammatory bowel disease and colorectal cancer. Preventive strategies could focus on targeting *pks*+ *E. coli* directly, exploiting inhibitors of colibactin biosynthesis, or preserving mucus integrity through diet and microbiome modulation. At the same time, host-directed interventions such as NAD⁺ repletion with nicotinamide riboside illustrate opportunities to restore epithelial resilience, complementing established anti-

inflammatory therapies. Finally, patient-derived organoids provide a versatile platform for testing these approaches in a personalized manner, bridging mechanistic discoveries with clinical application.

In conclusion, this work not only defines how colibactin leaves a lasting mutational signature in human cells but also establishes SBS88 as a functional and clinically relevant biomarker of microbial genotoxicity. By integrating microbiology, cancer genomics, and regenerative biology, it contributes to a broader effort to translate mutational signatures into mechanistic insight, prevention strategies, and precision medicine for colorectal cancer.

Publications and Manuscripts Arising from This Thesis

Published

Quevedo-Olmos, A., Wang, S., & Meyer, T. F. (2025). Decoding microbe-diet-host synergy in colorectal cancer. *Nat Microbiol*, 10(4), 819-820. <https://doi.org/10.1038/s41564-025-01971-3>

Under review

Wang*, S., Holthaus*, D. et al. Generation of Matched Intestinal Organoids and Organoid-derived Monolayers from Pathological and Non-Pathological Regions of IBD and Polyp Patients. *Bioprotocol*, submitted

In Preparation

Welz L., et al. A metabolic constraint in de novo NAD⁺ synthesis drives mucosal inflammation in IBD

Wang*, S., Quevedo-Olmos*, A., Arellano-Ayala*, K., et al. Validation of colibactin target sequences on DNA to re-evaluate its mutagenesis.

Wang*, S., Quevedo-Olmos*, A., et al. SBS88 as a mutational footprint of colibactin in inflamed patient-derived tissues

*Shared first authors

6. List of References

1. Bray, F., *et al.* Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **74**, 229-263 (2024).
2. Sedeta, E., Sung, H., Laversanne, M., Bray, F. & Jemal, A. Recent Mortality Patterns and Time Trends for the Major Cancers in 47 Countries Worldwide. *Cancer Epidem Biomar* **32**, 894-905 (2023).
3. Bray, F., *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **68**, 394-424 (2018).
4. Alzahrani, S.M., Al Doghaither, H.A. & Al-Ghafari, A.B. General insight into cancer: An overview of colorectal cancer (Review). *Mol Clin Oncol* **15**(2021).
5. Hanahan, D. & Weinberg, R.A. The hallmarks of cancer. *Cell* **100**, 57-70 (2000).
6. Idikio, H.A. Human cancer classification: a systems biology- based model integrating morphology, cancer stem cells, proteomics, and genomics. *J Cancer* **2**, 107-115 (2011).
7. El Tekle, G. & Garrett, W.S. Bacteria in cancer initiation, promotion and progression. *Nat Rev Cancer* **23**, 600-618 (2023).
8. Blackadar, C.B. Historical review of the causes of cancer. *World J Clin Oncol* **7**, 54-86 (2016).
9. Pradhan, R., Kundu, A. & Kundu, C.N. The cytokines in tumor microenvironment: from cancer initiation-elongation-progression to metastatic outgrowth. *Crit Rev Oncol Hematol* **196**, 104311 (2024).
10. Lynch, H.T., Snyder, C.L., Shaw, T.G., Heinen, C.D. & Hitchins, M.P. Milestones of Lynch syndrome: 1895-2015. *Nat Rev Cancer* **15**, 181-194 (2015).
11. Olkinuora, A.P., Peltomäki, P.T., Aaltonen, L.A. & Rajamäki, K. From APC to the genetics of hereditary and familial colon cancer syndromes. *Human Molecular Genetics* **30**, R206-R224 (2021).
12. Sandoval, R.L., *et al.* PREMM5 distinguishes sporadic from Lynch syndrome-associated MMR-deficient/MSI-high colorectal cancer. *Fam Cancer* **22**, 459-465 (2023).
13. Pourmasoumi, P., Moradi, A. & Bayat, M. BRCA1/2 Mutations and Breast/Ovarian Cancer Risk: A New Insights Review. *Reprod Sci* **31**, 3624-3634 (2024).
14. Calabrese, A., von Arx, C., Tafuti, A.A., Pensabene, M. & De Laurentiis, M. Prevention, diagnosis and clinical management of hereditary breast cancer beyond BRCA1/2 genes. *Cancer Treat Rev* **129**(2024).
15. Fitzmaurice, C., *et al.* The Global Burden of Cancer 2013 Global Burden of Disease Cancer Collaboration. *Jama Oncol* **1**, 505-527 (2015).
16. Soerjomataram, I. & Bray, F. Planning for tomorrow: global cancer incidence and the role of prevention 2020-2070. *Nat Rev Clin Oncol* **18**, 663-672 (2021).
17. Hatta, M.N.A., Hanif, E.A.M., Chin, S.F. & Neoh, H.M. Pathogens and Carcinogenesis: A Review. *Biology-Basel* **10**(2021).
18. Schiller, J.T. & Lowy, D.R. An Introduction to Virus Infections and Human Cancer. *Recent Results Cancer Res* **217**, 1-11 (2021).
19. Chen, Y.C., *et al.* The Updated and Secular Trend of Global Prevalence of Helicobacter Pylori Infection: A Systematic Review and Meta-Analysis. *Gastroenterology* **164**, S370-S370 (2023).
20. Tomasetti, C., Li, L. & Vogelstein, B. Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science* **355**, 1330-1334 (2017).
21. Demir Cetinkaya, B. & Biray Avci, C. Molecular perspective on targeted therapy in breast cancer: a review of current status. *Med Oncol* **39**, 149 (2022).
22. Huang, L.M., Guo, Z.X., Wang, F. & Fu, L.W. KRAS mutation: from undruggable to druggable in cancer. *Signal Transduct Tar* **6**(2021).
23. Gurel, B., *et al.* Nuclear MYC protein overexpression is an early alteration in human prostate carcinogenesis. *Modern Pathol* **21**, 1156-1167 (2008).

24. Dallafavera, R., *et al.* Human C-Myc Onc Gene Is Located on the Region of Chromosome-8 That Is Translocated in Burkitt-Lymphoma Cells. *P Natl Acad Sci-Biol* **79**, 7824-7827 (1982).
25. Adams, J.M., *et al.* The C-Myc Oncogene Driven by Immunoglobulin Enhancers Induces Lymphoid Malignancy in Transgenic Mice. *Nature* **318**, 533-538 (1985).
26. Levine, A.J. Spontaneous and inherited genetic alterations. *Oncogene* **40**, 5975-5983 (2021).
27. Alvarado-Ortiz, E., *et al.* Mutant p53 Gain-of-Function: Role in Cancer Development, Progression, and Therapeutic Approaches. *Front Cell Dev Biol* **8**(2021).
28. Zhang, L. & Shay, J.W. Multiple Roles of APC and its Therapeutic Implications in Colorectal Cancer. *J Natl Cancer I* **109**(2017).
29. Chinnam, M. & Goodrich, D.W. RB1, development, and cancer. *Curr Top Dev Biol* **94**, 129-169 (2011).
30. Li, G.M. Mechanisms and functions of DNA mismatch repair. *Cell Res* **18**, 85-98 (2008).
31. Nolano, A., *et al.* Hereditary Colorectal Cancer: State of the Art in Lynch Syndrome. *Cancers* **15**(2023).
32. Bugter, J.M., Fenderico, N. & Maurice, M.M. Mutations and mechanisms of WNT pathway tumour suppressors in cancer. *Nat Rev Cancer* **21**, 5-21 (2021).
33. Michaudel, C. & Sokol, H. The Gut Microbiota at the Service of Immunometabolism. *Cell Metab* **32**, 514-523 (2020).
34. Boleij, A. & Tjalsma, H. Gut bacteria in health and disease: a survey on the interface between intestinal microbiology and colorectal cancer. *Biol Rev* **87**, 701-730 (2012).
35. Gomaa, E.Z. Human gut microbiota/microbiome in health and diseases: a review. *Antonie Van Leeuwenhoek* **113**, 2019-2040 (2020).
36. Naito, Y. & Takagi, T. Role of gut microbiota in inflammatory bowel disease pathogenesis. *J Clin Biochem Nutr* **75**, 175-177 (2024).
37. Gomes, S., Rodrigues, A.C., Pazienza, V. & Preto, A. Modulation of the Tumor Microenvironment by Microbiota-Derived Short-Chain Fatty Acids: Impact in Colorectal Cancer Therapy. *International Journal of Molecular Sciences* **24**(2023).
38. Scarminio, V., Fruet, A.C., Witaicenis, A., Rall, V.L. & Di Stasi, L.C. Dietary intervention with green dwarf banana flour (*Musa sp AAA*) prevents intestinal inflammation in a trinitrobenzenesulfonic acid model of rat colitis. *Nutr Res* **32**, 202-209 (2012).
39. Almeida, L.D., Curimbaba, T.F.S., Chagas, A.S., Quaglio, A.E.V. & Di Stasi, L.C. Dietary intervention with green dwarf banana flour (*sp AAA*) modulates oxidative stress and colonic SCFAs production in the TNBS model of intestinal inflammation. *J Funct Foods* **38**, 497-504 (2017).
40. Curimbaba, T.F.S., *et al.* Prebiotic, antioxidant and anti-inflammatory properties of edible Amazon fruits. *Food Biosci* **36**(2020).
41. Mills, S., Stanton, C., Lane, J.A., Smith, G.J. & Ross, R.P. Precision Nutrition and the Microbiome, Part I: Current State of the Science. *Nutrients* **11**(2019).
42. Perse, M. Oxidative stress in the pathogenesis of colorectal cancer: cause or consequence? *Biomed Res Int* **2013**, 725710 (2013).
43. Artemev, A., Naik, S., Pougno, A., Honnavar, P. & Shanbhag, N.M. The Association of Microbiome Dysbiosis With Colorectal Cancer. *Cureus J Med Science* **14**(2022).
44. Waskito, L.A., Salama, N.R. & Yamaoka, Y. Pathogenesis of *Helicobacter pylori* infection. *Helicobacter* **23 Suppl 1**, e12516 (2018).
45. Hooi, J.K.Y., *et al.* Global Prevalence of *Helicobacter pylori* Infection: Systematic Review and Meta-Analysis. *Gastroenterology* **153**, 420-429 (2017).
46. Abdun, M.A., *et al.* Global Prevalence of *Helicobacter pylori* Infection-Associated Gastric Preneoplastic Lesions in Pediatric Patients: A Systematic Review and Meta-Analysis. *Helicobacter* **30**, e70021 (2025).
47. Montecucco, C. & Rappuoli, R. Living dangerously: How survives in the human stomach. *Nat Rev Mol Cell Bio* **2**, 457-466 (2001).

48. Fischer, W. Assembly and molecular mode of action of the Cag type IV secretion apparatus. *Febs J* **278**, 1203-1212 (2011).
49. Hatakeyama, M. Oncogenic mechanisms of the CagA protein. *Nat Rev Cancer* **4**, 688-694 (2004).
50. Akazawa, Y., *et al.* Endoplasmic Reticulum Stress Contributes to Helicobacter Pylori VacA-Induced Apoptosis. *Plos One* **8**(2013).
51. Greenfield, L.K. & Jones, N.L. Modulation of autophagy by and its role in gastric carcinogenesis. *Trends Microbiol* **21**, 602-612 (2013).
52. Guo, B., Zhang, J., Zhang, W., Chen, F. & Liu, B. Gut microbiota-derived short chain fatty acids act as mediators of the gut-brain axis targeting age-related neurodegenerative disorders: a narrative review. *Crit Rev Food Sci Nutr*, 1-22 (2023).
53. Li, S., *et al.* Tumorigenic bacteria in colorectal cancer: mechanisms and treatments. *Cancer Biol Med* **19**, 147-162 (2021).
54. Barbara, G., *et al.* Inflammatory and Microbiota-Related Regulation of the Intestinal Epithelial Barrier. *Frontiers in Nutrition* **8**(2021).
55. Abdolmaleky, H.M. & Zhou, J.R. Gut Microbiota Dysbiosis, Oxidative Stress, Inflammation, and Epigenetic Alterations in Metabolic Diseases. *Antioxidants-Basel* **13**(2024).
56. Dougherty, M.W. & Jobin, C. Shining a Light on Colibactin Biology. *Toxins (Basel)* **13**(2021).
57. Rondepierre, F., *et al.* Preclinical and clinical evidence of the association of colibactin-producing with anxiety and depression in colon cancer. *World J Gastroentero* **30**(2024).
58. Prorok-Hamon, M., *et al.* Colonic mucosa-associated diffusely adherent afaC+ Escherichia coli expressing IpfA and pks are increased in inflammatory bowel disease and colon cancer. *Gut* **63**, 761-770 (2014).
59. Tomkovich, S., *et al.* Locoregional Effects of Microbiota in a Preclinical Model of Colon Carcinogenesis. *Cancer Res* **77**, 2620-2632 (2017).
60. Putze, J., *et al.* Genetic Structure and Distribution of the Colibactin Genomic Island among Members of the Family. *Infect Immun* **77**, 4696-4703 (2009).
61. Kim, M., *et al.* Structural study on human microbiome-derived polyketide synthases that assemble genotoxic colibactin. *Structure* **33**(2025).
62. Nougayrede, J.P., *et al.* Escherichia coli induces DNA double-strand breaks in eukaryotic cells. *Science* **313**, 848-851 (2006).
63. Tripathi, P., Mousa, J.J., Guntaka, N.S. & Bruner, S.D. Structural basis of the amidase ClbL central to the biosynthesis of the genotoxin colibactin. *Acta Crystallogr D Struct Biol* **79**, 830-836 (2023).
64. Mousa, J.J., *et al.* MATE transport of the E. coli-derived genotoxin colibactin. *Nat Microbiol* **1**, 15009 (2016).
65. Zhang, G.J. & Sun, D.Q. The synthesis of the novel toxin-colibactin and its mechanisms of tumorigenesis of colorectal cancer. *Front Microbiol* **15**(2024).
66. Wernke, K.M., *et al.* Structure and bioactivity of colibactin. *Bioorg Med Chem Lett* **30**, 127280 (2020).
67. Wilson, M.R., *et al.* The human gut bacterial genotoxin colibactin alkylates DNA. *Science* **363**, eaar7785 (2019).
68. Bossuet-Greif, N., *et al.* The Colibactin Genotoxin Generates DNA Interstrand Cross-Links in Infected Cells. *mBio* **9**(2018).
69. Li, Z.R., *et al.* Macrocyclic colibactin induces DNA double-strand breaks via copper-mediated oxidative cleavage. *Nat Chem* **11**, 880-889 (2019).
70. Harnack, C., *et al.* Short-term mucosal disruption enables colibactin-producing E. coli to cause long-term perturbation of colonic homeostasis. *Gut Microbes* **15**, 2233689 (2023).

71. Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C. & Stratton, M.R. Signatures of mutational processes in human cancer. *Nature* **500**, 415-421 (2013).
72. Alexandrov, L.B., Ju, Y.S., Haase, K. & Stratton, M.R. Mutational signatures associated with tobacco smoking in human cancer. *Science* **354**, 618-622 (2015).
73. Ren, P.J., Zhang, J. & Vijg, J. Somatic mutations in aging and disease. *Geroscience* (2024).
74. Dziubanska-Kusibab, P.J., *et al.* Colibactin DNA-damage signature indicates mutational impact in colorectal cancer. *Nat Med* **26**, 1063-1069 (2020).
75. Pleguezuelos-Manzano, C., *et al.* Mutational signature in colorectal cancer caused by genotoxic. *Nature* **580**, 269-+ (2020).
76. Turner, J.R. Intestinal mucosal barrier function in health and disease. *Nature Reviews Immunology* **9**, 799-809 (2009).
77. Volk, N. & Lacy, B. Anatomy and Physiology of the Small Bowel. *Gastrointest Endosc Clin N Am* **27**, 1-13 (2017).
78. Das, D., Fletcher, R.B. & Ngai, J. Cellular mechanisms of epithelial stem cell self-renewal and differentiation during homeostasis and repair. *Wires Dev Biol* **9**(2020).
79. Gehart, H. & Clevers, H. Tales from the crypt: new insights into intestinal stem cells. *Nat Rev Gastro Hepat* **16**, 19-34 (2019).
80. Okumura, R. & Takeda, K. Roles of intestinal epithelial cells in the maintenance of gut homeostasis. *Exp Mol Med* **49**(2017).
81. Takahashi, T. & Shiraishi, A. Stem Cell Signaling Pathways in the Small Intestine. *International Journal of Molecular Sciences* **21**(2020).
82. Zihni, C., Mills, C., Matter, K. & Balda, M.S. Tight junctions: from simple barriers to multifunctional molecular gates. *Nat Rev Mol Cell Bio* **17**, 564-580 (2016).
83. Lee, S.H. Intestinal permeability regulation by tight junction: implication on inflammatory bowel diseases. *Intest Res* **13**, 11-18 (2015).
84. Garcia, M.A., Nelson, W.J. & Chavez, N. Cell-Cell Junctions Organize Structural and Signaling Networks. *Csh Perspect Biol* **10**(2018).
85. De Maio, A., Vega, V.L. & Contreras, J.E. Gap junctions, homeostasis, and injury. *J Cell Physiol* **191**, 269-282 (2002).
86. Chelakkot, C., Ghim, J. & Ryu, S.H. Mechanisms regulating intestinal barrier integrity and its pathological implications. *Exp Mol Med* **50**(2018).
87. Martini, E., Krug, S.M., Siegmund, B., Neurath, M.F. & Becker, C. Mend Your Fences The Epithelial Barrier and its Relationship With Mucosal Immunity in Inflammatory Bowel Disease. *Cell Mol Gastroenter* **4**, 33-46 (2017).
88. Toumi, R., Samer, A., Soufli, I., Rafa, H. & Touil-Boukoffa, C. Role of Probiotics and Their Metabolites in Inflammatory Bowel Diseases (IBDs). *Gastroenterol Insign* **12**, 56-66 (2021).
89. Zuo, T. & Ng, S.C. The Gut Microbiota in the Pathogenesis and Therapeutics of Inflammatory Bowel Disease. *Front Microbiol* **9**(2018).
90. Macura, B., Kiecka, A. & Szczepanik, M. Intestinal permeability disturbances: causes, diseases and therapy. *Clin Exp Med* **24**(2024).
91. Nishida, A., *et al.* Gut microbiota in the pathogenesis of inflammatory bowel disease. *Clin J Gastroenterol* **11**, 1-10 (2018).
92. Lupp, C., *et al.* Host-mediated inflammation disrupts the intestinal microbiota and promotes the overgrowth of Enterobacteriaceae (Vol 2, pg 119, 2007). *Cell Host Microbe* **2**, 204-204 (2007).
93. Arthur, J.C., *et al.* Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science* **338**, 120-123 (2012).
94. Cho, Y.H., Renouf, M.J., Omotoso, O. & McPhee, J.B. Inflammatory bowel disease-associated adherent-invasive Escherichia coli have elevated host-defense peptide resistance. *FEMS Microbiol Lett* **369**(2022).
95. Zheng, L., Duan, S.L., Dai, Y.C. & Wu, S.C. Role of adherent invasive Escherichia coli in pathogenesis of inflammatory bowel disease. *World J Clin Cases* **10**, 11671-11689 (2022).

96. Khan, S., Giri, S. & Naseri, S. Familial Adenomatous Polyposis: An Uncommon. *J Clin Diagn Res* **18**, Ej1-Ej2 (2024).
97. Petersen, G.M., Slack, J. & Nakamura, Y. Screening Guidelines and Premorbid Diagnosis of Familial Adenomatous Polyposis Using Linkage. *Gastroenterology* **100**, 1658-1664 (1991).
98. Bisgaard, M.L., Fenger, K., Bulow, S., Niebuhr, E. & Mohr, J. Familial Adenomatous Polyposis (Fap) - Frequency, Penetrance, and Mutation-Rate. *Hum Mutat* **3**, 121-125 (1994).
99. Giardiello, F.M., *et al.* Phenotypic Variability of Familial Adenomatous Polyposis in 11 Unrelated Families with Identical Apc Gene Mutation. *Gastroenterology* **106**, 1542-1547 (1994).
100. Dejea, C., Wick, E. & Sears, C.L. Bacterial oncogenesis in the colon. *Future Microbiol* **8**, 445-460 (2013).
101. Johnson, C.H., *et al.* Metabolism Links Bacterial Biofilms and Colon Carcinogenesis. *Cell Metab* **21**, 891-897 (2015).
102. Dejea, C.M., *et al.* Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. *Science* **359**, 592-+ (2018).
103. Wu, S.G., *et al.* A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses. *Nature Medicine* **15**, 1016-U1064 (2009).
104. Terlouw, D., *et al.* Recurrent APC Splice Variant c.835-8A>G in Patients With Unexplained Colorectal Polyposis Fulfilling the Colibactin Mutational Signature. *Gastroenterology* **159**, 1612-1614 e1615 (2020).
105. Bartfeld, S., *et al.* In vitro expansion of human gastric epithelial stem cells and their responses to bacterial infection. *Gastroenterology* **148**, 126-136 e126 (2015).
106. Heijmans, J., *et al.* ER Stress Causes Rapid Loss of Intestinal Epithelial Stemness through Activation of the Unfolded Protein Response. *Cell Rep* **3**, 1128-1139 (2013).
107. Warschkau, D., *et al.* From 3D to 2D: Harmonization of Protocols for Two-dimensional Cultures on Cell Culture Inserts of Intestinal Organoids from Various Species. *Bio-Protocol* **12**(2022).
108. Gu, Z.G., Gu, L., Eils, R., Schlesner, M. & Brors, B. implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811-2812 (2014).
109. Holthaus, D., *et al.* Dissection of Barrier Dysfunction in Organoid-Derived Human Intestinal Epithelia Induced by. *Gastroenterology* **162**, 844-858 (2022).
110. Hollstein, M., Alexandrov, L.B., Wild, C.P., Ardin, M. & Zavadil, J. Base changes in tumour DNA have the power to reveal the causes and evolution of cancer. *Oncogene* **36**, 158-167 (2017).
111. Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Campbell, P.J. & Stratton, M.R. Deciphering Signatures of Mutational Processes Operative in Human Cancer. *Cell Rep* **3**, 246-259 (2013).
112. Sato, T., *et al.* Single Lgr5 stem cells build crypt-villus structures without a mesenchymal niche. *Nature* **459**, 262-U147 (2009).
113. Zhang, H., *et al.* Global research trends on the relationship between IBD and CRC: a bibliometric analysis from 2000 to 2023. *J Health Popul Nutr* **43**, 83 (2024).
114. Pleguezuelos-Manzano, C. & Clevers, H. Mutational signature in colorectal cancer caused by genotoxic pks+ *E. coli*. *Nature* **580**, 269-273 (2020).
115. Bossuet-Greif, N., *et al.* The Colibactin Genotoxin Generates DNA Interstrand Cross-Links in Infected Cells. *Mbio* **9**(2018).
116. Xue, M.Z., *et al.* Structure elucidation of colibactin and its DNA cross-links. *Science* **365**, 1000-+ (2019).
117. Eckstein, F. Phosphorothioates, Essential Components of Therapeutic Oligonucleotides. *Nucleic Acid Therapeutics* **24**, 374-387 (2014).
118. Nookaew, I., *et al.* Detection and Discrimination of DNA Adducts Differing in Size, Regiochemistry, and Functional Group by Nanopore Sequencing. *Chem Res Toxicol* **33**, 2944-2952 (2020).

119. Wang, Y.H., Zhao, Y., Bollas, A., Wang, Y.R. & Au, K.F. Nanopore sequencing technology, bioinformatics and applications. *Nat Biotechnol* **39**, 1348-1365 (2021).
120. Rand, A.C., *et al.* Mapping DNA methylation with high-throughput nanopore sequencing. *Nat Methods* **14**, 411-+ (2017).
121. Baranello, L., *et al.* RNA Polymerase II Regulates Topoisomerase 1 Activity to Favor Efficient Transcription. *Cell* **165**, 357-371 (2016).
122. Pleguezuelos-Manzano, C., *et al.* Mutational signature in colorectal cancer caused by genotoxic pks(+) *E. coli*. *Nature* **580**, 269-273 (2020).
123. Pérez-Berezo, T., *et al.* Identification of an analgesic lipopeptide produced by the probiotic strain Nissle 1917. *Nature Communications* **8**(2017).
124. Tronnet, S., *et al.* Iron Homeostasis Regulates the Genotoxicity of That Produces Colibactin. *Infect Immun* **84**, 3358-3368 (2016).
125. Chagneau, C.V., *et al.* The Polyamine Spermidine Modulates the Production of the Bacterial Genotoxin Colibactin. *Mosphere* **4**(2019).
126. Oliero, M., Alaoui, A.A., McCartney, C. & Santos, M.M. Colorectal cancer and inulin supplementation: the good, the bad, and the unhelpful. *Gastroenterol Rep* **12**(2024).
127. Oliero, M., *et al.* Oligosaccharides increase the genotoxic effect of colibactin produced by strains. *Bmc Cancer* **21**(2021).
128. Thakur, B.K., *et al.* Dietary fibre counters the oncogenic potential of colibactin-producing in colorectal cancer. *Nature Microbiology* **10**(2025).
129. Sato, T. & Clevers, H. Single Lgr5 stem cells build crypt–villus structures in vitro without a mesenchymal niche. *Nature* **459**, 262-265 (2009).
130. Sato, T., Stange, D.E., Ferrante, M. & Clevers, H. Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. *Gastroenterology* **141**, 1762-1772 (2011).
131. Tindle, C., *et al.* A living organoid biobank of patients with Crohn's disease reveals molecular subtypes for personalized therapeutics. *Cell Rep Med* **5**, 101748 (2024).
132. Olafsson, S., *et al.* Somatic Evolution in Non-neoplastic IBD-Affected Colon. *Cell* **182**, 672-684 e611 (2020).
133. Zhang, H.B., *et al.* NAD repletion improves mitochondrial and stem cell function and enhances life span in mice. *Science* **352**, 1436-1443 (2016).
134. Colotta, F., Allavena, P., Sica, A., Garlanda, C. & Mantovani, A. Cancer-related inflammation, the seventh hallmark of cancer: Links to genetic instability. *Carcinogenesis* **30**, 1073-1081 (2009).

7. Acknowledgements

First and foremost, I would like to express my deepest gratitude to Prof. Thomas F. Meyer, who not only envisioned this project but also provided unwavering guidance and support throughout it. His expertise, mentorship, and encouragement were instrumental in the successful completion of this work. I would also like to thank Prof. Thomas Bosch, my official supervisor, for his support with all formal aspects of my doctoral process within the faculty.

I am equally grateful to Dr. Saskia F. Erttmann, Dr. Karina Arellano Ayala, and Dr. David Holthaus, whose mentorship and constructive discussions significantly shaped the direction of this research. Their guidance and willingness to share their knowledge enriched both the project and my scientific development.

Special thanks go to Dr. Karina Arellano Ayala and Alvaro Quevedo-Olmos for their substantial contributions. Karina played a key role in accelerating the project's progress, while Alvaro expertly handled all bioinformatics analyses. Their dedication, encouragement, and availability made a lasting impact on my academic journey.

I would like to particularly acknowledge Dr. David Holthaus, who inspired me to delve deeply into scientific exploration. His continuous support, insightful guidance, and collaborative spirit were invaluable. I greatly appreciate our discussions, the effective division of tasks, and his readiness to contribute even amidst a demanding schedule. I am proud that this collaborative spirit is reflected in our joint publication.

My sincere appreciation goes to Prof. Dr. Konrad Aden, who generously provided all patient biopsy samples for this project. Despite his busy clinical and research commitments as a doctor and group leader at IKMB, he was always available for scientific discussions and support. I also thank Prof. Dr. Mark Ellrichmann for kindly providing all polyp samples used in this study.

I am grateful to all those who contributed to the writing of this dissertation. I thank Prof. Thomas F. Meyer and Dr. David Holthaus for their valuable feedback and comments. A special thanks to David for his thorough proofreading of the manuscript.

Finally, I wish to extend my heartfelt thanks to all my colleagues like Xiaochen Zhang whose support made this doctoral work possible. Your help, both scientific and personal, has been deeply appreciated.

8. Abbreviations

2D	two-dimensional
3D	three-dimensional
ANOVA	analysis of variance
ADF	Advanced DMEM
APC	Adenomatous polyposis coli
bp	Base pair
BSA	bovine serum albumin
BFT	B. fragilis toxin
BLAST	Basic Local Alignment Search Tool
BMP	Bone morphogenetic protein
BCA	Bovine serum albumin
CDM	Colibactin damage motif
cDNA	complementary DNA
CFUs	Colony forming units
CHIR	CHIR99021, GSK inhibitor
CIN	Chromosomal instability
CNVs	Copy number variations
CRC	Colorectal cancer
DAPI	4',6-diamidino-2-phenylindole
DMEM	Dulbecco's modified eagle medium
DNA	Deoxyribonucleic acid
DSBs	Double-strand breaks
DSS	Dextran sulfate sodium
DTT	DL-Dithiotreitol
ECL	Extracellular matrix
E. coli	Escherichia coli
EDTA	ethylenediaminetetraacetic acid
EdU assay	5-ethynyl-2'-deoxyuridine assay, marks proliferating cells
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
ETBF	Enterotoxigenic Bacteroides fragilis

FAP	Familial adenomatous polyposis
Fig.	Figure
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GFP	Green fluorescent protein
HCL	Hydrochloric acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HNPCC	Hereditary nonpolyposis colorectal cancer
HR	Homologous recombination
IBD	Inflammatory bowel disease
IFA	immunofluorescence assay
KCL	Potassium chloride
KH ₂ PO ₄	Potassium dihydrogenphosphate
LOH	Loss of heterozygosity
MMR	Mismatch repair
MOI	Multiplicity of infection
MSI	Microsatellite instability
Na ₂ PHO ₄	Disodium hydrogen phosphate
NaCl	Sodium chloride
OD	Optical density
ODMs	organoid-derived monolayers
P/S	penicillin/streptomycin
PBS	phosphate-buffered saline
PAGE	Polyacrylamide gel electrophoresis
PGE ₂	Prostaglandin E ₂
PFA	Paraformaldehyde
PKS	Polyketide synthases
PPARs	Peroxisome proliferator-activated receptors
RT-qPCR	reverse transcriptase quantitative polymerase chain reaction
ROCK	Rho-associated protein kinase
Rspo	R-spondin
RT	Room temperature
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SCFA	short chain fatty acid

SD	standard deviation
TEER	transepithelial electrical resistance
TEMED	Tetramethylethylenediamine
TGF β	Transforming growth factor beta
Tris	Tris(hydroxymethyl)aminomethane
WGS	Whole genome sequencing
WRN	Conditioned medium with of WNT3A, R-Spondin1/3, and Noggin
WT	wild type

Units

°C	Degree(s) Celsius
μ F	Microfarad(s)
μ g	Microgram(s)
μ l	Microlitre(s)
μ m	Micrometre(s)
μ M	Micromolar
cm	Centimetre(s)
g	Gram(s)
Gy	Gray(s)
kD	Kilodalton(s)
kV	Kilovolt(s)
l	Litre(s)
M	Molar
mA	Milliampere(s)
mg	Milligram(s)
ml	Millilitre(s)
mm	Millimetre(s)
mM	Millimolar
ng	Nanogram(s)
nm	Nanometre(s)
U	Unit(s)
V	Volt(s)
x g	Relative centrifugal force