ADHESION AND INVASION DYNAMICS OF *FUSOBACTERIUM NUCLEATUM* IN COLORECTAL CANCER

by

Onalenna Neo

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY



UNIVERSITY OF BIRMINGHAM

College of Medical and Dental Sciences

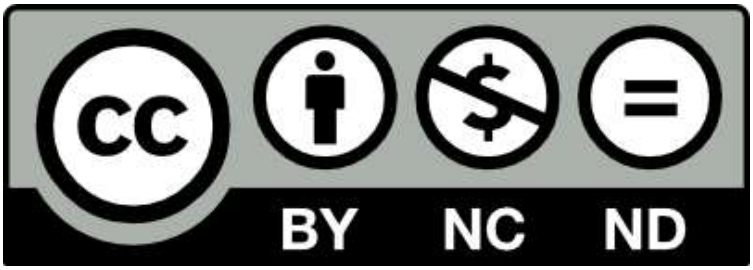
School of Dentistry

University of Birmingham

January 2024

UNIVERSITY OF BIRMINGHAM

University of Birmingham Research Archive
e-theses repository



This unpublished thesis/dissertation is under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence.

You are free to:

Share — copy and redistribute the material in any medium or format

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

NonCommercial — You may not use the material for commercial purposes.

NoDerivatives — If you remix, transform, or build upon the material, you may not distribute the modified material.

No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

Notices:

You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable exception or limitation.

No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as publicity, privacy, or moral rights may limit how you use the material.

Unless otherwise stated, any material in this thesis/dissertation that is cited to a third-party source is not included in the terms of this licence. Please refer to the original source(s) for licencing conditions of any quotes, images or other material cited to a third party.

ABSTRACT

The role of microbial dysbiosis in colorectal cancer (CRC), which ranks amongst the top five deadliest cancers, has gained a lot of research interest recently. This study focused on the bacterium *Fusobacterium nucleatum*, which is often enriched in CRC patients compared to healthy controls. The overarching aim of the study was to investigate the adhesion and invasion dynamics of three *F. nucleatum* subspecies, namely subspecies (ssp.) *nucleatum* (FNN23), ssp. *polymorphum* (FNP) and ssp. *animalis* (FNA), to improve understanding of the mechanism of action of *F. nucleatum* in CRC. Adhesion and invasion assays were performed using a monolayer model of a CRC cell line, Caco-2. Subspecies-specific differences in the levels of adhesion to CRC cells were observed, with FNP having higher adhesion levels than the other two subspecies. The adhesins, FadA and Fap2 are important *F. nucleatum* virulence factors in CRC. Therefore, their contribution to adherence and invasion, in FNN23 was investigated using deletion mutants of the adhesins. The mutations resulted in decreased adherence and invasion of Caco-2 cells by FNN23. The decrease was greater with the Fap2 mutant than the FadA mutant strain. Additionally, this study determined the impact on adhesion and invasion of Tspan6, a protein belonging to the tetraspanin family of proteins which have been implicated in various aspects of cancer. While Tspan6 expression improved the adherence of FNN23 to Caco-2 cells, it decreased that of FNA and made no difference to FNP adherence. Interestingly, the invasion of FNN23 to Tspan6 expressing cells was decreased, but no significant impact on FNA and FNP invasion was observed. Furthermore, the impact of the bacterial growth phase on adhesion dynamics was also investigated. FNN23 and FNA samples experienced over 2-fold increases in their levels of adherence at the exponential compared to the stationary phase, while FNP adherence remained unaffected. Then, bacterial RNA sequencing was performed to

establish whether there were differences in the expression of adhesins and other virulence factors resulting in the observed differences in adherence and invasion of the subspecies. Global gene expression differences between stationary and exponential-phase bacteria were observed. In addition, differences in the expression of adhesion genes at the exponential and stationary phases were discovered. This study also demonstrated differential induction of cytokine and chemokines by the *F. nucleatum* subspecies compared to controls using the Olink target 48 cytokine and chemokine assay service. Lastly, paired metagenomics and transcriptomics analyses were performed on 57 CRC samples and their controls to determine the impact of the metagenomic profile on gene expression. Metagenomic differences led to differences in immune cell composition and the expression of Hallmark pathways. Understanding the adherence and invasion dynamics of *F. nucleatum* subspecies is key to uncovering key mechanisms important in *F. nucleatum* virulence CRC, which can help inform diagnostics and therapeutic interventions.

I dedicate this thesis to my nephew, Leatile Neo.

Against the battle, I pray you win.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank the Government of Botswana for sponsoring my PhD. Thanks to my supervisors, Dr Sarah Kuehne and Dr Fedor Berditchevski for their invaluable support during my PhD. Both of you made an incredibly brave choice by agreeing to work with me on a new project with only two years left on the PhD. For that I will forever remain grateful. Thank you, Dr Kuehne, for your unwavering support, incredible scientific insights and amazing positivity that made all the hard parts about doing science bearable. Thank you, Dr Berditchevski, for challenging me to push myself harder, and for your guidance in the project. I am also very grateful to Professor Steven Van Laere and Professor Andrew Beggs for their collaboration and assistance. Thanks to Dr Michelle Buckner for her support and guidance during the first years of my PhD.

My sincere gratitude to all the amazing people I worked alongside at the T102 lab, G7 lab and at the Birmingham Dental hospital labs. Thanks to the technical team at the school of Dentistry for their assistance. Thank you to the students and postdocs at the Dental hospital for always being willing to help and for being friendly and supportive. Thank you to Sarah Element, Ruth Maganga, John Hey and Lizzy Darby for having been an incredible support system and for being there for me especially during the tougher parts of the journey. To the TT'ers at the Dental hospital, thank you for the help, the friendship, and the laughter.

I am grateful to my parents, for their steadfast support, unconditional love and for always keeping me in their prayers. Special thanks to my brother, Modise, for being supportive and for inspiring me to always push boundaries. I am grateful to God, for seeing me through one of the toughest periods of my life, both academically and personally. Indeed, "Modimo o nale nna".

CONTENTS

1 GENERAL INTRODUCTION 1

1.1 COLORECTAL CANCER (CRC) 2
1.2 CRC INCIDENCE AND GLOBAL DISTRIBUTION 3
1.3 CRC RISK FACTORS 5
1.3.1 Sex 5
1.3.2 Age 5
1.3.3 Race/ethnicity 6
1.3.4 Nutrition and diet 6
1.3.5 Physical activity 7
1.3.6 Smoking and alcohol consumption 7
1.4 DEVELOPMENT OF CRC 9
1.4.1 Adenoma-carcinoma model 9
1.4.2 Serrated pathway 11
1.4.3 Consensus molecular subtyping (CMS) 12
1.4.4 CRC diagnostics 14
1.4.5 CRC therapeutics 14
1.5 CRC AND THE MICROBIOTA 15
1.6 PROPOSED MODELS FOR THE ROLE OF BACTERIA IN COLORECTAL CANCER 18
1.6.1 Drivers' vs passengers' model 18
1.6.2 The alpha bug hypothesis 18
1.7 *FUSOBACTERIUM NUCLEATUM* (*F. NUCLEATUM*) 19
1.7.1 F. nucleatum subspecies 19
1.7.2 Role of F. nucleatum in oral biofilms 20
1.7.3 F. nucleatum in oral infections 21
1.7.4 F. nucleatum in extraoral infections 21
1.8 ROLE OF *F. NUCLEATUM* IN CANCER 24
1.8.1 Mechanism of F. nucleatum carcinogenesis 25
1.8.2 The role of F. nucleatum in colorectal cancer 25
1.9 *F. NUCLEATUM* VIRULENCE MECHANISMS IN CRC 26
1.9.1 FadA driven promotion of tumorigenesis 26
1.9.2 Modulation of the immune system 29
1.9.3 Modulation of autophagy 30
1.10 TETRASPANINS 32
1.10.1 Structure of Tetraspanis 32
1.10.2 Role of tetraspanins in viral and parasitic infections 34
1.10.3 Role of tetraspanins in bacterial infections 35
1.10.4 Role of tetraspanins in cancer 36
1.10.5 Tetraspanin 6 (Tspan6) 38
1.10.6 Role of tetraspanin 6 in colorectal cancer 39
1.11 IMPORTANCE 40
1.11.1 Aims and objectives. 41

2 GENERAL MATERIALS AND METHODS 43

2.1 MICROBIOLOGICAL METHODS 44
2.1.1 Bacterial strains 44
2.1.2 Bacterial culture and maintenance 44
2.1.3 Gram staining 46
2.1.4 16S rRNA polymerase chain reaction (PCR) 46
2.1.5 Gel electrophoresis 47
2.1.6 16S rRNA sequencing 48
2.1.7 Growth kinetics and Miles and Misra for determining number of viable cells 48
2.1.8 RNA extraction from F. nucleatum 49

2.1.9 Genomic DNA removal and sample clean-up of RNA *50*
2.2 MAMMALIAN CELL CULTURE METHODS 51
2.2.1 Cell lines *51*
2.2.2 Cell culture and maintenance *52*
2.2.3 Recovery of frozen cells *53*
2.2.4 Cell seeding *53*
2.2.5 Cryopreservation *54*
2.2.6 Protein extraction from Caco-2 cells *55*
2.2.7 Western Blotting and SDS-PAGE *55*
2.2.8 Mycoplasma testing *56*
2.2.9 Caco-2 cells polarisation *57*
2.2.10 Transepithelial electrical resistance (TEER) measurement *57*
2.2.11 Scanning Electron Microscopy (SEM) *58*
2.3 COMBINED CELL CULTURE AND MICROBIOLOGICAL METHODS 58
2.3.1 Adhesion/Invasion assay *58*
2.3.2 Adhesion and invasion for stationary vs exponential phase assays *60*
2.3.3 Chemokine and cytokine profiling *61*
2.4 METAGENOMICS AND TRANSCRIPTOMICS 64
2.4.1 RNA sequencing *64*
2.4.2 Bioinformatic analysis *64*
2.4.3 Whole metagenomic sequencing data *65*
2.4.4 Paired metagenomic and transcriptomic analysis *66*
2.5 STATISTICAL ANALYSIS 66

3 INVESTIGATING THE DYNAMICS OF ADHESION AND INVASION OF *F. NUCLEATUM* SUBSPECIES ON CRC CELLS 67

3.1 INTRODUCTION 68
3.1.1 Studying bacterial adhesion and invasion into eukaryotic cells *72*
3.1.2 Aims and objectives *73*
3.2 SUBSPECIES CHARACTERISATION 74
3.3 BACTERIAL GROWTH KINETICS 76
3.4 CACO-2 CELLS GROWTH AND MAINTENANCE 78
3.5 SEM IMAGES OF CACO-2 CELLS GROWN ON MEMBRANE INSERTS 79
3.6 WESTERN BLOTTING 80
3.7 MONOLAYER FORMATION 81
3.8 ADHESION AND INVASION OF SUBSPECIES TO CRC CELLS 82
3.9 *F. NUCLEATUM* SUBSPECIES DIFFER IN ADHESION AND INVASION TO CACO-2 CELLS 82
3.10 TSPAN6 EXPRESSION IMPACTS THE ASSOCIATION AND INTERNALISATION OF *F. NUCLEATUM* SUBSPECIES TO CACO-2 CELLS. 87
3.11 *F. NUCLEATUM* SUBSPECIES SHOW REDUCED INVASION RELATIVE TO ADHESION 88
3.12 THE ROLE OF THE *FADA* AND *FAP2* MUTANTS ON THE ADHESION OF FNN23 TO CACO-2 CELLS 89
3.13 DETERMINATION OF THE IMPACT OF MUTANTS ON THE *GROWTH* RATE OF FNN23 90
3.14 MUTANTS IN *FADA* AND *FAP2* IMPACT THE ASSOCIATION AND THE INTERNALISATION OF FNN23 TO CACO-2 CELLS 92
3.15 TSPAN6 EXPRESSION IMPACTS THE ADHERENCE OF THE *FADA* AND THE *FAP2* MUTANTS IN FNN23 TO CACO-2 CELLS. 94
3.16 TSPAN6 EXPRESSION IMPACTS THE INVASION OF THE *FADA* AND THE *FAP2* MUTANT STRAINS IN FNN23 TO CACO-2 CELLS 96
3.17 *S. GALLOLYTICUS* AND *S. EPIDERMIDIS* CONTROL FOR ADHESION AND INVASION TO T6 AND CONTROL CELLS 98
3.18 KEY FINDINGS OF THIS CHAPTER 100
3.19 DISCUSSION 101

4 GROWTH PHASE IMPACT ON THE ADHESION, INVASION, AND INFLAMMATORY MEDIATOR SECRETION OF *F. NUCLEATUM* SUBSPECIES UPON CACO-2 CELL INFECTION. 106

4.1 INTRODUCTION 107
1.1.1 Cytokine and chemokine expression in the tumour microenvironment (TME) *108*
4.1.1 Hypothesis *110*

4.1.2 Aims and Objectives *111*
4.2 IMPACT OF GROWTH PHASE ON THE ADHESION OF THE SUBSPECIES TO CACO-2 CELLS 112
4.3 IMPACT OF GROWTH PHASE ON THE ADHESION OF THE SUBSPECIES TO TSPAN6 EXPRESSING CACO-2 CELLS. 113
4.4 THE ADHESINS FAP2 AND FADA ARE IMPORTANT FOR THE ADHESION OF FNN23 TO CACO-2 CELLS AT THE EXPONENTIAL PHASE 115
4.5 IMPACT OF GROWTH PHASE ON THE INVASION OF THE SUBSPECIES TO CACO-2 CELLS, INCLUDING T6 EXPRESSING CACO-2 CELLS 117
4.6 KEY FINDINGS 120
4.7 EVALUATING THE IMPACT OF *F. NUCLEATUM* ON CYTOKINE EXPRESSION USING THE O-LINK TARGET 48 PANEL 121
4.8 GENERAL CYTOKINE EXPRESSION TRENDS ON THE O-LINK TARGET 48 CYTOKINE PANEL 123
4.9 TOP CYTOKINES ALL STRAINS 125
4.10 IMPACT OF GROWTH PHASE ON CYTOKINE EXPRESSION 128
4.11 IMPACT OF TSPAN6 EXPRESSION ON CYTOKINE EXPRESSION 132
4.12 KEY FINDINGS 135
4.13 DISCUSSION 136

5 RNA SEQUENCING TO UNCOVER IMPACT OF THE GROWTH PHASE AND STRAIN BACKGROUND ON GENE EXPRESSION OF *F. NUCLEATUM* SUBSPECIES 142

5.1 BACKGROUND 143
5.2 AIMS 144
5.3 RNA SEQUENCING QUALITY METRICS 145
5.4 RNA SEQUENCING QUALITY CHECK 146
5.5 GLOBAL GENE EXPRESSION PATTERNS 147
5.6 SUBSPECIES-SPECIFIC TRANSCRIPTOMIC DIFFERENCES 148
5.7 GLOBAL GROWTH PHASE IMPACT ON THE TRANSCRIPTOMIC PROFILE OF *F. NUCLEATUM* SUBSPECIES. 149
5.8 FUNCTIONAL ANALYSIS OF GLOBAL GENE EXPRESSION PATTERNS 151
5.9 NETWORK MAPPING AND ENRICHMENT ANALYSIS OF GLOBAL RNA EXPRESSION DIFFERENCES 152
5.10 STRAIN SPECIFIC DIFFERENCES IN GENE EXPRESSION 156
5.11 FUNCTIONAL ANALYSIS OF GROWTH PHASE-DEPENDENT STRAIN-SPECIFIC DIFFERENCES IN GENE EXPRESSION 159
5.12 FNN23 COMPARED TO MUTANTS 163
5.13 EXPRESSION LEVELS OF ADHESINS ACROSS THE DIFFERENT CONDITIONS 168
5.14 EXPRESSION LEVELS OF VIRULENCE GENES ASSOCIATED WITH ADHESION 170
5.15 DISCUSSION 172

6 UNCOVERING THE METAGENOMIC AND TRANSCRIPTOMIC PATTERNS OF CRC TUMOURS AND CONTROLS, THE ROLE OF TSPAN6 AND *F. NUCLEATUM* 177

6.1 INTRODUCTION 178
6.1.1 Tumour staging *179*
6.1.2 Background *179*
6.1.3 Aims *180*
6.2 DESCRIPTIVE STATISTICS OF THE DATASETS (CLINICOPATHOLOGICAL INFORMATION) 181
6.3 OVERVIEW OF OVERALL METAGENOMIC DIFFERENCES OF TUMOUR VS NORMAL SAMPLES 181
6.4 SAMPLE GROUPING BASED ON THE METAGENOMIC PROFILE 184
6.5 SAMPLE CLUSTERS HAVE SIGNIFICANT DIFFERENCES IN SOME OF THE BACTERIAL SPECIES' COMPOSITION. 185
6.6 DISTRIBUTION OF KEY PATHOGENS ACROSS THE LOUVAIN CLUSTERS 189
6.7 EVALUATION OF THE IMMUNE PROFILE OF THE SAMPLES REVEALS SIGNIFICANT DOWNREGULATION OF THE M2 MACROPHAGE POPULATIONS IN NORMAL SAMPLE. 191
6.8 COMPARISON OF HALLMARK ACTIVITY BETWEEN THE METAGENOMIC (LOUVAIN) CLUSTERS 195
6.9 COMPARISON OF HALLMARK ACTIVITY BETWEEN THE CLUSTERS AS A FUNCTION OF TSPAN6 198
6.10 SAMPLE GROUPING BASED ON THE GENE EXPRESSION PROFILE OF THE SAMPLES 205
6.11 DIFFERENTIAL PATHWAY ANALYSIS OF EXPRESSION-BASED CLUSTERS 207
6.12 DIFFERENCES IN GENE EXPRESSION AND PATHWAY ENRICHMENT BETWEEN SAMPLES WITH HIGH AND LOW NUMBERS OF *F. NUCLEATUM* 212

6.13 CORRELATION BETWEEN LOUVAIN CLUSTERS WITH TUMOUR STATUS AND CMS PREDICTION 214
6.14 DISCUSSION 217
6.15 CONCLUSION 223

7 GENERAL DISCUSSION 224

7.1 ADHESION AND INVASION DYNAMICS OF *F. NUCLEATUM* 225
7.2 IMPACT OF ADHERENCE AND INVASION ON THE RELEASE OF INFLAMMATORY MEDIATORS 229
7.3 TRANSCRIPTIONAL ANALYSIS OF *F. NUCLEATUM* SUBSPECIES REVEALS DIFFERENCES IN THE EXPRESSION OF ADHESION GENES 231
7.4 BIOINFORMATIC ANALYTICS TO UNDERSTAND THE IMPACT OF MICROBIAL DYSBIOSIS ON CRC 232
7.5 *F. NUCLEATUM*, CAUSE OR CONSEQUENCE 234

8 REFERENCES 236

9 APPENDICES 269

9.1 APPENDIX 6.1 270

LIST OF FIGURES

FIGURE 1.1: GLOBAL INCIDENCE AND MORTALITY OF CRC FOR BOTH SEXES. 4
FIGURE 1.2: MODIFIABLE AND NON-MODIFIABLE RISK FACTORS FOR CRC 8
FIGURE 1.3: CONSENSUS MOLECULAR SUBTYPE (CMS) CLASSIFICATIONS....13
FIGURE 1.4: ROLE OF F. NUCLEATUM IN DISEASE AT DIFFERENT BODY SITES 23
FIGURE 1.5 : ROLE AND FUNCTION OF FADA IN CRC. MFADA FROM F. NUCLEATUM EMBEDS IN THE INNER MEMBRANE AND ANCHORS PRE-FADA WHICH EXTENDS OUT OF THE MEMBRANE. 28
FIGURE 1.6: ROLE OF F. NUCLEATUM IN AUTOPHAGIC IMPAIRMENT....31
FIGURE 1.7: THE STRUCTURE OF TETRASPANINS. 33
FIGURE 2.1:SCHEMATIC OF CELL LINE TRANSFECTION. 52
FIGURE 2.2:SCHEMATIC OF THE ADHESION AND INVASION PROTOCOL. 60
FIGURE 3.1:MICROSCOPIC IMAGE OF GRAM-STAINED BACTERIA. 74
FIGURE 3.2:PCR CONFIRMATION OF SUBSPECIES ON ELECTROPHORESIS GEL 75
FIGURE 3.3: *OD600NM GROWTH KINETICS OF FNN23, FNA AND FN* 76
FIGURE 3.4: BACTERIAL QUANTIFICATION OF SUBSPECIES 77
FIGURE 3.5:CACO-2 CELL GROWTH. 78
FIGURE 3.6: SEM IMAGES OF CACO-2 CELLS....79
FIGURE 3.7: WESTERN BLOTS OF T6 EXPRESSION....80
FIGURE 3.8:SUBSPECIES-SPECIFIC DIFFERENCES IN ADHERENCE RATES OF FNP, FNA AND FNN23 TO T6 CELLS AND CONTROL CELLS 84
FIGURE 3.9:SUBSPECIES-SPECIFIC DIFFERENCES IN THE INVASION RATES OF FNP, FNA AND FNN23 TO T6 CELLS AND CONTROL CELLS....86
FIGURE 3.10: IMPACT OF TSPAN6 EXPRESSION ON THE ADHESION AND INVASION OF F. NUCLEATUM SUBSPECIES TO CACO-2 CELLS....88
FIGURE 3.11:INVADED CELLS RELATIVE TO ADHERED CELLS FOR T6 AND CONTROL CELLS....89
FIGURE 3.12:*GROWTH KINETICS OF SUBSPECIES OVER 24 HOURS* 90
FIGURE 3.13: BACTERIAL QUANTIFICATION OF THE FADA AND FAP2 MUTANTS AND WILD-TYPE FNN23 STRAIN. .91
FIGURE 3.14:IMPACT OF THE FADA AND FAP2 MUTANTS ON THE ADHESION AND INVASION OF FNN23 TO CACO-2 CELLS. 93
FIGURE 3.15: IMPACT OF A MUTATION IN THE FADA GENE AND THE FAP2 GENE ON THE ADHESION OF FNN23 TO T6 EXPRESSING CACO-2 CELLS. 95
FIGURE 3.16:IMPACT OF THE TSPAN6 EXPRESSION ON INVASION OF MUTANTS. 97
FIGURE 3.17: ADHESION AND INVASION OF SGG AND S. EPIDERMIDIS 99
FIGURE 4.1: ROLE OF INFLAMMATORY MEDIATORS IN THE TME 109
FIGURE 4.2: ADHERENCE TO CACO-2 CELLS OF THE SUBSPECIES (FNN23, FNP AND FNA AT EXPONENTIAL PHASE VS ADHERENCE AT STATIONARY PHASE. 113
FIGURE 4.3: EXPONENTIAL VS. STATIONARY PHASE ADHERENCE OF FNN23, FNP AND FNA TO T6 CELLS. Y- ...114
FIGURE 4.4: EXPONENTIAL VS. STATIONARY PHASE ADHERENCE OF FNN23ΔFAP2 AND FNN23 ΔFADA COMPARED TO WT FNN23....116
FIGURE 4.5: EXPONENTIAL VS. STATIONARY PHASE INVASION OF F. NUCLEATUM SUBSPECIES AND FNN23 MUTANTS IN FADA AND FAP2 TO CACO-2 CELLS....118
FIGURE 4.6: EXPONENTIAL VS. STATIONARY PHASE INVASION OF F. NUCLEATUM SUBSPECIES AND FNN23 ADHESIN MUTANTS TO T6 CELLS. REPLICATES FOR EACH EXPERIMENT. ASTERISK (*) MEANS $p<0.05$, NS = NOT SIGNIFICANT AT ALPHA=0.05....119
FIGURE 4.7: SAMPLE DISTRIBUTION PLOT SHOWING NPX (Y-AXIS) AND SAMPLE IDS (X-AXIS).....122
FIGURE 4.8: HEATMAP SHOWING THE CYTOKINE SECRETION PATTERNS OF 45 CYTOKINES FROM THE TARGET 48 O-LINK CYTOKINE PANEL. LEVELS OF EXPRESSION OF A PARTICULAR CYTOKINE FOR A CORRESPONDING STRAIN. STRAIN ID ARE COLOUR CODED ON THE LEGEND....124
FIGURE 4.9: BOXPLOTS OF THE COMPARISONS BETWEEN THE BACTERIA FOR CXCL10 AND IL17C EXPRESSION.127
FIGURE 4.10: PCA PLOT (TOP PANEL) AND HEATMAP (BOTTOM PANEL) OF SIX SIGNIFICANTLY DIFFERENTIALLY EXPRESSED CYTOKINES BASED ON THE GROWTH PHASE....130
FIGURE 4.11: BOXPLOTS OF THE EXPRESSION OF IL6 FOR EXPONENTIAL PHASE SAMPLES COMPARED TO STATIONARY SAMPLES 131

FIGURE 4.12: PCA PLOT BASED ON THE CYTOKINES THAT ARE DIFFERENTIALLY SECRETED BASED ON CELL TYPE. 133
FIGURE 4.13: HEATMAP OF FOUR SIGNIFICANTLY DIFFERENTIALLY EXPRESSED CYTOKINES BASED ON THE CELL TYPE. 134
FIGURE 5.1:EXPRESSION PLOT SHOWING THE EXPRESSION LEVEL(LOG CPM) BY LIBRARY. 146
FIGURE 5.2:MDS AND VARIANCE BY MDS DIMENSION OF THE STRAINS.. 148
FIGURE 5.3:HEATMAP SHOWING THE GENE EXPRESSION PATTERN OF THE STRAINS (-1 > LOG_2 FOLDCHANGE > 1, FDR = 0.05) RELATIVE TO AVERAGE EXPRESSION. 149
FIGURE 5.4: HEATMAP SHOWING THE GENE EXPRESSION PATTERN OF EXPONENTIAL PHASE SAMPLES RELATIVE TO STATIONARY PHASE SAMPLES (-1>LOG_2 FOLDCHANGE>1, FDR=0.05). 150
FIGURE 5.5: COG FUNCTIONAL GROUPS OF DIFFERENTIALLY EXPRESSED GENES OF STATIONARY VS EXPONENTIAL PHASE GENERATED FROM DEGS OF EXPONENTIAL PHASE RELATIVE TO STATIONARY PHASE. 152
FIGURE 5.6: CLUSTERS OF PROTEINS DERIVED FROM DEGS OF STATIONARYS VS EXPONENTIAL SAMPLES. 153
FIGURE 5.7: VENN DIAGRAM SHOWING THE NUMBER OF SHARED DEGS BETWEEN THE STAINS. THE NUMBER OF I158
FIGURE 5.8: COG FUNCTIONAL ANALYSIS OF DIFFERENTIALLY EXPRESSED GENES OF STATIONARY VS EXPONENTIAL FOR ALL STRAINS, INCLUDING THE THREE SUBSPECIES AND THE TWO MUTANTS. 162
FIGURE 5.9: MDS PLOT OF FNN23 COMPARED TO ADHESIN MUTANT STRAINS.....163
FIGURE 5.10: VENN DIAGRAM SHOWING THE NUMBER OF SHARED DEGS BETWEEN THE WILD TYPE FNN23 STRAIN AND THE ADHESIN MUTANTS. 164
FIGURE 5.11:HEATMAP OF DIFFERENTIALLY EXPRESSED GENES OF THE WILD TYPE FNN23, THE FADA AND FAP2 MUTANTS AT STATIONARY PHASE RELATIVE TO AVERAGE EXPRESSION.....167
FIGURE 5.12:HEATMAP OF THE LOG_2 FOLDCHANGES OF THE EXPRESSION ACROSS ALL STRAINS OF ADHESION-ASSOCIATED GENES. 170
FIGURE 6.1:SAMPLE GROUPING BASED ON THE METAGENOMIC PROFILE 185
FIGURE 6.2:VOLCANO PLOTS OF DIFFERENCES IN THE BACTERIAL COMPOSITION OF THE 5 LOUVAIN CLUSTERS..187
FIGURE 6.3:HEATMAP OF THE BACTERIAL SPECIES GROUPS AND THEIR DISTRIBUTION ACROSS THE LOUVAIN CLUSTERS. 188
FIGURE 6.4: DISTRIBUTION OF SOME CRC-ASSOCIATED BACTERIA ACROSS THE SAMPLES. 190
FIGURE 6.5:IMMUNE PROFILE OF THE SAMPLES (COMBINES NORMAL AND TUMOUR SAMPLES). 192
FIGURE 6.6:HEATMAP OF HALLMARK ACTIVITY AS A FUNCTION OF TSPAN6 EXPRESSION LEVEL 200
FIGURE 6.7: PANEL SHOWING CORRELATION OF TETRASPANIN EXPRESSION WITH HALLMARK ACTIVITIES FOR ALL SAMPLES, NORMAL AND TUMOURS. 201
FIGURE 6.8: PANEL SHOWING THE TOP (+VE) CORRELATIONS OF TETRASPANIN EXPRESSION WITH HALLMARK ACTIVITIES FOR TUMOUR (A) AND NORMAL (B) SAMPLES. 203
FIGURE 6.9: HEATMAP OF PATHWAY ACTIVITY AS A FUNCTION OF TSPAN6 EXPRESSION INTUMOURS AND NORMAL SAMPLE. 204
FIGURE 6.10: T-SNE CLUSTERS OF SAMPLES BASED ON THEIR EXPRESSION PROFILE.....205
FIGURE 6.11:T-SNE CLUSTERS OF NORMAL SAMPLES (LEFT PANEL) AND TUMOUR SAMPLES (RIGHT PANEL) BASED ON THEIR EXPRESSION PROFILE……..…. 209
FIGURE 6.12: ASSOCIATIONS OF LOUVAIN CLUSTERS WITH CLINICOPATHOLOGICAL FEATURES. 216

LIST OF TABLES

TABLE 2.1:SUMMARY OF BACTERIAL STRAINS SHOWING STRAIN CODE, GROWTH MEDIA AND WHERE THE STRAINS WERE SOURCED 45
TABLE 2.2:16S UNIVERSAL PRIMERS PCR CYCLING CONDITIONS 47
TABLE 2.3: LIST OF CYTOKINES ON THE OLINK TARGET 48/40 CYTOKINE PANEL....63
TABLE 3.1:OPTIMAL SEEDING DENSITY OF CACO-2 CELLS. 81
TABLE 4.1:BIOLOGICAL SAMPLE IDS AS USED WITH O-LINK AND THEIR CORRESPONDING STRAIN NAME AND GROWTH PHASE 121
TABLE 4.2: CYTOKINES DIFFERENTIALLY EXPRESSED FOR EXPONENTIAL VS. STATIONARY PHASE CULTURES 128
TABLE 5.1: RNA SEQUENCING DATA QUALITY....145
TABLE 5.2: KEGG ENRICHMENT IDS OF PROTEIN CLUSTERS OF THE DEGS OF EXPONENTIAL VS STATIONARY SAMPLES....154
TABLE 5.3:INTERPRO TERMS FOR PROTEINS IN CLUSTER 3 155
TABLE 5.4:GENES DIFFERENTIALLY EXPRESSED IN FNN23 STRAIN AND FAP2 AND FADA GENE MUTANT STRAINS 166
TABLE 6.1:TOP TEN OVER AND UNDER-ABUNDANT SPECIES FOR TUMOUR VS NORMAL BACTERIAL METAGENOME 183
TABLE 6.2: NUMBER OF DIFFERENTIALLY EXPRESSED SPECIES BETWEEN THE CLUSTERS 186
TABLE 6.3: COMPARISON OF IMMUNE CELL SCORES BETWEEN LOUVAIN METAGENOMIC CLUSTERS 194
TABLE 6.5:PATHWAYS DIFFERENTIALLY ENRICHED FOR COMBINED NORMAL AND TUMOURS (N+T) AND FOR TUMOUR(T) SAMPLES ONLY 197
TABLE 6.6:PATHWAYS DIFFERENTIALLY ENRICHED FOR NORMAL SAMPLES ONLY 198
TABLE 6.7:NUMBER OF DIFFERENTIALLY EXPRESSED GENES FOR THE EXPRESSION-BASED CLUSTERS. 206
TABLE 6.8: DIFFERENTIALLY ENRICHED PATHWAYS BETWEEN EXPRESSION-BASED CLUSTERS 208
TABLE 6.9:NUMBER OF DIFFERENTIALLY EXPRESSED GENES BETWEEN THE CLUSTERS FOR NORMALS AND FOR TUMOURS 211

ABBREVIATIONS

Aap	Accumulation associated protein
Agr	Accessory gene regulator
AJCC	American Joint Committee on Cancer
ANI	Aveerage nucleotide identity
ANOVA	Analysis of Variance
APC	Adenomatous polyposis coli
APP	Amyloid Precursor Protein
ATCC	American Type Culture Collection
ATG7	Autophagy related 7
ATM	Ataxia telangiectasia mutated
BHI	Brain heart infusion
BRAF	(B-Raf proto-oncogene, serine/threonine kinase
Caco-2	Cancer coli-2
CDT	Cytolethal distending toxin
CFU	Colony-forming units
CHF	Cytotoxic heme factor
CIMP	CpG island methylator pathways
CIN	Chromosomal instability
CMS	Consensus molecular subtyping
CMS	Consensus molecular subtype
COG	Clusters of Orthologous Genes
CPM	Counts per million
CRAC1	Colorectal adenoma and carcinoma 1
CRC	Colorectal cancer
CT	computerised tomography
CTLA4	Cytotoxic T-lymphocyte associated protein 4
DEG	Differentially expressed genes
DMSO	Dimethyl sulfoxide
DNA	Deoxy ribonucleic acid
EGFR	Epidermal growth factor receptor
EMT	Epithelial to mesenchymal transition
EMVI	Extramural venous invasion
EPEC	Enteropathogenic E. coli
ESCC	Esophageal squamous cell carcinoma
FAP	Familial adenomatous polyposis

FBS	Fecal bovine serum
FDR	False discovery rate
FISH	Fluorescent In Situ Hybridisation
Gal-GalNAc	D-galactose and N-acetyl-D-galactosamine
GCA	Gastric cardia adenocarcinoma
gDNA	Genomic DNA
GGD	Genome-to-genome distance
GLOBOCAN	Global cancer
GVSA	Gene set variation analysis
GWAS	Genome-wide association studies
HCV	Hepatitis C virus
HER2	Human epidermal receptor 2
HIV	Human immune deficienty virus
HPV	Human papilloma virus
IBD	Inflammatory bowel disease
IEC	Intestinal epithelial death
IEC	Intestinal epithelial cell
IgA	Immunoglobulin A
INFg	Interferron gamma
KEGG	Kyoto Encyclopedia of Genes and Genomes
KRAS	Kirsten rat sarcoma
LEL	Large extracellular loop
LOX-1	Lectin-like oxidised LDL receptor-1
LPS	Lipopolysaccharide
LPS	Lipopolysaccharide
LXR	Liver X receptor
MAVS	Mitochondrial antiviral-signalling protein
MDS	Multidimensional scaling
MHC	Major histocompatibility complex
MMR	Mismatch repair
MNGC	Multinucleated giant cells
MR1	Magnetic resonance imaging
MSCRAMMs	microbial surface components recognising adhesive matrix molecules
MSI	Microsatellite instability
MVB	Mulltivesicular bodies
NCBI	National Center for Biotechnology Information
NCTC	National Collection of Type Cultures
NES	Normalised enrichment score
NF-κB	Nuclear factor kappa B
NLR3	Nucleotide-binding domain, leucine-rich–containing family, pyrin domain–containing-3

NPX	Normalised protein expression
OD	Optical density
OSCC	Oral squamous cell carcinoma
PBS	Phosphate buffered saline
PCA	Principal component analysis
PD	Periodontal disease
PD-1	Programmed cell death protein 1
PET	Positron emission tomography
PIK3CA	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha
PTEN	Phosphatase and tensin homolog
qPCR	Quantitave polymerace chain reaction
RAS	Ras sarcoma virus
RLRs	Retinoic acid-inducible gene I-like receptors
rDNA	Ribosomal Deoxy ribonucleic acid
RNA	Ribo nucleic acid
ROS	Reactive oxygen species
RPMI	Roswell Park Memorial Institute
SDC4	Syndecan 4
SDS-PAGE	Sodium dodecyl sulphate – polyacrylamide gel electrophoresis
SEL	Small extracellular loop
SEM	Scanning Electron Microscopy (SEM)
SMAD4	Mothers against decapentaplegic homolog 4
SMAD7	Mothers against decapentaplegic homolog 7
SNP's	Single nucleotide polymorphisms
SPAR	*Sgg* pathogenicity-associated region
t-SNE	T-distributed Stochastic Neighbourhood Embedding
TEER	Transepithelial electrical resistance (TEER) measurement
TEM	Tetraspanins enriched microdomain
TFG-b	Transforming growth factor beta
TIGIT	T-cell immunoreceptor with Ig and ITIM domains
TLR-2	Toll like receptor 2
TLR-4	Toll-like receptor 4
TLR4-MYD88	Toll like receptor 4 - Myeloid differentiation primary response 88
TNM	Tumour Node Metastasis
TSA	Traditional serrated adenoma
UC	Ulcerative colitis
ULK1	Unc-51 like autophagy activating kinase 1
WT	Wild-type
5-FU	5-Fluorouracil

1 GENERAL INTRODUCTION

1.1 Colorectal cancer (CRC)

Colorectal cancer refers to a heterogeneous group of colon and rectum cancers resulting from the transformation of glandular epithelial cells (Yamauchi et al., 2012;Ewing et al., 2014). The majority of colorectal cancer cases are sporadic, and their development can span over a decade owing to a sequential acquisition of carcinogenic mutations that lead to disease (Alves Martins et al., 2019). A third of colorectal cancers have a familial or germline association (Jasperson et al., 2010). About 5% of the non-sporadic cancers have been associated with well-characterised clinical manifestations related to syndromes (Jasperson et al., 2010). These syndromes, including Lynch syndrome, Juvenile polyposis syndrome, and familial adenomatous polyposis (FAP) increase lifetime colorectal cancer risk and can also contribute to other extracolonic cancers (Gryfe, 2009;Jasperson et al., 2010). The syndromes are driven by autosomal disorders, and often lead to the expression of adenomatous polyps that have varying colorectal cancer risks (Jasperson et al., 2010).

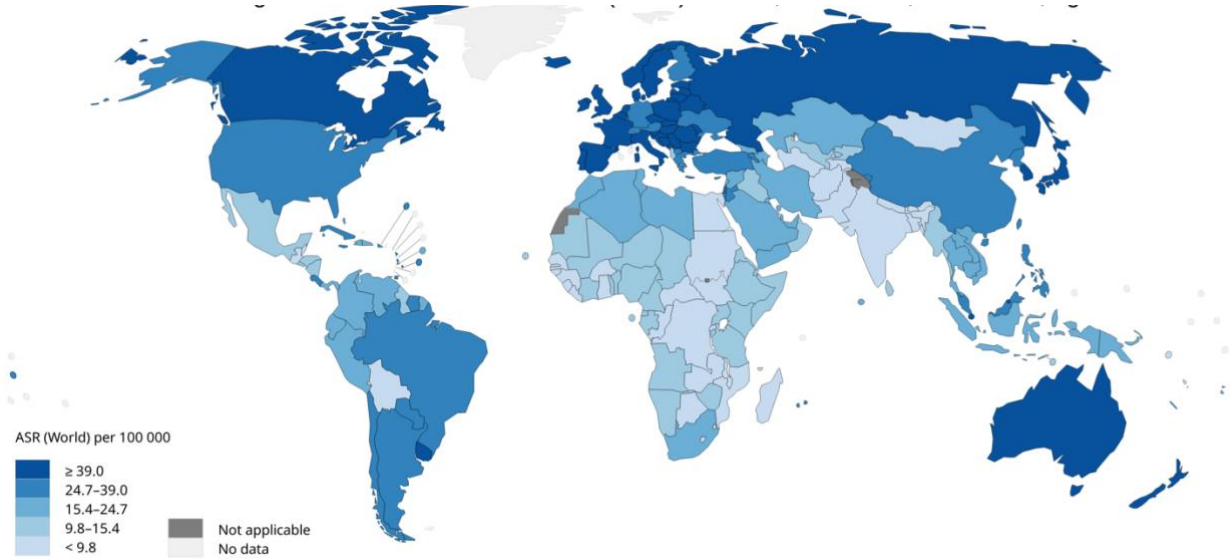
The exact aetiology of the other 20-25% of non sporadic cancers is still not understood. However, an association of the majority of these cases with family history, points to the existence of unidentified genetic contributions (Rustgi, 2007). Genome-wide association studies (GWAS) have identified several genetic mutations predisposing to CRC risk (Schumacher et al., 2015). Single nucleotide polymorphisms (SNPs) in alleles of genes such as SMAD7 (Mothers against decapentaplegic homolog 7) and CRAC1 (colorectal adenoma and carcinoma 1) have been linked to CRC risk. CRAC1 has been reported to function as a tumour suppressor gene (Popat et al., 2003), while SMAD7 is involved in biological processes such as

tumorigenesis and inflammation through the negative regulation of transforming growth factor beta (TFG-β) (Zhu et al., 2011). Functionally both SMAD7 and CRAC1 play important roles in regulation of gene expression in colorectal cancer. In total about 50 independent CRC susceptibility loci have been identified across ethnicities and counting (Broderick et al., 2007;Jaeger et al., 2008;Dunlop et al., 2012;Law et al., 2019). The aetiology of CRC is multifactorial, as genetic and environmental factors both can contribute to increased CRC risk.

1.2 CRC incidence and global distribution

CRC ranks 3rd in incidence and 2nd in mortality for all solid cancers worldwide, according to the GLOBOCAN statistics of 2020 (Sung et al., 2021). Around 2 million new cases of CRC were registered, with over 900 000 deaths observed due to CRC in 2018 alone (Dekker et al., 2019;Rawla et al., 2019). These figures are expected to reach upwards of 3.5 million new cases with nearly 2 million deaths by 2040 (Morgan et al., 2023). Furthermore, there has been an alarming increase in cases of early-onset CRC, affecting people of ages 50 and below (Patel et al., 2022). CRC has historically been more prevalent in more developed countries (Figure 1.1). However, an observed increase in the incidence of CRC in Asia and countries in Africa has been linked to an adaptation to a more Westernised lifestyle (Rawla et al., 2019). Limitations in resources for diagnosis and treatment in middle-income countries results in high mortality rates in these countries compared to high-income countries (Khan and Lengyel, 2023).

A)



B)

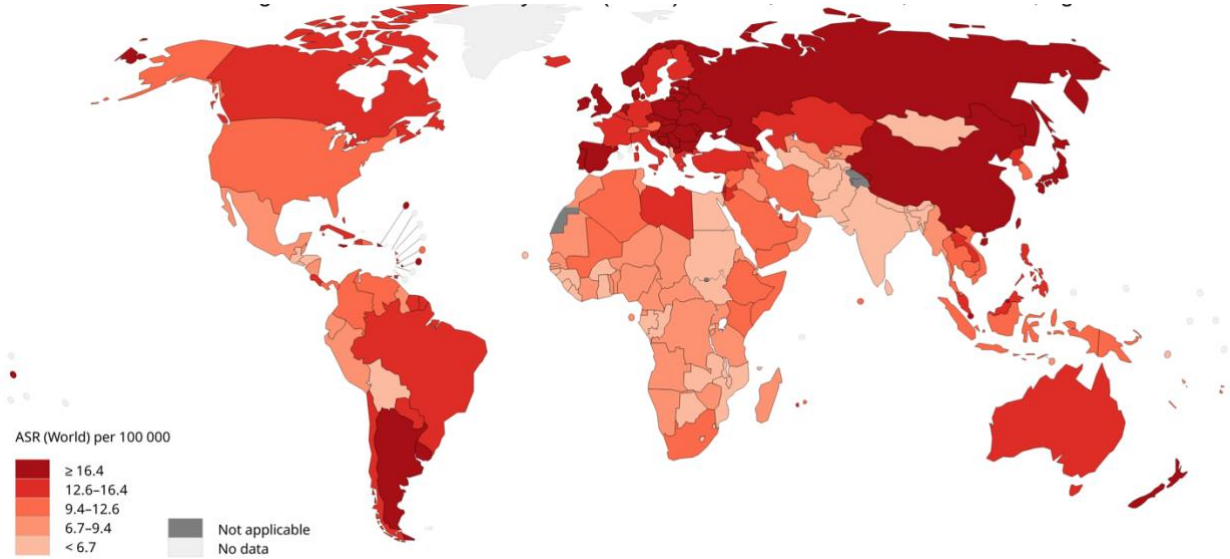


Figure 1.1: Global incidence and mortality of CRC for both sexes. *A) Age-standardised rates (ASR) for incidences in ages 15+ show high CRC incidence in Western countries such as Australia, Canada, and Europe. Lower incidence rates are observed for most African and Asian countries). B) ASR for global mortality due to CRC, with relatively higher mortality observed in high-income countries and lower rates observed in middle-income countries. Figures reprinted from IARC GLOBOCAN 2020 data. Maps generated from selecting "global", "colorectal" for cancer sites, "Inc." for incidence map and "mort" for mortality map from the tool at (https://gco.iarc.fr/today/online-analysis-map). Copyright 2024 (Accessed January 2024).*

1.3 CRC risk factors

Other than familial and hereditary risk factors that contribute to CRC, there are other risk factors, environmental and otherwise, that have been appreciated for their role in CRC initiation and progression (Figure 1.2). Some factors are non-modifiable, while others are considered modifiable and their prevention can help lower the incidence of CRC. Non-modifiable risk factors include sex, age, and ethnicity, whereas modifiable risk factors include diet, exercise, alcohol consumption and smoking (Wang et al., 2019c).

1.3.1 Sex

The lifetime risk for CRC is not very different in men (4.4%) and women (4.1%); however, after standardising for age, men have over a 30% risk for CRC compared to women (White et al., 2018;Lewandowska et al., 2022). Furthermore, there are higher mortality rates in men than women of the same age group (Majek et al., 2013). This has been attributed to the hormone oestrogen, which regulates some signalling pathways such as the WNT/ß catenin pathway (Abancens et al., 2020).

1.3.2 Age

Age is a significant risk factor for CRC, with many CRC cases being in the elderly population over the age of 50. The accumulation of mutations that lead to CRC happens over time, leading to diagnoses of CRC in people under 40 to be rare (Lewandowska et al., 2022). An increase in the number of CRC cases in people between the ages of 20 and 49 has been reported in several

countries, with more aggressive disease manifestations being observed in the younger population. Though some cases are due to heredity and familial background, the majority is sporadic and linked to an increase in obesity and diet-related risk factors (Brenner et al., 2017;Siegel et al., 2017;Movsisyan et al., 2021).

1.3.3 Race/ethnicity

There are observed racial and/or ethnic disparities in CRC risk based on race/ethnicity. There is a higher CRC incidence and mortality in African Americans compared to Asians and non-Hispanic whites, especially in non-Hispanic black men (Siegel et al., 2020). These disparities have been observed even after correction for risk factors (Siegel et al., 2014;Rutter et al., 2022). Furthermore, non-Hispanic black people were shown to have early diagnosis and more advanced disease at diagnosis with reduced survival after diagnosis. These disparities have been reported to be partly due to differences in socioeconomic status, genetic predisposition, environmental factors, and reduced access to screening in African Americans (Rahman et al., 2015;May et al., 2017).

1.3.4 Nutrition and diet

Diet constitutes one of the modifiable risk factors for CRC. High consumption of red and processed meat and low consumption of food high in dietary fibre has been shown to increase CRC risk. (Bradbury et al., 2020). Red meat has been classed as a group 1 carcinogen and can contribute to colon cancer by producing cytotoxic heme factor (CHF), which leads to intestinal

epithelium damage. Other mechanisms include reactive oxygen species (ROS) generation and contribution to proliferation through growth components such as arginine (Vernia et al., 2021). Other foods associated with increased CRC risk include highly processed, high-fat, and high-sugar foods. In contrast, vegetables and mineral-rich foods have an inverse association with CRC (Azeem et al., 2015).

1.3.5 Physical activity

Other modifiable risk factors of CRC are excess weight, and a sedentary lifestyle. Physical activity has resulted in over 20% of risk reduction in both men and women (Wolin et al., 2009;Boyle, 2012). Furthermore, observational studies, showed that increase in physical activity led to a decrease in mortality following diagnosis in a male-only cohort. (Meyerhardt et al., 2009). Several molecular processes, including upregulation of tumour suppressor genes such as P53, downregulation of anti-apoptotic mechanisms and angiogenesis upregulation, are some of the mechanisms that the preventative nature of the exercise is based upon (Leung et al., 2004).

1.3.6 Smoking and alcohol consumption

Other environmental risk factors for CRC risk include smoking and alcohol consumption. Increased alcohol consumption in mid-adult life has been linked to increases in CRC later in life (Hur et al., 2021). Another study showed an association between high alcohol consumption with early onset CRC (Chen et al., 2022a). Significant, dose-dependent associations between

smoking and CRC risk have been reported, and smoking is associated with increases in genetic mutations which can contribute to heterogeneity in CRC (Tsoi et al., 2009;Amitay et al., 2020)

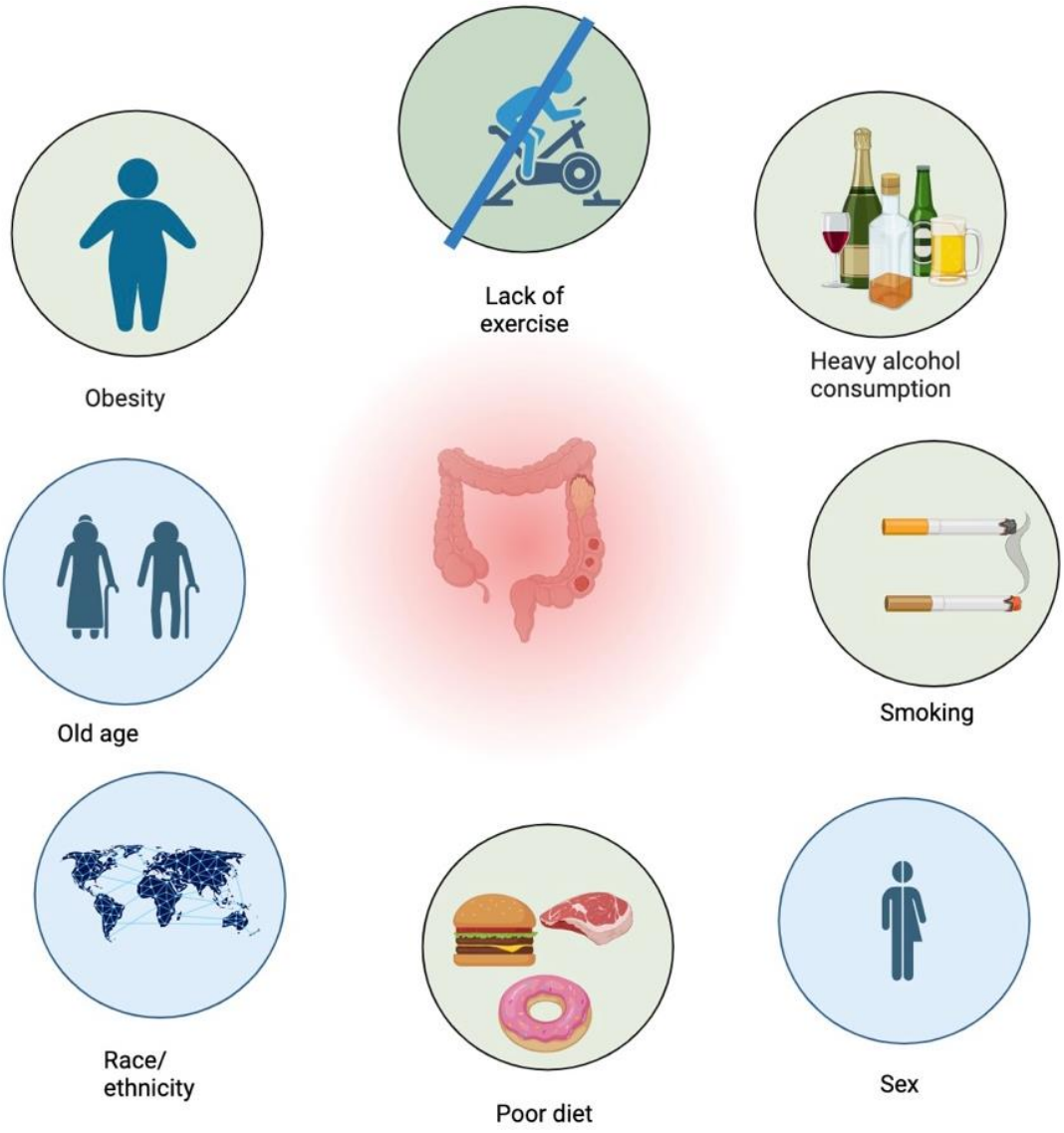


Figure 1.2: Modifiable and non-modifiable risk factors for CRC*. Diagrammatical summary of some modifiable (green fill) and non-modifiable (blue fill) risk factors associated with CRC. Modifiable risk factors include lack of exercise, heavy alcohol comsumption, smoking, and poor diet. Non – modifiable risk factors linked to CRC include race/ethnicity, sex and old age. Figure created with Biorender.*

1.4 Development of CRC

Only 15-20 % of CRC has been attributed to familial or genetic reasons. The remaining cases (around 75-80 %) are sporadic and are believed to be mainly due to environmental and lifestyle factors (Dougherty and Jobin, 2023). Genomic instability is one of the critical factors in CRC development (Grady and Carethers, 2008;Yamagishi et al., 2016). Three pathways have been recognised to be necessary for genomic instability resulting in CRC. These pathways are; microsatellite instability (MSI), chromosomal instability (CIN) and the CpG island methylator pathways (CIMP) (Pino and Chung, 2010). The CIN pathway, which is common in over 50% of sporadic CRC, is characterised by gain or loss of chromosomes in cells leading to aneuploidy (Pino and Chung, 2010;Tijhuis et al., 2019). MSI on the other hand, which occurs in around 10% of sporadic CRC, is due to defects in the DNA mismatch repair (MMR) genes (Nojadeh et al., 2018). MMR genes play a role in the repair of microsatellites, which are short (1-6 nucleotides) repeating DNA sequences that are sometimes generated due to replication errors (Li et al., 2020). Finally, the CIMP pathway is characterised by epigenetic silencing of genes such as tumour suppressor and DNA repair genes through hypermethylation of their promoters (Advani et al., 2019;Tapial et al., 2019).

1.4.1 Adenoma-carcinoma model

The conventional adenoma-carcinoma model, first proposed in 1990 by Fearon and Vogelstein, describes the sequential development of CRC from an initiating event leading to the development of adenomas which progresses to carcinoma. The authors purported that, initially, mutations in oncogenes and tumour suppressor genes are required to initiate carcinogenesis.

Further mutations, around five, irrespective of the mutation sequence, are needed to establish malignancy (Fearon and Vogelstein, 1990). This is also commonly known as the first-hit, second-hit model, where mutations in oncogenes and tumour suppressor genes constitute the 1st hit and further mutations, often driven by environmental and other predisposing factors, contribute to the 2nd hit that leads to tumorigenesis. A key mutation in CRC is a mutation in the oncogene *Kirsten rat sarcoma (KRAS)* which occurs in 40% of cases and results in a poor prognosis. The *KRAS* gene is a GTPase in the RAS (Ras sarcoma virus) pathway which is involved in cell survival, proliferation, and differentiation through relaying messages to the nucleus from the cytoplasm (Sameer, 2013;Zhu et al., 2021a). Another protooncogene important in CRC development is the *BRAF (B-Raf proto-oncogene, serine/threonine kinase)* gene involved in the extracellular cellular regulated kinase (ERK) signalling pathway. It is involved in around 12 % of metastatic colorectal cancer and, like *KRAS* mutations, has a poor prognostic profile (Grothey et al., 2021). BRAF and KRAS mutations are mutually exclusive, whereas BRAF mutation often occur with APC (adenomatous polyposis coli) gene mutations (Morkel et al., 2015). APC is one of the primary tumour suppressor genes for CRC initiation. It is mutated in around 80 % of sporadic CRC and involved in the early stages of CRC (Kwong and Dove, 2009). APC is involved in eliminating beta-catenin, which is a part of the growth and proliferation promoting WNT signalling pathway and in a range of other pathways (Hankey et al., 2018). Germline mutations in *APC* lead to the FAP syndrome (Crabtree et al., 2003). Another important gene in CRC development is the TP53 gene, which is mutated in over 60% of CRC cases (Hassin et al., 2022). For later stages of the sequence, mutations in *SMAD4* (Mothers against decapentaplegic homolog 4), which is a regulator of the TFG-β signalling pathway have been documented. In addition, activating mutations in *PIK3CA*

(phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha) also contribute to the late stages of this sequence resulting in increased proliferation (Nguyen et al., 2020).

1.4.2 Serrated pathway

The serrated pathway, responsible for about a tenth of CRC cases, is an alternate pathway to the classical adenoma-carcinoma pathway, and it is characterised by morphologically distinct polyps with a "saw-toothed" crypt architecture hence the name Serrated. In this model, serrated polyps, instead of adenomatous polyps, form the primary lesion that will result in CRC. There are different types of serrated polyps, namely, hyperplastic polyps, sessile serrated adenoma, traditional serrated adenoma (TSA) and mixed polyps. These differ in their frequency of detection, the location where they form and their histopathology (Yamane et al., 2014). This pathway is mainly driven by activating mutations in BRAF, with a few cases having KRAS mutations and by CIMP. It is mainly characterised by microsatellite instability (MSI-high) , though a few cases have had low MSI and chromosomal stability, like the adenoma-carcinoma sequence (O'Brien et al., 2008). BRAF and KRAS encode kinases that are involved in pathways that are important for proliferation, cytokine production, cell differentiation and in the activation of epidermal growth factor receptor (EGFR). Mutations in BRAF and KRAS, have dismal prognoses and lead to aberrant growth, metastasis and resistance to EGFR targeted therapies (Meng et al., 2021). Furthermore, methylation of mismatch repair genes (e.g. MHL1 promoter) which is often associated with BRAF mutations in this pathway, leads to microsatellite instability (Grassi et al., 2021).

1.4.3 Consensus molecular subtyping (CMS)

Gene expression data can provide relevant information on the molecular profile of tumours and help in their stratification. Most stratification of tumours has traditionally relied on histological samples. With the advent of high throughput and widespread genomics, this can be improved and completed by gene expression-based typing. Aneuploidy, chromosomal and microsatellite instability in CRC means that CRC is the heterogenous response to treatment. Molecular profiling of colorectal cancer can have consequences on the increasingly popular biomarker-guided therapies (Sveen et al., 2020). Currently, the most robust classification of colorectal cancer is the Consensus Molecular subtyping (CMS) (Figure 1.3). CMS was developed by a consortium of experts who independently analysed available gene expression-based data for colorectal cancer. Four CMS subtypes, CSM1-CMS4 were developed based on gene expression data from several meta- transcriptomic data and studies on CRC (Guinney et al., 2015). CMS1 is characterised by the dysregulation of genes involved in microsatellite instability. CMS1 tumours are hypermutated, hypermethylated and are characterised by overexpression of DNA repair genes. Immune activation is another identifying feature of this subtype. CMS2 is also known as the canonical pathway and is characterised by the dysregulation of genes involved in transformation, such as WNT genes and it is the most common presentation of the four subtypes. CMS3, with an incidence of over 10%, is characterised by dysregulation in the metabolic pathways affecting the cell cycle, glutaminogenesis and lipidogenesis. Tumours with this profile have a low copy number, high chromosomal instability and are often accompanied by mutations in the DNA repair mechanisms (Dienstmann et al., 2017). The final subtype, CMS4, constitutes tumours with epithelial-mesenchymal transformation characterised by invasion of the tumour into nearby tissues (stromal invasion), TGF-B activation, and

angiogenesis. It is associated with *APC, KRAS* and *TP53* mutations (Guinney et al., 2015). About a tenth of tumours did not fall into any of the four subtypes and had a combination of features of multiple CMS types (Guinney et al., 2015).

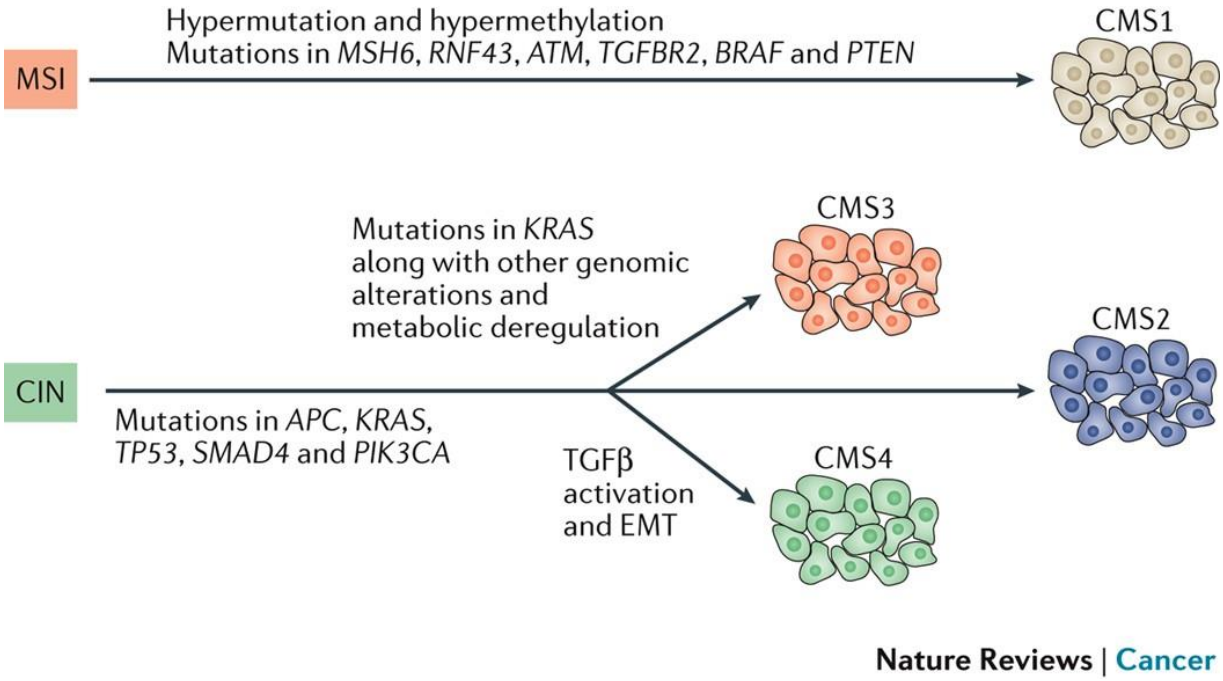


Figure 1.3: Consensus molecular subtype (CMS) classifications. *Mutations that potentially contribute to the different CMS types are shown. Consensus molecular subtype 1 (CMS1) is characterised by microsatellite instability (MS1) and mutations in the following genes; MSH6 (mutS homolog 6), RNF43(Ring finger protein 43), ATM (Ataxia telangiectasia mutated), TGFBR2 (Transforming growth factor receptor beta 2), BRAF(v-Raf Murine Sarcoma Viral Oncogene Homolog B) and PTEN (Phosphatase and tensin homolog). The other consensus molecular subtypes are characterised by chromosomal instability (CIN). Consensus molecular subtype 2 is characterised by mutations in APC (Adenomatous polyposis coli), KRAS (Kirsten ras sarcoma), TP53(Tumour protein 53), SMAD4(mothers against decapentaplegic homolog 4) and PIK3CA (phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha) whereas common mutations for consensus molecular subtype 3 (CMS3) are mutations in KRAS together with metabolic pertubations. Consensus molecular subtype 4 (CMS4) is mainly characterised by epithelial to mesenchymal transition (EMT) and transforming growt factor Beta (TGFβ) activation. Taken from (Dienstmann et al., 2017). Reproduced with permission from Springer Nature.*

1.4.4 CRC diagnostics

Common symptoms of CRC include rectal bleeding which is often accompanied by abdominal pains, weight loss, anaemia and changes in bowel habits (Biller and Schrag, 2021). Diagnosis of CRC is through colonoscopies and endorectal imaging for staging of rectal cancers. In addition, biopsies are taken for histological profiling and further differential diagnoses. Magnetic resonance imaging (MRI), computerised tomography (CT) and positron emission tomography (PET) scans can also be used in diagnostics to get high-resolution images to help detect metastases and detect recurrence of tumours (Weitz et al., 2005). CRC tumours are staged based mainly on the TNM nomenclature, which stages tumours based on the site of the primary tumour (T), spread to lymph nodes (N) and metastasis (M) (Buccafusca et al., 2019). Furthermore, molecular pathological tests can also be done alongside other diagnostic processes to determine the BRAF and KRAS status of the cancer. Screening for mutations in these genes is an important determinant for some CRC therapeutic options (Hummel et al., 2021).

1.4.5 CRC therapeutics

Treatment of CRC includes surgical excision of polyps and tumours, radiation, and chemotherapy. Some metastatic tumours cannot be resected, and in such cases, different treatment options, such as immunotherapies, are used. For example, Cetuximab (a monoclonal antibody) treats metastatic CRC by targeting the often upregulated EGFR receptor (Jonker et al., 2007). Other monoclonal antibodies function as immune checkpoint inhibitors, leading to the activation of immune cells such as T-cells and potentiation of their action against tumour cells. Checkpoint inhibitors that target PD-1 (programmed cell death protein 1) and CTLA4

(Cytotoxic T-lymphocyte associated protein 4) have been approved in the use against CRC (Johdi and Sukor, 2020). Inhibitors against KRAS-positive CRC are also in development and showing promise in clinical trials (Strickler et al., 2022). Even though advances in diagnosis and treatment options have improved overall survival rates for CRC, early detection of CRC remains a crucial part of tackling CRC. It can help prevent adverse outcomes with increased survival associated with pre-malignant stages. Screening campaigns can help reduce the burden of CRC through early detection.

1.5 CRC and the microbiota

The human gut contains over 100 trillion microorganisms whose composition and diversity change over time and is impacted by, among other things, diet, lifestyle and age (Wu et al., 2019a). Bacteria in the gut are essential for priming the gut and helping to strengthen the gut immune system. The gut provides the bacteria with a nutritionally dense environment, and a balanced relationship is important for immune system development and defence against infection. “Good bacteria” provide colonisation resistance, and some produce metabolites that are important for the efficient digestion of food (Zhang et al., 2015). However, microbial dysbiosis can occur in the gut, characterised by changes and imbalances in the diversity and composition of the gut resulting in illness. Microbial dysbiosis has been associated with several diseases, including inflammatory bowel disease and ulcerative colitis. Bacteria have been linked to the development and progression of cancer in the past e.g., chronic *Helicobacter pylori* infection was shown to increase the risk of developing gastric cancer (Wroblewski et al., 2010). It is widely accepted that microbial dysbiosis has a role in CRC. As the gut microbiota is

comprised of trillions of bacteria, and colon cancer has a multifactorial aetiology, understanding the role of the gut in CRC is both complex and requires a multifaceted approach. Several studies have shown an association between CRC and various microbiota members, including viruses, fungi, and bacteria (Li et al., 2021b;Wang et al., 2021;Marongiu and Allgayer, 2022). Differences in the diversity, taxonomical composition and abundance of bacteria have been observed among normal, early stage and CRC groups (Fang et al., 2021). Metagenomic analysis have revealed that this is characterised by an enrichment of certain genera and a depletion of others in CRC groups compared to healthy controls. Bacteria such as *Fusobacterium nucleatum* (*F. nucleatum), Streptococcus gallolyticus (S. gallolyticus), Parvimonas micra (P. micra),* and *Porphyromonas gingivalis (P.gingivalis)* are some of the most enriched bacteria in CRC tumours compared to control samples (Li et al., 2021b). Bacteria such as *Clostridium butyricum* produce metabolites that can have inhibitory consequences on colorectal cancers. One such metabolite is butyrate which is a short-chain fatty acid produced by bacteria following digestion of fibres. Butyrate improves barrier function in epithelial cells and inhibits tumour growth by downregulating pro-carcinogenic enzymes (Hajjar et al., 2021).

Even though histological evidence of the enrichment of certain bacterial species in CRC has been shown, most of the findings rely heavily on sequencing studies such as 16S and short read metagenomic sequences due to the high throughput nature of the methods (Osman et al., 2018). However, there is a limitation in using sequencing data in that although it is informative, establisihing causation from corelation is challenging . More complex experimental setups, including longitudinal studies/clinical trials combined with advanced models would be required to begin to unpick causation from correlation. Additionally, metagenomic data is

limited by the fact that it is often greatly contaminated with host DNA and possible contamination from kits and reagents used. This can affect downstream analyses as well as estimations of bacterial abundance (Chrisman et al., 2022).

Within fusobacterial strains, FNA has been reported to be the subspecies that is mostly commonly isolated in tumours compared to normal cells (Morsi et al., 2022). This was also reported by Zepeda Rivera, Bullman and colleagues who found that most of the isolates in the tumours were from FNA. Interestingly, on analysis of the FNA isolates, a birfucation into two clades was identified within the subspecies. Two clades named FNA C1 and FNA C2 were identified with the latter being more enriched in CRC in the colon and the former being more common in the oral cavity. Morphological differences were observed between the clades with FNA C2 found to be longer and thinner than FNA C1 a feature which is believed to increase the invasiness of FNA C2 to HCT116 cells compared FNA C1 (Zepeda-Rivera et al., 2024). Spatial difference in FNA isolates signifies the need for spatial transcriptomics which can allow for tumour microbiota heterogeneity to be observed (Galeano Niño et al., 2022).

1.6 Proposed models for the role of bacteria in colorectal cancer

1.6.1 Drivers' vs passengers' model

In the bacterial drivers-passengers model, certain bacteria are thought to play a role in the initiation of CRC, while other drivers colonise the tumour microenvironment following transformational changes that result from tumour initiation. "Driver" bacteria include bacteria such as *Bacteroides fragilis*, and *F. nucleatum*. These bacteria can adhere to the colonic epithelium and contribute to processes that promote carcinogenesis. The "driver" bacteria are then slowly replaced by "passenger" bacteria such as *Escherichia coli* (*E.coli*) which may have virulence factors that help them out-compete "passengers" (Avril and DePaolo, 2021).

1.6.2 The alpha bug hypothesis

The alpha bug hypothesis purports that there is a select group of bacteria that are capable of initiating carcinogenesis and then act to remodel the colonic epithelia to contribute to tumour progression. One of the bacteria that was suggested to be an "alpha bug" is *B. fragilis*, which can induce tumorigenesis in the multiple intestinal neoplasia mouse model (Sears and Pardoll, 2011). This hypothesis ties in with the adenoma-carcinoma sequence, where somatic mutations provide the first hit, and the environment adds the second hit by contributing to the accumulation of mutations leading to adenomas and carcinomas. An adaptation of this model to include bacteria, proposes that following genetic mutations and initial first-hit somatic mutations, bacteria contribute to the accumulation of subsequent mutations through DNA

damage, aberrant inflammation and through fostering an environment that increases the risk for further mutations (Drewes et al., 2016;Cheng et al., 2020).

1.7 *Fusobacterium nucleatum* (*F. nucleatum*)

F. nucleatum is a rod-shaped, indole-producing Gram-negative obligate anaerobe of the genus fusobacteria in the Bacteroidaceae family (Bennett and Eley, 1993;Brennan and Garrett, 2019). First observed as early as 1898, there are currently at least 12 known species of fusobacteria. (Citron, 2002;Hofstad, 2006). *F. nucleatum* is a common commensal in the oral cavity and is among the most extensively studied species of this genus.

1.7.1 *F. nucleatum* subspecies

Based on protein electrophoresis and DNA hybridisation analysis, *F. nucleatum* was initially thought to comprise of three subspecies; *F. nucleatum* ssp. *nucleatum*, ssp. *polymorphum* and ssp. *vincentii* (DZINK et al., 1990). The list of subspecies was further expanded to include ssp. *fusiforme* and ssp. *animalis*. This was based on the use of techniques such as restriction enzyme-based electrophoretic separation, improved DNA-DNA hybridisation studies, and electrophoretic separation of glutamate dehydrogenase and 2-oxoglutarate reductase (Gharbia and Shah, 1992). However, due to similarities in their *rpoB* and zinc protease genes, the ssp. *fusiforme* and ssp *vincentii* are indiscriminate from each other and are now considered the same subspecies (Kim et al., 2010).

Recently, a reclassification of *F. nucleatum* subspecies into species has been proposed based on results from improved bioinformatic analysis of genomic sequences (Kook et al., 2017). The phylogenetic relationship between the subspecies was evaluated using average nucleotide identity (ANI), genome-to-genome distance (GGD) and 16S rDNA analysis. The ANI and GGD threshold for sub-speciation are (95–96%) and 70%, respectively. In their study, Kook *et al*. found that the type strains of the four subspecies had ANI values that fell between 89.80% to 92.97%. Additionally, the GGD scores ranged from 42.3 to 46.0%, which is below the threshold for subspecies delineation. A more recent phylogenomic analysis of sequenced *F. nucleatum* strains and strains on the NCBI database revealed that they sub speciate into one of the four widely accepted subspecies except for one, which could be a new subspecies (Ma et al., 2023b).

1.7.2 Role of *F. nucleatum* in oral biofilms

F. nucleatum is key to the structural and mechanical composition of both healthy and diseased oral biofilms. Its elongated shape enables it to co-aggregate with other bacteria through its surface adhesins (Aruni et al., 2015). *F. nucleatum* binds early colonisers of the oral biofilm, which comprise facultative anaerobes, including Streptococci and Actinomyces species (Aruni et al., 2015). Together with bacteria such as *Prevotella intermedia*, *F. nucleatum* also binds late colonisers, which mainly comprise obligate anaerobes such as *P. gingivalis* and *Tannerella forsythia* (Aruni et al., 2015;Wu et al., 2021). Several adhesins, such as RadD, FadaA and Fap2 play a role in the co-aggregation of *F. nucleatum* with other members of the oral biofilm (Lima et al., 2017b).

1.7.3 *F. nucleatum* in oral infections

F. nucleatum plays a role in periodontal disease (PD), a dental plaque-associated chronic inflammation of the structures and cells surrounding the tooth (Marcotte and Lavoie, 1998;Kwon et al., 2021). It has been implicated in gingivitis, a milder form of PD, which affects gingival tissues and can progress to PD if left untreated (Pihlstrom et al., 2005;Chen et al., 2022b). Periodontitis, a more advanced form of PD, is characterised by structural damage to the tooth resulting in alveolar bone loss and the formation of pockets between the gingiva and the root of the tooth (Pihlstrom et al., 2005;Usui et al., 2021). The oral biofilm (dental plaque) is central to the aetiology of PD as pathogenic bacteria in it led to the persistent inflammation characteristic of PD.

1.7.4 *F. nucleatum* in extraoral infections

Even though it is primarily an oral commensal, *F. nucleatum* has been implicated in several infections outside the oral cavity (Figure 1.4). *F. nucleatum* has been implicated in systemic infections such as pre-term birth. Pregnant mice infected with *F. nucleatum* suffered stillbirths and pre-term births following a hematogenous transmission of the bacteria to the placenta resulting in a localised uterine infection (Han et al., 2004). In a recent article, *F. nucleatum* was isolated from the lungs of a preterm foetus (Bonasoni et al., 2022). In addition, strains isolated from pre-term births were shown to be identical to isolates of *F. nucleatum* from mothers and partners of preterm babies implying that oral health may affect pregnancy outcomes (Gauthier et al., 2011).

Microbial dysbiosis has been implicated in the pathogenesis of inflammatory bowel disease (IBD) - a disease characterised by an immune-driven chronic inflammation of the intestinal mucosal lining (Li et al., 2014). Through analysis of longitudinal metagenomic data, an increased abundance of *F. nucleatum* in IBD patients compared to controls was observed (Huh and Roh, 2020). *F. nucleatum* was found to promote intestinal epithelial death (IEC) and autophagy in ulcerative colitis. *F. nucleatum* lipopolysaccharide (LPS) was reported to facilitate *F. nucleatum*-mediated IEC (Su et al., 2020). In the same study, *F. nucleatum* was found to be more abundant in UC patient samples than in controls, as detected through Fluorescent In Situ Hybridisation (FISH).

More recently, the bacterium was enriched in over 50% of endometriosis patients, but in less than 10% of controls. It was shown to increase the weight and number of endometriotic lesions with incidence and number of lesions decreasing upon antibiotics treatment in a mouse model (Muraoka et al., 2023). Case reports have shown the isolation *F. nucleatum* in patients suffering from infective endocarditis, with bacteraemia from dental procedures and PD being possible sources of origin (Afra et al., 2013;Dhaliwal et al., 2023). Retrospective studies have shown an often ignored role of this bacterium in bacteraemia which can play a role in infections of other body sites and contribute to sepsis (Eykyn, 1989;Afra et al., 2013;Goldberg et al., 2013). *F. nucleatum* was also shown to play a role in atherosclerosis (Lee et al., 2012;Wang et al., 2023b;Zhou et al., 2023a). It invades aortic tissues and exacerbates atherosclerotic lesions in mouse models. Furthermore, the bacterium destabilised atherosclerotic plaques by increasing cell apoptosis and upregulating the activity of pro-atherosclerotic mediators such as interleukin-6 (IL-6) (Zhou et al., 2022). *F.nucleatum* has been isolated in several cases of liver abscesses

(Nagpal et al., 2015;Jayasimhan et al., 2017) and brain abscesses (Han et al., 2003;Heckmann et al., 2003;Hischebeth et al., 2014;Franceschi et al., 2020) often linked with dental procedures and/or periodontitis (Nagpal et al., 2015). Other diseases and infections that *F. nucleatum* has been associated with are respiratory tract infections, including tonsillitis, urinary tract infections and osteomyelitis (Brennan and Garrett, 2019).

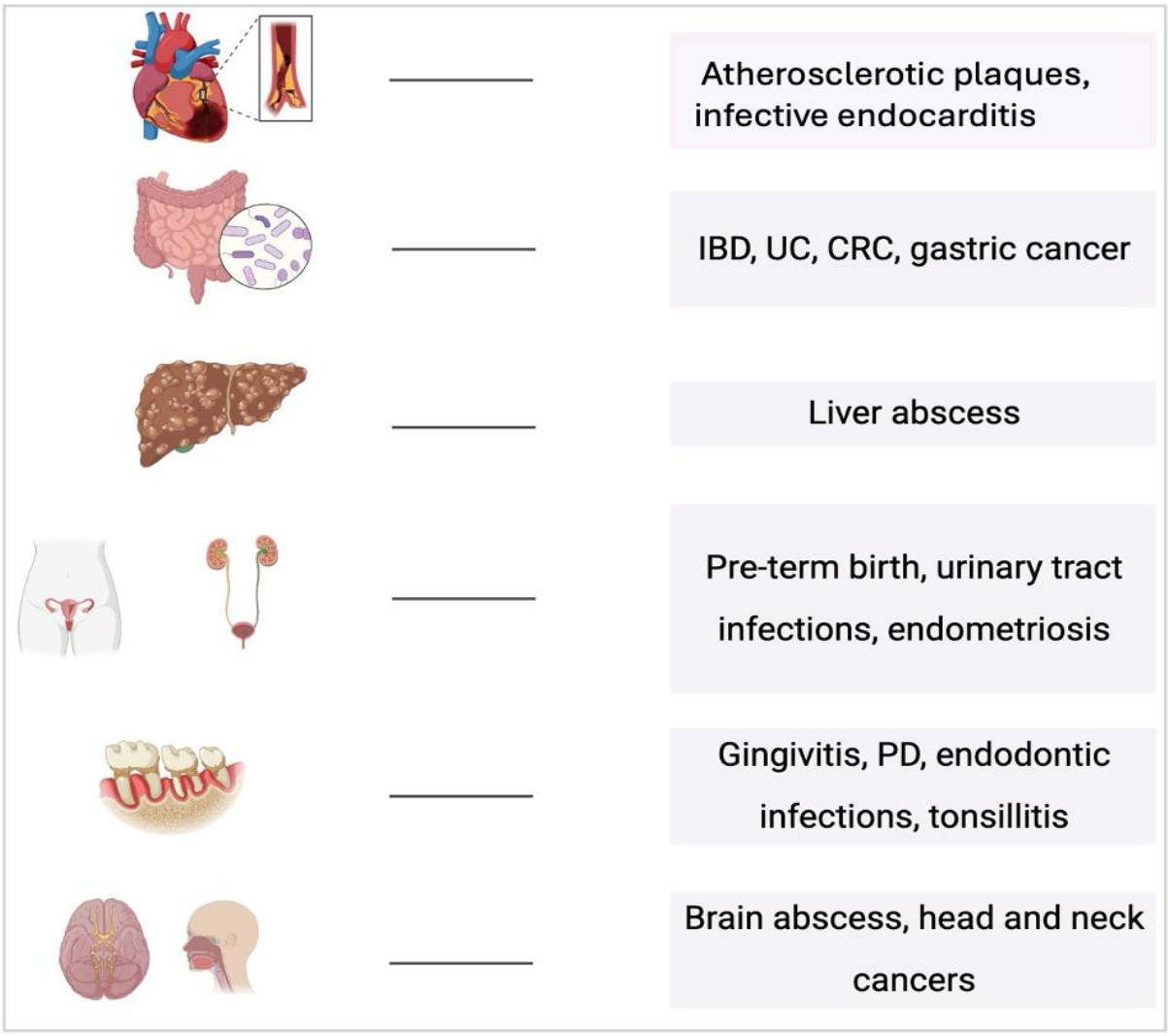


Figure 1.4: Role of F. nucleatum in disease at different body sites. *Associative to moderate evidence exists for the potential role of F. nucleatum in several infections and diseases across body sides* (Brennan and Garrett, 2019). *Some of the infections and diseases that F. nucleatum has been linked with are shown. These include the heart, leading to infective endocarditis, and plaques. In the stomach and colon it is linked to inflammatory bowel disease (IBD), ulcerative colitis (UC), colorectal cancer (CRC) and gastric cancer. In the liver and the brain, it is associated with abscesses. Pre-term birth, urinary tract infections and endometriosis are associated with infection of the reproductive system by F. nucleatum. In the mouth it is associated with Periodontitis (PD) and endodontic infections. Figure created with Biorender.*

1.8 Role of *F. nucleatum* in cancer

F. nucleatum has gained interest as an emerging oncopathogen and is associated with several cancers. The bacterium was found to be more prevalent in biofilm samples from tumours of oral squamous cell carcinoma (OSCC) patients compared to controls. It was isolated together with bacteria such as Prevotella, Clostridium and Porphyromonas species in OSCC (Nagy et al., 1998;Zhang et al., 2019). Next-generation sequencing studies have corroborated these studies, with more taxa such as Parvimonas, Peptostreptococcus, Catonella, and other periodontitis-associated bacteria showing enrichment in OSCC lesions compared to paired controls (Zhao et al., 2017). Differences in taxa were observed between deep tissue and surface tissue samples in some studies (Gopinath et al., 2021). Using qPCR and 16S ribosomal RNA gene sequencing, *F. nucleatum* was reported to be overabundant in two further malignancies: Esophageal squamous cell carcinoma (ESCC) and gastric cardia adenocarcinoma (GCA). Enrichment of the bacteria in the tumours was associated with poor prognosis and poor response to chemotherapy in these cancers (Yamamura et al., 2016;Shao et al., 2019). Other cancers that *F. nucleatum* has been associated with include head and neck, cervical, breast, lung, and pancreatic cancers, where the bacterium is often found to be more prevalent in tissues from tumours than adjacent normal tissues or in healthy controls (He et al., 2022). Its role in CRC is the most extensively researched. This work will contribute to this, looking at the dynamics of the interactions of *F. nucleatum* with colorectal cancer cells.

1.8.1 Mechanism of *F. nucleatum* carcinogenesis

The exact mechanism of action of *F. nucleatum* in various cancers remains poorly understood and is an area being actively researched. In OSCC, CCL20-mediated inflammation is reported as one of the ways in which *F. nucleatum* contributes to carcinogenesis (Mandal et al., 2021). CCL20 is associated with increased metastasis and poor prognosis in several tumours (Hippe et al., 2020). In addition, it binds to toll like receptor 2 (TLR2,) which results in the nuclear translocation of nuclear factor kappa B (NF-κB) and activation of the NLR3 (nucleotide-binding domain, leucine-rich–containing family, pyrin domain–containing-3) inflammasome resulting in increased inflammation (Bui et al., 2016). *F. nucleatum* further contributes to OSCC through DNA damage resulting in unregulated cellular proliferation and by contributing to the epithelial-mesenchymal transmission (Geng et al., 2020;Zhang et al., 2020). In breast cancer, *F. nucleatum* was shown to be positively associated with Gal-GalNAc (D-galactose and N-acetyl-D-galactosamine) in mammary tissues. The bacterium binds to Gal-GalNAc using the Fap2 adhesin facilitating the invasion of the bacterium into eukaryotic cells and suppressing T-cell mediated immune responses (Parhi et al., 2020).

1.8.2 The role of *F. nucleatum* in colorectal cancer

The involvement of *F. nucleatum* in CRC has recently gained a lot of interest in the scientific community. Several studies have reported the enrichment of *F. nucleatum* in CRC patient tumours compared to paired normal controls (Castellarin et al., 2012;Kostic et al., 2013;Li et

al., 2016). An overabundance of *F. nucleatum* has also been isolated from faeces of patients compared to healthy controls (Aitchison et al., 2022). Following infection of C57BL/6 $Apc^{Min/+}$ with *F. nucleatum*, a marked increase in the number and the size of polyps in treated mice compared to those inoculated with streptococcus species was observed. Furthermore, an increase in immune cells of the $CD11b^{+}$ myeloid lineage was also observed (Kostic et al., 2013). Another study investigated *F. nucleatum* colonisation and tumorigenesis using germ-free, specific pathogen-free and C57BL/6 $Apc^{Min/+}$ mice. *F. nucleatum* has been found to contribute to metastasis, with *F. nucleatum* abundance being found to be higher in metastatic lesions than in controls. Infecting the mice with *F. nucleatum* led to more aggressive tumours (Chen et al., 2020). Colonisation of tumours with *F. nucleatum* has been linked to reduced survival in CRC patients and can lead to chemoresistance (Yu et al., 2017a). Delineating whether the presence of *F. nucleatum* in the tumour is a consequence of the tumour or if it is a cause of the tumorigenic environment remains an area of active research.

1.9 *F. nucleatum* virulence mechanisms in CRC

1.9.1 FadA driven promotion of tumorigenesis

FadA, an adhesin conserved in fusobacteria, is essential for adhesion to eukaryotic cells and for aggregation. FadA exists in two forms, pre-FadA which has 129 amino acids and mature FadA (mFadA) which is made up of 111 amino acids (Rubinstein et al., 2013b). Pre-FadA is the intact form of the protein and contains an 18 amino acids signal peptide sequence. At neutaral

pH pre-FadA is hydrophobic wheareas the mature form is soluble, however solubility is attained when pre-FadA is mixed with mFDA or when it is in acidic conditions. The signal peptide in pre-FadA contributes to the hydrophobicity of this form of the adhesin. mFadA is the secreted form of the protein wheras pre-FadA is found embedded in the inner membrane.The two forms associate to form the FadAc complex which facilitates binding to and invasion of host cells. Association of the two forms leads to the formation of different sized oligomers. It has beedn proposed that while pre-FadA anchors in the inner membrane, mFadA monomers are continoulsly added to the complex all the way to the outer membrane in a fashion similar to how filamentous appendages such as pili assemble. Binding to host cells appeared to require the complex of the two forms as no binding was observed when only mFadA was used in binding experiements to (Xu et al., 2007). Resolution of the crystal structure of mFadA showed that it is made up of two antiparallel alpha helices that are joined by a loop comprising of eight immune acid residues including tyrosine and leucine residues. The leucine residues were shown to have important structural and functional properties for the FadA protein. They interact to form a leucine chain, arising due to formation of hydrophobic interactions between the residues which are important for the stability of the protein and for filamentation and binding. (Nithianantham et al., 2009b). Levels of the FadA gene have been reported to be higher in CRC tumours than in controls and throughout the different stages of the adenoma to carcinoma stages (Rubinstein et al., 2013a). Annexin A1, a modulator of cyclin D1 is overexpressed in carcinomas compared to normal tissue. It was proposed that a positive feedback loop mechanism occurs in *F. nucleatum*-mediated CRC pathogenesis between annexin1 and *F. nucleatum*. In this mechanism, *F. nucleatum* through the FadA adhesin, binds to e-cadherin (CDH1) stabilised by annexin 1 (Rubinstein et al., 2019a). This then activates the WNT/ ß-catenin signalling pathway resulting in the upregulation of oncogenes such as *myc* genes and

cyclin D1 (Rubinstein et al., 2019b) (Figure 1.5). Wnt signalling activation is one of the pathways that are crucial for tumorigenesis as it has a role in differentiation, proliferation and cancer stem cell formation (Zhao et al., 2022a). Others have shown that FadA from *F. nucleatum* associates with checkpoint kinase 2 (chk2) - whose expression is upregulated in the ß-catenin pathway leading to DNA damage (Guo et al., 2020). In addition, significantly higher levels of anti-FadAc IgA levels but not of IgG were found in colorectal cancer patients compared to non sufferers. Based on this, a mucosal immune response is likely induced by *F. nucleatum* translocation through the digestive tract (Baik et al., 2022).

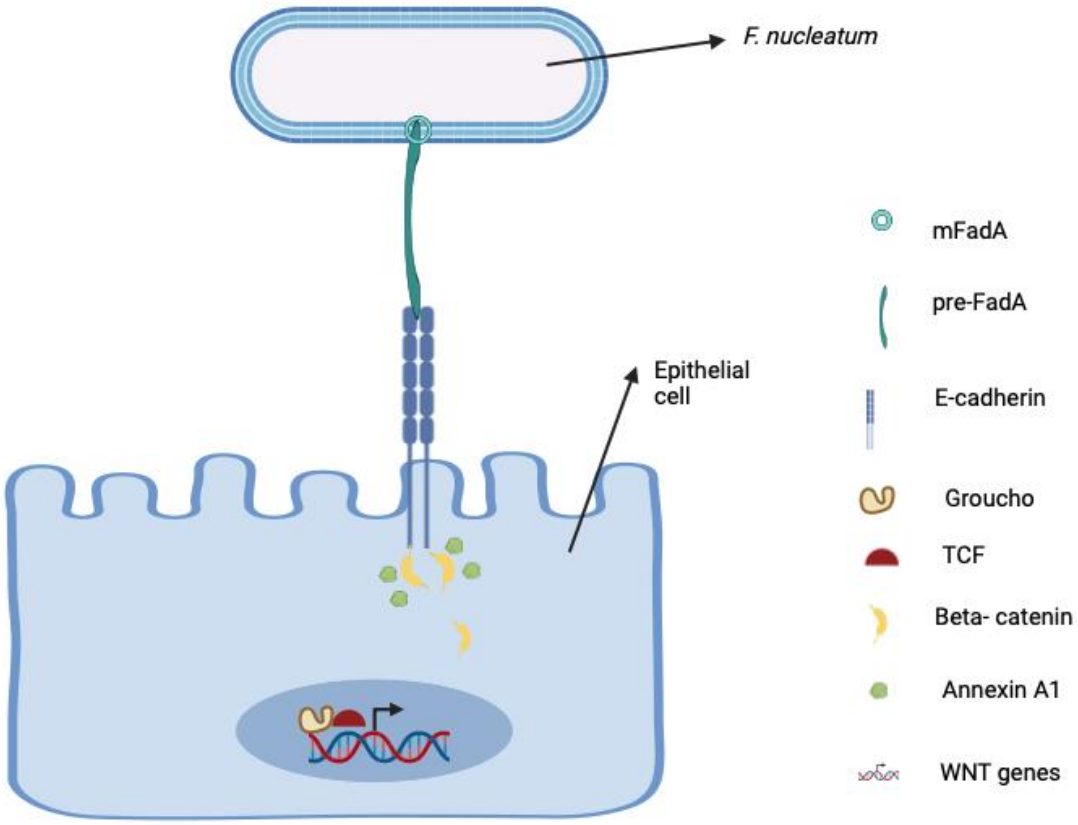


Figure 1.5 : Role and function of FadA in CRC. mFadA from F. nucleatum embeds in the inner membrane and anchors pre-FadA which extends out of the membrane*. This forms the FadAc complex which is important for the adherence and the invasion of FadA into epithelial cells, Once attached, FadAc binds to E-cadherin which then associates with beta catenin, stabilised by Annexin A1. Beta catenin translocated to the nucleus wehere it activates the WNT signalling genes through the displacement of groucho from TCF . Figure created using Biorender.*

1.9.2 Modulation of the immune system

CRC tissues have been found to overexpress the disaccharide Gal-GalNac compared to normal intestinal tissues. The fusobacterial haemagglutination leptin Fap2 can bind to Gal-Gal-Nac, facilitating the adhesion and/or invasion of the bacteria to CRC cells (Abed et al., 2016a). Strains of *F. nucleatum* lacking Fap2 expression have been shown to have a reduced ability to colonise mice models of CRC (Abed et al., 2016a). Inhibition of Gal-Gal-Nac with O-glycanase significantly reduced the binding of *F. nucleatum* to CRC tissues. It has been postulated that Fap2 plays a role in immune regulation in the CRC tumour microenvironment by associating with TIGIT (T-cell immunoreceptor with Ig and ITIM domains), an inhibitory receptor on natural killer and T-cells leading to the inhibition of its cytotoxicity (Gur et al., 2015). In addition, *F. nucleatum*, through the FomA adhesin, can activate B- cells through interactions with MHC class II and CD86 (a co-stimulatory molecule on antigen-presenting cells) via TLR2 (Toussi et al., 2012). Another study using outer membrane vesicles from the gut showed that vesicles activated NF- κB responses in a TLR2-FomA-dependent mechanism (Martin-Gallausiaux et al., 2020). Furthermore, its arginine-inhibitable adhesin RadD and Fap2 were shown to induce the death of Jurkhat cells - a $CD4^{+}/CD8^{-}$ T-cell lymphoma cell line. Levels of cell death were attenuated upon inactivation of the adhesins demonstrating that *F. nucleatum* can potentially suppress the adaptive immune response (Kaplan et al., 2010).

1.9.3 Modulation of autophagy

Autophagy is a biological process, conserved in all eukaryotes which entails lysosomal mediated degradation of proteins, organelles and other macromolecules preceded by the recyclying of the degradation products (Parzych and Klionsky, 2014). In cancer the role of autophagy is often reffered to as a "double-edged sword", as it can function as a tumour promoter and can also help curb tumour growth (Lim et al., 2021). In CRC, autophagy has been reported to inhibit tumour growth in the early stages while increasing tumour resistance in the later stages of the disease (Ma et al., 2023a). The latter effect is relevant to the action of *F. nucleatum* in CRC through autophagic dysregulation. *F. nucleatum*, through the TLR4-MYD88 (toll like receptor 4 - Myeloid differentiation primary response 88) signalling pathway, leads to the loss of the small miRNAs, miRNA-18a and miRNA-4802, which negatively regulate autophagy related *7(ATG7)* and *Unc-51 like autophagy activating kinase 1 (ULK1*), respectively (Sun et al., 2019). ULK1 is an autophagy-activating kinase (Lin and Hurley, 2016), and ATG7 has multifaceted roles in p53-driven apoptosis, cell cycle arrest and phagophore expansion in the autophagy pathway (Komatsu et al., 2007;Collier et al., 2021). Loss of these miRNAs leads to dysregulated autophagic responses, inhibition of apoptosis and results in chemoresistance (Figure 1.6) (Yu et al., 2017b). By suppressing autophagy, *F. nucleatum* leads to resistance to the drugs oxaliplatin and 5-fluorouracil (5-FU) (Yu et al., 2017a). It was demonstrated that this autophagic dysregulation led to increased inflammatory cytokine (IL-8 and TNF-α) release, a result of *F. nucleatum* driven oxidative damage. Furthermore, impairment of the autophagic flux pathway has contributed to increased reactive oxygen species

(ROS) generation in cell lines and mouse models. An increase in reactive oxygen species can lead to DNA damage resulting in the acquisition of somatic mutations in the colonic epithelial. ROS can further contribute to tumorigenesis as it upregulates the inflammatory response. It was proposed that an increase in ROS production following *F. nucleatum* infection leads to an increase in the production of pro-inflammatory cytokines such as IL-8 and Il-1ß (Tang et al., 2016). Duan *et al* showed that *F. nucleatum*-induced autophagic impairment could be mediated by treatment with the bacterium *Lactobacillus rhamnosus*, a probiotic bacterium from the clade of *Lactobacillus casei* (Duan et al., 2021).

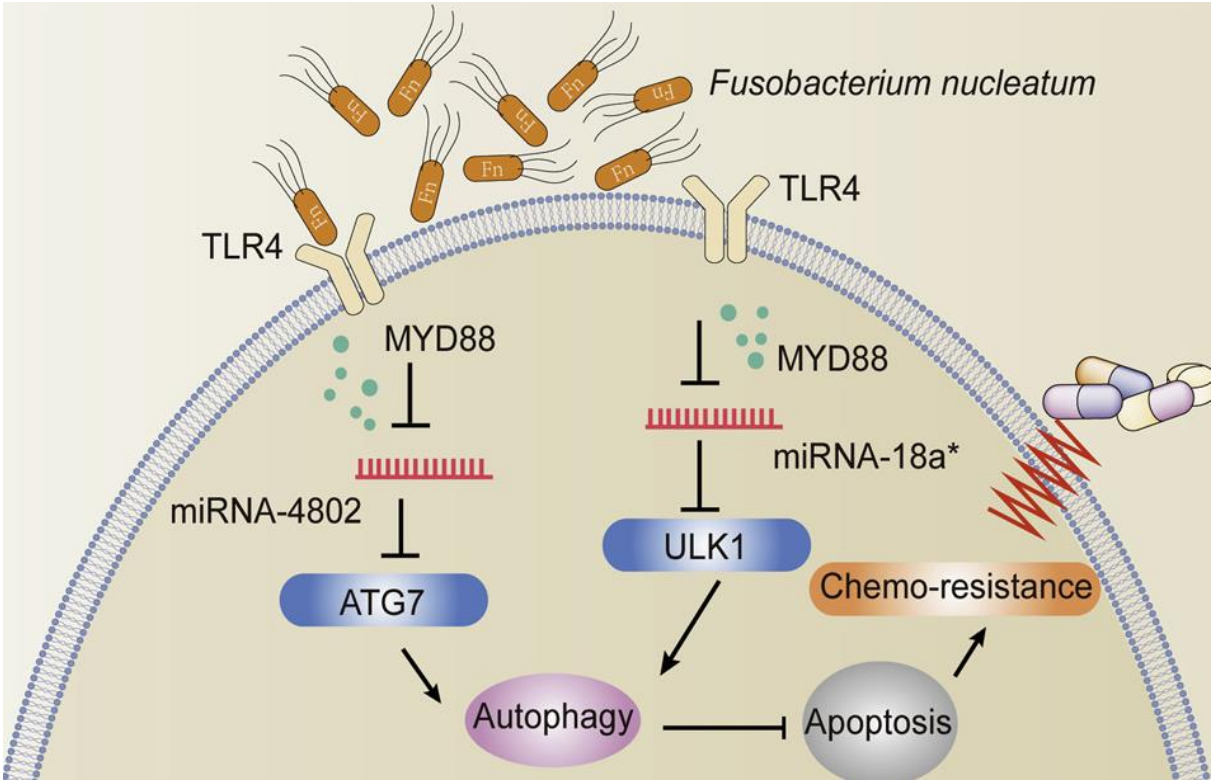


Figure 1.6: Role of F. nucleatum in autophagic impairment*. F. nucleatum (labelled Fn) binds Toll like receptor 4 (TLR4) and inhibits ATG7(autophagy related 7) and ULK1(Unc-51-like kinase 1) through Myeloid differentiation primary response 88 (MYD88) signalling resulting in dysregulated autophagy. Inhibition of ULK1 and ATG7 is due to the loss of the micro RNA's miRNA-4802 and miRNA-18A* which are released upon interaction of the bacteria with TLR4. This leads to a down-regulation of apoptosis and subsequently leads to chemoresistance. Figure taken from* (Yu et al., 2017a). *© 2017 Elsevier Inc.*

1.10 Tetraspanins

Tetraspanins are a family of transmembrane 20-30 kDa scaffold glycoproteins, of which 33 have been characterised in humans (Zou et al., 2018;New et al., 2021). They are present in several organisms, including *Drosophila melanogaster, Caenorhabditis elegans*, in insects and some plants (Huang et al., 2005;Albers et al., 2016). Tetraspanins interact with various partner proteins and are involved in cell migration, cell adhesion, fusion, and cellular communication (Huang et al., 2005;Rubinstein, 2011). Generally, they function as molecular "organisers", and can form Tetraspanin enriched microdomains (TEMs) consisting of tetraspanin dimers, signalling ligands, adaptor proteins and receptors (Hemler, 2008). Some tetraspanins such as CD9 and CD81 are distributed in a broader range of tissues whereas others, such as CD53 and CD37 are not (Wright et al., 2004). They have been implicated in many diseases such as cancer, diabetes, neurological diseases, viral and bacterial diseases and are important modulators of signalling pathways through interactions with partner proteins (Karam et al., 2020;McLaughlin et al., 2020;Becic et al., 2021;Huang et al., 2022).

1.10.1 Structure of Tetraspanis

Tetraspanins comprise 4 conserved membrane-spanning helical domains (Figure 1.7). Two transmembrane domains link extracellularly to form a large extracellular loop (LEL), also known as EC2. The LEL contains 3 helical domains and a hypervariable variable region stabilised by disulphide bonds (Umeda et al., 2020). Tetraspanin LEL sequence diversity is the highest of all the domains, and lateral interactions with partner proteins e.g., integrins, are

mediated through this loop (Mazurov et al., 2007). The other two tetraspanin domains are linked by a small extracellular loop (SEL) or EC1 (Hemler, 2014). Short N and C terminus tails and palmitoylation sites sit in the intracellular region exposed to the cytoplasm of the cell. Although other four domain transmembrane proteins exist, they are distinguished from tetraspanins due to the conserved cysteine and polar region in the tetraspanins (Stipp et al., 2003).

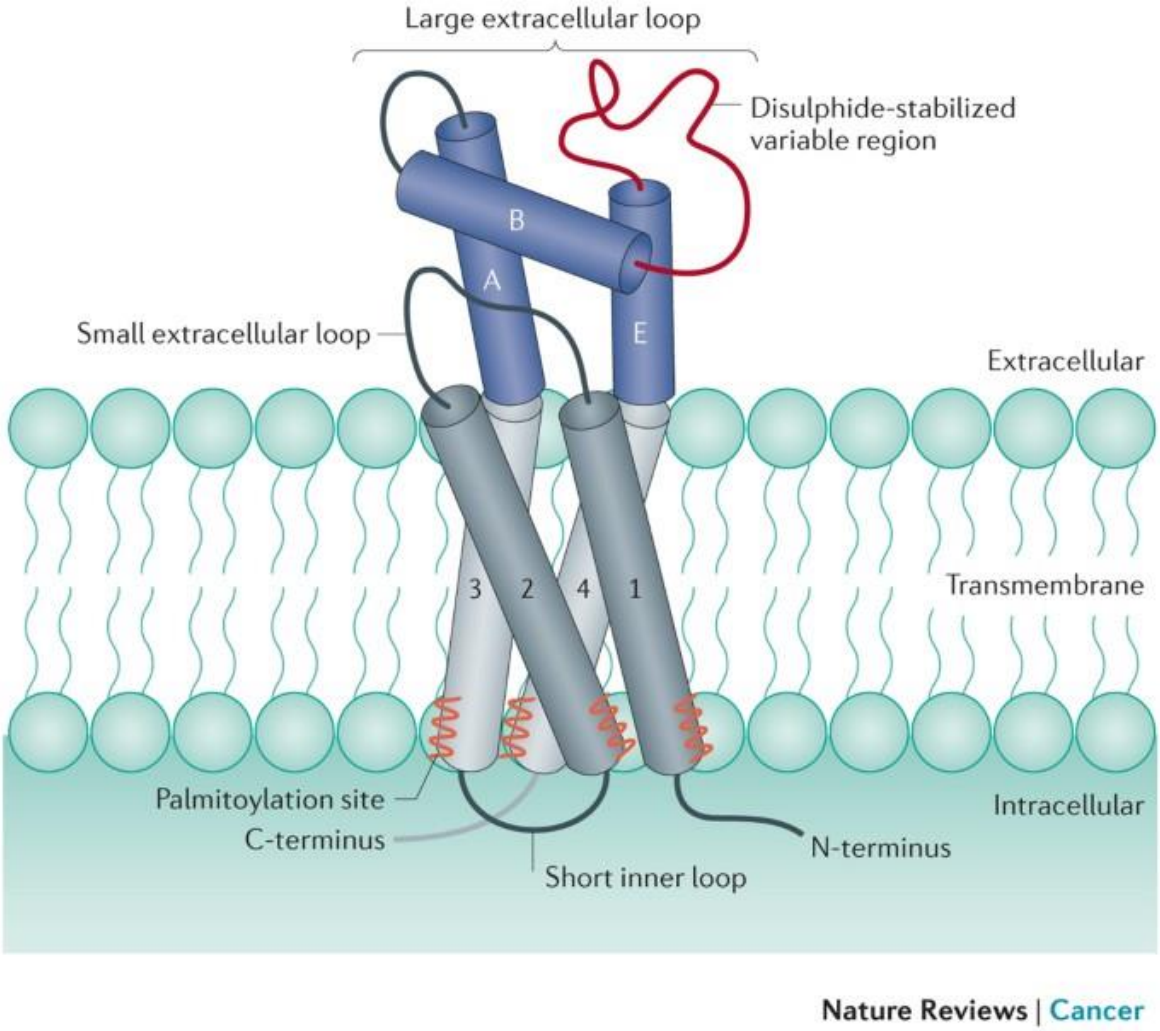


Figure 1.7: The structure of tetraspanins. *The large extracellular loop (LEL) and the small extracellular loop link four transmembrane domains, tm3(3)-tm4(4) and tm2(2) -tm1(1), respectively. The LEL has a variable region stabilised by disulphide bonds. A palmitoylation site and N and C terminus domains are located in the intracellular region.. There is also a short inner loop located in the intracellular region. Figure reproduced from* (Hemler, 2014). *Figure reproduced with permission from Springer Nature*.

1.10.2 Role of tetraspanins in viral and parasitic infections

In viral infections, tetraspanins are known to increase susceptibility to infection as they can facilitate virus entry and egress in host cells (Hantak et al., 2019). For example, CD81 is important for hepatitis C virus (HCV) entry into host cells and for plasmodium infection of hepatocytes (Fénéant et al., 2014;Manzoni et al., 2017) . Several tetraspanins including CD9, CD81 and CD82 are enriched at the binding site of the HIV-1 virus but their exact role in HIV infection if any, remains unclear. In 1995, it was reported that CD4 in T-cells can bind to CD81 and CD82 (Imai et al., 1995) which has consequences for the diffusion of HIV-1 receptors on the cell membrane (Suárez et al., 2018). It has been reported that these tetraspanins may act as both inhibitors and promoters of infection and can reduce transmission.

The tetraspanins CD151 and CD63 are important for the entry of the virus's cytomegalovirus and HPV. CD63 positive endosomes containing HPV viral particles interact with syntenin-1 adaptor facilitating viral assembly, transport and spread (Gräßel et al., 2016). CD9 and CD81 tetraspanin rich domains have facilitated proteins that promote the fusion and entry of coronaviruses and influenza A viruses (Earnest et al., 2015). Tetraspanins have been reported to also interact with fungal receptors, affecting the inflammatory response to fungal infections. CD63 and CD37 interacted with Dectin-1 of *Candida albicans*, affecting the immune response against infection. Mice deficient in CD37 showed IL-6 dependent increase in IgA fungal infection clearance (van Spriel and Figdor, 2010).

1.10.3 Role of tetraspanins in bacterial infections

The tetraspanins CD9, CD63 and CD151 are important for adhesion of several bacteria such as *Streptococcus pneumoniae*, *E. coli*, *Neisseria meningitidis* and *Salmonella typhymirium* (Green et al., 2011;Hassuna et al., 2017). Enteropathogenic *E. coli* utilises FimH, a fimbrae protein, to directly bind to the uroplakin complex of Tspan 20 and Tspan 21 to facilitate adhesion to the urothelium (Zhou et al., 2001). Tetraspanins have also been shown to facilitate ligand-receptor binding, which is important for the pathogenesis of certain bacteria. CD9 associates with the Diphtheria toxin receptor, facilitating the ligand-receptor interaction of the Diphtheria toxin (Hasuwa et al., 2001). *Staphylococcus aureus* α-hemolysin (α-toxin, Hla) mediated spread of *S. aureus* utilises TspanC8 tetraspanin to traffic a protease called ADAM 10 to the surface of the cells (Dornier et al., 2012;Matthews et al., 2017;Liao et al., 2023). The protease binds to its receptor, which is activated by *S. aureus* toxin binding leading to cleavage of cadherin molecules and resulting in bacterial spread (Karam et al., 2020).

Together with its role in toxin mediated *S. aureus* pathogenesis, the ADAM 10 receptor is important in sepsis (Liao et al., 2023). Another tetraspanin, CD63 helps sustain chlamydia colonisation by providing nutrients and promoting maturation of endosomal organelles known as multivesicular bodies (MVB) that contain chlamydia (Beatty, 2006). In addition, CD63 found on the surface of phagosomes containing *Mycobacterium tuberculosis*, promotes survival of the bacterium by recruiting lysozyme which provides nutrients to the bacteria (Seto et al., 2010). The tetraspanin CD81 on the other hand, has been shown to play a role in the survival of enteropathogenic *E. coli* (EPEC). Furthermore, it is implicated in the invasion of *Listeria*

monocytogenes (L. monocytogenes), an intracellular obligate food pathogen that causes listeriosis to host cells (Tham et al., 2010;Ramachandran et al., 2018). Additionally, CD81 and CD9 have been found to impact the formation of multinucleated giant cells (MNGC) of *Burkholderia thailandensis (B. thailandensis)*. MGNC formation, which is thought to be important for the pathogenicity of these bacteria, was increased in infection of macrophages derived from CD9 null mice with *B. thailandesis* and with treatment with antibodies against CD9 and CD81 (Elgawidi et al., 2020).

1.10.4 Role of tetraspanins in cancer

Due to their ability to interact with various ligands, receptors, and other protein partners, tetraspanins have been shown to impact the growth, progression and spread of various tumours. The role of different tetraspanins in malignancies has been reviewed comprehensively (Romanska and Berditchevski, 2011;Zhou et al., 2023b), showing the involvement of tetraspanins in cancers of the breast, prostate, colon and in other epithelial malignancies such as in pancreatic and lung cancers. Tetraspanins have been shown to play a role in tumour metastasis through interaction with partners involved in invasion, cytoskeletal remodelling, extravasation, and other mobility processes.

CO-029 was found to promote the migration of cancer cells in *vitro* experiments. Silencing this tetraspanin led to altered expression of some integrin and fibronectin molecules which are important in cell motility and adhesion (Guo et al., 2012). CD151 is another tetraspanin whose role in tumour initiation, progression and metastasis has been extensively reported (Franco et

al., 2010;Hayward et al., 2020;Zhu et al., 2021b). It has a role in various stages of tumorigenesis, including metastasis and de novo carcinogenesis, and is linked to poor prognostic outcomes (Yang et al., 2008;Kumari et al., 2015).

TSPAN8 was shown by several studies to be tumour-promoting. It is overexpressed in breast cancer cells, promotes cancer stemness through upregulating stemness genes, and correlates with chemoresistance (Vences-Catalan and Levy, 2022). TSPAN8 plays a role in over eight malignancies, including liver, colon, breast, lung, and gastric cancers and is associated with a more invasive phenotype (Heo and Lee, 2020). In pancreatic cancer, TSPAN8 promotes metastasis through an epidermal growth factor (EGF)-dependent pathway and is modulated by the transcription factor SOX9 (Li et al., 2021a).

CD9 has an inhibitory effect on metastasis and the spread of gastric cancer to lymph nodes (Chen et al., 2011). However, discordant findings have been reported where CD9 has been shown to promote gastric cancer invasiveness (Hori et al., 2004). The discordance exists also for other malignancies such as ovarian, lung, colon, and pancreatic cancer, where both tumour-promoting and inhibiting activities have been reported (Detchokul et al., 2014;Lorico et al., 2021).

Contradictory reports also exist on the role of CD63 on cancer, with some studies reporting a role as a tumour suppressor and others reporting tumour-promoting properties. CD63 expression profile can vary at different stages for several cancers e.g., high expression in early melanoma with expression decreased in later stages. In the later stages of carcinogenesis, CD63 acts as a negative regulator of metastasis and silencing it increases cancer motility and invasiveness (Lupia et al., 2014) . This late stage suppression of CD63 with consequences on invasiveness has also been reported for oesophageal cancer (Dey et al., 2023). Conversely,

Kaprio *et al* showed an association of CD63 with poor prognosis, poor differentiation and metastasis in colorectal cancer samples (Kaprio et al., 2020). In breast cancer, it was demonstrated that CD63 was overexpressed in human epidermal receptor 2 (HER2) positive cancer, an aggressive form of breast cancer (Iwabuchi et al., 2021).

CD82 has been reported to have tumour-suppressive roles in breast cancer, and low expression is associated with tumorigenesis, and high expression is associated with improved survival (Viera et al., 2021). CD82 inhibits 2D cell migration in breast cancer and fibronectin driven EMT in prostate cancer cells (Lee et al., 2017;Ordas et al., 2021). Furthermore, CD82 has been hailed as a tumour suppressor in other cancers such as non-small-cell lung carcinoma (Ci et al., 2018), pancreatic cancer (Friess et al., 1998), melanoma (Tang et al., 2014) and colon cancer (Zhu et al., 2022).

1.10.5 Tetraspanin 6 (Tspan6)

Tspan6 has charged residues in the transmembrane regions which are commonly conserved in Tspans. Lysine (K) is present in Tspan6 at TM1, while asparagine(N) is not maintained by Tspan6 at TM1 , a glutamic acid between the TM2 and TM3 domain (Todd et al., 1998). One N-linked gylycosylation site at postion Asn134 and an O- linked glysosylation site have been reported for Tspan6 (York et al., 2020). Tspan6 has 6 cysteines which allow it to form three disulphide bonds which are important for the folding of the large extracellular loop (Susa et al., 2024).Tspan6, or TM4SF6, is an understudied member of the tetraspanin protein family

and is ubiquitously expressed in the colon, ovaries , liver, oesophagus and other tissues in the body (Fagerberg et al., 2014). Tspan6 negatively regulates exosome production by degrading syntenin associated with syndecan 4 (SDC4) (Ghossoub et al., 2020). Furthermore, it is overexpressed in Alzheimer's disease, where it upregulates the production of amyloid beta proteins and associates with syntenin for exosome-mediated secretion of the Amyloid Precursor Protein (APP) C-terminal fragments (APP-CTF) (Guix et al., 2017). In addition, Tspan6 is deleted together with four other genes in epilepsy patients and was shown to be a regulator of synaptic function and to potentially have a role in motor learning (Salas et al., 2017). Tspan6 was also found to negatively regulate retinoic acid-inducible gene I-like receptors (RLRs) through interactions with mitochondrial antiviral-signalling protein (MAVS) during viral infections (Wang et al., 2012). Recently, Molostov *et al* showed that TSPAN6 localised in breast cancer cells derived EVs potentiates the migration of B-cells towards tumours in a liver X receptor (LXR) pathway (Molostvov et al., 2023).

1.10.6 Role of tetraspanin 6 in colorectal cancer

Tetraspanin 6 is often downregulated in colorectal cancer samples compared to normal controls (Druliner et al., 2018). The role of TSPAN6 as a negative regulator of EGF-signalling pathway-driven tumorigenesis was recently demonstrated. This study revealed insights into TSPAN6-mediated negative regulation of EGF-signalling pathway-mediated CRC carcinogenesis. Knockdown of TSPAN6 in an intestinal neoplasia mouse model resulted in increased frequency and size of adenomas. Mechanistically it is alluded that TSPAN6 forms a complex with syntenin-1 and modulates the recruitment of transmembrane TGF-a (tmTGF-A) into

extracellular vesicles. In addition, TSPAN6 expression was positively correlated with better responses to epidermal growth factor receptor (EGFR therapy and improved survival in patients from the COIN Medical Research Council (MRC) COIN trial (Andrijes et al., 2021)

1.11 Importance

A current appreciation of the potential role of bacteria in CRC is met by a paucity in knowledge of the exact mechanism of how bacteria contribute to different aspects of carcinogenesis. *F. nucleatum* is one of the most studied bacteria to its role in CRC however, most studies use the subspecies *nucleatum*. *F. nucleatum* is made up of five subspecies which have been shown to have differential behaviours e.g., in biofilm formation (Muchova et al., 2022). The main hypothesis of this study is that F. *nucleatum* susbpecies will differ in their ability to attach to and invade CRC cells due to potential differences in their expression of adhesins which are key to their interactions with host cells. Understanding how different subspecies of *F. nucleatum* adhere to and invade CRC cells can help understand important factors for pathogen-host interactions in CRC.

1.11.1 Aims and objectives.

The aim of this study was to understand the adhesion and invasion dynamics of *F. nucleatum* subspecies to CRC cells and to determine the role of microbial dysbiosis in CRC.

The objectives of this study were to:

1. Compare the differences in adhesion to CRC cells of different subspecies of *F. nucleatum through* adhesion assays using a Caco-2 cell model.
2. Investigate the role of the adhesins Fap2 and FadA in the adhesion and invasion of *F. nucleatum* subspecies by utilising mutants in the *Fap2* and f*adA* genes of an *F. nucleatum* strain.
3. Determine growth conditions that affect the adhesion and invasion of *F. nucleatum* subspecies to Caco-2 cells by using both exponential phase and stationary phase bacteria for adhesion assays.
4. Explore through RNA sequencing the differences in gene expression between exponential and stationary phase *F. nucleatum* subspecies.
5. Investigate the role of tetraspanin 6 in the adhesion and invasion of *F. nucleatum* subspecies to Caco-2 cells by performing adhesion assays with Caco-2 cells that express Tspan6.
6. Determine the impact of *F. nucleatum* infection on the release of cytokines from Caco-2 cells through profiling of the cytokines released when *F. nucleatum* attaches Caco-2 cells.

7. To investigate the role of microbial dysbiosis in CRC through the analysis of metagenomic and RNA sequencing data from CRC samples and their paired controls.

2 GENERAL MATERIALS AND METHODS

2.1 Microbiological methods

2.1.1 Bacterial strains

Four American Type Culture Collection (ATCC) strains were obtained from the Oral Microbiology group (Birmingham, Dental Hospital) -80° C freezer glycerol stocks (Table 2.1). Three were the *F. nucleatum* subspecies; *F. nucleatum* ssp. *nucleatum* ATCC 23726 (FNN23), ssp. *animalis* ATCC 51191 (FNA) and ssp. *polymorphum* ATCC 10953 (FNP). The fourth was the ATCC-type strain, *Staphylococcus epidermidis* (ATCC 14990). Dr Daniel Slade kindly provided FNN23 adhesin mutant strains in *FadA* and *Fap2* (Virginia Polytechnic Institute and State University, Blacksburg, VA) generated from the ATCC 23726 FNN23 strain. *Streptococcus gallolyticus (S. gallolyticus)* (NCTC 13773) was purchased from the National Collection of Type Cultures (NCTC).

2.1.2 Bacterial culture and maintenance

All *F. nucleatum* subspecies were grown on Schaedler anaerobic agar plates (SAA) (Sigma Aldrich) in an anaerobic chamber (80% N_2, 10% CO_2 and 10% H_2; Don Whitley Scientific, UK) at 37 ° C for 48 hrs. Liquid cultures were prepared by inoculating reduced Schaedler anaerobic broth (SAB) (Sigma-Aldrich) with two to three bacterial colonies and incubating in the anaerobic chamber at 37 ° C overnight. *S. gallolyticus* and *S. epidermidis* were grown on brain heart infusion (BHI) agar (Sigma Aldrich) and liquid cultures made in BHI broth (Sigma

Aldrich). *S. gallolyticus* agar plates were incubated overnight in a humidified 5% CO_2 incubator (Heracell 150i CO2 incubator), and liquid cultures were placed in a shaking incubator (Biotek) at 37 ° C overnight. *S. epidermidis* agar plates and broth cultures were maintained in the anaerobic chamber.

Table 2.1:Summary of bacterial strains showing strain code, growth media and where the strains were sourced

Strain	Code	Source	Growth media
FNN23	ATCC 23726	Oral Microbiology group -80 freezer	SAA and SAB
FNA	ATCC 51191	Oral Microbiology group -80 freezer	SAA and SAB
FNP	ATCC 10953	Oral Microbiology group -80 freezer	SAA and SAB
FNN23FadA	ATCC 23726	The Slade lab	SAA and SAB
FNN23Fap2	ATCC 23726	The Slade lab	SAA and SAB
Streptococcus gallolyticus	NCTC 13773	NCTC	BHI agar & broth
Staphylococcus epidermidis	ATCC 14990	Oral Microbiology group -80° C freezer	BHI agar & broth

2.1.3 Gram staining

Gram staining was performed to visualise the bacteria and observe colony morphology. Each bacterial colony was smeared on a microscope glass slide, and a drop of water was used to facilitate even smearing. Then, crystal violet (Sigma-Aldrich) was splashed across the slide and washed off with water after a minute. Next, a few drops of iodine were added before washing off with water. After shaking off excess water from the slide, the sample was decolourised with a few drops of alcohol for around five seconds. Slides were washed with water, and safranin (Sigma-Aldrich) was added to counterstain for a minute. This was washed off with water, blot dried, and slides visualised using a light microscope (Zeiss).

2.1.4 16S rRNA polymerase chain reaction (PCR)

To confirm the identity of the *F. nucleatum* subspecies used, the 16s rRNA gene was amplified using 16S bacterial universal primers. A colony of the bacteria was resuspended in 200 µl of sterile water and was placed on a heating block (100°C) for 7-10 minutes to lyse the cells. A master mix containing 10 µl of Q5 Master mix (New England Biolabs), 2 µl each of the forward and the reverse primers diluted 1:10 to a 10uM concentration and 5 µl nuclease-free water was then made. For each sample, 19 µl of the master mix was used in a PCR tube, and 1 µl of the DNA (sample) was added for the reaction. For the negative control, sterile water was added. Samples were then put in a thermocycler (VeritiTM 96-Well Fast Thermal Cycler) and run using the PCR condition below (Table 2.2). The sequence of the primers used were 5- 'AGAGTTTGATCATGGCTCAG' -3' and 5- 'CGGTTACCTTGTTACGACTT' – 3 for the

forward (27F) and the reverse (1492R) primers, respectively. The prepared samples and negative controls were amplified for 80 minutes in a thermocycler.

Table 2.2: 16S universal primers PCR cycling conditions

Temperature (° C)	Time (seconds)	No of cycles
98	30	1
98	15	40
55	30	40
72	30	40
72	5	1
15	∞	1

2.1.5 Gel electrophoresis

Gel electrophoresis was done to allow for sample visualisation following PCR amplification of the 16S rRNA gene. Agarose powder was dissolved in appropriate volumes of 1 % Tris-acetate-EDTA (TAE) buffer (ThermoFisher Scientific) to make 1.2 % agarose gel (Bioline, Meridian Bioscience) and heated in a microwave until transparent. Once the gel was cooled, SYBR green nucleic gel stain (Invitrogen) was added and mixed. The samples and negative controls were mixed with 4 µl of DNA loading buffer (Bioline, Meridian Bioscience) for easy visualisation while loading into the gel. Once set, the gel was transferred to a tank filled with 1x TAE buffer. 10 µl of the 1kb Hyperladder (Bioline) was added into the first well, followed by 10 µl of

negative controls and the samples. The gel was run for 40 minutes at 400 amps and 120 volts. Finally, the gel was imaged on the G: BOX series (Syngene) and the corresponding Gene Snap software (Syngene) to visualise the presence and/or absence of bands on the PCR gel.

2.1.6 16S rRNA sequencing

About 5 μl of the PCR product from the PCR run was sent together with the forward and reverse primers to Source Biosciences (Cambridge, UK) for 16S rRNA sequencing to allow for confirmation of the identity of the bacteria. The results were fed into the NCBI website's Standard Nucleotide BLAST platform to identify the bacteria strains. The platform allowed the identification of the strains at the subspecies level.

2.1.7 Growth kinetics and Miles and Misra for determining number of viable cells

Bacterial growth over 24 hours was monitored at 2-hour intervals for 8 hours and at 24 hours. To do this, a 1:100 dilution of the overnight culture was made in 20 ml broth. At each time point, including at the start (time = 0), 1 ml of culture was transferred into a micro cuvette and measured in a spectrophotometer (Jenway 7315) at OD_{600} nm; blanked with a sterile broth-only sample. OD600 nm reading >1 were diluted and re-measured to improve accuracy. Simultaneous to the OD readings, the Miles and Misra method was used to quantify the number of colony forming units (Miles et al., 1938). For this, 10-fold dilutions of the bacterial culture

were performed in a 96-well plate by transferring 20 µl of culture into 180 µl of PBS and serially diluting to achieve dilutions up to 10^7. Then, 20 µl of each dilution was drop spotted on agar plates in triplicate, and plates were left to incubate for 48 hours. Colony-forming units (CFU) in the culture were determined after incubation using the following formula;

CFU/ ml = average number of colonies x 50 x dilution factor

2.1.8 RNA extraction from *F. nucleatum*

RNA was extracted from *F. nucleatum* subspecies using a using a phenol/chlroform extraction method. Bacteria were grown in 30 ml of Schaedler broth in the anaerobic chamber. Upon reaching the required OD, bacteria were spun down in a centrifuge (Universal 320R, Hettich) at 2800 x g for 10 minutes to pellet the cells. The pellet was washed once by centrifuging at the same speed and time in PBS. 1 ml of 1:1 cold acetone: ethanol mixture was added to the pellet and frozen for 30 minutes at -80°C. Then the bacteria were thawed and centrifuged again for 3 minutes at 17000 x g. The supernatant was discarded, and the pellet was washed by centrifuging at 17000 x g in 500 µl of pH 7.8 Tris-EDTA (TE) buffer (10mM Tris-HCL, 1 mM EDTA) for 3 minutes. Cells were mixed by briefly centrifuging at 8000 x g for 3 minutes. The supernatant was removed, and 0.4 M sodium acetate (pH 5.2) was added to the cell suspension with 200 µl of 1% zwittergent (Millipore). Cells were briefly vortexed until the mixture appeared homogenous. Next, cells were pelleted at 8000 x g for 3 minutes. The supernatant was

discarded, and cells were washed by resuspending in 400 µl of water and centrifuging at 8000 x g. After discarding the supernatant, 200 µl of 0.04 M sodium acetate (pH 5.2), 160 µl of water and 40 µl of 10% SDS were added, and samples vortexed for 20 minutes. The samples were then transferred to a heating block (95°C) for 5 minutes. Next, 400 µl of each sample was mixed with 400 µl of phenol:chloroform:isoamyl alcohol (125:24:1, pH 4.5) (Sigma-Aldrich) in a phase lock tube (5 PRIME), vortexed and centrifuged at maximum speed for 3 minutes. The upper aqueous phase was gently removed into a fresh phase lock tube and mixed with one volume of chloroform (Sigma-Aldrich). This was followed by centrifugation at 8000 x g for 2 minutes. The upper aqueous phase was transferred to a microcentrifuge tube, and 0.1 volume of 3M sodium acetate and 2.5 volumes of 95% ethanol were added. Samples were chilled on ice for about 3 minutes and spun down for 10 minutes at top speed. The supernatant was discarded, and samples were washed with 70% ethanol and centrifuged for 10 minutes at maximum speed. Then, the supernatant was removed, and samples were left to air dry at room temperature. Samples were resuspended in 100 µl nuclease-free water and stored at -80°C.

2.1.9 Genomic DNA removal and sample clean-up of RNA

Genomic DNA removal following RNA extraction from *F. nucleatum* subspecies was done using the gDNA eliminator column of the RNAeasy plus minikit plus (Qiagen) as instructed by the manufacturer. Following gDNA removal, RNA was cleaned up using the RNAeasy mini kit following the manufacturer's instruction to remove any residual phenol and chloroform from the RNA extraction. The RNA concentration was determined using a nanodrop (JENWAY),

noting the RNA concentration and the 260/280 values. Samples were taken forward when they had a 260/280 value of ~2.0.

2.2 Mammalian cell culture methods

2.2.1 Cell lines

Caco-2 cell line expressing human Tspan6 (Caco-2/Tspan6, also referred to as T6 cells) and control Caco-2 cells (Caco-2/pLVX, as referred to as pLVX or control cells) were kindly provided by Dr Fedor Berditchevski from the Institute of Cancer and Genomic Sciences at the University of Birmingham. Cells were originally sourced from ATCC, and they are transformed epithelial cell lines initially isolated from the colon of a 72-year-old male CRC patient (Fogh et al., 1977). T6 cells were transfected using lentiviral vectors that express pLVX-Flag-Tspan6 plasmid, while pLVX cells (control cells) were transfected with an empty lentiviral vector as illustrated in schematic (Figure 2.1) (Andrijes et al., 2021).

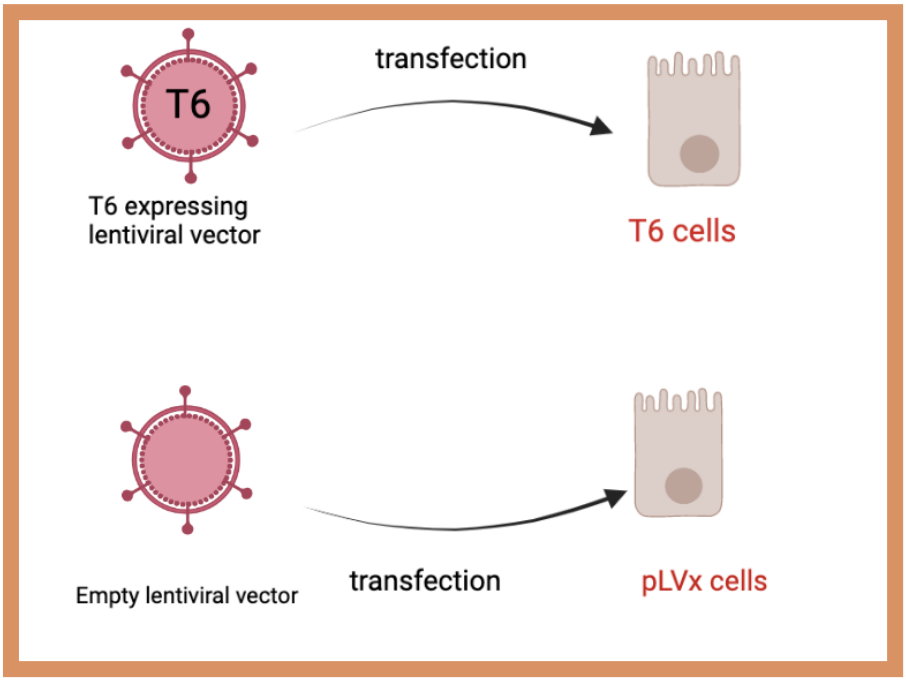


Figure 2.1:Schematic of cell line transfection*. T6 cells were created by transfecting with Tspan6 expressing lentiviral vector, and pLVx cells transfected with empty lentiviral vector. Figure created with Biorender.*

2.2.2 Cell culture and maintenance

Cells were grown in Roswell Park Memorial Institute (RPMI 1640) media (Sigma Aldrich) supplemented with 10% (v/v) FBS (Sigma Aldrich) and 5% (v/v) penicillin/streptomycin (Sigma-Aldrich) and kept in a humidified 5% CO_2 incubator at 37°C (Heracell 150i CO_2 incubator). Cellular growth was monitored using a Zeiss light microscope and a camera (AxioCam ERc5s Zeis). Cells were subcultured upon reaching 70 - 80% confluency. To subculture, media was removed from the flask and cells were washed twice with PBS to remove serum. Then, cells were incubated for 5 mins at 37°C with the enzyme TrypLE Express Enzyme (Gibco) to dissociate them from the flask. Volumes of media equal to those of the TrypLE enzyme were added to the flask, and cells were centrifuged (Universal 320R, Hettich) at 180

x g and 37°C for 3 minutes. Media was removed, and the cell pellet was resuspended in fresh pre-warmed media at a volume dependent on the required plaiting density. Cells were monitored daily, and media changed every three days.

2.2.3 Recovery of frozen cells

Frozen Caco-2 cells were recovered by briefly immersing the cryovial containing the cells in warm water at 37°C for about a minute. Thawed cells were transferred into 9 ml of RPMI media containing 10% (v/v) FBS and centrifuged (Universal 320R, Hettich) for 3 minutes at 147 x g. The supernatant was removed, and cells were resuspended in fresh media in culture flasks for incubation.

2.2.4 Cell seeding

To seed cells, they were detached and centrifuged (Universal 320R, Hettich) at 140 x g for 3 minutes. Next, cells were resuspended in a known volume of media in a microcentrifuge tube. Then, 10 µl of this was added to 10 µl of trypan blue (0.4%) (Gibco) for live-dead stain. To count the cells, the cell-trypan blue mixture was added to a haemocytometer (Neubauer) (depth 0.1 mm). Cells were quantified by counting the number of cells in 8 big squares of the haemocytometer, 4 on each of the two chambers of the haemocytometer. To calculate the cell density, the following equations were used.

$$1.\ \text{No of viable cells} = \frac{\text{total number of cells across 8 squares}}{8} * \text{dilution factor}$$

$$2.\ \text{Cell density} = \frac{\text{Number of viable cells}}{\text{Volume of cell suspension}}$$

From the above equations, the suspension can be diluted or concentrated depending on the desired seeding density.

2.2.5 Cryopreservation

Stocks of Caco-2 cells were preserved at different passage numbers for future experiments. To do this, cells were collected and detached from flasks as described previously (section 2.2.2) and added to freezing media comprised of 90% (v/v) FBS and 10% (v/v) of dimethyl sulfoxide (DMSO) (Sigma). A cell density of $1x10^6$ cells was preserved in the freezing media and stored in 1 ml cryovials. Cells were stored short-term at -80°C or in liquid nitrogen for longer-term storage.

2.2.6 Protein extraction from Caco-2 cells

Whole protein lysates were extracted from Caco-2 cells for western blot and SDS-PAGE (sodium dodecyl sulphate – polyacrylamide gel electrophoresis). Lysates were extracted from confluent cells using 25 µl of 4x Laemmli buffer (Bio-Rad) and then heated at 95°C for 5 minutes to denature them. Samples were centrifuged at 16000 x g for 10 minutes and either used immediately or stored at -80°C for future use.

2.2.7 Western Blotting and SDS-PAGE

Western blot and SDS-PAGE were done to confirm the expression of Tspan6 in T6 cells and the absence of Tspan6 in pLVX (control) cells. Protein lysates (~ 40µg of total protein) from the -80°C freezer were first heated in a heating block at 95°C for 3-4 minutes. Then, 0.1% tracking dye, bromophenol blue (Sigma-Aldrich), was added to the lysates. Samples were then flicked to mix, briefly centrifuged at 180 x g and reduced with 5% (v/v) of 2-mercaptoethanol (Sigma). Lysates (20 µl) were loaded into SDS-PAGE gels (10% v/v) that had been made using established protocols. Protein resolution was performed at 13mAmp with 1 x SDS/Tris/Glycine running buffer (GeneFlow) in an electrophoresis chamber (Bio-Rad) for 45 minutes. Following protein resolution, 100 ml of 10x /SDS/ Tris/Glycine buffer and 100 ml of methanol were added to 800 ml of water for the wet transfer of proteins onto nitrocellulose membranes. This was left to run at 40 V and 4°C overnight. After 24 hours, the membrane was washed with distilled water and stained with Ponceau S staining solution (Sigma-Aldrich) to detect and locate the

protein bands. The membrane was washed with Tris-buffered saline with 0.1% Tween® 20 detergent (TBST) and then incubated with the rabbit Tspan6 (Invitrogen) primary antibody. Antibodies to be used (primary and secondary) were prepared by adding 10ml of 5% w/v of skimmed milk in TBST. A 1:1000 dilution of the primary antibody was added to the membranes, and samples were placed on a shaking incubator for 24 hours. Membranes were washed with TBST (3x10 mins) and incubated on a shaker for an hour with a secondary antibody (goat anti-rabbit Ab IRDye 800CW (LI-COR), which was diluted 1:10000. This was followed by washing (3x10 minutes) with TBST followed by distilled water (1x 5 minutes). Lastly, visualisation and signal quantification were done using the Odyssey Infrared Imaging System (LI-COR) and the Odyssey image processing software.

2.2.8 Mycoplasma testing

Cells were regularly tested for the presence of mycoplasma using the EZ-PCR kit (Sartorius). For this, 1ml of media from cells in culture for at least 48 hours was used for testing. Testing was performed following the manufacturer's instructions. The PCR bands from the test were visualised using gel electrophoresis as described previously (2.1.5).

2.2.9 Caco-2 cells polarisation

Caco-2 cells were cultured on 12-well transwell, 0.4 μm 12 mm polyester membrane corning plates (Appleton Woods). To promote cell adhesion, the membranes were first coated with 200 μl of collagen 1 (Sigma) diluted 1:100 in distilled water for 3 hours and left to dry overnight. The plates were seeded with Caco-2 cells at a density of $1x10^5$ cells/ml. The bottom compartment was filled with 2 ml of media, and plates were put in the 5% humidified CO_2 incubator at 37°C. Media was changed every other day.

2.2.10 Transepithelial electrical resistance (TEER) measurement

On days 18, 20 and 21, TEER measurements were taken for cells grown on transwell inserts using a Voltohmmeter (World precision instruments). Culture media was removed and replaced with sterile PBS. Measurements were taken by measuring the resistance between the short probe placed inside the upper compartment of the transwell and the longer probe placed in the lower compartment of the well. A well containing just PBS was used to calibrate the machine. TEER values subtracted from reading of the blank measurement are calculated as shown below.

$$Teer(W\ cm2) = Resistance\ (W)\ x\ surface\ area\ of\ insert(cm2)$$

2.2.11 Scanning Electron Microscopy (SEM)

Cells grown on transwell inserts were visualised using SEM to determine that they had the morphology of polarised cells. Firstly, cells were washed three times in PBS and fixed using 2.5 % molecular-grade glutaraldehyde in 0.1M cacodylate buffer (BioWorld, Ireland) for 10 minutes. Glutaraldehyde was removed, and dehydration with increasing ethanol concentration from 20% up to 100% at 10-minute intervals followed. Alcohol was then removed, and hexamethyldisilane (HDMS) (Sigma-Aldrich) was added to samples as an alternative to critical point drying. This was left to dry and evaporate overnight. Samples were then mounted on specimen stubs and gold-coated in a vacuum before imaging on the Zeiss EVO MA10 scanning electron microscope. Images were captured at 8000 X magnification with a working distance (WD) of 9 mm.

2.3 Combined cell culture and microbiological methods

2.3.1 Adhesion/Invasion assay

Monolayers of two-day-old Caco-2 cells seeded at a 3.2 x 10^5 cells/ml in 24 well plates were used for adhesion/invasion assays. Two plates, one for adhesion and another for invasion, were seeded. Culture media was removed from the Caco-2 cells, and they were washed twice with PBS. Simultaneously, bacterial overnight cultures (section 2.1.2), were prepared for infection of the Caco-2 cells. The stationary phase cultures were pelleted, washed twice with PBS, and

resuspended in 1.5 ml RPMI media after OD correction to OD600nm of 1. Then, 100 µl of bacteria was added to the Caco-2 cells and incubated for four hours at 37°C in a 5% CO_2 humidified incubator. Following incubation, cells in the adhesion plates were washed twice with PBS, lysed with sterile distilled water, serially diluted in RPMI media and plated Miles and Misra style (2.1.7) on agar plates to enumerate the number of adhered bacteria. For invasion, media containing bacteria was removed after 4 hours of incubation. Cells were washed twice with PBS, and antibiotics were added (300 mg/ml metronidazole (Sigma-Aldrich), 50 mg/ml gentamicin (ThermoFisher Scientific) and 5 mg/ml penicillin (Sigma-Aldrich). After 2 hours of incubation in antibiotic media, cells were washed, lysed, and plated Miles and Misra style to get the colony-forming units of invaded bacteria. Plates were then incubated for 48 hours in the anaerobic chamber at 37°C before counting. The adhesion and invasion protocols are illustrated (Figure 2.2).

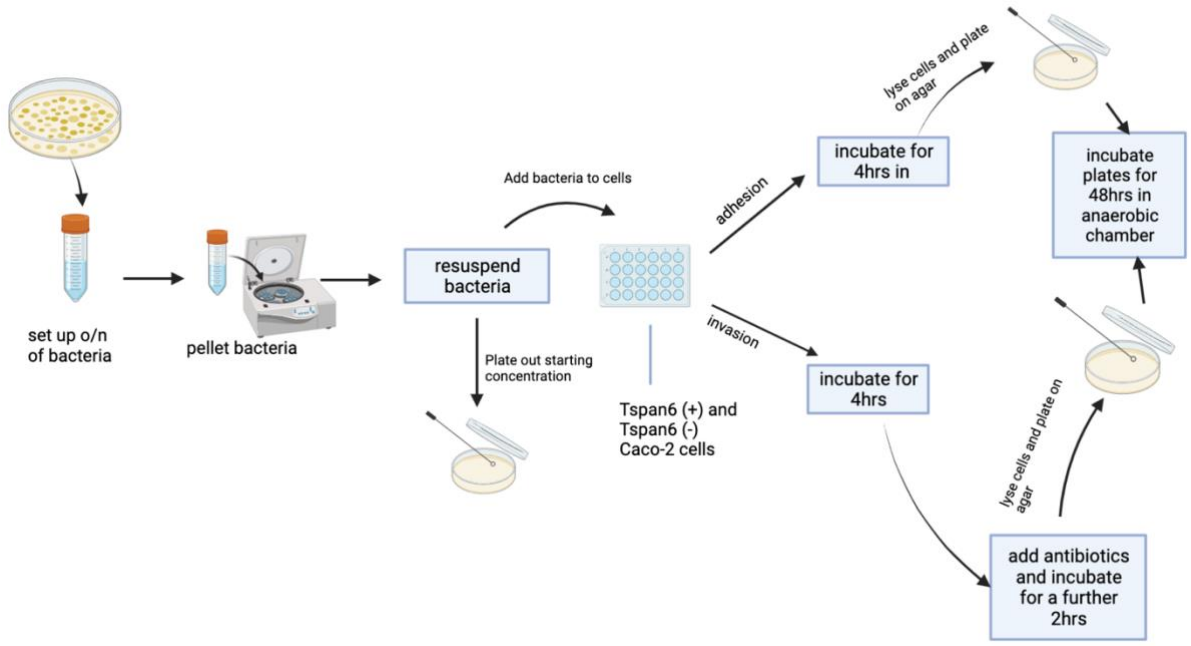


Figure 2.2:Schematic of the adhesion and invasion protocol*. Bacteria at required OD are added to Caco-2 cells. For invasion, cells are treated with antibiotics after 4 hours of incubation and incubated for a further 2 hours while incubation for adhesion is for 4 hours without antibiotic treatment. Image created with Biorender.*

2.3.2 Adhesion and invasion for stationary vs exponential phase assays

Adhesion and invasion assays for stationary vs exponential phase bacteria were performed as described in section 2.3.1 with modifications. First, exponential phase cultures (EXP) were produced from overnight cultures by inoculating 5 ml of fresh broth with 50 µl of overnight culture. Samples were left to grow for approximately 5 hours or until an OD_{600}nm of between 0.2 and 0.3 was obtained, corresponding to *F. nucleatum's* early exponential phase. At the same time, stationary phase cultures were obtained from the overnight cultures that had been growing for 24 hours. The incubation time after adding samples for adhesion was 1 hour compared to

the 4-hour incubation time of section 2.3.1. Time adjustments were also applied to invasion (gentamicin protection) assays; incubation was for 2 hours instead of 6 hours.

2.3.3 Chemokine and cytokine profiling

To evaluate the impact of infection of Caco-2 cells with *F. nucleatum* on the secretion of cytokines and chemokines, bacteria-conditioned RPMI media was prepared for chemokine profiling. To prepare the conditioned media, $3x10^5$ cells/ml of Caco-2 cells (T6 and pLVX) were seeded as described previously (2.2.4). Caco-2 cells were infected with bacteria at both the stationary and exponential phases (see section 2.3.2). Stationary phase bacteria had been allowed to grow for 24 hours, and OD was corrected to 1 for use in infection. Exponential cultures were subcultured from overnight cultures and allowed to grow for about five and half hours until OD600 nm of about 0.2-0.4 was reached. The bacteria and Caco-2 cells were washed and prepped as described for adhesion and invasion experiments above (Section 2.3.1). However, co-incubation of cells with bacterial cells was for 24 hours at 37°C in a humidified incubator, 5%CO_2, 37°C. Next, 1 ml of media from the infection assay was put in a microcentrifuge tube and spun down at 8000 x g for 10 minutes to sediment any bacteria and cell debris in the media. Then, 40 µl of media for each sample was sent for cytokine profiling to an Olink service provider (University of Oxford technology platforms group). The Olink® Target 48 Cytokine panel, which tests for 45 cytokines (Table 2.3) was used. Cytokine profiling with the Olink panel was carried out on media from FNA, FNN23, and FNP stationary and

exponential phase samples. *Sgg* samples were included as a control. Two stationary samples from the FNN23 Δ*fadA* mutant strain were also included to explore if the FNN23 chemokine and cytokine response could potentially be FadA dependent.

Table 2.3: List of cytokines on the Olink target 48/40 cytokine panel[1]

Target 48 cytokine and chemokines	
C-C motif chemokine 13 (CCL13)	Interleukin-17C (IL17C)
C-C motif chemokine 19 (CCL19)	Interleukin-17F (IL17F)
C-C motif chemokine 2 (CCL2)	Interleukin-18 (IL18)
C-C motif chemokine 3 (CCL3)	Interleukin-2 (IL2)
C-C motif chemokine 4 (CCL4)	Interleukin-27 (EBI3_IL27)
C-C motif chemokine 7 (CCL7)	Interleukin-33 (IL33)
C-C motif chemokine 8 (CCL8)	Interleukin-4 (IL4)
C-X-C motif chemokine 10 (CXCL10)	Interleukin-6 (IL6)
C-X-C motif chemokine 11 (CXCL11)	Interleukin-7 (IL7)
C-X-C motif chemokine 9 (CXCL9)	Interleukin-8 (CXCL8)
Eotaxin (CCL11)	Interstitial collagenase (MMP1)
Fms-related tyrosine kinase 3 ligand (FLT3LG)	Lymphotoxin-alpha (LTA)
Granulocyte colony-stimulating factor (CSF3)	Macrophage colony-stimulating factor 1 (CSF1)
Granulocyte-macrophage colony-stimulating factor (CSF2)	Macrophage metalloelastase (MMP12)
Hepatocyte growth factor (HGF)	Oncostatin-M (OSM)
Interferon gamma (IFNG)	Oxidized low-density lipoprotein receptor 1 (OLR1)
Interleukin-1 beta (IL1B)	Pro-epidermal growth factor (EGF)
Interleukin-10 (IL10)	Protransforming growth factor alpha (TGFA)
Interleukin-13 (IL13)	Stromal cell-derived factor 1 (CXCL12)
Interleukin-15 (IL15)	Thymic stromal lymphopoietin (TSLP)
Interleukin-17A (IL17A)	Tumor necrosis factor (TNF)
Tumor necrosis factor ligand superfamily member 12 (TNFSF12)	Tumor necrosis factor ligand superfamily member 10 (TNFSF10)
Vascular endothelial growth factor A (VEGFA)	

[1] Cytokines and chemokine list taken from Olink website: Olink® Target 48 Cytokine Panels(https://olink.com/products-services/target/48-cytokine-panel/)

2.4 Metagenomics and transcriptomics

2.4.1 RNA sequencing

RNA sequencing was performed using three *F. nucleatum* subspecies (FNA, FNP and FNN23) and the *fadA* and *fap2* adhesin mutants in FNN23. To do this, 20 samples were sequenced; half were grown to the early exponential phase and the other half to the stationary phase. For each strain, four samples were sequenced, with two biological replicates for the exponential phase and two for the stationary phase. Standard RNA sequencing was carried out by Azenta life sciences (Genewiz) on the llumina NovaSeq 6000 instrument at 2x150 bp sequencing configuration at 10M read pairs per sample. rRNA depletion was prepared with QIAGEN FastSelect rRNA HMR kit, and sequencing library preparation done using the NEBNext Ultra II RNA kit (New England Biolabs). Sequencing was validated on the Agilent technologies Tapestation 4200, and quantification was performed with the ThermoFisher Qubit 2.0 Fluorometer and quantitative pCR.

2.4.2 Bioinformatic analysis

RNA sequencing raw reads were processed and visualised using the DEGUST program (Powell, 2019). Differential expression analyses were performed using EdgeR(quasi-likelihood). Further analysis of RNA sequencing data was performed in R (version 2022.07.2

Build 576). Venn diagrams of differentially expressed genes were performed using jvenn (Bardou et al., 2014). Genes were considered differentially expressed using the log_2fold change cut-off value of 1 and the False discovery rate (FDR) of <0.05. Genes were annotated using Eggnog mapper (Cantalapiedra et al., 2021). Eggnog output gave the Clusters of Orthologous Genes (COG) annotations (Tatusov et al., 1997), which were used to perform functional analysis of the differentially expressed genes using ggplot2 in R. Additionally, network mapping and enrichment analysis were performed using the String (Szklarczyk et al., 2019) database to visualise protein networks enriched in the dataset.

2.4.3 Whole metagenomic sequencing data

Sequencing and preliminary data analysis of paired whole metagenomic and transcriptomic results of 57 CRC tumours and controls was performed by Professor Andrew Beggs (Institute of Cancer and Genomic Sciences, University of Birmingham, United Kingdom) and team. The control samples were taken feom the opposite endof the resection specimen as far away from the tumour as possible. Using Parvian software, bacterial count data was obtained for in-house analysis. A paired analysis of whole metagenomic data and transcriptomic data was performed with the assistance and guidance of Professor Steven Van Laere at the University of Antwerp, Belgium.

2.4.4 Paired metagenomic and transcriptomic analysis

Deseq2 package was used to normalise the count matrices for metagenomics data, and the Limma package was used to normalise the gene expression data. Sample clustering for both metagenomics and transcriptomics data was performed using the T-distributed Stochastic Neighbourhood Embedding (t-SNE) program in R. The TME package was used for differential immune cell expression analysis of the determination of immune scores. CMS typing of samples was queried, and pathway analysis was performed using 50 predetermined cancer pathways.

2.5 Statistical analysis

Unless stated otherwise, one sample t-tests and one way ANOVA with Tukey's multiple comparisons tests were used. Data analysis was done in Microsoft Excel (Version 16.74), and data was graphically visualised and presented using GraphPad prism 10. R studio was also used for bioinformatic RNA sequencing data and cytokine profiling data analysis.

3 INVESTIGATING THE DYNAMICS OF ADHESION AND INVASION OF *F. NUCLEATUM* SUBSPECIES ON CRC CELLS

3.1 Introduction

Genomic and pathogenic differences have been reported for *F. nucleatum* subspecies; the ssp. *nucleatum* is mainly associated with diseased sites, while other subspecies, such as *fusiforme*, are associated with healthy sites (Gharbia et al., 1990;Lourenço et al., 2014). Additionally, studies have reported inequalities in the subspecies impact on the phagocytic function of neutrophils, with greater phagocytosis observed for the subspecies *polymorphum* than the other subspecies (Kurgan et al., 2017). Others reported subspecies-specific differences in the structure of different models of *in vitro* biofilms. They found differences in the establishment of *F. nucleatum* ssp. in biofilm models and their impact on the structure of the biofilm, with members of the biofilm such as *Candida albicans* being decreased when specific subspecies of *F. nucleatum* are present (Thurnheer et al., 2019). More recently, Muchova *et al*. showed that *F. nucleatum* subspecies displayed variations in their *in vitro* biofilm-forming abilities in single-species biofilms grown on different surfaces. They also illustrated that the ssp. *polymorphu*m did not form a stable biofilm compared to the other subspecies. Furthermore, this study showed through bioinformatic analyses that, at least for the ATCC-type strains, some of the major adhesins are not conserved across the subspecies (Muchova et al., 2022).

Most studies on the detection and isolation of *F. nucleatum* from patients do not classify the bacteria to subspecies level as some of the detection methods used are not sensitive to such an extent. In a study that achieved subspecies level resolution, a novel PCR method was used to investigate faecal and tumour isolates from CRC tumours and the ssp. *animalis* was the most isolated from patients, while ssp. *nucleatum* was not as prevalent (Bi et al., 2021). Knowledge

of the individual contributions and interactions of the subspecies with epithelial cells in CRC is still in infancy. One key step in bacterial interactions with epithelial cells is adhesion and invasion to said cells. Bacterial adhesion is critical in bacterial colonisation, persistence and establishment of symbiotic and infectious interactions with host cells (Stones and Krachler, 2016). Understanding the dynamics of adhesion and invasion of key oncopathogens such as *F. nucleatum* is important for uncovering the mechanism of these bacteria in CRC. The adhesion of *F. nucleatum* to different epithelial cell lines has been looked at by others (Rubinstein et al., 2019b;Wang et al., 2019a;Zhang et al., 2022). However, most studies do not mention what subspecies they used and/or use only one (ssp. *nucleatum*). Comparative analysis of the adhesion and invasion dynamics of the subspecies of *F. nucleatum* is still lacking. Understanding the differences in how the subspecies adhere to and invade colon cancer cells can help facilitate understanding of specific factors important to the adhesion and invasion of these bacteria.

This chapter evaluates the differences in adhesion and invasion of three *F. nucleatum* subspecies to Caco-2 cells; subspecies *nucleatum* (FNN23), subspecies *animalis* (FNA) and the subspecies *polymorphum* (FNP). The impact of mutations in FadA and Fap2 adhesins of FNN23 on adhesion and invasion are also evaluated. Furthermore, host-pathogen interactions are explored by looking at the impact of Tspan6 on the adhesion and invasion of *F. nucleatum* subspecies. Tetraspanins have been implicated in bacterial infections, cancer progression, initiation, and spread (see general introduction, section 1.10). As mentioned earlier, Tspan6 has been reported to be under- expressed in CRC tissues and has been found to impact EGFR-dependent signals in CRC. (Andrijes et al., 2021). Control experiments, using the bacterium

Staphylococcus epidermidis (*S. epidermidis*) and *Streptococcus gallolyticus* subspecies *gallolyticus* (*Sgg*), were performed to determine whether any observed impacts of Tspan6 on adhesion and invasion were specific to *F. nucleatum*.

Like *F. nucleatum*, *Sgg* has also gained interest over the past decade as an emerging CRC pathogen. This Gram-positive bacterium is a normal commensal of the gut of several mammals. However, it is involved in a number of opportunistic infections, such as infective endocarditis and sepsis in immunocompromised individuals (Pasquereau-Kotula et al., 2018). Patients with bacteraemia and endocarditis caused by *Sgg* have an increased risk of getting a colorectal cancer diagnosis compared to those whose infections are due to other bacteria (Kumar et al., 2018). *Sgg* strains contribute to CRC by upregulating the beta-catenin signalling pathway in a strain-specific manner (Kumar et al., 2018). Kumar *et al.* further demonstrated that *Sgg* in CRC modulates collagen, which contributes to adherence of the bacteria and increased proliferation in target cells (Kumar et al., 2022). Additionally, *Sgg* has a type VII secretion system, which is important for its adherence to epithelial cells and its tumour-promotive properties (Taylor et al., 2021). More recently, the same authors showed that deletion of the locus named *Sgg* pathogenicity-associated region (SPAR) led to a reduction in adherence, colonisation and tumour promotion both *in vitro* and *in vivo* in a mechanism that possibly impacts the effectors of the T7SS they previously identified (Taylor et al., 2023). The role of secretion systems in the virulence of oncogenic bacteria has also been reported in *F. nucleatum*. For instance, type 5 secretion systems have been identified in *F. nucleatum* genomes. Some of these type 5 secretion systems are the secreted autotransporters that have been shown to have a role in adhesion (e.g., Fap2 adhesin) (Umaña et al., 2019). *Sgg* was included as a control as it is a

known oncopathogen and was used to help determine whether the interactions of *F. nucleatum* with Caco-2 cells and with Tspan6 expressing cells were pathogen specific in CRC associated bacteria.

S. epidermidis, the other control organism used in the current study, is a common constituent of the skin microbiome. Opportunistic infections caused by this bacterium are mainly associated with catheters and implantation devices, resulting in biofilm-rich, difficult-to-treat infections. Furthermore, it also causes bloodstream infections which can lead to sepsis, endocarditis and many complications in neonatal infections (Severn and Horswill, 2023). *S. epidermidis* has several virulence factors which help it in biofilm formation, antibiotic resistance, immune evasion and adherence to the extracellular matrix and other surfaces. It has the accumulation associated protein (Aap) which is important for host adhesion and biofilm formation (Schaeffer et al., 2015). Several surface proteins aid its adherence to abiotic materials and some biomaterials and ligands. These include SdrG/Fbe, a fibrinogen-binding protein, the collagen-binding proteins SdrF and Geh, and the plasminogen-binding protein SesJ (Foster, 2020). Some of these molecules, e.g., SesJ, are a part of what is known as microbial surface components recognising adhesive matrix molecules (MSCRAMMs), common in the virulence repertoire of Gram-positive bacteria (Arora et al., 2020). To our knowledge, no studies currently report on the interaction of *F. nucleatum*, *Sgg* and *S. epidermidis* with tetraspanin 6. Recently, interactions of a bacterium related to *S. epidermidis*, *Staphylococcus aureus (S. aureus)* with the tetraspanin *CD9* were reported. CD9 indirectly decreased the adherence of *S. aureus* in a syndecan-1, fibronectin, and integrins-mediated mechanism (Green et al., 2023b). Additionally, peptides from CD9 inhibited the adherence of *Pseudomonas aeruginosa* to corneal epithelial

cells and in a mouse model of corneal keratitis (Jadi et al., 2023). Similarly, in this thesis, the impact on adherence and invasion to Tspan6 by *F. nucleatum*, with *Sgg* and *S. epidermidis* controls is evaluated. *S. epidermidis* is not a known oncopathogen, it therefore served as a control to determine if there was a difference in the interactions with Caco-2 cells between *F.nucleatum* and a commensal non CRC bacterium.

3.1.1 Studying bacterial adhesion and invasion into eukaryotic cells

Many *in vitro* experiments looking at the adhesion of bacteria onto colonic epithelial cells utilise monolayers of cell lines grown on plastic or discs. Caco-2, T84 and HT29 cells- are some of the most used cell lines in CRC-related *in vitro* studies (Hekmati et al., 1990;McCool et al., 1990;Devriese et al., 2017). Caco-2 cells are capable of spontaneous differentiation into polarisation when grown on membranes. This can be used to study bacterial adhesion but is also useful for looking at nutrient transport across the apical and basal layers of the cell (Kauffman et al., 2013). More complex models, such as 3D organoids and tumour spheroids, have also been used to study the dynamics of malignancy (Badder et al., 2020;Kasper et al., 2020). In this thesis, a monolayer model of Caco-2 cells grown on flat bottom plastic plates and trans well inserts were used to investigate the dynamics of adhesion and invasion of *F. nucleatum* in CRC. The subspecies FNA, FNN23 and FNP and Δ*FadA* and Δ*Fap2* mutants in the FNN23 strains were used. The role of Tspan6 was investigated in the association of the bacteria with Caco-2 cells transfected with a Tspan6 expressing lentiviral vector referred to here as T6 cells. pLVx cells, sometimes referred to as control cells, were transfected with an empty lentiviral vector.

The hypotheses of this study are:

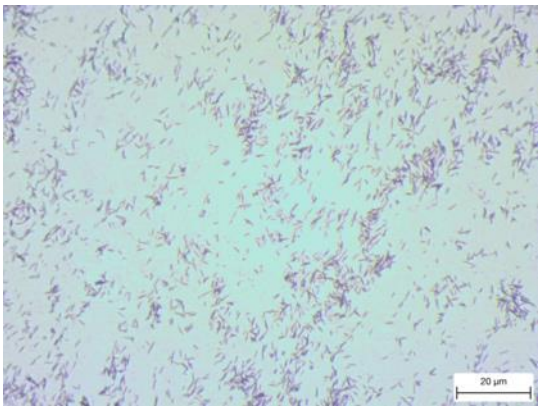
1. *F. nucleatum* subspecies differ in their adhesion and invasion rate to Caco-2 cells.
2. Expression of Tspan6 reduces the adhesion and invasion of the subspecies into Caco-2 cells.
3. Mutations in the adhesins FadA and Fap2 will decrease subspecies' adhesion and invasion rates to Caco-2 cells, especially to Tspan6 expressing cells.

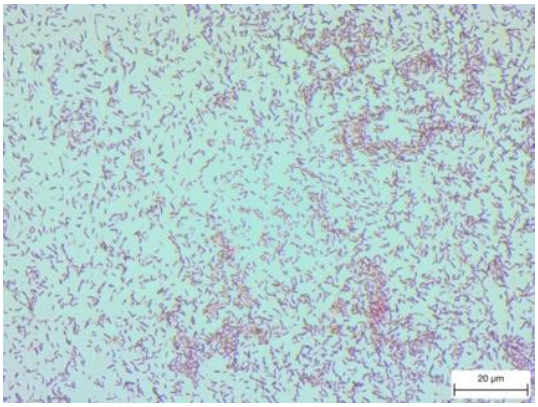
3.1.2 Aims and objectives.

1. To determine the rate of adhesion to colorectal cancer cells of the three subspecies of *F. nucleatum.*
2. To determine the impact of Tspan6 on the adhesion and invasion of the *F. nucleatum* strains in the CRC model.
3. To investigate the role of the mutants in FNN23 of the adhesins Fad A and Fap2 on the adhesion and invasion of the subspecies to Caco-2 cells and the impact of Tspan6 expression on their interactions with the bacterium.

3.2 Subspecies characterisation

Phenotypic characterisation of the subspecies was done using Gram staining to visualise the morphology of the subspecies and to confirm their Gram status (Figure 3.1). As expected of Gram-negative bacteria, the three subspecies stained pink and had a rod-shaped morphology.

A) 

B) 

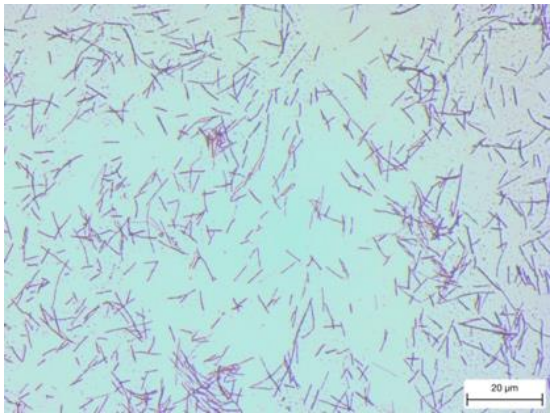
C) 

Figure 3.1*:*Microscopic image of Gram-stained bacteria. *The subspecies FNA(A), FNN23(B), and FNP(C) all stained pink on the Gram stain, and all have a rod-like morphology.Images taken on a Zeiss microspe at x100 magnification.*

16S rDNA PCR followed by 16S sequencing was done to confirm the identity of the subspecies before use for experiments, and results were visualised on an electrophoresis gel (Figure 3.2). All three subspecies had a 1500 bp band on the PCR gel, which is expected when using universal primers. The order of the subspecies was Lane 1-FNA, Lane 2- FNN23, Lane 3- FNP and Lane 4 contained the subspecies *vincentii* (FNV). There was no band on the negative control. All the samples were confirmed by 16s sequencing to be the correct subspecies.

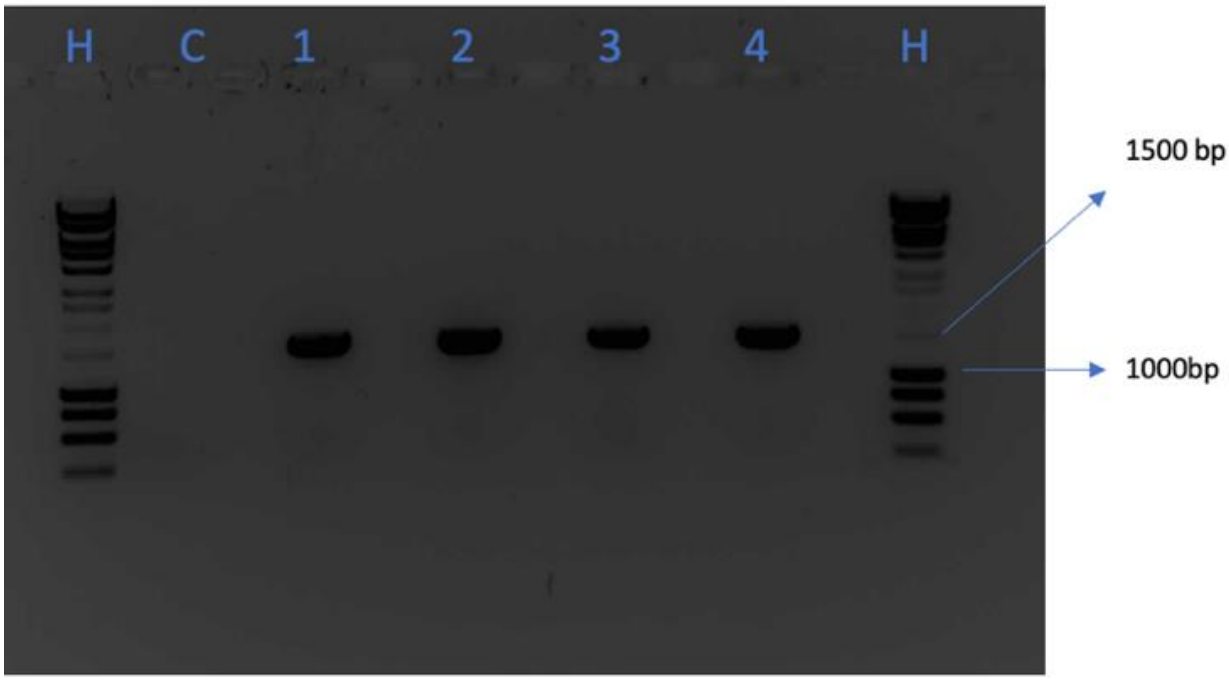


Figure 3.2*:*PCR confirmation of subspecies on electrophoresis gel:*Agarose gel images of show a band at 1500 bp for all the subspecies bp. The subspecies were loaded in the order of; FNA in lane 1, FNN23 in lane 2 FNP in lane 3 and FNV in lane 4. H-hyper ladder, C- control (no template). The expected band size is 1500 bp.A band is visible for all the lanes (1,2,3,4) corresponsing to the 1500 bp mark.*

3.3 Bacterial growth kinetics

Following PCR and 16s sequencing confirmation of the subspecies, more phenotypic characterisation was carried out by determining their growth kinetics. *F. nucleatum* subspecies, OD_{600} nm values measured with a spectrophotometer were plotted on a graph (Figure 3.3). In a 24-hour time, the highest recorded OD for all the three subspecies was at 24 hours. Generally, the three subspecies show similar a similar growth trend.

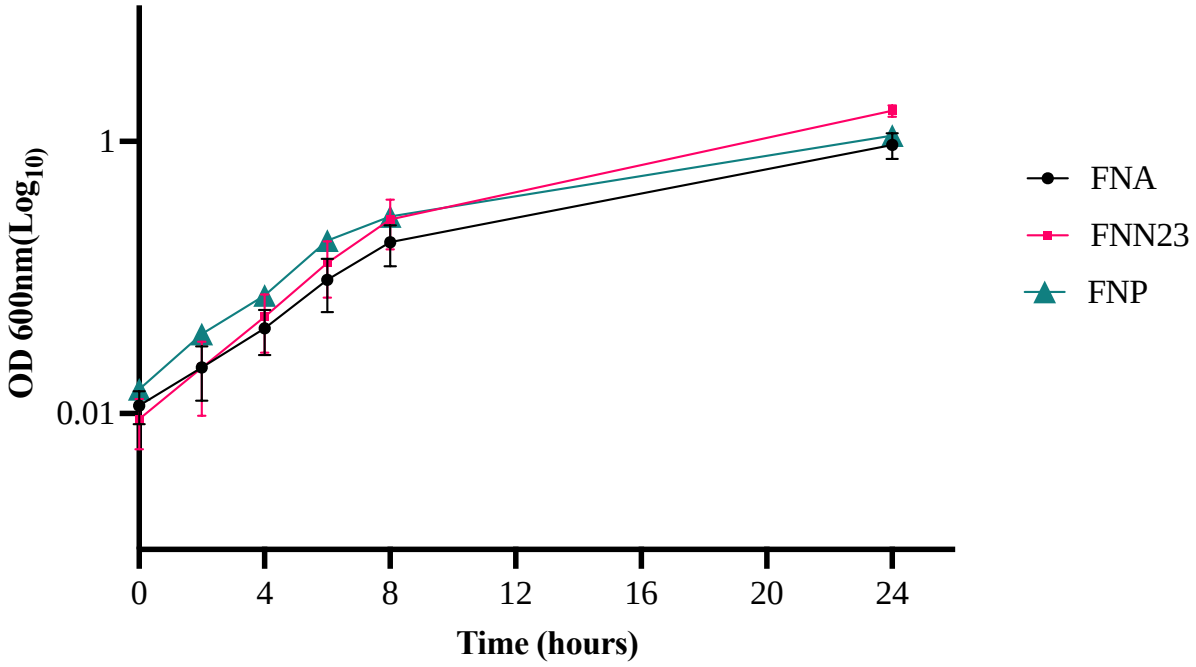


Figure 3.3: ***OD600nm growth kinetics of FNN23, FNA and FNP***. *OD600nm was measured every two hours from the starting time point (time=0) for 8 hours and at 24 hours. Data points represent the mean and standard deviation of 3 independent experiments, each comprised of 3 technical replicates. The y-axis show the time in hours of incubation and the y-axis shows the OD600nm(log10) of each subspecies at the different time points.*

Along with the OD growth curve, bacterial numbers at every time point were determined in Miles and Misra style. FNN23 had the highest CFU/ml(log_{10}) of all the subspecies for most of the time points (Figure 3.4). FNP and FNA appear to have similar bacterial numbers for most time points except at t = 6 hours. Consistent with OD600nm values (Figure 3.3), the highest CFU/ml(log_{10}) at t= 24hours.

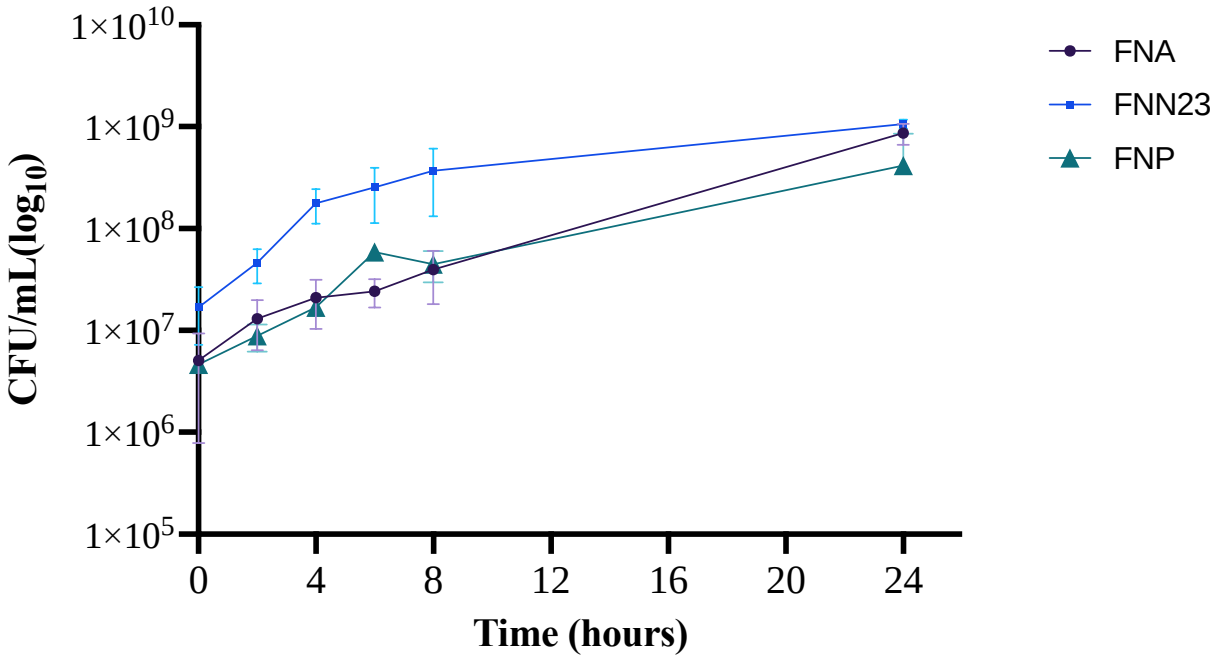
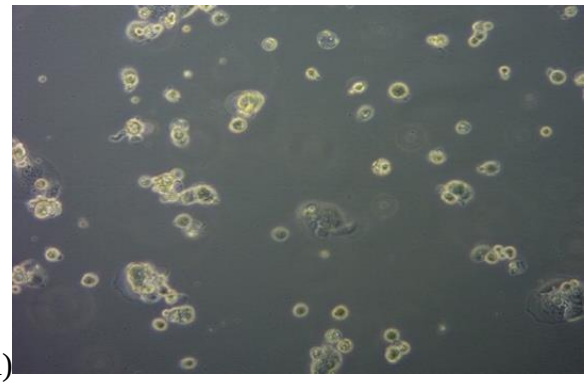
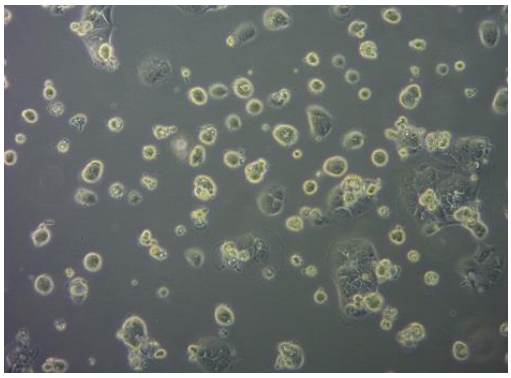


Figure 3.4: Bacterial quantification of subspecies. *Number of colony forming unitsfor each subspecies over a 24 hour time period is shown. Number of colonies were determined using the Miles and misra method (Section 2.1.7). FNN23 shows higher bacterial numbers than the other subspecies for at least 4 of the time points. Data points represent the mean and standard deviation of the mean of 3 independent experiments, each comprised of 3 technical replicates. X-axis shows time of incubation in hours for 24 hours and y-axis show the log colony forming units in CFU/ml (log_{10}).*

3.4 Caco-2 cells growth and maintenance

Caco-2 cells were recovered and maintained for use in experiments. Sub cultured cells grow to form monolayers on plates in a timeline dependent partly on the initial seeding density. The progress of Caco-2 cells from one day following initial subculturing (Figure 3.5A) to three (Figure 3.5B), five (Figure 5C) and seven days (Figure 3.5D) after initial seeding, where cells have reached up to 95% confluency is shown. Cells were mycoplasma tested periodically to ensure they were good to continue using for experiments.

A)  B) 

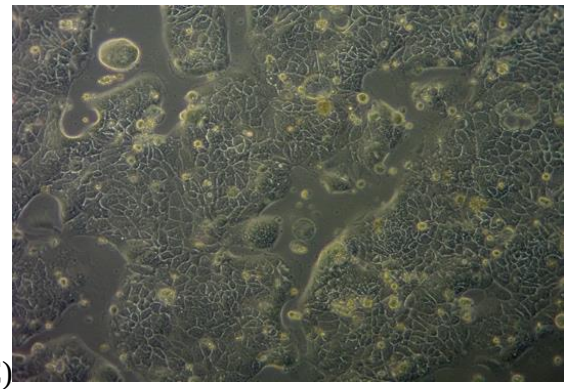
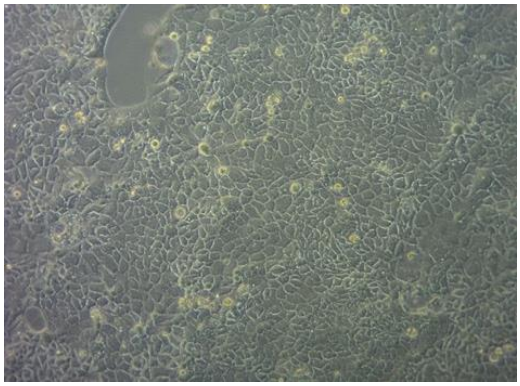
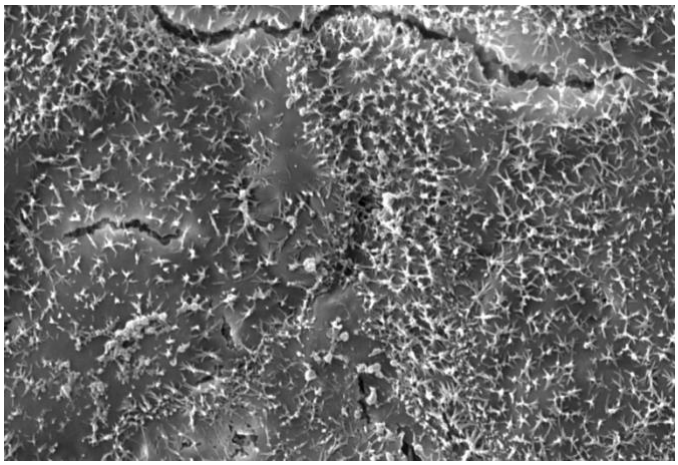
C)  D) 

Figure 3.5:Caco-2 cell growth*. Caco-2 cells growth a day after passaging (A), two days after passaging (B), 5 days after passaging and 6 days after passaging (C) at 95% confluency and fully confluent (D).Images show growth on plastic plates. Image taken using a Zeiss light microscope at x10 magnification. Cells shown here were grown on plastic.*

3.5 SEM images of Caco-2 cells grown on membrane inserts

Caco-2 cells were also grown on membrane inserts to allow for polarisation into basal and apical surfaces. This was done to visualise the cells and confirm that they behave as expected, i.e. that cells are capable of spontaneous differentiation as is known of Caco-2 cells (Kucki et al., 2017). Cells were polarised at around 21 days of incubation and visualised using scanning electron microscopy (SEM). T6 cells (Figure 3.6A) and pLVx cells (Figure 3.6B) showed a brush border appearance.

A)



B)

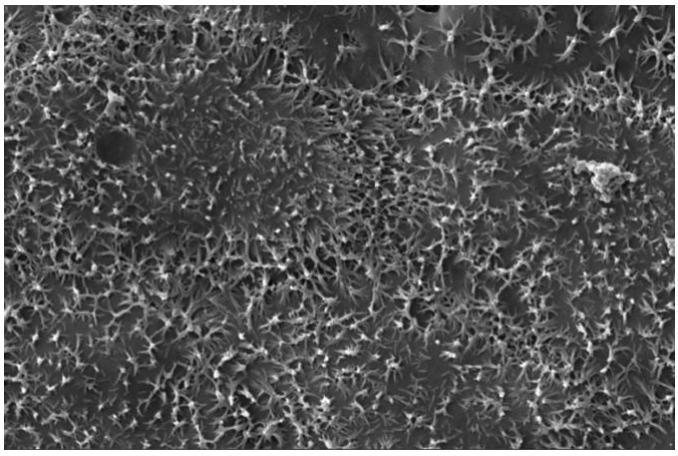


Figure 3.6: SEM images of Caco-2 cells*. pLVx (A) and Tspan 6 expressing cells (B). SEM was performed on cells that had been growing on transwell membrane inserts plates for 21 days. All images taken at magnification x 8.00 K, electron high tension value(EHT) of 20.00 kV and working distance(WD) of 9.0 mm. Microvilli protruding from the cell creating a brush border appearance.*

3.6 Western Blotting

Western blotting was performed to verify the expression of Tspan6 in T6 cells. T6 cells (labelled as T6 1 and T6 2) show a prominent signal between 180 and 250kDa, while the signal is absent in the control pLVx cells (Figure 3.7). The predicted molecular weight of T6 is around 26 kDa, howevever, the signal observed in T6 cells was almost ten times the expected molecular weight. The unusual molecular weight signal could be due to post-translational modifications such as glycosylation or palmitoylation. The accompanying ponceau stain shows that the wells were all loaded with relatively similar amounts of lysate.

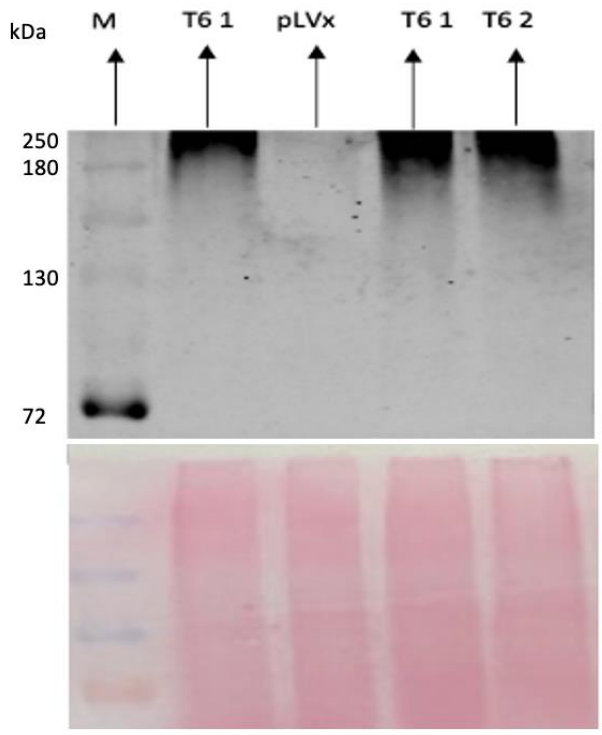


Figure 3.7: Western blots of T6 expression. *Image showing western blot of tetraspanin 6 expressing Caco-2 cells (T61, T62) and Tetraspanin 6 negative control cells (pLVx) using anti Tetraspanin 6 antibodies on a sodium dodecyl sulphate – polyacrylamide gel. On the upper panel (black and white), bands show a signal in wells loaded with T6 lysates. No bands are present on the control pLVx wells. The lower gel panel shows the loading of the ponceau stain which gives an indication of the amount of lysate loaded to each well. M represents pre-stained marker (Molecular weight 11-250kDa). T61 and T62 are both technical replicate of the lysates from cells that have been transfected with the lentiviral vector that express Tetraspanin 6.*

3.7 Monolayer formation

Before adhesion/invasion assays were performed, the optimal Caco-2 cell density that forms monolayers after 48 hours was determined for both T6 cells and pLVx (control) cells. Different seeding densities were used in a 24-well plate, and plates were observed under the microscope at 24 and 48 hours to monitor growth. The seeding density that leads to the formation of a monolayer after 48 hours was determined to be around 3.2 $x10^5$ cells/ml in a 24-well plate (Table 3.1).

Table 3.1:Optimal seeding density of Caco-2 cells.

Optimal seeding density					
			Volume = 1 ml		
Number of cells	90 000	190000	320000	500000	700000
24 hours confluency	x	x	x	√	√
48 hours confluency	x	√	√	√	√
Volume = 1.5 ml					
Number of cells	90 000	190000	320000	500000	600000
24 hours confluency	x	x	x	√	√
48 hours confluency	x	x	x	√	√

*x represents <100 % confluence, and 100 % confluence is denoted by √. The experiment was repeated 3 times with 3 technical replicates each time.

3.8 Adhesion and invasion of subspecies to CRC cells

To uncover *F. nucleatum* subspecies' critical interactions with CRC cells, adhesion and invasion assays were performed using monolayers of CaCo-2- cells to determine the number of adhered and internalised bacteria. Bacterial adherence refers to the cells bound on the surface of the Caco-2 cells and cells internalised into Caco-2 monolayers. In contrast, bacterial invasion refers to bacteria internalised into Caco-2 cells following antibiotic protection. The initial inoculum for adhesion assays was determined using the spectrophotometer to ensure the same MOI (20:1) was used for all strains. However, bacterial counts on agar plates of initial bacteria showed intra and inter-experiment variation in the number of input bacteria. This is because *F. nucleatum* sometimes clumps in liquid culture and it is difficult to obtain a homogenous culture and achieve accuracy using the spectrophotometer. Therefore, for comparison between subspecies, the percentage of initial inoculum is used to account for differences in input bacteria.

3.9 *F. nucleatum* subspecies differ in adhesion and invasion to Caco-2 cells

The three subspecies show differential adherence to Tspan6 expressing cells and control Caco-2 cells (Figure 3.8). The mean adherence (+/- standard deviation) was calculated using an ordinary one-way ANOVA and Tukey's multiple comparisons test used to test for significant differences between the mean adherence of the subspecies. When comparing subspecies mean adherence to control cells, FNP had the highest adherence (0.720 ±0.165 %) compared to

FNN23 (0.102 ±0.0771 %) and FNA (0.1767 ±0.0603%). FNP also had the highest mean adhesion to T6 cells (0.6767±0.2040), followed by FNN23 (0.1760±0.1191), with FNA having the lowest mean percentage adhesion (0.0533±0.0153). Tukey's multiple comprisons tests found no statistically significant difference in the mean adherence of FNA vs FNN23 for adherence to both T6 (adjusted p-value = 0.5304) and control cells (adjusted p- value = 0.898). Statistically significant differences were found only between FNP adherence rates and the adherence rates of FNA and FNN23 to both T6 and control caco-2 cells. The adjusted p values were < 0.0001 for all comparisons of FNP adherence rates vs adherence rates of FNA and of FNN23 to both cells (Figure 3.8).

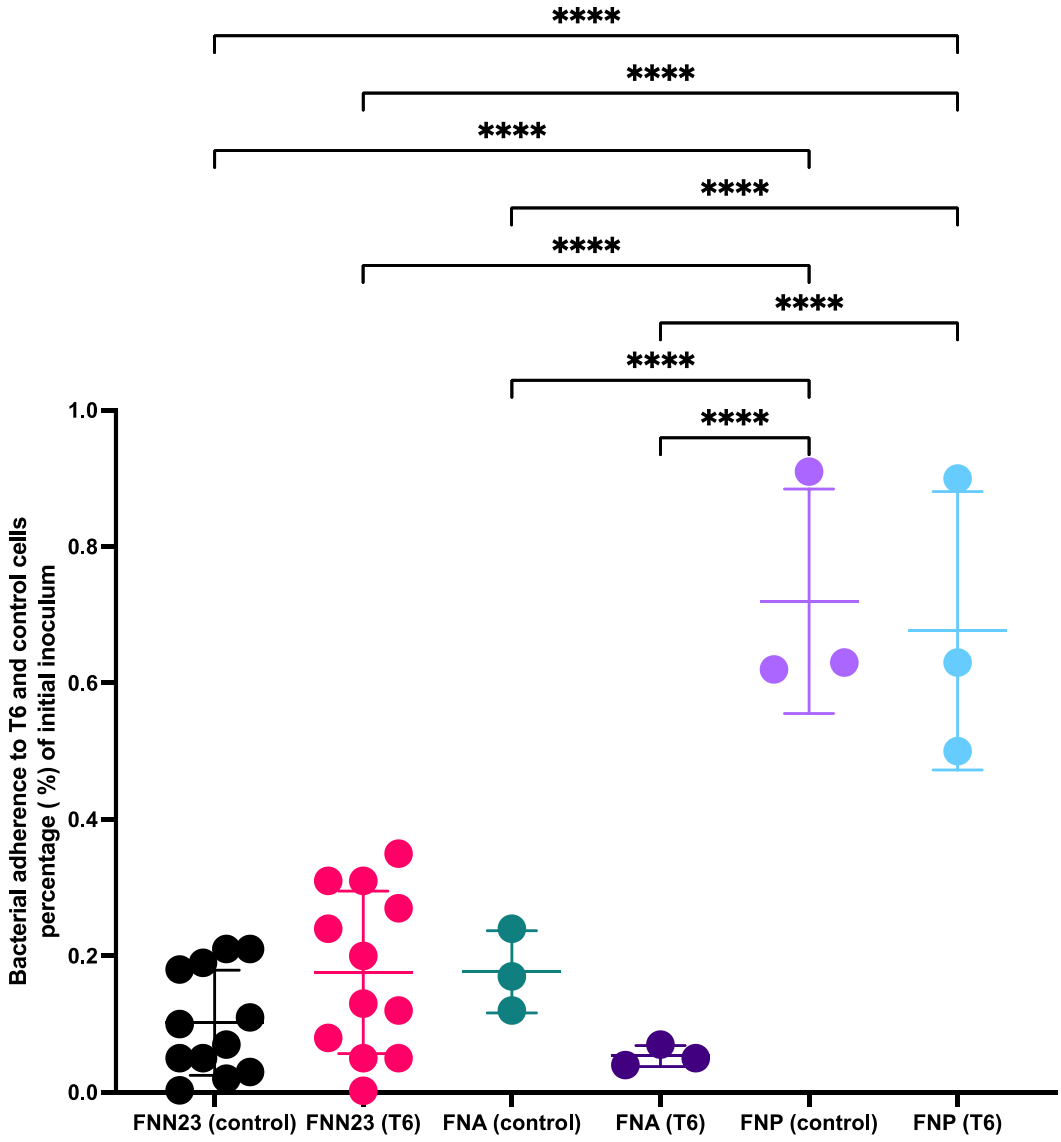


***Figure 3.8*:Subspecies-specific differences in adherence rates of FNP, FNA and FNN23 to T6 cells and control cells.** *Cells seeded at MOI 20:1 and incubated for 4 hours for adhesion.The experiment carried out as decribed in section 2.3.1. Ordinary one-way ANOVA with Tukey's multiple comparisons tests show statistically significant differences in the mean adhesion of FNP (T6) compared to FNN23 (T6), FNA (T6), FNA(control), FNN23 (control) and significant differences in the mean adherence of FNP(control) compared to FNA(control), FNN23(control) and FNA (T6) and FNN23(T6), P< 0.0001. Lines represent mean adhesion, and error bars represent the standard deviation.* Asterisks (****) means p<0.0001).

The invasion rates of the subspecies to Tspan6 expressing cells and to control cells were also determined (Figure 3.9). A one-way ANOVA test with Tukey's test for multiple comparisons was used to determine if there were significant differences between the invasion of the subspecies to caco-2 cells. FNP mean invasion to T6 cells was the highest (0.130 ± 0.0755) followed closely by FNP mean invasion to control cells (0.113 ± 0.0568). FNN23 invasion rated were the second highest, with FNN23 invasion to control cells being higher (0.109± 0.129) than invasion to control cells (0.0525±0.053). The lowest invasion rates were observed for FNA for both control (0.04±0.02) and T6 cells (0.0233±0.0231). Interestingly, differences between the percentage invasion of the subspecies to T6 cells and control cells were not statistically significant. The mean difference between the invasion to T6 cells for FNP and FNA was the highest (-0.1067, adjusted p-value =0.718).

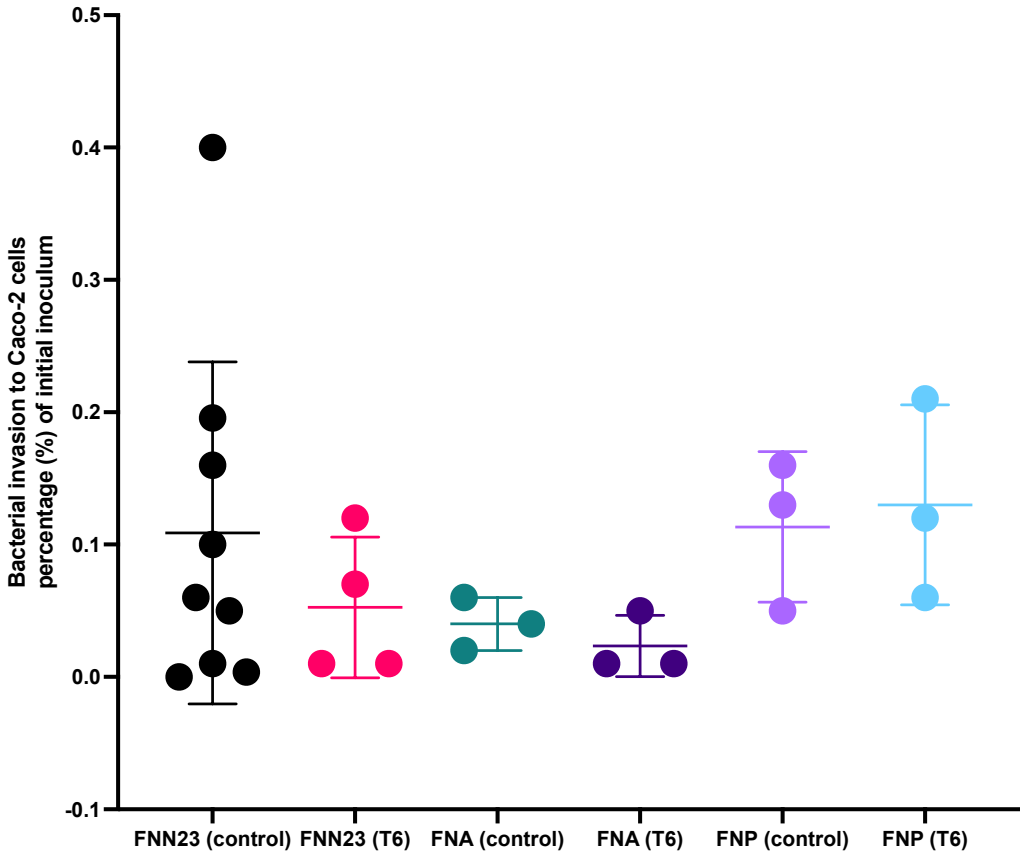


Figure 3.9:Subspecies-specific differences in the invasion rates of FNP, FNA and FNN23 to T6 cells and control cells. *Cells seeded at MOI 20:1 and incubated for a total of 6 hours for invasion. The experiment was carried out as decribed in section 2.3.1. Ordinary one-way ANOVA with Tukey's multiple comparisons tests did not show statistically significant differences in the comparison of the mean invasion between the subspecies. Lines represent mean mean invasion, and error bars represent the standard deviation.*

3.10 Tspan6 expression impacts the association and internalisation of *F. nucleatum* subspecies to Caco-2 cells.

The role of Tspan6 in the interaction of the subspecies with CRC cells was also investigated. Adhesion and invasion assays were performed using Tspan6 expressing cells (T6 cells). One sample t-test compared the mean adhesion and invasion differences of subspecies to T6 cells compared to control cells. Subspecies-specific differences in interactions with T6 cells were observed (Figures 3.10A and B). Tspan6 expression decreases the mean adherence of FNA (34.39, P= 0.0363) but increases that of FNN23 to Caco-2 cells (141, P = 0.0218). Conversely, Tspan6 expression does not impact the average adherence of FNP to Caco-2 cells (mean = 99.96 P=0.9989). Unlike for adhesion, FNN23 invasion was significantly decreased by Tspan6 expression (55.84, P=0.0155). FNA invasion was also reduced by Tspan6 expression (46.03 P = 0.0896), but the difference was not statistically significant at alpha = 0.05. For FNP, a slight increase in mean invasion to Caco-2 cells was observed (114.4 P= 0.3783) with Tspan6 expression. Using Ordinary one-way ANOVA with Tukey's multiple comparisons test, the differences between the rates of adhesion and invasion of the subspecies to T6 cells were statistically evaluated (Figure 3.10)

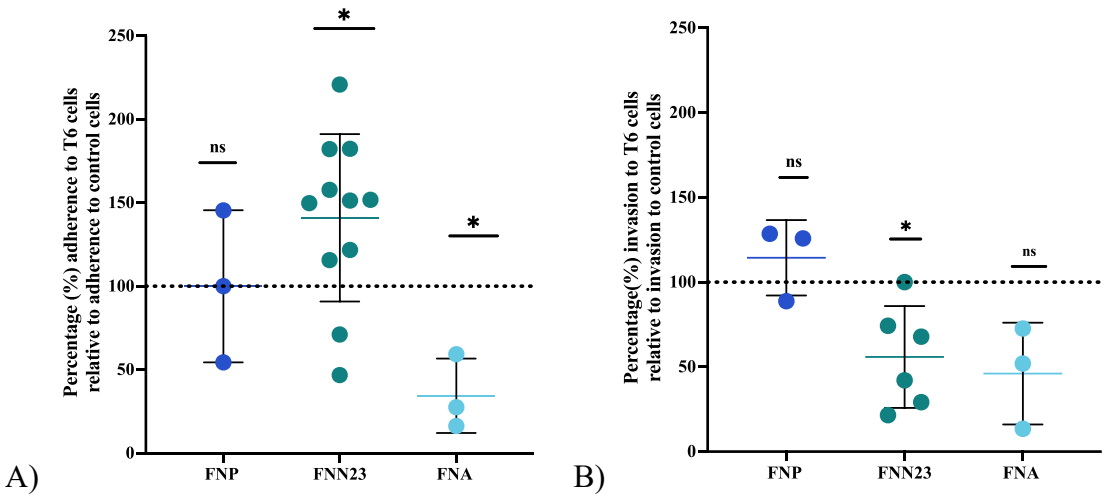


Figure 3.10: Impact of Tspan6 expression on the adhesion and invasion of F. nucleatum subspecies to Caco-2 cells*. (A) Mean adhesion by FNN23 is increased, and FNA adherence is reduced by Tspan6 expression with no impact for FNP (B) Tspan6 expression decreases mean invasion of FNN23 and FNA but increases that of FNP. Lines represent the mean adhesion and invasion, and error bars represent the standard deviation of the discrepancy from the mean. At least three experimental replicates with three technical replicates per experiment were done. Asterisk (*) represent p<0.05 and (ns) means "not significant".*

3.11 *F. nucleatum* subspecies show reduced invasion relative to adhesion

Bacteria can induce changes in eukaryotic cells by interacting with pattern recognition receptors and other partners on the surface of eukaryotic cells. However, internalisation and/or invasion is also important in the infection process of certain bacteria. The number of invaded or internalised bacteria relative to the number of adhered bacteria can shed light on the virulence of bacteria depending on their mechanism of action. The percentage invasion relative to adhesion was determined for the subspecies for interactions with T6 cells and control cells. Less than 50% of adhered bacteria are generally internalised at the conditions tested (Figure 3.11). For each subspecies, there is no significant difference in the number of bacteria invaded relative to adhered T6 cells compared to control cells. Of the three subspecies, FNN23 shows higher invasion relative to adhesion for both T6 cells and control cells.

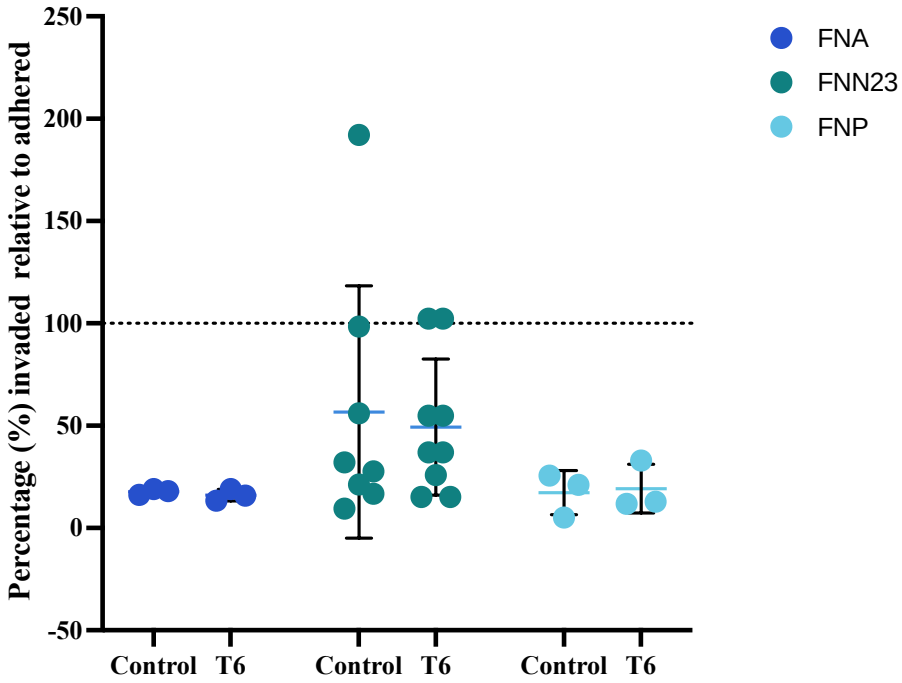


Figure 3.11:Invaded cells relative to adhered cells for T6 and control cells. *For all the subspecies, the percentage of invaded cells relative to adhered does not significantly vary for T6 and control cells. The invasion rate compared to adhesion does not vary significantly between the subspecies. Data shows the mean of at least 3 independent experiments (with three technical replicates), and error bars show the standard deviation from the mean.*

3.12 The role of the *FadA* and *Fap2* mutants on the adhesion of FNN23 to Caco-2 cells

FadA and Fap2 adhesins have been studied for their different roles in adhesion, immune modulation, and co-aggregation in *F. nucleatum*. *F. nucleatum* currently has only one genetically tractable subspecies: the ssp. *nucleatum* (FNN23). Deletion mutant strains of *FadA* and *Fap2* genes were used in adhesion/invasion to see their impact on the adhesion of FNN23 and determine how they affect its interaction with T6-expressing cells.

3.13 Determination of the impact of mutants on the *growth* rate of FNN23

Prior to adhesion/invasion assays with FNN23 mutant strains, the growth kinetics of the strains were determined to check the impact of the mutations on the growth of the bacteria compared to the wild-type strain (Figure 3.12) every 2 hours for 8 hours and at 24 hours. No significant differences in the growth of subspecies were observed at all time points in the 24-hour time interval, suggesting that the mutations do not affect the growth of FNN23.

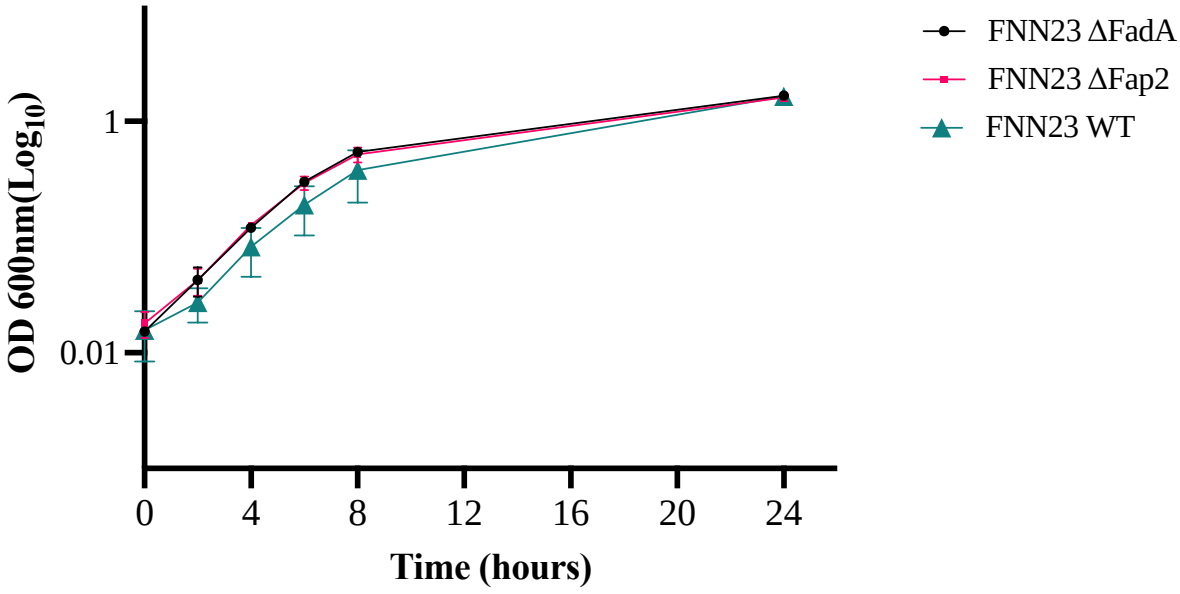


Figure 3.12:***Growth kinetics of subspecies over 24 hours***. *No significant growth difference was observed, and error bars overlap for all the time points. The curves represent the mean and standard deviation of the mean of 3 independent experiments, each comprised of 3 technical replicates.*

Numerical quantification of the bacteria was done simultaneously with OD measurements (Figure 3.13). The wild type has slightly higher numbers of bacteria across all time points. However, the error bars overlap for all time points for all the strains except for at t=4, where none of the error bars overlap. Overall, the mutations do not appear to have a significant impact on the multiplication of bacteria over 24 hours.

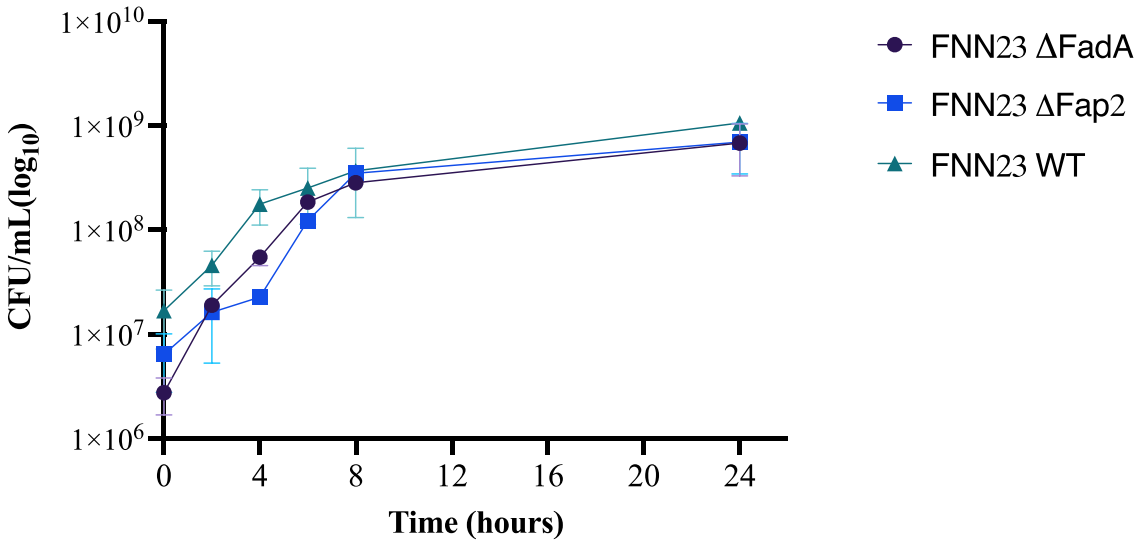
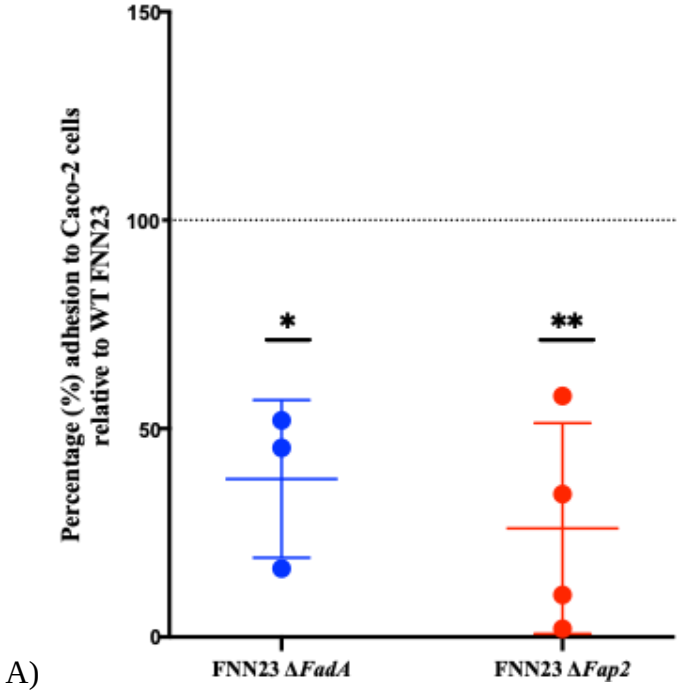


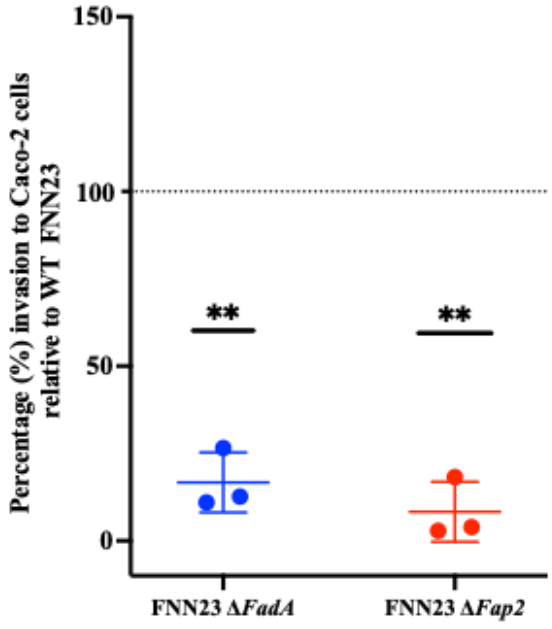
Figure 3.13: Bacterial quantification of the FadA and Fap2 mutants and wild-type FNN23 strain*. The WT shows higher bacterial numbers than the other subspecies for all the time points. The curves represent the mean and standard deviation of the mean of 3 independent experiments, each comprised of 3 technical replicates.*

3.14 Mutants in *FadA* and *Fap2* impact the association and the internalisation of FNN23 to Caco-2 cells

FNN23 *Fap2* and *FadA* mutant strains were used in adhesion and invasion assays, and their adhesion and invasion rates were compared to the wild-type (WT) FNN23 strain. The mutants had decreased adhesion to Caco-2 cells (Figure 3.14) compared to the WT strain. The reduction in adhesion with the *FadA* mutant is nearly 3-fold (2.64) while nearly 4-fold (3.84) with the *Fap2* mutant strain. One sample t-test with the mean set to 100 for comparison with adhesion by the wild-type strain showed statistically significant differences (P= 0.0296 for FNN23 Δ*FadA and P*= 0.0099 for FNN23 Δ*Fap2)* in adhesion of the mutant strains to Caco-2 cells relative to the WT strain. Similar to adhesion, the mutant strains led to a significant reduction in the invasion of FNN23. Over 5-fold reduction was observed for the FadA mutant strain, and a dramatic 11-fold reduction in invasion was observed for the Fap2 mutant strain. Two-tailed one-sample t-test showed a significant difference p= 0.0035 and p= 0.0029 for the *FadA* and the *Fap2* mutant strains, respectively.



A)



B)

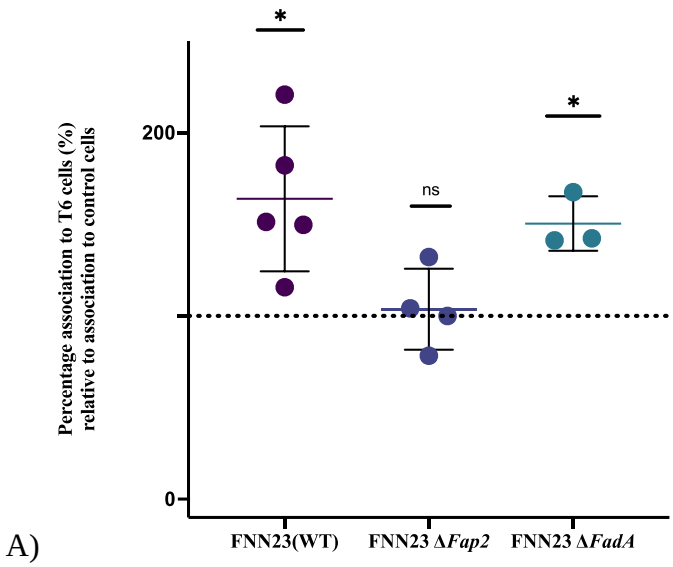
Figure 3.14*:*Impact of the FadA and Fap2 mutants on the adhesion and invasion of FNN23 to Caco-2 cells. *Mutations in FadA and Fap2 reduce the adhesion of the FNN23 strain(A) and its subsequent invasion (B). Bars represent average percentage adhesion(A) and average percentage invasion (B) relative to FNN23 WT strain standardised to 100%. Asterisk (*) indicates P < 0.05, (**) means p<0.001 at alpha of 0.05.*

3.15 Tspan6 expression impacts the adherence of the *FadA* and the *Fap2* mutants in FNN23 to Caco-2 cells.

Tetraspanins have been reported to interact with partner proteins which can include adhesion molecules in their tetraspanin-enriched domains. It is possible that Tspan6 associates with either of the adhesins in their interactions with Caco-2 cells. As a result, Tspan6 expression was evaluated for its impact on the adhesion of the mutants to Caco-2 cells. A one-sample t-test was performed to analyse the difference in adhesion to T6 cells compared to control cells standardised to 100% and to analyse differences between the mutant strains compared to the wild-type strain. For the FadA mutant strain, an improvement in adhesion (mean = 150, P= 0.0278) to Caco-2 cells was observed in T6-expressing cells compared to control cells Figure 3.15A). Though the increase in adhesion to T6 cells compared to control cells was significant, this did not differ significantly from the wild-type strain's interaction with the same cells (Figure 3.15B), suggesting that FadA might not have a role in the interaction with T6 cells.

Contrary to the FadA mutant strain, no difference in adhesion to T6 cells compared to control cells was observed for the Fap2 mutant strain (mean = 103.7, P= 0.7621). However, when compared to the adhesion of the wild-type strain, the adhesion to T6 cells by Fap2 is

significantly lower (mean= 24.66, P= 0.0069) than the adhesion to the same cells by the wild-type strain.



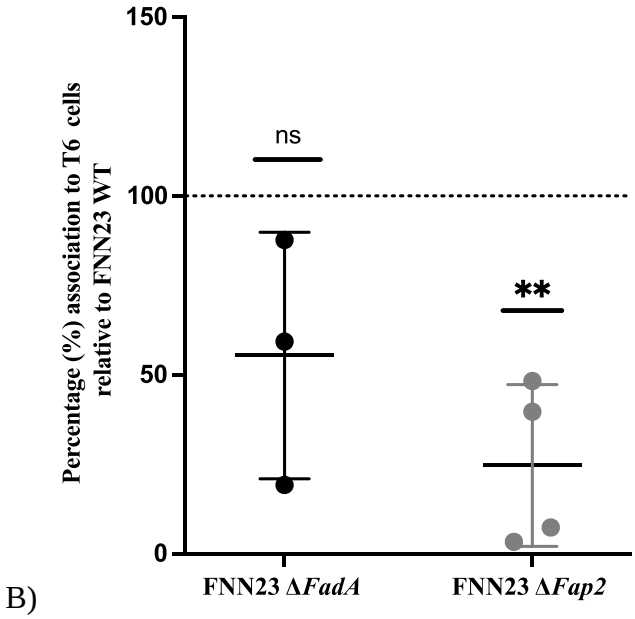
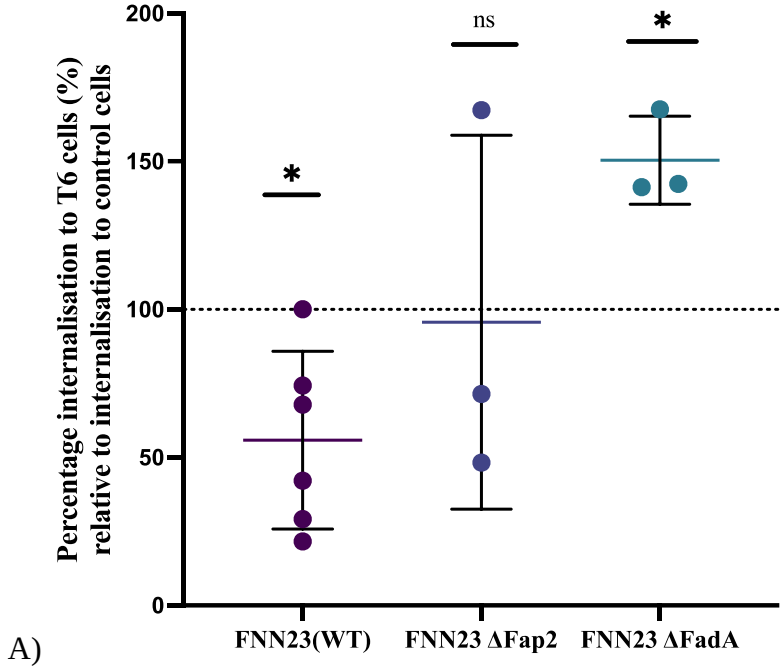


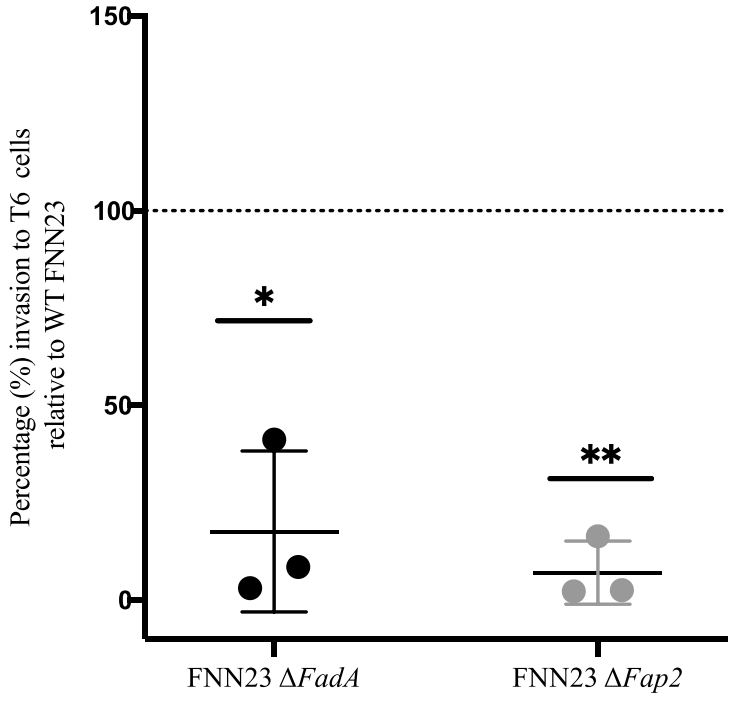
Figure 3.15: Impact of a mutation in the FadA gene and the Fap2 gene on the adhesion of FNN23 to T6 expressing Caco-2 cells*. A). Percentage adherence of the wild type FNN23 and the mutant strains to Tspan6 expressing cells compared to adherence to control cells. Lines represent the percentage of adhesion with control cells standardised to 100%. (B). Shows adherence of the FadA and Fap2 mutant strains in comparison to adherence of the wild type strain. Lines represent the percentage of adhesion with adherence of wild type strain standardised to 100% . Error bars represent the standard deviation from the mean. Experiments were repeated at least three times, with three technical replicates for each experiment. Asterisks (*) means p<0.05, and (ns) means "not significant".*

3.16 Tspan6 expression impacts the invasion of the *FadA* and the *Fap2* mutant strains in FNN23 to Caco-2 cells.

The impact of Tspan6 expression on the invasion of the mutant strain was also investigated. Tspan6 expression significantly improves the invasion of the *FadA* mutant strain (mean = 150.4, P= 0.0278) to T6 cells compared to control cells. Conversely, it has no significant impact on the invasion of the *Fap2* mutant strain to the same cells (mean = 95.65, P= 0.9160). Interestingly, the invasion of the wild-type strain is lower in T6 cells than in control cells, despite the improved adhesion mentioned earlier (Figure 3.16A). Though T6 expression decreases the invasion of the wild-type strain, its invasion was still greater than that of both mutants to T6 cells (Figure 3.16B). A greater decrease was observed for the *Fap2* mutant strain (mean = 7.072 %, P= 0.0025) than for the FadA mutant strain (mean = 17.60 %, P= 0.0203) compared to a standardised mean percentage adhesion of 100% for the WT strain.



A)



B)

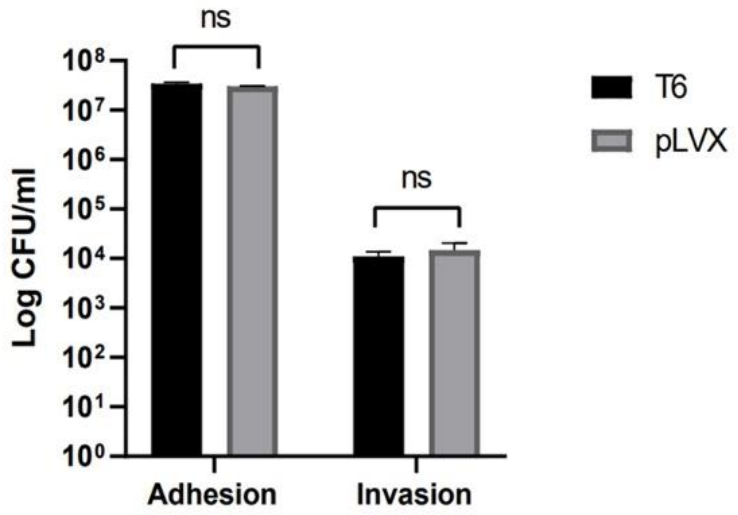
Figure 3.16:Impact of Tspan6 expression on invasion of mutants. *A) Comparison of invasion to T6 cells compared to control cells for the mutants and the wild type. One sample t-test used to compare differences in invasion for the wild type strains to T6 cells compared to control cells which has been standardised to 100 %. Significant differences for invasion of FNN23 and Fad A. B) Comparison of the levels of invasion of the mutant strains to T6 cells relative to the wild type strain. Wild type invasion levels have been standardised to 100 %. One sample t- test performed to compare the mean percentage invasion of the mutant strains compared to the wild type strain. The error bars represent the standard deviation from the mean. Experiments were repeated three times*

*with three technical replicates for each experiment. asterisk (*) means p<0.05, ** p<0.005, ns = not significant at alpha=0.05.*

3.17 *S. gallolyticus* and *S. epidermidis* control for adhesion and invasion to T6 and control cells

To determine whether the impact of Tspan6 on adhesion and invasion was specific to *F. nucleatum*, *Sgg* and *S. epidermidis* were used in a control experiment to check if the observed impacts of Tspan6 on adhesion and invasion are *F. nucleatum* specific. *S. epidermidis* showed no difference in adhesion to T6 cells compared to control cells (pLVx) (student t-test, (p = 0.06736916)). This was also the case for the difference in the bacteria's invasion of T6 cells compared to control cells (Figure 3.17A). Similarly, *Sgg* adhesion to T6 compared to control cells was not statistically significant at an alpha of 0.05 (Figure 3.17B). However, Tspan6 expression improved *Sgg* invasion to Caco-2 cells resulting in a statistically significant mean invasion compared to control cells (student t-test, p = (0.00687159)).

A)



B)

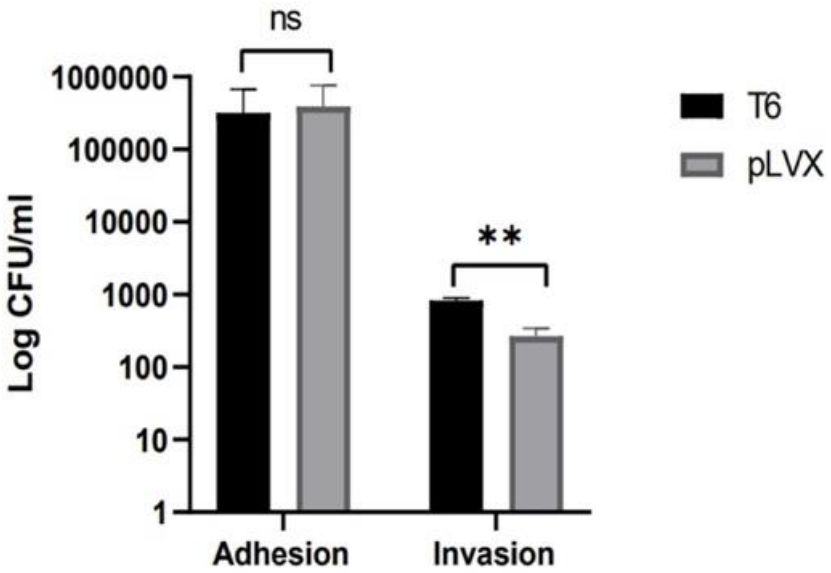


Figure 3.17*: *Adhesion and invasion of Sgg and S. epidermidis. *(A) The mean difference of the adhesion of S. epidermidis to T6 cells compared to control cells (B) The mean difference of adhesion of Sgg to T6 cells compared to control cells(pLVx). Adherence and invasion experiments were performed according to section (2.3.1). Student t-test,performed to check if there are significant differences In the adherence and invasion of the control strains. Bars show mean adhesion and invasion from three independent experiments with three technical repeats. Error bars indicate the standard deviation from the mean. Asterisk (**) means p<0.005, and (ns) means "not significant". Data was collected and analysed by Gabriela Naia with my supervision for her Master by research degree and used here with permission.*

3.18 Key findings of this chapter

1. FNP shows comparably higher adhesion rates to Caco-2 cells (T6 and control cells) than the other subspecies.
2. Tspan6 expression improves the adhesion of FNN23, reduces that of FNA and makes no difference to the adhesion of FNP to Caco-2 cells.
3. Subspecies show no significant differences in invasion to Caco-2 cells; however, Tspan6 expression decreased the invasion rates of FNN23 significantly and FNA (insignificantly).
4. Mutations in the *FadA* and the *Fap2* genes lead to decreased adhesion and invasion of Caco-2 cells compared to the WT strain.
5. Tspan6 expression improves FadA mutant strain adhesion but decreases Fap2 mutant strain adhesion to Caco-2 cells.
6. Invasion to Caco-2 cells is reduced for both mutant strains when Tspan6 is expressed.
7. Tspan6 expression does not impact the adherence of both *Sgg* and *S. epidermidis* control strains.
8. The expression of Tspan6 led to an increase in invasion for *Sgg* but not for *S. epidermidis.*

3.19 Discussion

The role of *F. nucleatum* in CRC has been extensively reported in the literature. However, there is a paucity of knowledge on the exact mechanism of action of this bacterium in CRC. Data on the interactions of *F. nucleatum* subspecies with CRC cells can help unravel the mechanisms that drive its contribution to CRC. This work aimed to determine whether there are differences in the adherence and internalisation of three subspecies of *F. nucleatum* to CRC cells. Furthermore, the role of the Tspan6, a transmembrane protein often overexpressed in CRC tissues, was also evaluated for its potential impact on the adhesion of the subspecies. Additionally, the role of the adhesion mutant in FNN23, *FadA*, and *Fap2* was determined to ascertain the adhesins' contribution to adherence and invasion.

Initial work characterising the bacteria using growth kinetics and bacterial quantification for the subspecies and adhesin mutants revealed no significant growth differences between the subspecies. Adhesion and invasion assays revealed a subspecies-dependent difference in adherence of the subspecies to Caco-2 cells. Generally, FNP showed significantly higher adherence to Caco-2 cells than other subspecies. The adherence levels of FNN23 and FNA were not significantly different. Adhesion was slightly reduced with FNA; however, the subspecies did not significantly differ in their levels of invasion. We postulate that differences in adhesion may be due to variations in the adhesins profile of the subspecies and in the expression levels of said adhesins for each subspecies. Earlier studies have shown heterogeneity in the adhesive abilities of some *F. nucleatum* subspecies to oral epithelial cells, and have also purported that differences may be due to variations in the expression of adhesins (Xie et al., 1991) *F.*

nucleatum subspecies heterogeneous behaviour is not surprising as it has been reported for other processes, such as biofilm formation, where the subspecies *polymorphum* was a poor biofilm former in single species biofilms (Muchova et al., 2022).

Previous reports have reported the role of tetraspanins in the adherence of bacteria such as *Neisseria meningitis, L. monocytogenes, Staphylococcus aureus* and *E. coli* (Tham et al., 2010;Green et al., 2011;Karam et al., 2020). Here, we investigated the role of the poorly studied tetraspanin, Tspan6, on adhesion and invasion of the subspecies into Caco-2 cells. Using cells that express Tspan6 and control cells, adhesion by FNP was the highest of all the subspecies, with over 50% more adherence to T6 cells than FNN23 and FNA. Invasion was also higher for FNP compared to the other two subspecies of T6 cells though the differences in invasion were not statistically significant. This result aligns with reports from the literature, where others found FNP to be more invasive compared to subspecies *nucleatum* and *vincenti*i (Dabija-Wolter et al., 2009). Tspan6 expression improved the adhesion of FNN23 but decreased its invasion of Caco-2 cells. Conversely, it reduced both the adherence and internalisation of FNA. Interestingly, it did not impact FNP adhesion and invasion of Caco-2 cells. These results indicate that FNP adhesion is independent of Tspan6 and that Tspan6 might impact the adherence of FNN23 and FNA in a subspecies-dependent mechanism. For both FNN23 and FNA, Tspan6 appears to decrease invasion, possibly because of its possible role as a tumour suppressor gene.

We then looked at the role of two major adhesins *FadA* and *Fap2,* in the adherence of FNN23 and aimed to determine their importance in the interaction between the bacteria and Caco-2 cells. Mutations in *FadA* and *Fap2* did not induce a significant growth impact on the strains compared to the wild-type FNN23 strain. However, both mutations led to a decrease in both the adhesion and the invasion of Caco-2 cells compared to WT FNN23. This was also true for the adhesion and invasion of the mutants to T6-expressing Caco-2 cells compared to the WT strain. Similar results have been reported by others, where a reduction in adhesion to placental cells by a *Fap2* mutant strain was observed (Coppenhagen-Glazer et al., 2015). The impact of *FadA* mutation on adherence might be strain and cell-type dependent. A *FadA* mutant in the strain reduced adherence and invasion to human oral keratinocytes (reversed by the gene complementation) using the *F. nucleatum* 12230 US1 strain (Ikegami et al., 2009) but had no consequence on adherence and invasion of *F. nucleatum* to a colon cancer cell line called HTCT11 (Casasanta et al., 2020). Even though adhesion and invasion of the mutants to T6 cells were reduced compared to the WT strain, T6 expression only slightly improved the adhesion and invasion of the *FadA* mutant to Caco-2 cells. Still, it made no difference for the *Fap2* mutant strain. These results suggest that *Fap2* but not *FadAs* role in the adhesion and the invasion of FNN23 to Caco-2 cells may be partially regulated by T6 expression.

Experiments with both mutants did not lead to a complete loss of adherence and invasive ability as they may each compensate in the absence of the other. However, a double mutant in *Fap2* and *FadA* did not further reduce adherence in experiments done with HCT116 cells. There was a reduction in adherence and invasion due to other transporter proteins but none to an extent like the *Fap2* reduction on adherence and invasion (Casasanta et al., 2020). Colorectal cancer

has been shown to overexpress Gal-GalNAc, a tumour marker recognised by Fap2 (Abed et al., 2016a). E- cadherin was found to be the receptor used by *FadA* to bind to cells resulting in an increased cellular permeability (Fardini et al., 2011). Tspan6 potentially interacts, directly or indirectly, with molecules on the surface of the bacteria and impacts adherence and invasion.

Sgg and *S. epidermidis* were used as controls for the adhesion invasion assays. *Sgg*'s role in CRC is well documented, so it served as a control to determine whether observed dynamics with *F. nucleatum* are also true for other CRC–linked pathogens. No difference was observed in adhesion to T6 cells compared to controls; however, a significant reduction in invasion to T6 cells was observed, an outcome like that observed for FNN23. Generally, Tspan6 expression appears to reduce the levels of invasion of the tested bacteria; it may affect similar processes in invasion for both FNN23, *Sgg* and FNA. Conversely, the expression of Tspan6 did not impact *S. epidermidis* adherence and invasion. This may be because it does not activate similar carcinogenic pathways directly or indirectly impacted by Tspan6 like the other tested bacteria. Both of these controls are Gram positive bacteria and have different virulence mechanisms and this may confound their utility as controls.

Infection and incubation during the invasion and adhesion assays were done in an aerobic environment, which may have impacted the growth of the bacteria and dampened their adhesive and invasive capabilities. Work to compare incubation in anaerobic conditions can be carried out to compare the impact of growth conditions on adhesion and invasion. It is known that *F. nucleatum* can survive under aerobic conditions (GURSOY et al., 2008); however, an optimum

environment would support the anaerobic requirements of the bacteria and the aerobic needs of the Caco-2 cells. Additionally, as Caco-2 cells are already transformed, it is impossible to determine and infer the interaction of the bacteria with normal cells to begin controlling for the effects of transformation on cellular adhesion by bacteria. In conclusion, we show differential adhesion to Caco-2 cells by *F. nucleatum* subspecies. Furthermore, we show that Tspan6 expression impacts the adhesion and the invasion of the subspecies to Caco-2 cells in a subspecies-specific manner. Adhesion and invasion of the FNN23 Caco-2 cells are impacted by the adhesins FadA and Fap2. Additionally, we demonstrate that Tspan6 expression has a marked impact on the FadA mutant strain, which suggests that it could interact with the bacteria in a Fap2-dependent manner.

4 GROWTH PHASE IMPACT ON THE ADHESION, INVASION, AND INFLAMMATORY MEDIATOR SECRETION OF *F. NUCLEATUM* SUBSPECIES UPON CACO-2 CELL INFECTION

4.1 Introduction

In Chapter 3, the adhesion and invasion dynamics of *F. nucleatum* subspecies to CRC cells were demonstrated. However, pathogen factors that impact the levels of adhesion and invasion into CRC cells are yet to be elucidated. This section investigates whether the rates of adhesion and invasion of *F. nucleatum* subspecies are influenced by the bacterial growth phase. Additionally, the impact on host cells of bacterial adhesion and invasion was evaluated by exploring the host secretion of inflammatory and immune modulators upon infection. Bacterial growth comprises four phases: the lag, exponential, stationary and death phases (Maier, 2009). Different morphological, transcriptomic, and metabolic signatures characterise the stages. In the lag phase, bacteria acclimatise to their environment, and the expression of transporter proteins, needed for substrate utilisation and nutrient acquisition, is prioritised (Rolfe et al., 2012;Veselovsky et al., 2022). The exponential phase, further split into early, mid, and late phases, is characterised by amino acid synthesis, increased metabolic demand, and rapid cell division. In the stationary phase, cell division declines, and most of the genes expressed in the exponential phase are downregulated while stress-related transcripts dominate (Pobre et al., 2019). The death phase involves increased loss of viable cells due to increased accumulation of toxins and prolonged starvation (Navarro Llorens et al., 2010).

The impact of the growth phase on adhesion and invasion has been reported in other bacteria. For example, *S. aureus* adhesion to mammary cells was improved during the exponential phase due to increased expression of the accessory gene regulator *(agr)*, resulting in upregulated fibrinogen expression (Lammers et al., 1999;Ji et al., 2020). Conversely, in *E. coli*, adhesion

was higher during stationary phase than during exponential phase due to variations in charge distribution at the two growth phases. A more uniform charge distribution, characteristic of exponential phase bacteria, increased repulsion to surfaces, reducing the adhesion of *E. coli* (Walker et al., 2005). In *Burkholderia cepacia*, adhesion was higher at the early exponential phase than in the late exponential phase (Mariappan et al., 2011). Expression levels of adhesins at different growth phases could influence the outcome of adhesin/invasion results depending on what growth conditions the experiments are carried out under. To determine whether the bacterial growth phase affects adhesion and invasion in *F. nucleatum*, adhesion at the stationary phase was compared to that of bacteria in the exponential phase.

1.1.1 Cytokine and chemokine expression in the tumour microenvironment (TME)

The tumour microenvironment (TME) comprises a complex interplay of immune cells, cancer cells, blood vessels and stromal cells (Anderson and Simon, 2020). Cytokines and chemokines are important in cancer biology as they can modulate the TME through the recruitment and activation of immune cells, resulting in either pro-tumorigenic or anti-tumour effects (Kondoh and Mizuno-Kamiya, 2022). Cytokines and chemokines are also important in the response to bacterial colonisation and infection. The pro-inflammatory cytokine and chemokine response functions to heighten the defence against offending agents and is often mounted by immune cells such as activated macrophages and neutrophils (Turner et al., 2014). On the other hand, the anti-inflammatory response acts to regulate and repress the activity of pro-inflammatory cytokines (Chen et al., 2018). Malignant transformation of cells has been associated with chronic inflammation, and cancer cells can hijack inflammatory responses to drive tumour growth and proliferation (Zhao et al., 2021a). During acute inflammation, immune cells such

as dendritic cells recognise tumour antigens and produce pro-inflammatory cytokines such as IL-1 and INFγ, which prime cytotoxic natural killer cells and T cells to attack and suppress the tumour (Figure 4.1). Furthermore, activated mature dendritic cells can also activate B-cells, leading to heightened anti-tumour responses, which include phagocytosis and CD4+ and CD8 + T cells mediated tumour inhibition. However, if the inflammatory state continues and chronic inflammation ensues, tumours tip the balance from tumour suppression to tumour promotion. Tumour cells can produce immunomodulatory cytokines such as IL6, IL10, and VEGF. These cytokines recruit immunoregulatory cells, such as regulatory T and B cells, which dampen the immune response against the tumours and lead to the upregulation of growth factors that promote tumour growth (Zhao et al., 2021a).

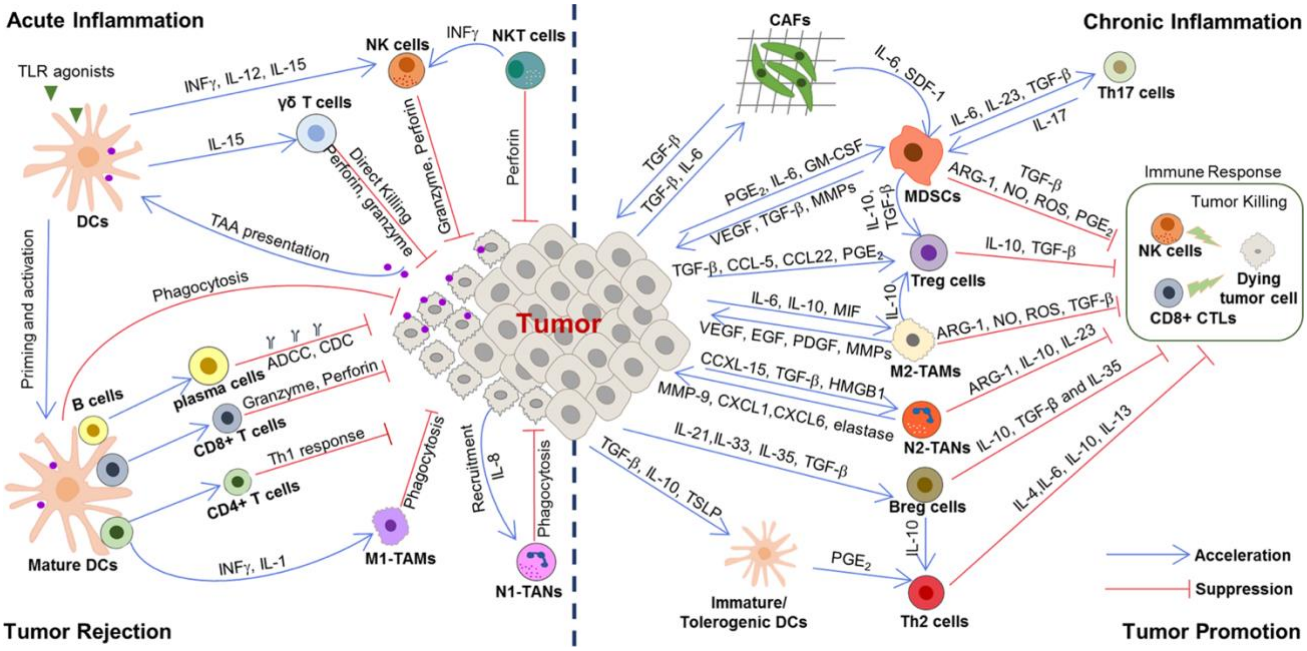


Figure 4.1: Role of inflammatory mediators in the TME. *Acute inflammation following tumour antigen recognition drives anti-tumour responses driven by dendritic cells activation of cytotoxic immune cells (CD4+, CD8+ T cells). Dendritic cells also activate natural killer cells, which produce toxic mediators such as perforin. Persistent inflammation leads to dysregulation in the anti-tumour response, characterised by the recruitment of regulatory T and B cells that repress anti-tumour immune responses. Figure reproduced from* (Zhao et al., 2021a). Copyright © 2021.

The association of CRC with bacteria means that it is imperative to understand how the CRC TME changes due to increased colonisation by certain bacteria. For example, *F. nucleatum* invasion of HCT116 cells, a colon cancer cell line, promoted the secretion of metastasis-associated cytokines IL8 and CXCL1. Induction of these cytokines increased tumour cell migration and immune cell recruitment in a *Fap2*-dependent manner (Casasanta et al., 2020). In this chapter, the Caco-2 cell response to infection by *F. nucleatum* is evaluated, and *Sgg* is used as a control. Additionally, the impact of Tspan6 expression and bacterial growth phase on the release of cytokines and chemokines was explored.

4.1.1 Hypothesis

Adhesion and invasion levels are elevated at exponential rather than stationary phase due to increased expression of adhesins. Higher levels of adhesion at the exponential phase result in an elevated cytokine and chemokine response by Caco-2 cells.

4.1.2 Aims and Objectives

1. To determine whether bacterial growth phase impacts the rate of adhesion and invasion of subspecies to CRC cells by performing adhesion/invasion assays with stationary and exponential phase bacteria.
2. To investigate whether there are differences in the impact of Tspan6 on the adhesion and invasion of *F. nucleatum* subspecies at different growth phases.
3. To evaluate the cytokine induction profile of *F. nucleatum* and *Sgg* (control) upon infection of Caco-2 cells and how this is affected by growth phase and cell type using the Olink target 48 cytokine panel

4.2 Impact of growth phase on the adhesion of the subspecies to Caco-2 cells

The impact of the growth phase on the adhesion of the three subspecies (FNA, FNN23 and FNP) was evaluated. FNN23 and FNA in the exponential phase adhered more to Caco-2 cells than those in the stationary phase (standardised to 100%) (Figure 4.2). A five-fold increase in adherence to Caco-2 cells was observed for FNN23, while FNA saw a 2.3-fold increase in adherence. A one-sample t-test with a theoretical mean set at 100 (for the mean adhesion at stationary phase) showed a statistically significant difference in the FNN23 exponential relative to stationary phase adhesion (actual mean = 526.8, two-tailed p value= 0.0114). Even though FNA also showed a higher actual mean (231.8) than the theoretical mean, the difference in adhesion for the two growth phases was insignificant, as determined by a one-sample t-test (two-tailed p-value = 0.2406). Conversely, FNP showed a slight but insignificant decrease in adhesion by exponential compared to stationary phase bacteria (mean =89.19, p=0.5908). These results suggest that FNN23 is the most amenable to growth phase changes, followed by FNA, whereas FNP is largely unaffected. One-way ANOVA was used to determine if the inter-species difference in the rate of adherence of exponential phase bacteria relative to adherence of stationary phase bacteria was significantly different between subspecies. A significant difference was found only for FNN23 vs. FNP (adjusted p value= 0.0178).

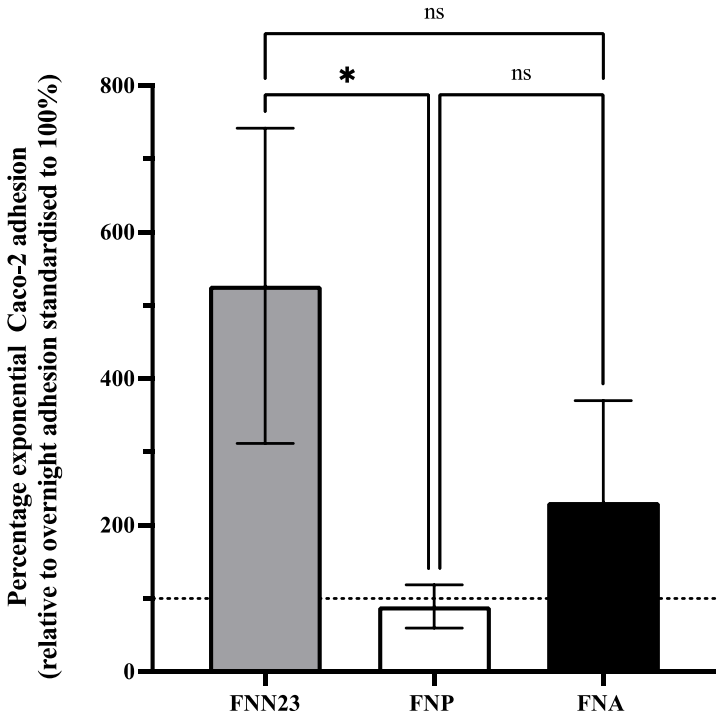


Figure 4.2: Adherence to caco-2 cells of the subspecies (FNN23, FNP and FNA at exponential phase vs adherence at stationary phase. *Y-axis show the percentage adherence of exponential samples relative to overnight samples. FNN23 shows higher percentage adherence at exponential relative to the stationary phase followed by FNA. FNN23 exponential adherence compared to stationary adherence significantly differs from FNP's (adjusted p value= 0.0178). Bars represent the mean exponential adherence relative to stationary adherence (standardised to 100 %), with error bars as standard deviation from the mean. Experiments were repeated three times with three technical replicates for each experiment. Asterisk (*) means p<0.05, ** p<0.005, ns = not significant at alpha=0.05.*

4.3 Impact of growth phase on the adhesion of the subspecies to Tspan6 expressing Caco-2 cells.

Next, the impact of Tspan6 expression on adherence at the two growth phases was also evaluated (Figure 4.3). Akin to interactions with control cells, FNN23 and FNA at the exponential phase showed a marked improvement in adhesion to T6 cells compared to the stationary phase cultures. Adherence was 3.5-fold higher for FNN23 with exponential than with

stationary phase cultures. Consistent with observations in control cells, exponential phase FNA showed improved adherence to T6 cells with a nearly a 2-fold (1.81) improvement in adhesion compared to stationary phase. Adherence to T6 cells for FNP was not affected by the bacterial growth phase, with only a slight 10% improvement in adhesion by exponential compared to stationary phase bacteria, which was not statistically significant. One-way ANOVA with Tukey's multiple comparisons was used to compare the differences in adherence to T6 cells of exponential phase relative to stationary phase cultures for the three subspecies. FNN23 exponential phase adherence was significantly higher than FNP and FNA, with p-values of 0.0074 and 0.0176 respectively.

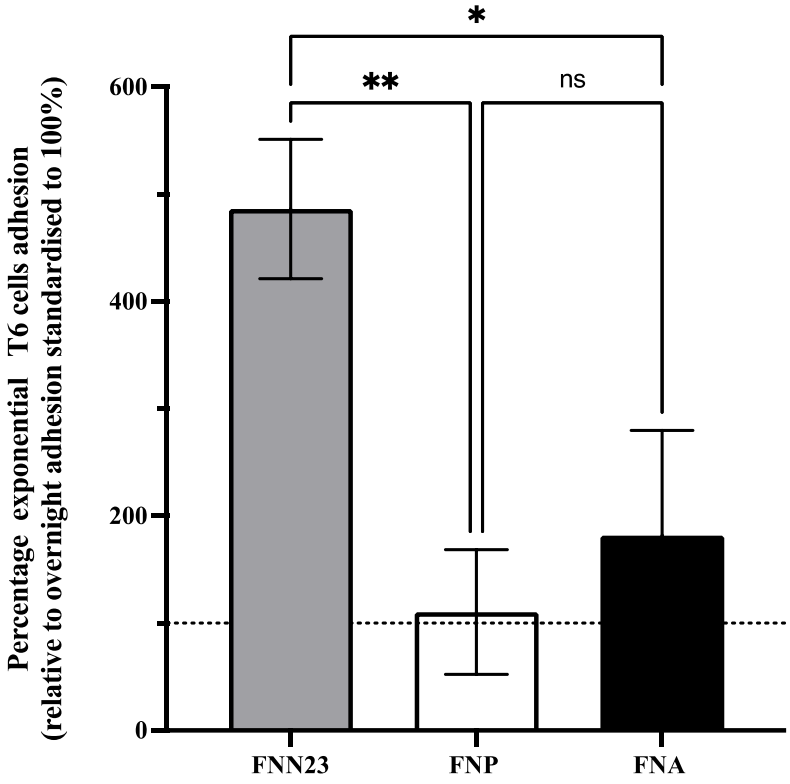
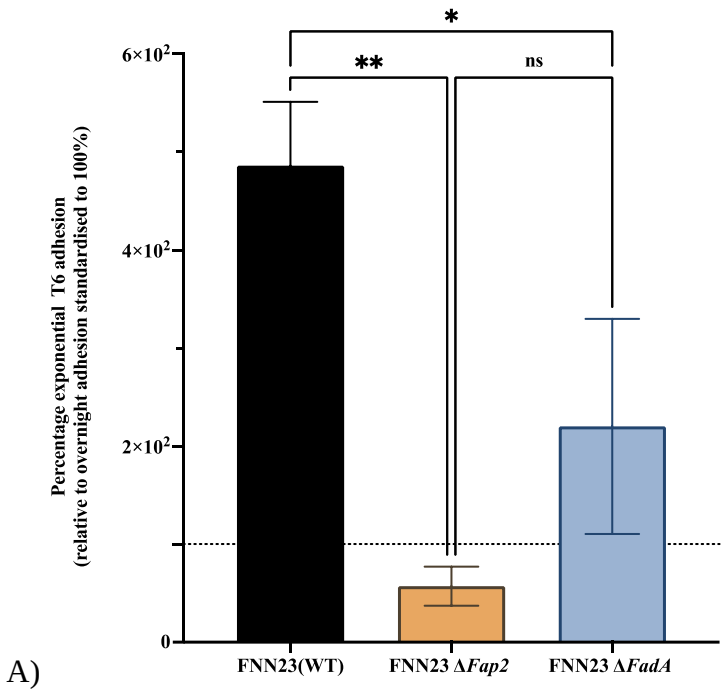


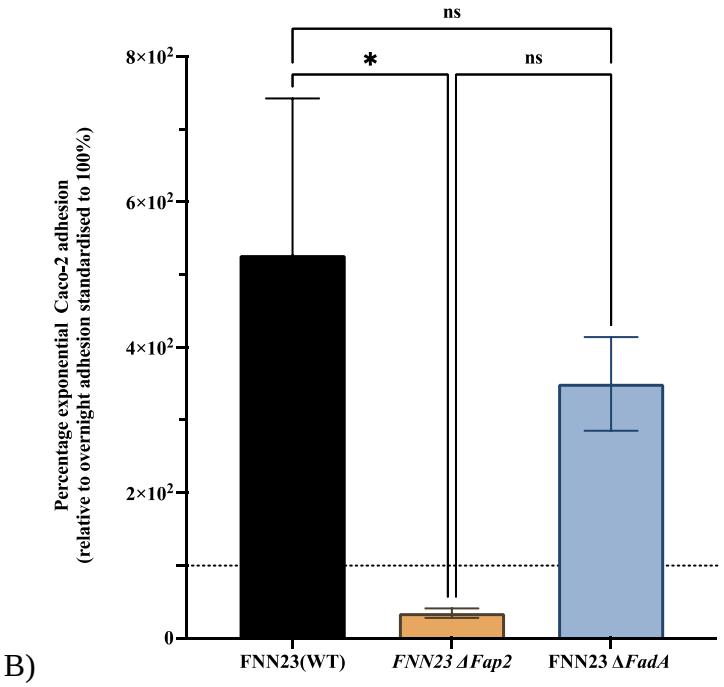
***Figure 4.3*: Exponential vs. stationary phase adherence of FNN23, FNP and FNA to T6 cells**. *Y-axis show percentage adherence of exponential cultures to T6 cells relative to adherence of stationary/overnight cultures. X-axis show the subspecies names Bars represent the mean exponential adherence relative to stationary adherence (standardised to 100 %), with error bars as standard deviation from the mean. Experiments were repeated three times with three technical replicates for each experiment. Asterisk (*) means p<0.05, ** p<0.005, ns = not significant at alpha=0.05.*

4.4 The adhesins Fap2 and FadA are important for the adhesion of FNN23 to Caco-2 cells at the exponential phase

FNN23 adhesion to Caco-2 cells was dramatically improved when exponential phase bacteria were used. This could be due to an upregulation in FNN23 adhesins during exponential phase. To investigate the possible contributions of two *F. nucleatum* adhesins (FadA and Fap2) to the observed differences, FNN23 mutant strains in these adhesins were used. Results for adhesion assays with T6 cells (A) and control cells (B) are shown (Figure 4.4). Stationary phase adhesion was set to 100 % and mean relative adherence by exponential phase cultures is shown (Figure 4.4). The wild-type strain has expected higher levels of exponential vs. stationary phase adherence than both adhesin mutant strains. One-way ANOVA with Tukey's multiple comparisons revealed significant differences ($p = 0.0071$ and $p = 0.0219$) in exponential vs. stationary adhesion to T6 cells between the wild-type, Fap2 and FadA mutant strains, respectively. Adhesion levels for the wild-type strain are reported above (section 4.2). Adherence to T6 cells by the Fap2 mutant strain at the exponential phase was nearly two-fold (1.73) lower than at the stationary phase. An even greater reduction (2.89-fold) in adherence of Fap2 mutant by exponential relative to stationary phase bacteria is observed for control cells (Figure 4.4B). Contrarily, the FadA mutant's adherence was improved with exponential compared to stationary phase cultures for both T6 and control cells. Improvement was more pronounced for adherence to control than for T6 cells (3.49 vs. 2.2-fold differences).



A)



B)

Figure 4.4: Exponential vs. stationary phase adherence of FNN23Δfap2 and FNN23 ΔfadA compared to WT FNN23*. Adherence of the Fap2 mutant is decreased at the exponential phase compared to the stationary phase, while that of the FadA mutant is higher at the exponential phase for both T6(A) and control cells(B). Experiments were repeated three times with three technical replicates for each experiment. Bars represent the mean exponential adherence relative to stationary adherence (standardised to 100), with error bars as standard deviation from the mean. Experiments were repeated three times with three technical replicates for each experiment. Asterisk (*) means p<0.05, ** p<0.005, ns = not significant at alpha=0.05.*

4.5 Impact of growth phase on the invasion of the subspecies to Caco-2 cells, including T6 expressing Caco-2 cells

The impact of the bacterial growth phase on the invasion of the subspecies was also determined. Invasion by stationary phase bacteria is standardised to 100%, and the relative invasion by exponential phase bacteria is shown (Figure 4.5). Overall, all the strains except FNN23 Δ*fap2* saw an improvement in the levels of invasion by exponential phase samples compared to the stationary phase samples. The decrease in invasion of FNN23 Δ*fap2* was significant (T= 24.33, two-tailed P value= 0.0262). However, the greatest increase in invasion, which was in FNN23 (6-fold increase) was not statistically significant (two-tailed p-value =0.4847). Furthermore, FNN23 had an R^2 value of 0.524, which suggests that only about half of the observed effect can be attributed to the tested variable; therefore, this result should be interpreted cautiously. FNN23 exponential phase invasion was followed closely by that of FNN23 Δ*fadA*, which had a 4-fold increase in invasion with exponential compared to stationary phase samples (R^2 value = 0.8264). FNA and FNP had over 2-fold improvement in invasion with exponential samples and R^2 values of 0.9381 and 0.973, respectively. The increase in invasion was not significant for the other subspecies either. However, the higher R^2 values indicate that a substantial amount of the observed outcome can be attributed to the tested variable (exponential vs. stationary).

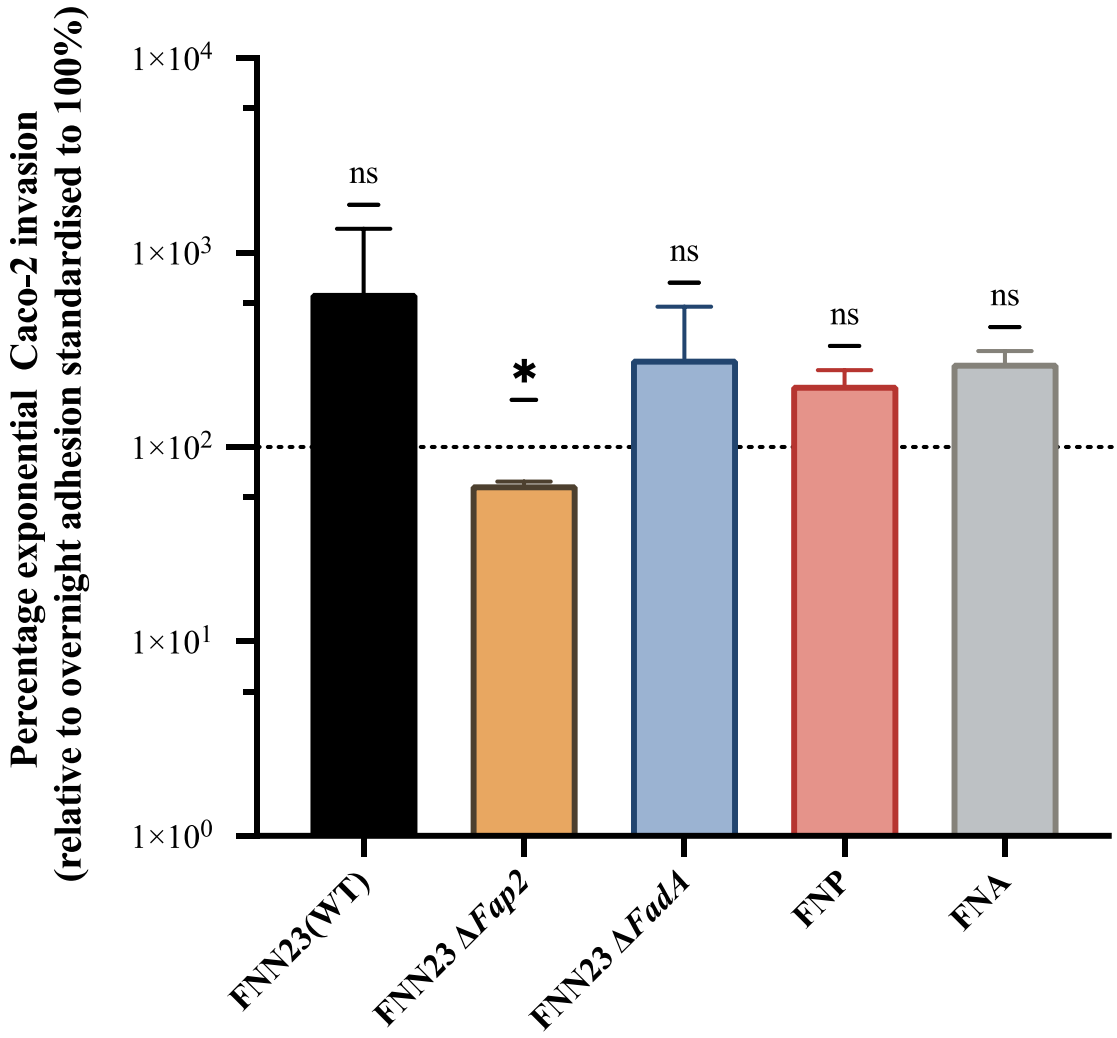


Figure 4.5: Exponential vs. stationary phase invasion of F. nucleatum subspecies and FNN23 mutants in FadA and Fap2 to Caco-2 cells. *Strain names are shown on the x-axis and they y-axis shows the percentage invasion to caco-2 control cells by exponential cultures relative to overnight/stationary cultures for all strains. One sample t-test (theoretical mean=100) was performed. Fap2 mutant strain showed a statistically significant decrease in invasion with exponential phase cultures (p=0.0262). Experiments were repeated three times with three technical replicates for each experiment. asterisk (*) means p<0.05, ** p<0.005, ns = not significant at alpha=0.05.*

In the evaluation of the impact of the growth phase on the invasion of Tspan6-expressing cells, invasion by exponential relative to stationary phase cultures followed a similar trend to that of invasion to control cells (Figure 4.6). FNN23, FNA, FNP and FNN23 Δ*fadA* all had over 2-fold increases in invasion by exponential phase cultures, while FNN23 Δ*fap2* exponential invasion to T6 cells was decreased. The difference in invasion was only significant for FNP (t=20, p-value = 0.0314). FNA had a large R^2 value (0.9654), while the FNN23 strains had R^2 values of under 0.5.

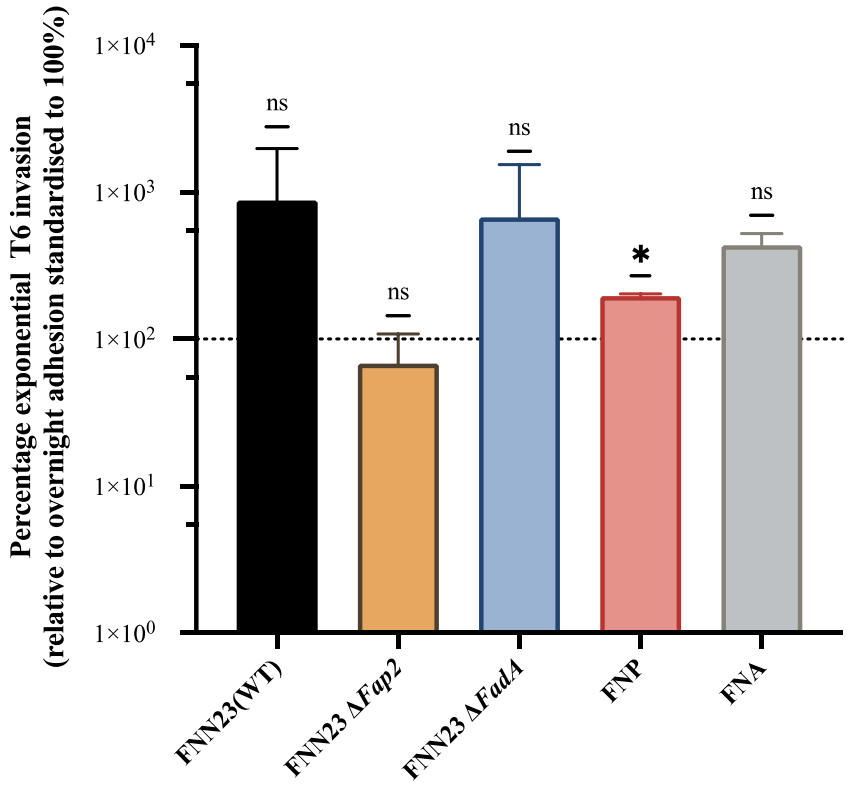


Figure 4.6: Exponential vs. stationary phase invasion of F. nucleatum subspecies and FNN23 adhesin mutants to T6 cells. *Strain names are shown on the x-axis and they y-axis shows the percentage invasion to T6 by exponential cultures relative to overnight/stationary cultures for all strains. Increased invasion with exponential phase was observed for FNN23, FNA, FNP and FNN23 ΔfadA. One sample t-test (theoretical mean=100) only showed a statistically significant difference for FNP (P=0.0314). FNN23 Δfap2 showed a decrease in invasion with exponential phase cultures, but the decrease was not statistically significant. Bars represent average exponential invasion, and error bars represent standard deviation from the mean. Experiments were repeated at least three times, with three technical replicates for each experiment. asterisk (*) means p<0.05, ns = not significant at alpha=0.05.*

Taken together, invasion of the subspecies to T6 and control cells show a trend toward increased invasion by exponential phase cultures relative to stationary phase ones. Even though invasion was largely elevated at the exponential phase relative to the stationary phase (> 2-fold changes), the differences were not statistically significant. Additionally, for some of the samples, especially FNN23 and FNN23 Δ*fadA*, variability spanning multiple orders of magnitude was observed in the experimental replicates. Therefore, some outliers were removed from the experimental replicates, resulting in small sample sizes (n=2) for most comparisons. The small sample size impacts the statistical power of the analysis and, as a result, makes it challenging to make any meaningful conclusions.

4.6 Key findings

1. The adherence of FNN23 and FNA to T6 cells and control cells was markedly improved when exponential phase samples were used, whereas that of FNP remained nearly unaffected.
2. The exponential growth phase leads to improved FNN23 Δ*fadA* and decreased adherence for FNN23 Δ*fap2* to T6 and control cells compared to the stationary growth phase.
3. Invasion of FNN23, FNA, FNP and FNN23 Δ*fadA* was improved at the exponential phase for both T6 and to control cells. Only FNN23 Δ*fap2* showed a decrease in invasion.

4.7 Evaluating the impact of *F. nucleatum* on cytokine expression using the O-LINK target 48 panel

To explore the host-pathogen interactions between *F. nucleatum* subspecies and Caco-2 cells, the cells were infected with *F. nucleatum* and *Sgg* at different growth phases and evaluated for the induction of inflammatory mediators. The sample IDs for the biological replicates used for the experiment and their corresponding strain and growth labels are shown in Table 4.1.

Table 4.1:Biological sample IDs as used with O-link and their corresponding strain name and growth phase

Biological replicates ID	Cell type	Strain	Growth phase
FNAT6_ON1, FNAT6_ON2	T6	FNA	Stationary
FNA_ON1, FNA_ON2	Control	FNA	Stationary
FNAT6_EXP1, FNAT6_EXP2	T6	FNA	Exponential
FNA_EXP1, FNA_EXP2	Control	FNA	Exponential
FNPT6_ON1, FNPT6_ON2	T6	FNP	Stationary
FNP_ON1, FNP_ON2	Control	FNP	Stationary
FNPT6_EXP1, FNPT6_EXP2	T6	FNP	Exponential
FNP_EXP1, FNP_EXP2	Control	FNP	Exponential
FNN23T6_ON1, FNN23T6_ON2	T6	FNN23	Stationary
FNN23_ON1, FNN23_ON2	Control	FNN23	Stationary
FNN23T6_EXP1, FNN23T6_EXP2	T6	FNN23	Exponential
FNN23_EXP1, FNN23_EXP2	Control	FNN23	Exponential
FadA2, FadA1	Control	FNN23 Δ*FadA*	Stationary
S. GALT6_ON1, S. GALT6_ON2	T6	*Sgg*	Stationary
S.GAL_ON1, S.GAL_ON2	Control	*Sgg*	Stationary
S. GALT6_EXP1, S. GALT6_EXP2	T6	*Sgg*	Exponential
S.GAL_EXP1, S.GAL_EXP2	Control	*Sgg*	Exponential

*S.GAL refers to *Sgg*

The sample distribution based on normalised protein expression (NPX) data for the tested cytokines is shown (Figure 4.7). NPX data is generated after correcting for the sample controls in the experiment. The sample distribution plot gives an overview of the mean and the interquartile range per sample. For this experiment, media and cell-only controls had values below the detection limit and were not included in the analysis. Generally, for all the samples, the median and the interquartile range of cytokine secretion lie between 0 and 5. For each sample, there are at least two cytokines with NPX values that lie above the limit of the upper whisker of the boxplot. It is worth noting that NPX values are not directly comparable between the cytokines due to calibration differences for each cytokine. NPX values for a single cytokine can still be compared such that a high NPX value would correspond to a greater protein concentration.

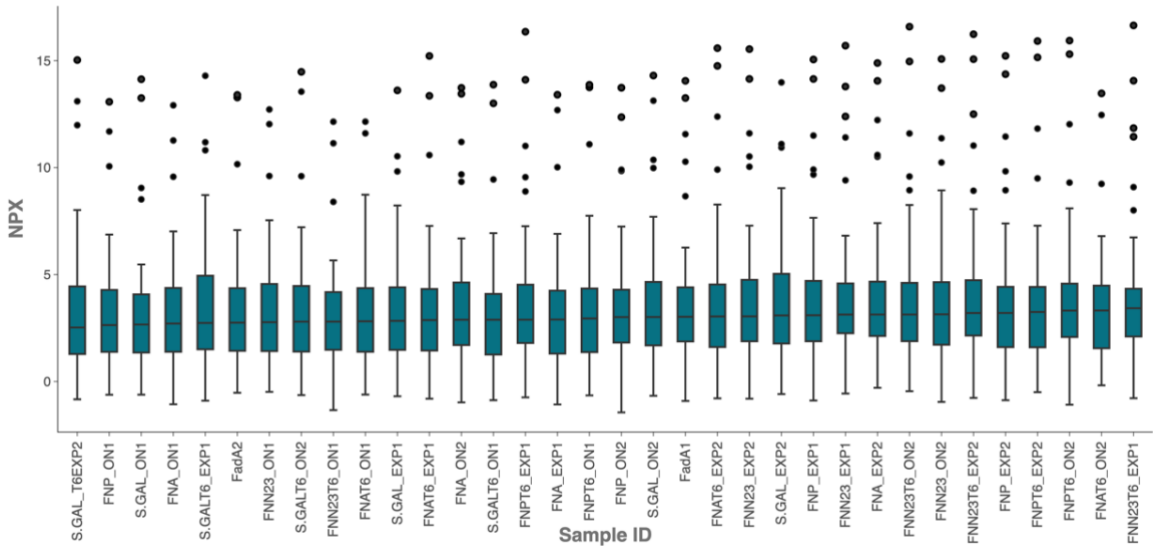


Figure 4.7: Sample distribution plot showing NPX (y-axis) and sample IDs (x-axis). *Sample plot shows box plots of the NPX values across all the tested cytokines, allowing visualisation of the median and the interquartile range of each sample. The box plots shows the interquartile range accounting for the central 50% of the log CPM values and the whiskers represent the minimum and maximum expression values.Sample names are shown on the x-axis and NPX calues are shown on the y-axis.*

4.8 General cytokine expression trends on the O-link target 48 cytokine panel

Next, the expression of the cytokines across the samples was visualised using heatmaps (Figure 4.8). Overall, when comparing the *F. nucleatum* subspecies to the control *Sgg* samples, *Sgg* appears to be associated more with patterns of dampened cytokine response. Five out of eight FNP and *Sgg* samples had patterns of downregulation of the following cytokines: IL2, IL1B, CCL13, MMP12, OSM, IL27, IL17A, CCL19, IL13, HGF, IFNG, CXCL12, TSLP, IL17C, IL10. These mediators have diverse functions, majority of which are pro-inflammatory e.g IL2, IL27, IL1B, IL17A and IL17C. Others such as CCL13 and CCL19 which both have functions in immune cell recruitment have been reported to act in concert to promoting resistance to apoptosis in some B cell leukaemia cells (Chunsong et al., 2006). In contrast, all but one of the FNN23 samples had an upregulation in the same cytokines. There is also a pattern of upregulation for the majority of the FNN23 and FNP strains for the cytokines VEGFA, CSF3, CXCL8, CSF1, OLR1, CXCL11, CXCL9, CXCL10, IL17C, MMP1, and CCL7. Most samples did not produce detectable levels of IL4, as indicated by gaps on the heatmap. Generally, most exponential phase samples appeared to have a higher cytokine expression profile than stationary samples.

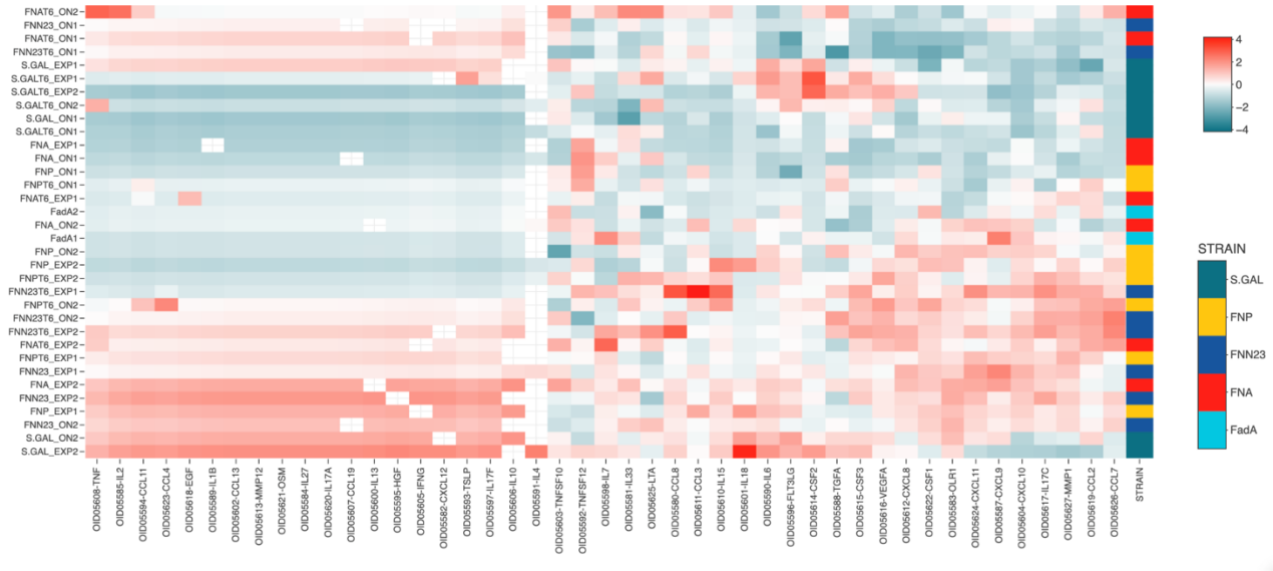


Figure 4.8: Heatmap showing the cytokine secretion patterns of 45 cytokines from the target 48 O-link cytokine panel*. Sample IDs are shown on the y-axis and cytokine name on the x-axis. The expression ranges from -4 to 4 as shown in the legend going from low to high expression.Samples with no expression of the cytokine have white bars. Each bar shows the levels of expression of a particular cytokine for a corresponding strain. Strain ID are colour coded on the legend.*

4.9 Top cytokines all strains

A one-way ANOVA was performed to evaluate the difference in cytokine expression for the strains. 9 out of 45 cytokines had a significant raw p-value at alpha =0.05 (Table 4.2) Adjusted p-values were also calculated to reduce the probability of false positives. Only CXCL10 (C-X-C motif chemokine 10) and IL17C (Interleukin 17C) were below the adjusted p-value threshold. The raw p-values were also considered because the power of the dataset was reduced due to there being 2 biological replicates per sample. So, based on the p-values the differences in CCL7, CCL8, CSF2, CXCL9, FLT3LG, CSF3 and MMP1 were considered as they may have biologically significance. CSF2 and CSF3 were more highly responsive to *Sgg* exponential culture infection than the other samples. CXCL9, MMP1, and CCL7 had an elevated expression level in *F. nucleatum* subspecies samples compared to *Sgg* samples. On the other hand, FLT3LG and CSF3 were elevated in *Sgg* samples and some exponential phase samples of the *F. nucleatum* subspecies.

Table 4.1 :Top nine significantly expressed cytokines

Assay	UniProt ID	sumsq	meansq	statistic	p value	Adjusted p value
CXCL10	P02778	227.52	56.88	17.03	2.74E-07	1.23E-05
IL17C	Q9P0M4	5.92	1.48	6.04	0.00116	0.0262
CCL7	P80098	7.66	1.91	2.95	0.0367	0.236
CCL8	P80075	0.63	0.16	3.07	0.0316	0.236
CSF2	P04141	3.89	0.97	3.51	0.0186	0.236
CXCL9	Q07325	4.26	1.06	3.33	0.0232	0.236
FLT3LG	P49771	9.32	2.33	3.05	0.0325	0.236
CSF3	P09919	9.32	2.33	2.71	0.0497	0.249
MMP1	P03956	30.62	7.66	2.79	0.0447	0.249

A Tukey's multiple comparisons test was used to compare differences between the two cytokines with the most significant differences in expression. CXCL10 expression significantly differed between *Sgg* and the subspecies FNN23, FNP and FNA with adjusted p-values of 4.29E-07, 3.05E-06 and 2.76E-05 respectively (Figure 4.9 top panel). There was also a significant difference between *Sgg* and the FadA adhesion mutant (adjusted p-value =0.00297). Significant differences in IL17C expression for *Sgg* compared to FNN23 (adjusted p-value = 0.000663) and to FNP (adjusted p-value = 0.0180) were also observed (Figure 4.9 bottom panel). Of the *F. nucleatum* subspecies, FNA had the lowest secretion of IL17C, though the differences were not statistically significant.

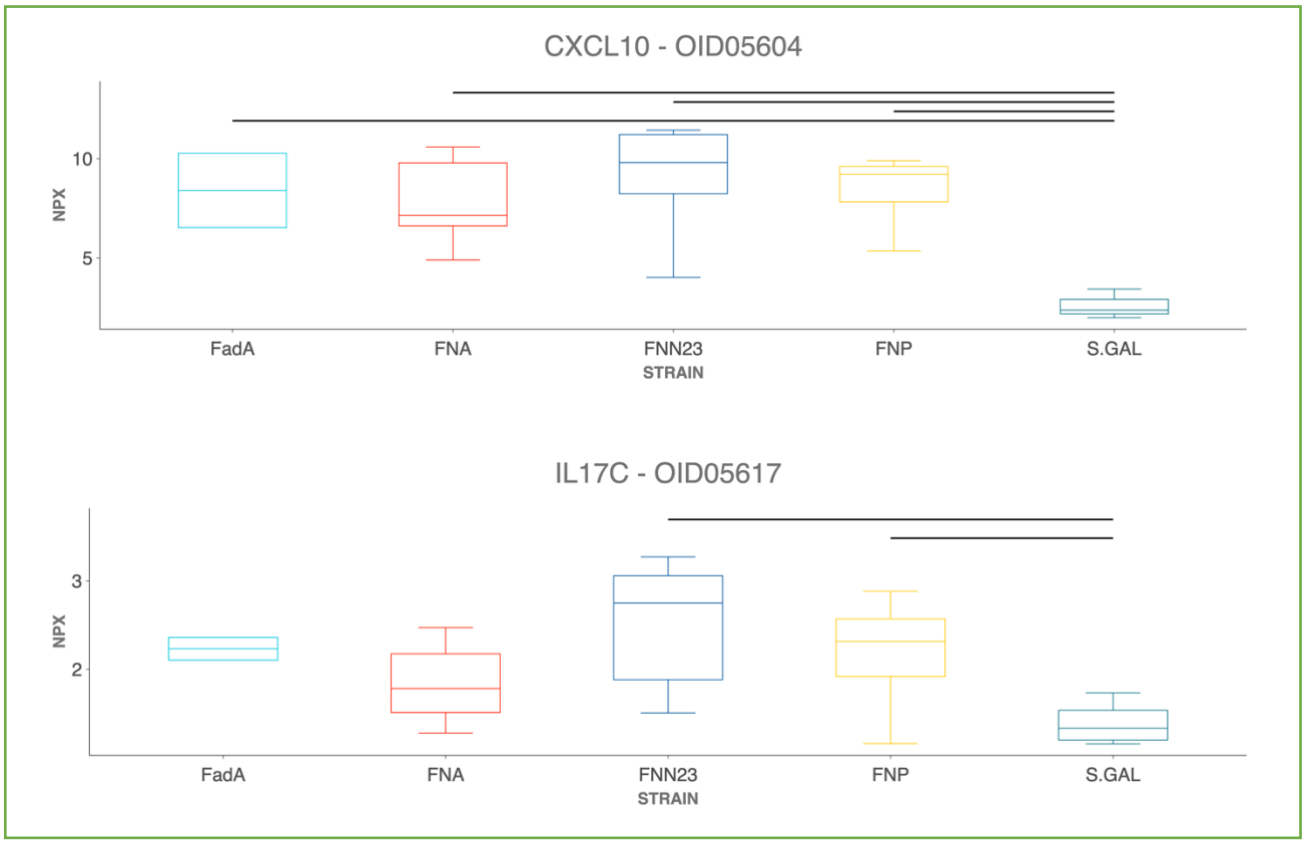


Figure 4.9: Boxplots of the comparisons between the bacteria for CXCL10 and IL17C expression. *The top panel shows a significant difference between Sgg (S.GAL) vs F. nucleatum subspecies expression of CXCL10. The horizontal line drawn on top of the box plots represents significance difference between any two points that it connects. The NPX values are shown on the y-axis and the strain names are shown on the x-axis. The bottom panel shows a comparison of the expression of IL17C for F. nucleatum strains and Sgg. The horizontal line drawn on top of the box plots represents significance difference between any two points that it connect*

4.10 Impact of growth phase on cytokine expression

As shown earlier, there were growth phase-dependent differences in the adherence and the invasion of Caco-2 cells by *F. nucleatum* subspecies. The impact on cytokine expression of the growth phase was also evaluated. A t-test comparing cytokine secretion of all stationary samples compared to all the exponential phase samples was performed, and six cytokines out of 45 had significant p-values (alpha <0.05). The cytokines with significant p-values were IL6, CXCL8, CSF3, FLT3LG, VEGFA and CSF2 (Table 4.5). Adjusted p-values were also calculated and only IL6 had a significant adjusted p-value (0.0284).

Table 4.2: Cytokines differentially expressed for exponential vs. stationary phase cultures

Assay	Estimate	EXP	O. N	statistic	P. value	conf.low	conf.high	Adjusted P value
IL6	1.54	6.96	5.42	3.80	0.000632	0.711	2.36	0.0284
CXCL8	1.13	11.4	10.3	3.23	0.00298	0.415	1.84	0.0671
CSF3	0.932	1.42	0.487	2.96	0.00588	0.289	1.58	0.0882
FLT3LG	0.843	4.27	3.42	2.86	0.00851	0.235	1.45	0.0958
VEGFA	0.6103	14.2	13.6	2.38	0.0238	0.0867	1.13	0.2141
CSF2	0.422	4.53	4.11	2.06	0.0528	-0.0057	0.849	0.350

A PCA plot based on the six cytokines was generated to visualise how they influence sample grouping. Samples on PC1, which captures 64.75 % of the variation, are clustered based on the growth phase with stationary phase samples (blue dots) on the positive side of the plot and exponential samples (red dots) on the negative scale of the plot (Figure 4.10 top panel). This

supports the observation that these cytokines are differentially expressed at different growth phases. Interestingly, some stationary samples from experimental replicate 2 clustered with the exponential phase, while two FNA exponential samples clustered with the stationary samples. Overall, the separation of samples based on the growth phase for the six cytokines is demonstrated, showing that the growth phase is potentially a significant factor in the expression of the cytokines. This is also clearly apparent in the accompanying heatmap (Figure 4.10 bottom panel) that shows the pattens of expression of the bacteria based on the six cytokines. Most cytokines show upregulation for exponential samples with some strain-specific differences. A striking upregulation of the CSF2 cytokine is noted for *Sgg* exponential phase samples on T6 cells.

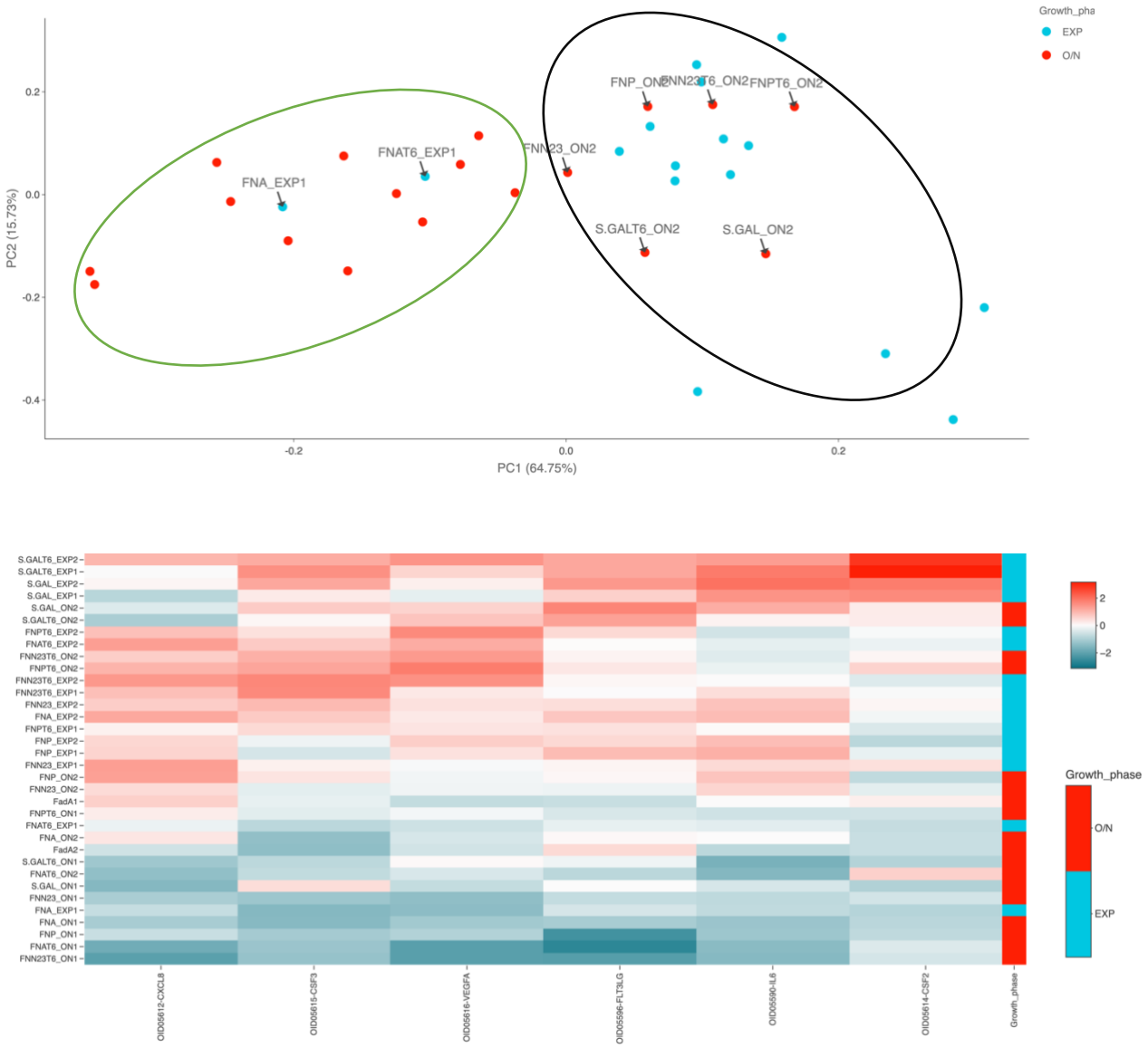


Figure 4.10: PCA plot (top panel) and heatmap (bottom panel) of six significantly differentially expressed cytokines based on the growth phase. *On the top panel most exponential phase samples are in the green circle, while most stationary samples are in the black circle .Exponential phase samples are indicated in blue dots while stationary pahse samples (o/n) are indicated in red dots. The x-axis shows the PC1 while the y-axis shows how the data is distributed across PC2. On the bottom panel exponential phase samples are indicated by a blue block for the growth phase whereas stationary phase (O/N) are indicated with a red block. The names of the strains are shown on the y-axis while the names of the cytokines that are significantly different based on growth phase are shown on the x-axis.*

When comparing cytokine expression based on the growth phase, interleukin 6 was the most significantly differentially expressed cytokine. Secretion was higher for exponential phase cultures than stationary cultures (adjusted p value=0.0284) (Figure 4.11). IL6 has important roles in CRC, such as promoting tumorigenesis, upregulating metastasis, and drives tumour stemness (Wang et al., 2019b;Lin et al., 2020)



Figure 4.11: Boxplots of the expression of IL6 for exponential phase samples compared to stationary samples*. The line indicates a significant adjusted p-value with exponential phase samples (EXP) having a significantly higher IL6 expression level than stationary phase(O/N) samples. The x-axis show the growth phase of the samples with EXP indicating exponential phase samples and O/N indicating stationeary phase samples. The x-axis show the NPX values whereas the x-axis show the growth rate of the sampels. The straight line running from the top of both plots shows statistically significant difference in the expression of 1L6 for the two growth phases tested.*

4.11 Impact of Tspan6 expression on cytokine expression

Conditioned media was obtained from infection of Caco-2 cells that expressed Tspan6 and from control cells (pLVX) to investigate whether the expression of Tspan6 impacts cytokine secretion**.** A two-sided Welch two-sample t-test of control cells vs T6 cells was performed, and four cytokines (CCL2, LTA, OLR1 and IL18) had significant p-values, albeit the adjusted p-values were not significant (Table 4.4). OLR1 and IL18 secretion was lower in T6 cells compared to control cells, whereas CLL2 and LTA were higher in T6 cells compared to the control.

Table 4.3: Cytokines with significant p values for control cells vs. T6 cells

Assay	Estimate	Control cells	T6 cells	statistic	P value	Conf low	Conf high	Adjusted pval
CCL2	-1.073	5.90	6.97	-3.21	0.00317	-1.76	-0.391	0.136
LTA	-0.301	3.12	3.43	-2.94	0.00645	-0.511	-0.0914	0.136
OLR1	1.42	8.75	7.32	2.78	0.00906	0.380	2.46	0.136
IL18	0.442	5.15	4.71	2.10	0.0488	0.00260	0.882	0.549

A PCA plot was generated to get a clear image of whether, for the four cytokines, the data shows any similarity or dissimilarity based on cell type (Figure 4.12). Principal component one accounted for (36.59%) of the variation. Clustering on this principal component appears to be based on cell type, with most of the T6 samples clustering on the negative scale of the PC1 scale while the control samples grouped on the positive scale of the principal component with

only two T6 samples grouping with the controls. Principal component 2 also captured a substantial variation (30.92%), which accounts for strain differences.

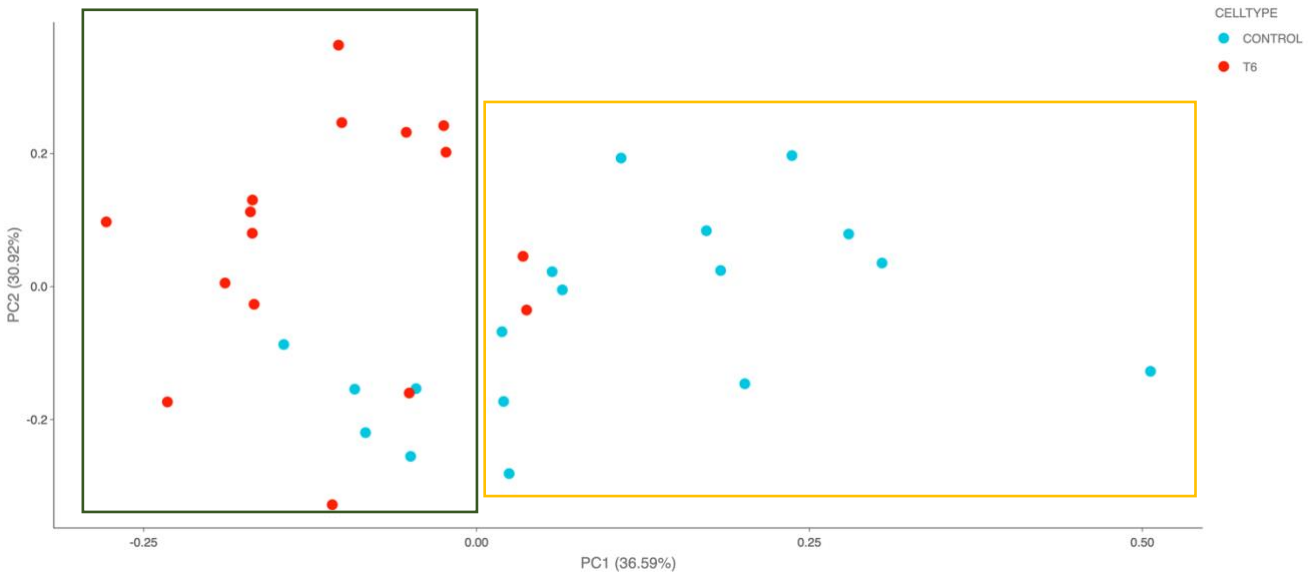


Figure 4.12: PCA plot based on the cytokines that are differentially secreted based on cell type. *Control samples are shown in blue dots and are mainly contained in the orange rectangular box while samples from T6 cells are shown in red dots and are mainly contained in the black rectangular box. The x-axis show the data variation represented by prinicipal component 1(PC1) and the y-axis shows data variation represented by principal component 2(PC2).*

A heatmap of all the samples for the four cytokines was generated to visualise the trends in the data. The heatmap (Figure 4.13) showed distinct patterns of upregulation and downregulation of cytokines dependent on the cell type. Some strain-specific profiles are also evident, e.g., most of the upregulation in IL18 being in *Sgg* and FNP exponential samples, suggesting strain-specific, growth phase dependent differences. The same strains also showed higher OLR1 secretion. The FadA mutant strain was mostly downregulated. There was evidence of biological replicate differences; one FadA mutant had an upregulation in IL18, while the other had an upregulation in CLL2 secretion.

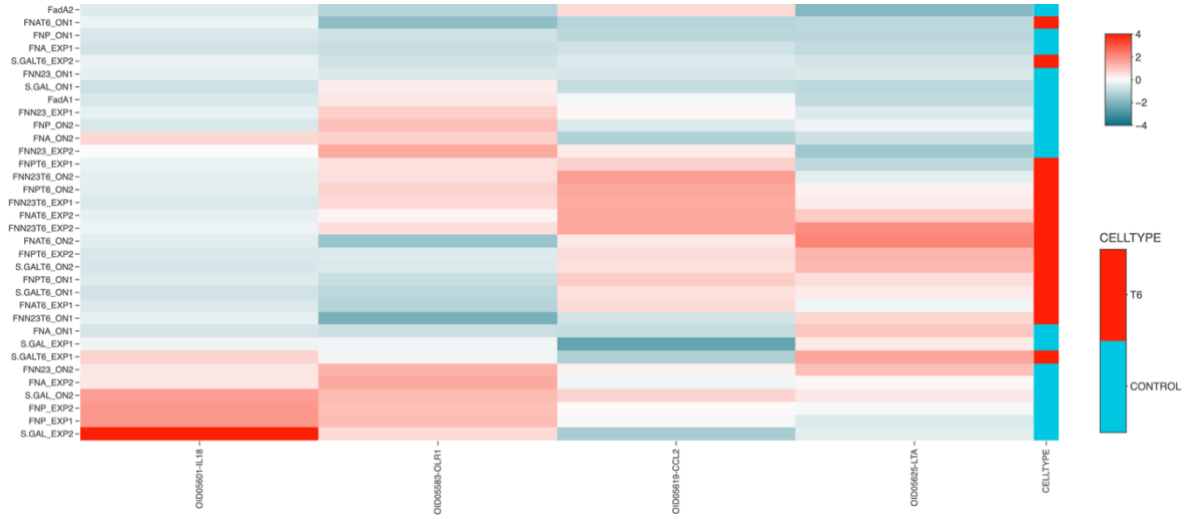


Figure 4.13: Heatmap of four significantly differentially expressed cytokines based on the cell type. T6 cells are coded red for cell type whereas control cells are coded blue. *The names of the strains are shown on the y-axis while the names of the cytokines that are significantly different based on growth phase are shown on the x-axis. The expression ranges from -4 to 4 as shown in the legend going from low to high expression. Samples with no expression of the cytokine have white bars. Each bar shows the levels of expression of a particular cytokine for a corresponding strain.*

4.12 Key findings

1. There are species-specific differences in the induction of cytokines between *F. nucleatum* subspecies and *Sgg.*

2. CXCL10 and IL17C are the only statistically significantly expressed cytokines between *Sgg* and *F. nucleatum* subspecies.

3. Most exponential phase samples were associated with cytokine overexpression compared to most stationary phase samples.

4. IL6 was the only statistically significantly expressed cytokine when comparing exponential to stationary phase samples.

5. Tspan6 expression increased the expression of CLL2 and LTA and reduced that of OLR1 and Il18 compared to control cells.

4.13 Discussion

The first part of this chapter aimed to determine whether the bacterial growth phase could impact levels of adhesion and invasion of the subspecies FNN23, FNA and FNP. FNN23 and FNA experienced higher adherence levels to T6 and control cells with exponential relative to stationary phase samples. Interestingly, FNP adherence levels to T6 and control cells remained unaffected by the growth phase. Based on these observations, it can be postulated that for FNA and FNN23, the expression of certain adhesins may be elevated at the exponential phase, resulting in improved adherence. This may also reflect the differences in the adhesin expression profiles for the *F. nucleatum* subspecies. The possibility of certain adhesins in *F. nucleatum* being differentially expressed at the exponential phase was also considered by Lima *et al*., who noticed improved *F. nucleatum* co-aggregation with *Streptococcus gordonii* at the exponential phase compared to the stationary phase (Lima et al., 2017a). To determine if the adhesins Fap2 and FadA could be playing a role in the increase in adherence of FNN23, adhesin mutant strains in these adhesins were used. At the exponential phase, FNN23 Δ*fadA* strain showed an improvement in adherence, whereas the adherence of FNN23 Δ*fap2* was reduced for both T6 and control cells. Taken together, these results suggest that Fap2 may contribute more to adherence of FNN23 at the exponential phase than FadA, as its expression (in the FadA mutant) improves adherence, and its absence (FNN23 Δ*fap2*) leads to a reduction in adherence. This suggests that at the exponential phase, the expression of Fap2 may be elevated, leading to higher adherence. The impact of the growth on the invasion of subspecies was also determined. All subspecies showed increased invasion at the exponential phase relative to the stationary phase except for the FNN23 Δ*fap2* strain, which had a statistically significant decrease in invasion at

the exponential phase. This highlights Fap2 as a key adhesin in the *F. nucleatum* adhesion and invasion to mammalian cells, as reported in the literature (Gur et al., 2015;Casasanta et al., 2020). It is worth noting that there was a lot of variation between experimental replicates for invasion. Future work will involve acquiring more experimental repeats to increase the sample size and improve the statistical power of the analysis.

The second part of the chapter evaluated the impact of infection of Caco-2 cells with *F. nucleatum* on the secretion of cytokines, chemokines, and other inflammatory mediators. When comparing the cytokine induction pattern for all the strains, two cytokines (CXCL10 and IL17C) had statistically significant differences in induction/secretion for some *F. nucleatum* subspecies compared to the control *Sgg.* All the *F. nucleatum* subspecies had relatively higher levels of CXCL10 expression compared to *Sgg*, whereas, for IL17C, only FNN23 and FNP had higher expression than *Sgg*. CXCL10 has a dual role in cancer; it binds CXCR3 and promotes tumour proliferation, migration and recurrence but can also contribute to killing tumour cells by recruiting anti-tumour immune cells (Zipin-Roitman et al., 2007;Bai et al., 2016). During infection CXCL10 is important for defence against pathogens (Herzig et al., 2014); therefore its elevation following *F. nucleatum* infection is not surprising. CXCL10, gets upregulated, during infection and recruits natural killer cells and T lymphocytes to the site of infection and acts as an effective antimicrobial (Herzig et al., 2014;Li et al., 2022). *F. nucleatum* subspecies mount a stronger CXCL10 response than *Sgg;* therefore, we hypothesise that *Sgg* may be suppressing CXCL10 to avoid immune-mediated elimination by the CXCL10 pathway. This has been observed in other bacteria, such as Chlamydia, which produces chlamydial protease activity factor (CPAF), to inhibit the activity of CXCL10 (Schott et al., 2020). IL17C, the other significantly differentially expressed cytokine, is often upregulated in CRC tumours, and this drives intestinal epithelial cell (IEC) survival, resulting in tumour promotion (Song et al., 2014).

IL17C is one of 6 structurally different IL17 cytokines that have functions in both cancer and infectious diseases (Ge et al., 2020). It can be produced by epithelial cells and is important for timely host response against epithelial cell damage (McGeachy et al., 2019). Additionally, it is involved in a signalling cascade through the Interleukin 17 receptor E (IL-17RE) that results in the production of antimicrobial peptides (Song et al., 2011;Nies and Panzer, 2020). IL17C acts through the toll-like receptor (TLR)-MyD88-dependent pathway. *F. nucleatum* also uses this pathway to promote tumorigenesis by binding to toll-like receptor 4 (TLR-4) (Yang et al., 2017). TLR-4 recognises many ligands in bacteria, but it mainly recognises LPS on Gram-negative bacteria. Due to this, it is possible that IL17C may not be interacting with *Sgg* in a similar way to *F. nucleatum* strains, hence the observed variation in IL17C secretion by the two species. Instead *Sgg* has lipoteichoic acid (LTA), which can interact with TLR-4, although with reduced affinity compared to LPS/TLR-4 interaction (Branger et al., 2004). More recently, it was demonstrated that LTA can inactivate and inhibit TLR-4 signalling, resulting in dampened inflammation (Watanabe et al., 2023); *Sgg*, may be downregulating IL17C secretion through LTA hence a lower IL17C response than *F. nucleatum* subspecies. Then, when comparing cytokine expression based on bacterial growth phase, IL6 was significantly statistically differentially secreted. This cytokine has an important functions both in CRC and in response to infection by microbes (Lin et al., 2020). Additionally, IL6, which is usually upregulated in CRC patients compared to the healthy population, is associated with reduced patient survival (Xu et al., 2016). Interestingly, pro-inflammatory and anti-inflammatory roles have been reported for IL6 (Scheller et al., 2011;Wu et al., 2016a). It had an immunomodulatory function in a *Yersinia enterocolitica* infection with mice deficient in IL6, resulting in hyperinflammation. In studies with *F. nucleatum*, higher levels of the bacteria were correlated with increased induction of IL6 and CXCL8 (Kostic et al., 2013;Proença et al., 2018). The

upregulation of IL6 for exponential phase samples suggests that they may be more immunogenic, triggering the secretion of the immunomodulatory IL6.

Other cytokines did not have significant adjusted p-values but had significant raw p-values. When evaluating the impact of strain differences in cytokine expression**,** the cytokines and chemokines that only had significant raw p-values were CCL7, CCL8, CSF2, CSF3, CXCL9, FLT3LG, and the metalloproteinase MMP1. Similarly, the cytokines and chemokines; CXCL8, CSF3, FLT3LG, VEGFA and CSF2 had insignificant adjusted p-values but significant raw p-values when evaluating the impact of growth-phase on cytokine and chemokine expression. Lastly, for the impact on cytokine secretion of T6 expression the cytokines and chemokines CCL2, LTA, OLR1 and IL18 had significant raw p values. These cytokines have many functions; some have pro-inflammatory while others have anti-inflammatory effects. CSF2 and CSF3, which showed both strain and growth phase-dependent differences, are often overexpressed in CRC tumours, have a role in immune cell production and release, and have been associated with reduced survival and chemoresistance (Morris et al., 2014;Xu et al., 2019;Saunders et al., 2021). Elevation of the matrix metalloproteinase 1(MMP1) with *F. nucleatum* is in line with previous reports that this metalloproteinase is often upregulated with *F. nucleatum* infection and is involved in tumour progression and migration in CRC (Wang et al., 2020;Salvucci et al., 2022). Fms-related tyrosine kinase 3 ligand (FLT3G), which had strain differences and was elevated for exponential phase cultured compared to stationary phase samples, functions to expand and activate dendritic cells. It helps to improve responses to chemotherapy and is also important in infection control (Guermonprez et al., 2013;Lin et al., 2018;Pol et al., 2020). Chemokine (C-C motif) ligand 7 (CCL7) and CXCL9 were elevated in

F. nucleatum compared to *Sgg*. CCL7 recruits immune cells such as eosinophils and increases cell invasion and metastasis in CRC (CHIDIMATSU et al., 2023). Additionally, others have shown that *F. nucleatum*-mediated upregulation in CCL7 can contribute to increased tumour metastasis (Li et al., 2019;Wu et al., 2023). CXCL9, on the other hand, contributes to improved survival in CRC patients by promoting the infiltration of anti-tumour CD4+ or CD8+ T cells (Wu et al., 2016b). Furthermore, CXCL9 is important in the defence against microorganisms, which has been demonstrated for bacteria such as *Citrobacter rodentium, Streptococcus pyogenes* and *Streptococcus pneumoniae* (Egesten et al., 2007;Eliasson et al., 2010;Reid-Yu et al., 2015). CXCL8 and VEGFA, both elevated at the stationary phase, have roles in angiogenesis (Shibuya, 2011;Pennel et al., 2022). Additionally, CXCL8 is strongly linked to tumour invasiveness and metastasis in CRC, while VEGFA is associated with poor prognosis and recurrence (Shen et al., 2017;Dakowicz et al., 2022). LTA, also known as tumour necrosis factor beta, was slightly elevated for T6 cells compared to controls, as was CCL2. CCL2 upregulation in this study was more evident in the *F. nucleatum* subspecies than in the *Sgg* controls. This aligns with reports of elevated CLL2 upon *F. nucleatum* infection in epithelial cells (Milward et al., 2007). Interestingly, CCL2 is a strong neoplastic agent in CRC with suppressive effects on some T cell populations. (Chun et al., 2015). This suggests that by upregulating CLL2 expression, *F. nucleatum* may contribute to the initial stages of CRC carcinogenesis. Contrarily, the other upregulated cytokine, LTA, positively correlates with better survival in several cancers (Shi et al., 2022). Under-expressed cytokines for T6 compared to control cells were OLR1 and IL18. In control cells, the IL18 upregulation was enriched in FNP and some *Sgg* exponential samples. IL18 is usually considered a tumour suppressor due to its inhibitory effect on tumour proliferation (Feng et al., 2020). IL18 is also important in bacterial defence against several pathogens, including group B *streptococcus* (Sahoo et al.,

2011). Unlike IL18, ORL1 has more cancer-promoting functions. It is important for tumour glycolytic metabolism and has been implicated in chemoresistance in colon cancer (Zhao et al., 2021b). Lectin-like oxidised LDL receptor-1 (LOX-1), the gene product for *OLR1*, is often overexpressed in aggressive and metastatic CRC (Murdocca et al., 2016).

In conclusion, this work has demonstrated that adhesion by *F. nucleatum* subspecies is impacted by the growth phase in a subspecies-specific manner. Furthermore, it shows species-specific, growth phase-dependent differences in cytokine and chemokine secretion, which are also impacted by Tspan6 expression. This demonstrates the complex interplay of the inflammatory mediators, host factors, pathogens and their characteristics in the tumour environment.

5 RNA SEQUENCING TO UNCOVER IMPACT OF THE GROWTH PHASE AND STRAIN BACKGROUND ON GENE EXPRESSION OF *F. NUCLEATUM* SUBSPECIES

5.1 Background

Adhesion and invasion experiments (Chapters 3 and 4) showed subspecies-dependent differences, which are impacted by the growth phase of bacteria. To understand pathogen dynamics that may lead to differences in adhesion and invasion, RNA sequencing was done at both exponential and stationary phases to uncover potential differences in the expression profile of genes, which could have implications for the subsequent interaction of bacteria with epithelial cells. As already established in the previous chapters, *F. nucleatum* has several adhesins that help in the invasion and colonisation of host cells (Umaña et al., 2019). We hypothesised that some differences in adhesion and invasion could be due to variations in the expression levels of adhesin proteins between the subspecies. Additionally, the impact of mutations in two major adhesins in *F. nucleatum*, FadA and Fap2, was evaluated. This allowed explorations of the impact of growth phase, strain background, and adhesin deletion on the transcriptomic landscape of *F. nucleatum*.

5.2 Aims

1. To determine the impact of the bacterial growth phase on the expression profile of *F. nucleatum subspecies* FNA, FNN23 and FNP.
2. To determine whether there are subspecies and strain-specific differences in gene expression of the bacteria.
3. To evaluate the impact of adhesion deletion on the expression profile of FNN23 mutants.

5.3 RNA sequencing quality metrics

After trimming, RNA sequencing libraries produced an average of 22,468,571 reads per sample. Reads were mapped to the reference genome (*Fusobacterium nucleatum* subspecies nucleatum 25586) (Table 5.1). An average of 18,144,539 reads per sample were mapped to the reference. Samples of the FNN23 strain and the FNN23 mutant strains mapped successfully (average of 91.7% mapped reads) to the reference. However, for FNA and FNP samples, the percentage of reads mapped to the reference was comparably lower (an average of 62.62%).

Table 5.1: RNA sequencing data quality

Sample ID	Description	Number of reads	Number of mapped reads	Reads mapped to reference (%)
FadA1EXP	FadA exponential replicate 1	20,512,888	19,056,753	92.9
FadA1O-N	FadA overnight replicate 1	23,683,112	21,471,197	90.66
FadA2EXP	FadA exponential replicate 2	18,709,038	17,266,653	92.29
FadA2O-N	FadA overnight replicate 2	21,715,276	18,777,701	86.47
Fap21O-N	Fap2 overnight replicate 1	17,823,034	16,399,304	92.01
Fap21EXP	Fap2 exponential replicate 1	17,437,914	15,672,434	89.88
Fap22EXP	Fap2 exponential replicate 2	42,037,472	38,747,699	92.17
Fap22O-N	Fap2 overnight replicate 1	23,614,046	22,434,251	95
FN231EXP	FNN23 exponential replicate 1	25,807,192	24,066,716	93.26
FN231O-N	FNN23 overnight replicate 1	25,352,232	23,274,528	91.8
FN232EXP	FNN23 exponential replicate 2	22,839,992	21,525,187	94.24
FN232O-N	FNN23 overnight replicate 2	17,574,148	15,773,304	89.75
FNA1EXP	FNA exponential replicate 1	21,582,292	14,759,292	68.39
FNA1O-N	FNA overnight replicate 1	22,691,104	12,749,863	56.19
FNA2EXP	FNA exponential replicate 2	32,995,664	21,912,270	66.41
FNA2O-N	FNA overnight replicate 2	16,818,978	9,505,262	56.52
FNP1EXP	FNP exponential replicate 1	24,934,926	15,804,939	63.38
FNP1O-N	FNP overnight replicate 1	18,285,022	10,688,080	58.45
FNP2O-N	FNP overnight replicate 2	17,934,704	11,514,499	64.2
FNP2EXP	FNP exponential replicate 2	17,022,394	11,490,852	67.5

5.4 RNA sequencing quality check

Data normalisation and statistics were done using EdgeR quasi-likelihood in Degust. The program normalised the reads into counts per million (CPM) values and performed statistical analysis to give the False discovery rate (FDR, Benjamini-Hochberg) and corresponding p-values. The expression plot (Figure 5.1) shows the gene expression patterns across all the samples. Overall, the range of expression is similar and consistent across the samples (x-axis), with the highest log CPM (y-axis) values being around 16. Most expression values fall between log CPM values of 4 and 8, with minor variations between the samples and between the biological replicates.

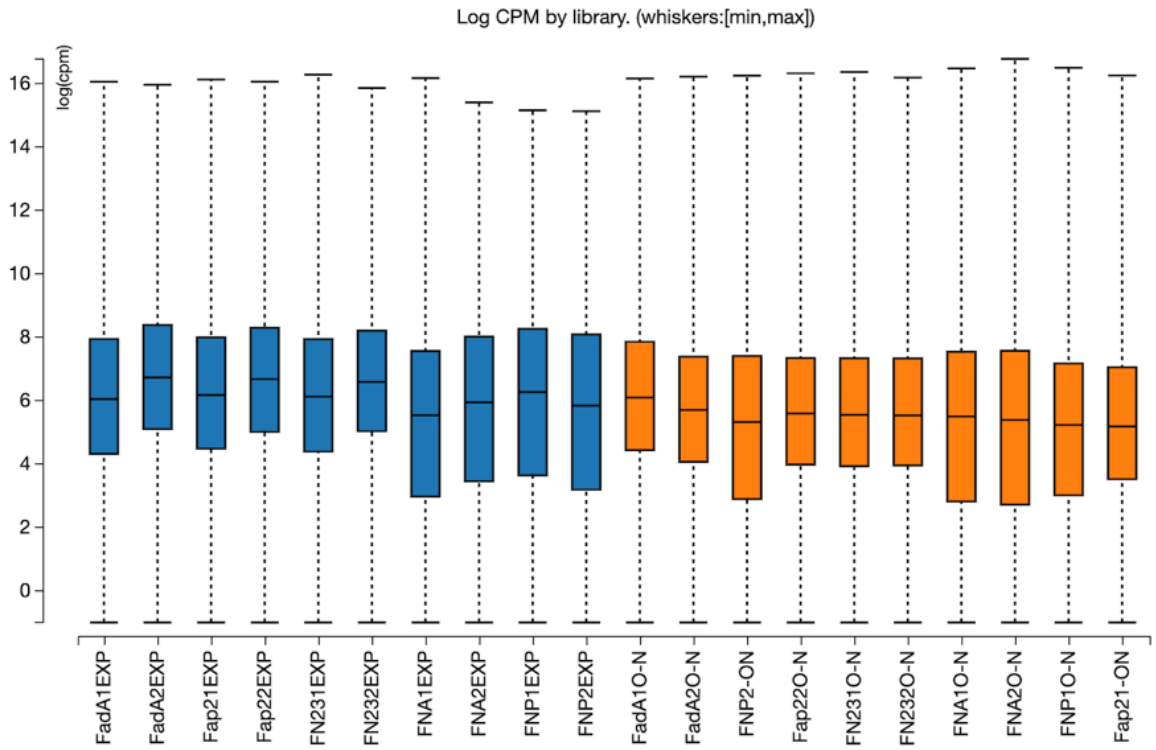
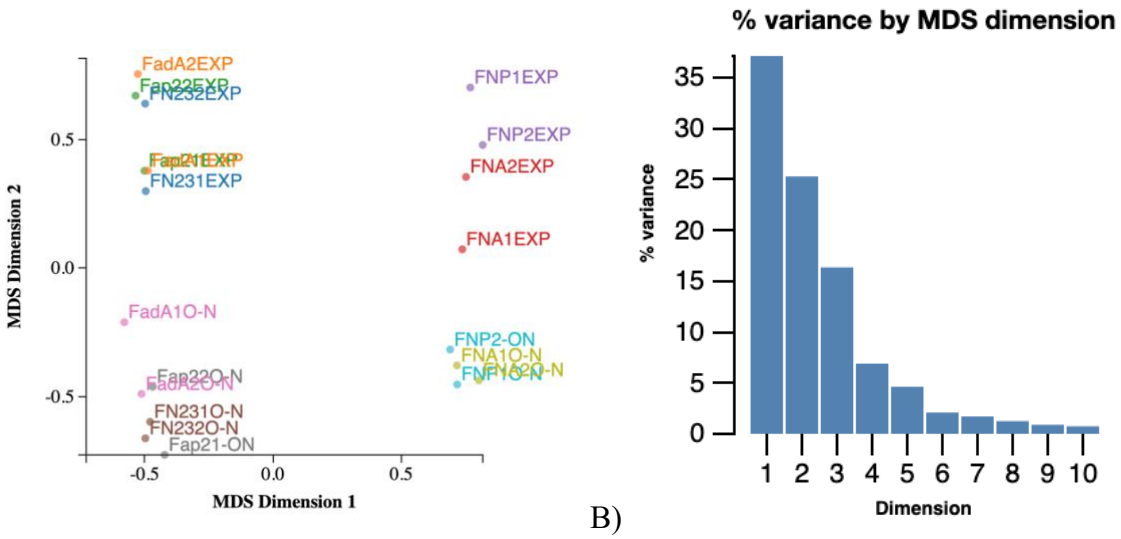


Figure 5.1:Expression plot showing the expression level(log CPM) by library. *The x-axis shows the sample names and replicate number od sample (see table 5.1) for sample description. The y-axis shows the log CPM*

values by library. The box plot shows the interquartile range accounting for the central 50% of the log CPM values and the whiskers represent the minimum and maximum expression values. Exponential samples are represented by blue coloured box plots and overnight samples are represented by orange coloured box plots.

5.5 Global gene expression patterns

To evaluate the overall impact on gene expression of the bacterial growth phase and compare transcriptome profiles of the strains, multidimensional scaling (MDS) was used (Figure 5.2). MDS is a dimensionality reduction technique that can help in the visualisation of similarities or dissimilarities between data points (Tzeng et al., 2008). An MDS plot was generated (Figure 5.3) showing data points along the MDS Dimensions 1 and 2, which show the coordinates of each data point. The MDS plot showed some strain-dependent differences in clustering. Samples from FNP and FNA clustered together, separate from samples of the FNN23 subspecies (inclusive of samples from mutants) in MDS dimension 1. FNA and FNP stationary samples showed tight clustering, while the exponential samples were slightly dispersed. Similarity, seen on the MDS dimension 2, was growth phase-dependent with exponential phase samples clustering to one end of the plot separate from the others. FNP and FNA samples clustered together, suggesting a similar transcriptomic profile. The accompanying variance plot shows that MDS dimension 1 captures around 35% of the variation in the data, and about 25% is captured by MDS dimension 2. Together, MDS dimensions 1 and 2 capture around 60% of the data variability and provide a comprehensive representation of the variation in the gene expression data.



A) B)

Figure 5.2*:*MDS and variance by MDS dimension of the strains. *A) MDS plot showing how the data is clustered/distributed along the two main MDS dimensions (MDS dimension 1 and MDS dimension 2). The x-axis shows distribution across MDS dimension 1 and the y-axis show distribution along the MDS dimension 2. (B) Variance plot showing percentage variance(x-axis) and corresponding MDS dimension (y-axis) from MDS dimension 1 to MDS dimension 10.*

5.6 Subspecies-specific transcriptomic differences

Next, the overall expression patterns of the subspecies were visualised using heatmaps (Figure 5.3). Thresholds of $\log_2$ fold-change ≤ -1 or ≥ 1 with FDR< 0.05 were used to plot significantly differentially expressed genes across the datasets relative to average expression. Overall, there was a subspecies-specific difference in gene expression patterns. All the FNN23 strains (WT and mutants) show overall consistent and similar patterns of downregulation and upregulation. However, subtle strain-specific differences between the mutants and the wild-type strain can be seen (left end of the heatmap); some genes are under-expressed in the mutants compared to the wild-type strain. The FNN23 type strains differ in expression patterns from FNP and FNA. Interestingly, FNP and FNA show similar expression patterns with minor subspecies-specific

differences (right end of the heatmap) where there are regions of contrasting patterns of downregulation and upregulation of certain genes.

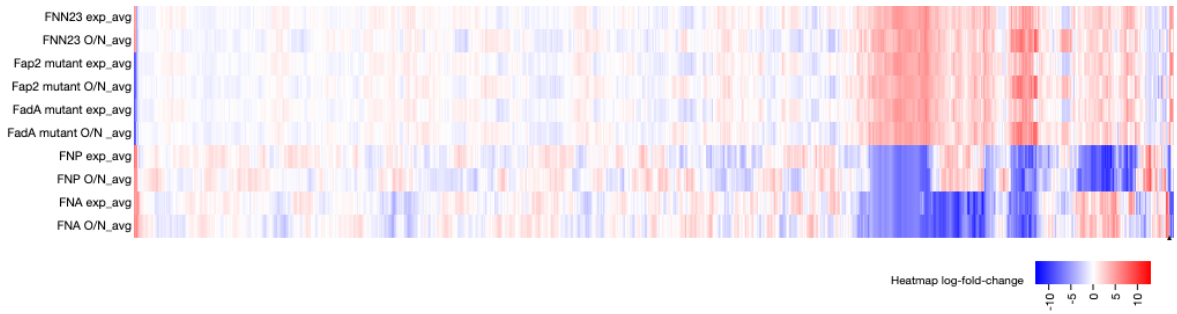


***Figure 5.3*:Heatmap showing the gene expression pattern of the strains (-1 > log_2 foldchange > 1, FDR = 0.05) relative to average expression**. *Mean expression values for stationary and exponential for each strain are shown. Samples containing the tag exp_avg are exponential samples while those with the tag O/N_avg are from the stationary samples. Each coloured bar represents a gene. Two biological replicates were used per sample and the data shown is the average of the two. Log fold change values range from -10 to 10 with colour gradient going from blue for low log-fold-change values < 0 to red for log fold changes > .*

5.7 Global growth phase impact on the transcriptomic profile of *F. nucleatum* subspecies.

The impact of the growth phase on the expression patterns of the subspecies was evaluated, and a heatmap of the exponential phase relative to stationary phase (stationary samples) was generated (Figure 5.4). Consistent intra-condition patterns of downregulation and upregulation were observed. This translated to a consistent contrast in the inter-condition patterns with regions of under-expression in exponential samples corresponding to regions of over-expression in stationary samples and vice versa.

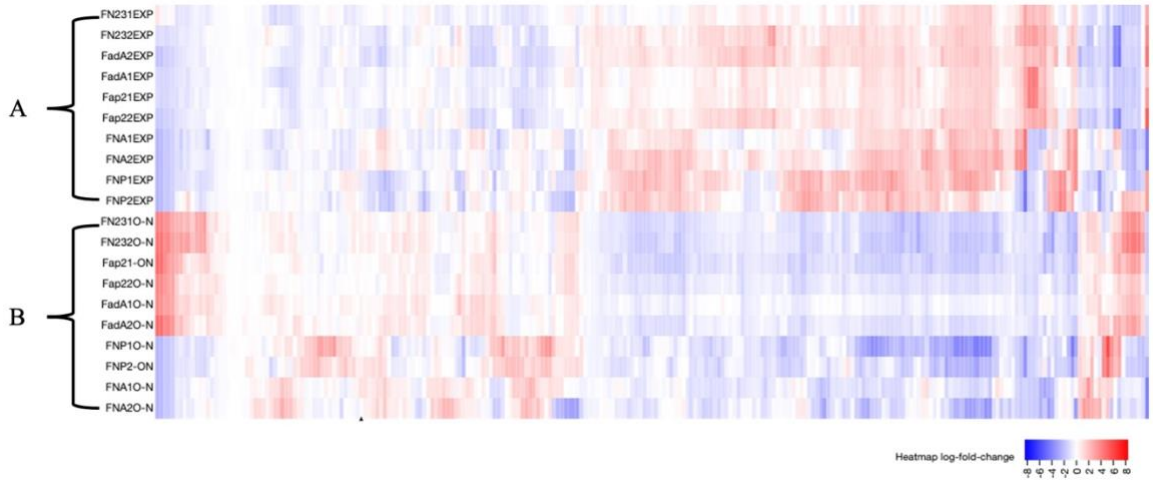


Figure 5.4: Heatmap showing the gene expression pattern of exponential phase samples relative to stationary phase samples (-1>log$_2$ foldchange>1, FDR=0.05)*. Exponential phase samples (A) and stationary samples(B) are shown in the brackets. The heatmap shows an expression of exponential samples relative to stationary samples. Each coloured bar represents a gene. Two biological replicates were used for each subspecies and for the mutant strains and are denoted by the number 1 or 2 tagged to the name of the subspecies. Log fold change values range from -8 to 8 with colour gradient going from blue for low log-fold-change values < 0 to red for log fold changes >0 .*

There were 251 differentially expressed genes (DEGs) when comparing stationary samples to exponential samples. Around 50 % (127) of genes were overexpressed, and 49 % (124) of genes were under-expressed at the exponential phase compared to the stationary phase for all samples combined.

5.8 Functional analysis of global gene expression patterns

To get a general idea of the functions of the differentially expressed genes of stationary compared to exponential phase samples, DEGs were fed into NCBI datasets to obtain their protein IDs and their amino acid sequences. COG functions were annotated in Eggnog and genes significantly upregulated and downregulated for each COG classification were determined and visualised using ggplot2. The functions of DEGs plotted for exponential relative to stationary are shown (Figure 5.5). Some genes did not have a COG classification, and some had unknown functions. Many genes of unknown function were significantly under-expressed (77.27%) at the exponential phase compared to the stationary phase. Genes in the signal transduction and defence categories were under-expressed (100%). Other categories that contained a higher proportion of under-expressed genes were genes involved in carbohydrate transport and metabolism (75%), amino acid transport and metabolism (68.42%), and energy production and conversion (67%). Contrarily, all genes in the cell motility category, the posttranslational modification, turnover, and chaperone categories were overexpressed (100%). Most genes involved in translation, ribosomal structure and biogenesis were also overexpressed (95.74%). Overexpression was also observed in genes involved in transcription (71.43%), intracellular trafficking, secretion, and vesicular transport (66.67%).

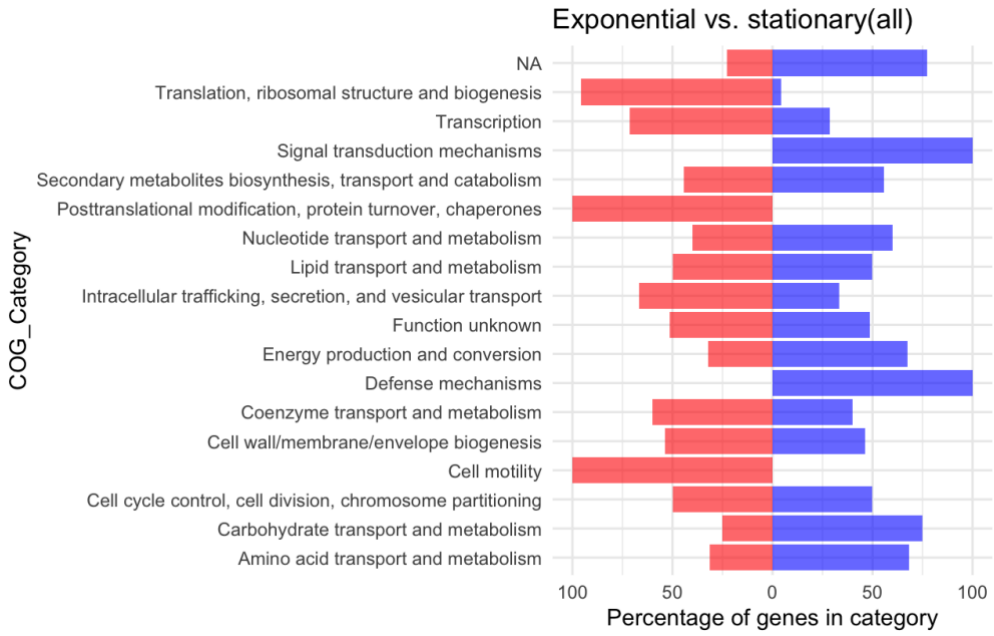


Figure 5.5: COG functional groups of differentially expressed genes of stationary vs exponential phase generated from DEGs of exponential phase relative to stationary phase*. The x-axis shows the percentage of DEGs underexpressed or overexpressed in that category, and the description of the COG_category is shown on the y-axis. Blue represent the percentage of genes downregulated and red repreent the percentage of the genes that have been upregulated.*

5.9 Network mapping and enrichment analysis of global RNA expression differences

To gain a better understanding of the relationship between the genes and determine enrichment patterns within the dataset, network mapping was performed using the string database. Amino acid sequences of the differentially expressed genes were fed into the database, and 242 proteins were identified from a list of 251 genes. The proteins they code for were used to create a protein network map (Figure 5.6). Clusters were created based on the highest interaction score (0.90), and k-mean clustering (k=3) was used to create three main clusters from the data. Edges on the network represent both physical and functional associations between the proteins. Most of the

representative proteins (104) fell into Cluster 1, followed by Cluster 2, which comprised 74 proteins and lastly, cluster 3, with 64 proteins.

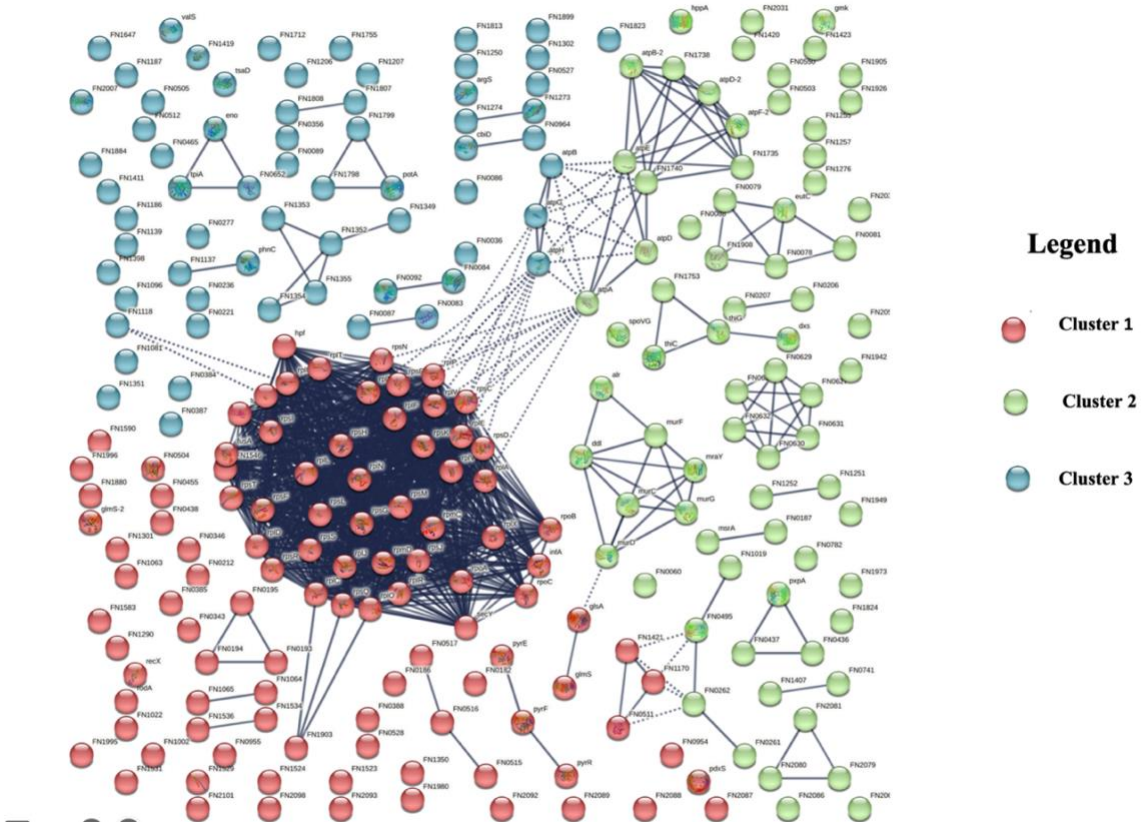


Figure 5.6: Clusters of proteins derived from DEGs of stationarys vs exponential samples. *Three clusters are shown, with proteins represented by the nodes. Edges show protein - protein interactions. Red nodes represent cluster 1 proteins, green represent cluster 2 and blue spheres represent cluster 3 proteins.*

Enrichment analysis revealed that the KEGG (Kyoto Encyclopedia of Genes and Genomes) classification for cluster 1 was the ribosome pathway (strength 1.15 and FDR= 2.5E-25). This cluster had many interconnected edges, suggesting stronger connections between proteins in the cluster. The cluster was enriched with proteins involved in assembling components of the small and large ribosome subunits. These proteins are also important for mRNA and rRNA

binding and play a role in translation and regulation processes, which are upregulated in exponential samples (see Figure 5.6). Cluster 2, on the other hand, has proteins that are involved in a diverse range of processes, with 6 KEGG classifications identified for the set of proteins in this cluster (Table 5.2). KEGG enrichment was significant for pathways involved in peptidoglycan synthesis (FDR=0.00018), vancomycin resistance (FDR=0.00034), oxidative phosphorylation (FDR=1.33E-05), and thiamine metabolism (FDR=0.030). Gene ontology terms for biological processes that cluster 2 proteins are involved in include ATP biosynthetic pathways and cell wall organisation.

Table 5.2: KEGG enrichment IDs of protein clusters of the DEGs of exponential vs stationary samples

Cluster	Protein ID	Description	Observed gene count	Background gene count	strength[2]	FDR[3]
1	Fnu03010	Ribosome	36	51	1.15	2.58E-25
2	Fnu00190	Oxidative phosphorylation	9	19	1.13	1.33E-05
	Fnu01100	Metabolic pathways	30	355	0.38	9.45E-05
	Fnu00550	Peptidoglycan biosynthesis	7	16	1.09	0.00018
	Fnu01502	Vancomycin resistance	5	6	1.37	0.00034
	Fnu02060	Phosphotransferase system (PTS)	5	13	1.04	0.0041
	Fnu00730	Thiamine metabolism	4	13	0.94	0.030

[2] Log_{10}(observed gene count / background gene count). Strength measures how large the enrichment effect is.
[3] FDR= False discovery rate.Shows p-values corrected for multiple testing within each category using the Benjamini-Hochberg procedure.

Finally, the third cluster contained membrane proteins key for signal transduction and ATP binding. The protein families in the clusters were queried for associated InterPro terms and compared to the background count, a reference for the total number of genes belonging to that ID in the database. Cluster 3 showed significant enrichment for 5 Interpro terms (Table 5.3), including ABC3 transport permease protein domain, MacB domains CheY-like superfamily and signal transduction response regulator receiver domain. Bacterial transport systems have a wide range of activity in bacteria. MaCB forms an efflux pump with TolC and MacA proteins which facilitates antimicrobial resistance and release of virulence effectors in Gram negative bacteria (Greene et al., 2018). CheY is signal regulator protein that has roles in signal transduction with possible effects on cell motility and gene expression alterations (Wolanin et al., 2003).

Table 5.3:Interpro terms for proteins in cluster 3

Term ID	Term description	Observed gene count	Background gene count	Strength[4]
IPR003838	ABC3 transporter permease protein domain	5	5	1.5
IPR025857	MacB-like periplasmic core domain	5	5	1.5
IPR001789	Signal transduction response regulator receiver domain	5	8	1.3
IPR011006	CheY-like superfamily	5	8	1.3
IPR017911	MacB, ATP-binding domain	4	4	1.5

[4] Strength refers to a measure of how large the enrichment is. It is the ratio of o $\log_{10}$(observed gene count / background gene count).

Generally, when comparing stationary samples to exponential phase samples for all tested strains, genes involved in transcription and translation are upregulated, and these are made up of genes found mainly in cluster 1. Genes involved in amino acid metabolism, mainly found in cluster 2, are downregulated. Taken together, the functional analysis data and the network analysis show genes involved in a wide range of key activities are significantly differentially expressed when exponential cultures are compared to stationary samples. There is an upregulation of protein synthesis, motility, and biogenesis at the exponential phase, while these processes are almost equally downregulated in stationary sample.

5.10 Strain specific differences in gene expression

The MDS plot and heatmap revealed some subspecies-specific differences in the transcriptomic profiles of the samples. To get a closer look at these differences, differential expression analysis was performed using thresholds of FDR= 0.05 and $\log_2$ fold-change $\leq$ -1 or $\geq$ 1 for the exponential phase relative to stationary samples for each subspecies. The number of DEGs for each strain when comparing stationary to exponential samples is shown (Figure 5.7). Fap2 mutant strain had the most (433) and FNA had the least (124) number of differentially expressed genes at the exponential phase compared to the stationary phase. A list of the differentially expressed genes was evaluated using a Venn diagram to see the shared elements and to observe if any genes are uniquely differentially expressed for the strains. A total of 31 genes were shared by the strains when evaluating stationary vs exponential. Most of these genes were in the ribosome pathways determined using the String database (data not shown), consistent with the

global gene expression patterns mentioned earlier (section 5.5). FNP had the greatest number of genes (220) “uniquely” differentially expressed compared to the other strains. In contrast, FNA and the FadA mutant strain had the least (18) number of genes that were “uniquely” differentially expressed. These strain-specific differences may possibly account for differences in virulence characteristics of the subspecies for processes such as adhesion, biofilm-forming ability, and responses to immune cells.

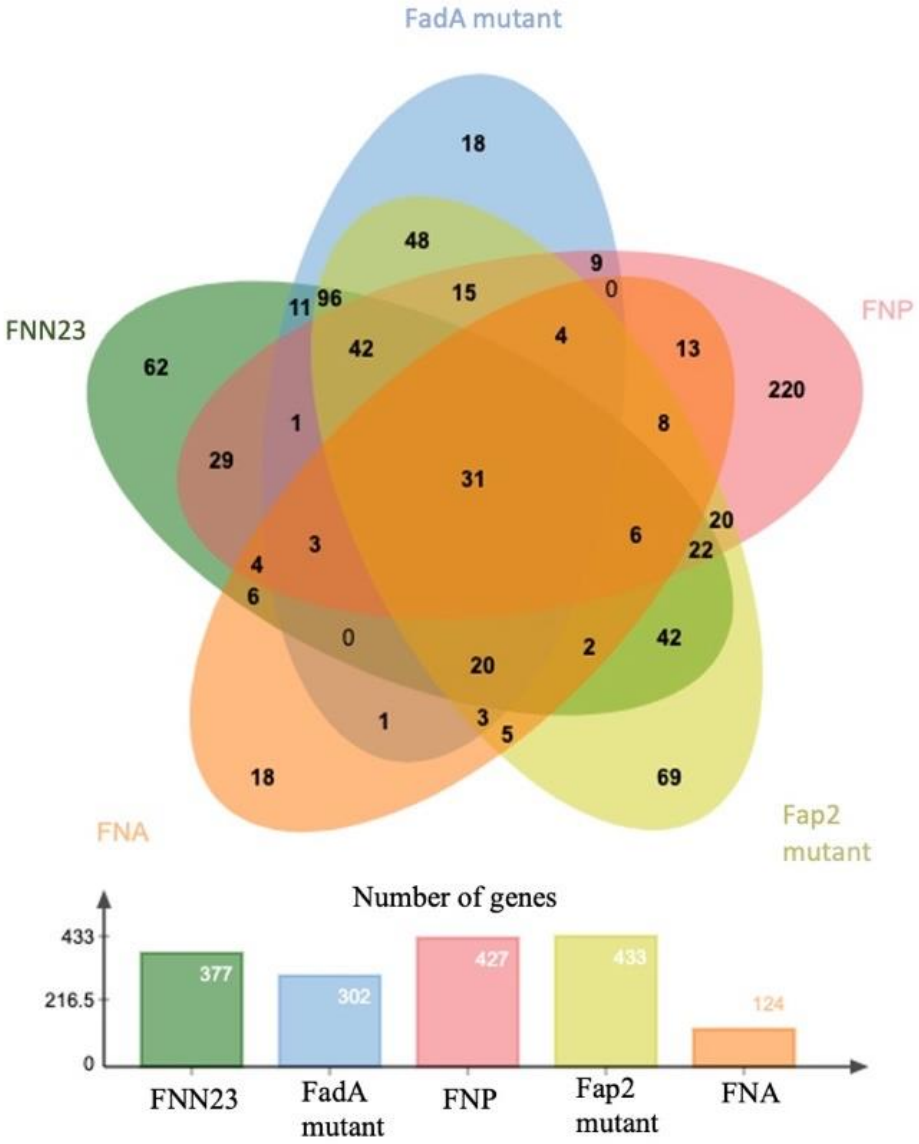
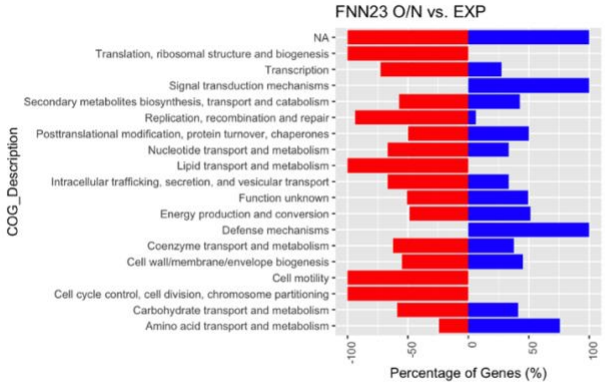


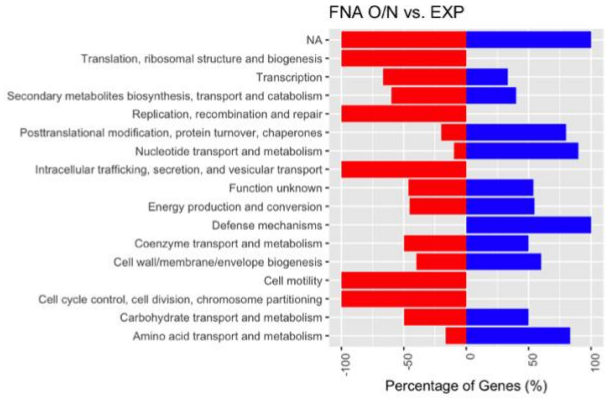
Figure 5.7: Venn diagram showing the number of shared DEGs between the stains. *The number of input DEGs from each strain are shown on the accompanying bar plot.On the bar plot, the strain names are on the x-axis while the y -axis shows the number of DEGs which is also shown in each bar of the bar chart. The different colours represent a single strain on both the venn diagram and the bar plot.*

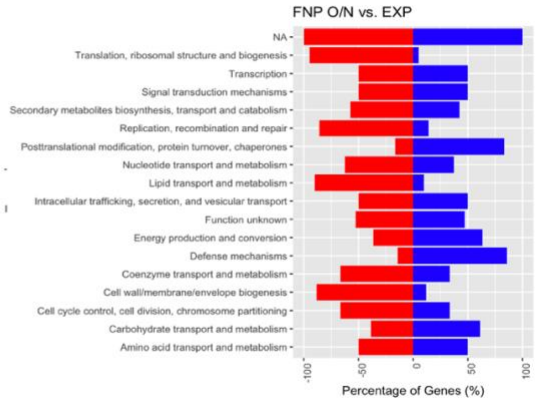
5.11 Functional analysis of growth phase-dependent strain-specific differences in gene expression

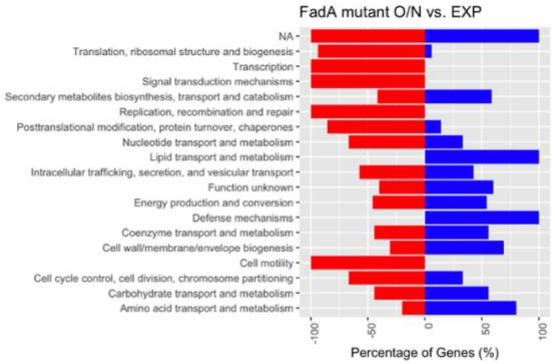
The general function of the differentially expressed genes was explored using COG analysis for each subspecies to get an idea of the functional variations of the expression profiles of the strains (Figure 5.8). Both intra-strain variations in exponential vs stationary expression and inter-strain variations were evaluated in this way. Most genes involved in translation, ribosomal structure and biogenesis were upregulated for all the strains. FNN23 and FNA had 100% genes in that category upregulated. All the transcription genes where upregulated for the FadA mutant strain, whereas for the Fap2 mutant strain and FNP, only half of the genes in that category were upregulated. All the genes in the signal transduction mechanisms category for FNN23 and Fap2 mutant were downregulated, but the same were upregulated in the FadA mutant strain. For all the strains, genes in the replication, recombination and repair category were upregulated, with a 100% upregulation observed for FNA, FadA mutant and Fap2 mutant strains. The category of lipid transport and metabolism had 100 % of the genes overexpressed for FNN23, while all the genes in that category were under-expressed for both mutants. FNA was the only strain with 100% overexpression in the genes for intracellular trafficking, secretion, and vesicular transport category. All the strains, except for FNP, had all the genes in the defence mechanisms category downregulated at the exponential phase, while FNP had a portion upregulated. Moreover, all the cell motility genes were upregulated for FNN23, FNA and FadA mutant strains. Fap2 mutant strain and FNP did not bring up a hit for this category. All the genes in the cell cycle control, cell division, chromosome partitioning category and all under-expressed in the Fap2

mutant strain. Overall, there is a clear strain-specific variation in the functional profile of the DEGs from the strains. There are some similarities in the general functions that are upregulated. However, the number of genes upregulated per category differs for every strain.


FNN23 O/N vs. EXP
COG_Description
NA
Translation, ribosomal structure and biogenesis
Transcription
Signal transduction mechanisms
Secondary metabolites biosynthesis, transport and catabolism
Replication, recombination and repair
Posttranslational modification, protein turnover, chaperones
Nucleotide transport and metabolism
Lipid transport and metabolism
Intracellular trafficking, secretion, and vesicular transport
Function unknown
Energy production and conversion
Defense mechanisms
Coenzyme transport and metabolism
Cell wall/membrane/envelope biogenesis
Cell motility
Cell cycle control, cell division, chromosome partitioning
Carbohydrate transport and metabolism
Amino acid transport and metabolism
-100
-50
0
50
100
Percentage of Genes (%)


FNA O/N vs. EXP
NA
Translation, ribosomal structure and biogenesis
Transcription
Secondary metabolites biosynthesis, transport and catabolism
Replication, recombination and repair
Posttranslational modification, protein turnover, chaperones
Nucleotide transport and metabolism
Intracellular trafficking, secretion, and vesicular transport
Function unknown
Energy production and conversion
Defense mechanisms
Coenzyme transport and metabolism
Cell wall/membrane/envelope biogenesis
Cell motility
Cell cycle control, cell division, chromosome partitioning
Carbohydrate transport and metabolism
Amino acid transport and metabolism
-100
-50
0
50
100
Percentage of Genes (%)


FNP O/N vs. EXP
NA
Translation, ribosomal structure and biogenesis
Transcription
Signal transduction mechanisms
Secondary metabolites biosynthesis, transport and catabolism
Replication, recombination and repair
Posttranslational modification, protein turnover, chaperones
Nucleotide transport and metabolism
Lipid transport and metabolism
Intracellular trafficking, secretion, and vesicular transport
Function unknown
Energy production and conversion
Defense mechanisms
Coenzyme transport and metabolism
Cell wall/membrane/envelope biogenesis
Cell cycle control, cell division, chromosome partitioning
Carbohydrate transport and metabolism
Amino acid transport and metabolism
-100
-50
0
50
100
Percentage of Genes (%)


FadA mutant O/N vs. EXP
NA
Translation, ribosomal structure and biogenesis
Transcription
Signal transduction mechanisms
Secondary metabolites biosynthesis, transport and catabolism
Replication, recombination and repair
Posttranslational modification, protein turnover, chaperones
Nucleotide transport and metabolism
Lipid transport and metabolism
Intracellular trafficking, secretion, and vesicular transport
Function unknown
Energy production and conversion
Defense mechanisms
Coenzyme transport and metabolism
Cell wall/membrane/envelope biogenesis
Cell motility
Cell cycle control, cell division, chromosome partitioning
Carbohydrate transport and metabolism
Amino acid transport and metabolism
-100
-50
0
50
100
Percentage of Genes (%)

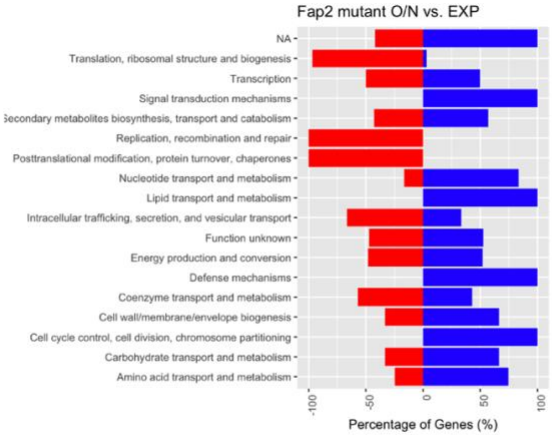

Fap2 mutant O/N vs. EXP
NA
Translation, ribosomal structure and biogenesis
Transcription
Signal transduction mechanisms
Secondary metabolites biosynthesis, transport and catabolism
Replication, recombination and repair
Posttranslational modification, protein turnover, chaperones
Nucleotide transport and metabolism
Lipid transport and metabolism
Intracellular trafficking, secretion, and vesicular transport
Function unknown
Energy production and conversion
Defense mechanisms
Coenzyme transport and metabolism
Cell wall/membrane/envelope biogenesis
Cell cycle control, cell division, chromosome partitioning
Carbohydrate transport and metabolism
Amino acid transport and metabolism
-100
-50
0
50
100
Percentage of Genes (%)

Figure 5.8: Cog functional analysis of differentially expressed genes of stationary vs exponential for all strains, including the three subspecies and the two mutants. *Blue bars show percentage of genes that have been downregulated and red bars shows percentage of genes that have been upregulated. The x-axis shows percentage of genes per category that have either been upregulated or downregulated for that category. The cog categories/ descriptions are shpwn on the y-axis of each graph.*

5.12 FNN23 compared to mutants

The expression profile of the wild-type FNN23 strain was more closely compared to that of the two mutants- FadA and Fap2. Visualised on an MDS plot, most of the variation in the data (70%) is explained by MDS dimension 1 (Figure 5.9). Samples clustered according to the growth phase. All the six stationary samples were more closely grouped with slight dispersion. For the exponential samples, the samples further split into two clusters, with samples from biological replicate one grouping separate from those from biological replicate two.

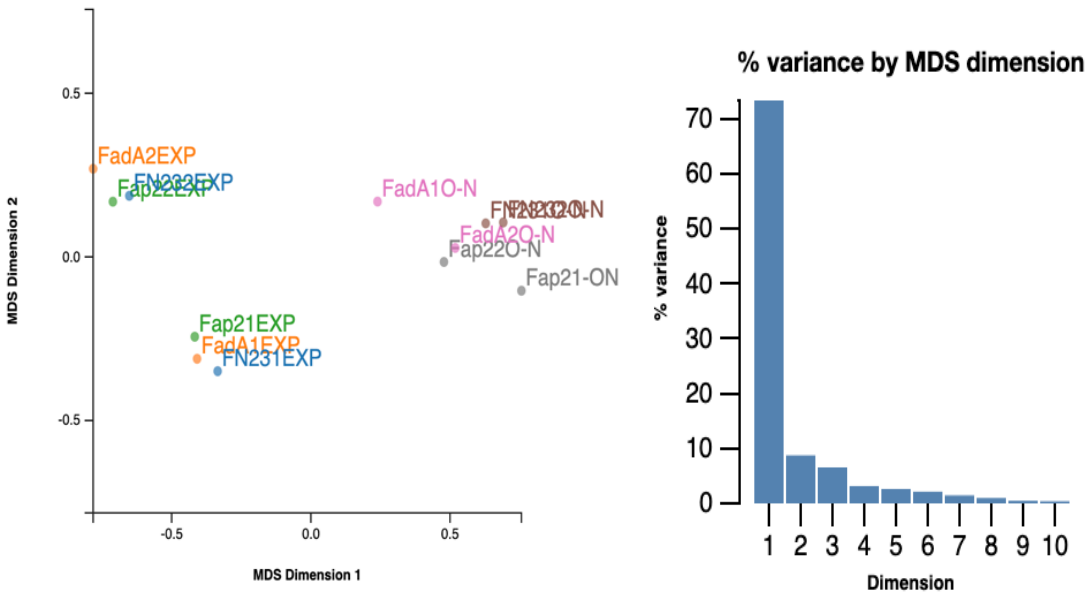


Figure 5.9: MDS plot of FNN23 compared to adhesin mutant strains. *(Left panel) MDS plot showing how the data is clustered/distributed along the two main MDS dimensions (MDS dimension 1 and MDS dimension 2). The x-axis shows distribution across MDS dimension 1 and the y-axis show distribution along the MDS dimension 2. (Right panel) Variance plot showing percentage variance (x-axis) and corresponding MDS dimension (y-axis) from MDS dimension 1 to MDS dimension 10.*

The DEGs for stationary vs exponential for the wild-type strain was more closely compared to that of the mutants. A combined 189 similar genes were significantly differentially expressed when pooling together the DEGs of stationary compared to exponential for the wild-type strain and the mutants (Figure 5.10). The FadA mutant strain shared a lot of its DEGs with the other strains, whereas the FNN23 and the Fap2 mutant strain had about 100 “uniquely” differentially expressed genes compared to the others. This suggests that mutations in these adhesins have significant impacts on the overall transcriptome of the strains other than the silencing of target genes.

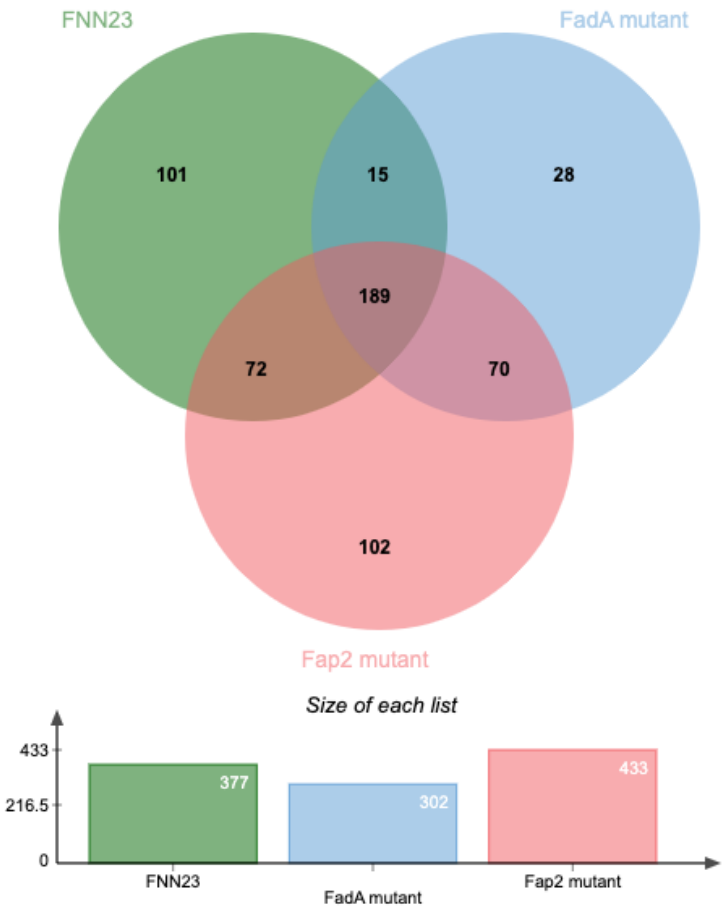


Figure 5.10: Venn diagram showing the number of shared DEGs between the wild type FNN23 strain and the adhesin mutants. *The number of input DEGs from each strain are shown on the accompanying bar plot.On the bar*

plot, the strain names are on the x-axis while the -axis shows the number of DEGs which is also shown in each bar of the bar chart. The different colours represent a single strain on both the venn diagram and the bar plot.

Next more specific differences in the transcriptomic profile of the mutants compared to the wild type at the exponential phase were determined. Differential expression analysis was performed and the log_2 foldchange of the expression profile of each mutant relative to that of the WT was determined. This was first done for exponential phase samples only. There was an expected difference in the expression of *fadA* in the FadA mutant strain as it has a *fadA* gene deletion (Table 5.4). This is reflected by the significant downregulation (14-fold) of *fadA* expression compared to the wild-type strain. The *fadA* gene was also slightly downregulated in the Fap2 mutant strain (log_2 foldchange = -0.0721) compared to the wild-type strain. Two other genes that were significantly differentially downregulated were the genes *C7Y58_RS02335* and *C7Y58_RS02340*, which are locus tags for the *UDP-glucose-hexose-1-phosphate uridyltransferase* and the *galactokinase* genes respectively. Creation of the adhesins in the FNN23 type strain involved a genetic system which required deletion of the *galK* and *galT* operons. (Casasanta et al., 2020). Therefore, the significant under-expression of the galactokinase gene was not surprising. For mutants, there was over a 14-fold decrease in the *UDP-glucose-hexose-1-phosphate uridyltransferase* gene. A 10-fold reduction in expression was observed for the FadA mutant strain, while a 12.88-fold decrease was observed for the Fap2 mutant strain for the *galactokinase* gene. The *galactokinase* gene is on the same locus ad the *UDP-glucose-hexose-1-phosphate uridyltransferase* gene, therefore this gene may have been deleted when the galactokinase gene was deleted hence its downregulation. Differential gene expression was also expected for the *fap2* gene. However, possible annotation errors in the reference sequence may have led to the gene not being properly represented in the dataset.

Table 5.4:Genes differentially expressed in FNN23 strain and Fap2 and FadA gene mutant strains

Gene	Name	FNN23 (WT)	Δ*FadA*	Δ*Fap2*	Average Expression	P value
C7Y58_RS04385	Adhesion protein *FadA*	4.70	-9.32	4.62	13.16	2.35E-06
C7Y58_RS02335	*UDP-glucose--hexose-1-phosphate uridylyltransferase*	9.46	-4.73	-4.73	6.06	7.85E-06
C7Y58_RS02340	*Galactokinase*	7.73	-2.57	-5.16	4.77	6.89E-06

Next, differences between the mutant and the wild-type strains at the stationary phase were also determined. More differences were observed when stationary samples were used compared to exponential, with over 40 genes differentially expressed when comparing the wild type to the mutants. Like with exponential phase samples, the *galactokinase* and *UDP-glucose--hexose-1-phosphate uridylyltransferase* genes were significantly under-expressed in the mutants compared to the wild-type strain (Figure 5.11). The $\log_2$ fold change of the wild type expression of UDP-*glucose--hexose-1-phosphate uridylyltransferase* relative to average expression was 9.32 while that of the Fap2 and FadA mutant strains was -4.66. Furthermore, the $\log_2$ fold change of the wild type expression of the *galactokinase* gene was 8 while that of the mutants was -4. Again, downregulation of *fadA* gene in the FadA mutant strain was observed with $\log_2$ fold changes of 3.71, 4.02 and -7. 73 for wild-type, Fap2 mutant strain and FadA mutant strains respectively. Interestingly, expression of the CAECAM binding protein gene *cbpf* was slightly higher in the mutants; log2 foldchanges of 0.798 and 0.848 for the *FadA* mutant and *Fap2* mutant, respectively

and -1.65 for the wild-type strain relative to average expression. Another gene (*C7Y58_RSO02630*) which codes for an autotransporter (YadA -like) was upregulated in FNN23 compared to *FadA* and *Fap2* mutants with respective log_2 foldchanges of 1.11, -0.703 and -0.410 relative to the average expression. These genes (*Cbpf, FadA, C7Y58_RSO02630*) are some of the genes that may have roles for adhesion and therefore, their differential expression may contribute to differences in adherence and invasion during host-pathogen interactions.

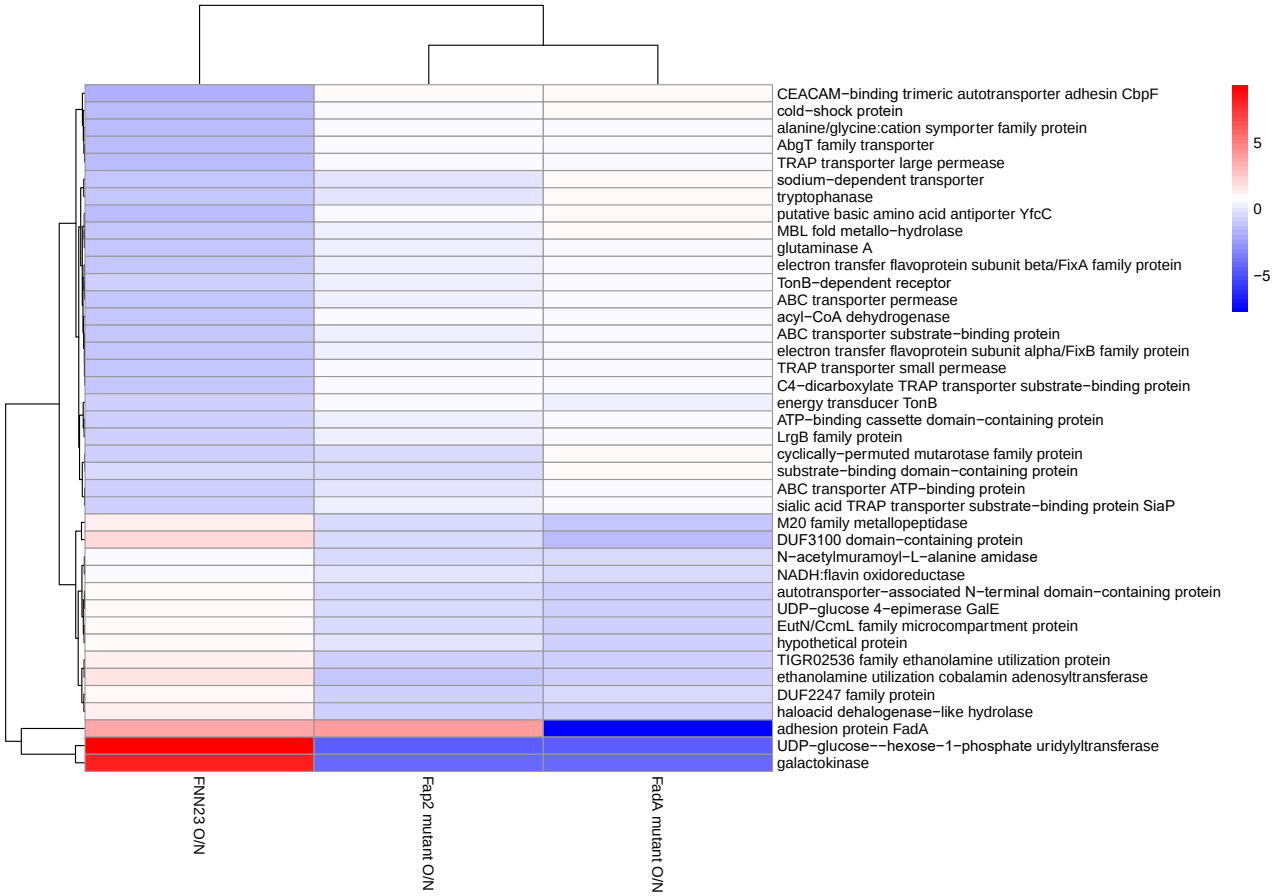


Figure 5.11*:*Heatmap of differentially expressed genes of the wild type FNN23, the FadA and Fap2 mutants at stationary phase relative to average expression.*The gene name is shown on the y-axis and the strain name is shown on the x-axis. Each bar shows the expression levels (Log_2fold change) of the gene at stationary phase relative to average expression of the same gene. The rows and columns have been subject to hierachichal clustering.*

Genes for *tonB* and *tonB*-dependent receptor which are important for hemin uptake in Gram-negative bacteria were both downregulated in the wild type compared to the mutants (Liu et al., 2021). Some of the genes are involved in carbohydrate and galactose metabolism. Other differentially expressed genes are involved in the N-acetylneuraminate metabolic processes important for sialometabolism.

5.13 Expression levels of adhesins across the different conditions

Next, the gene expression profile of the *F. nucleatum* subspecies was evaluated to identify strain-specific differences that could explain some variation in levels of adhesion and invasion observed in chapters 3 and 4. The expression levels of known adhesin genes and virulence genes associated with attachment were compared between the strains and at different growth phases (Figure 5.12). Across all the samples, expression patterns are not uniform, suggesting that there are subspecies-specific differences in the expression of the selected genes. FNP stationary samples show an overall overexpression of the selected genes relative to the other samples. *FadA* had a significant overexpression across all the samples except for in the FadA mutant strain samples. Interestingly, expression of the CAECAM 1 activator gene- *cbpf*, varied; both FNP samples and the FadA mutant stationary samples had very low expression of this gene, whereas FNN23 and Fap2 samples FadA mutant strains had higher expression levels. Analysis revealed that *radD,* another known adhesin of *F. nucleatum*, which is important for co-aggregation (Kaplan et al., 2009) was more highly expressed for FNA stationary samples than the others. Furthermore, its expression was slightly

elevated in the mutant samples, while FNP samples had the lowest expression of this gene. *RadA*, which has been shown to be important for The *C7Y58_RS04340* gene and *C7Y58_RS02630* genes were both positive hits for the aidA domain; another adhesin of the autotransporter type. *C7Y58_ RS04340* showed higher expression for FNN23 and the *Fap2* mutant at both growth phases, while expression is relatively lower for FNP and FNA subspecies at both. Sequences of these two genes aligned with the *fap2* gene sequence, *C7Y58_RS02630* had a 46.85 % identity with a query cover of 68%, while *C7Y58_ RS04340* covered 57% of the *fap2* gene and had 37% identity. This suggests that these genes may be part of a domain similar to *fap2* or are part of *fap2* as fap2 and other similar large proteins had been misannotated in the publicly available reference (Umaña et al., 2019).

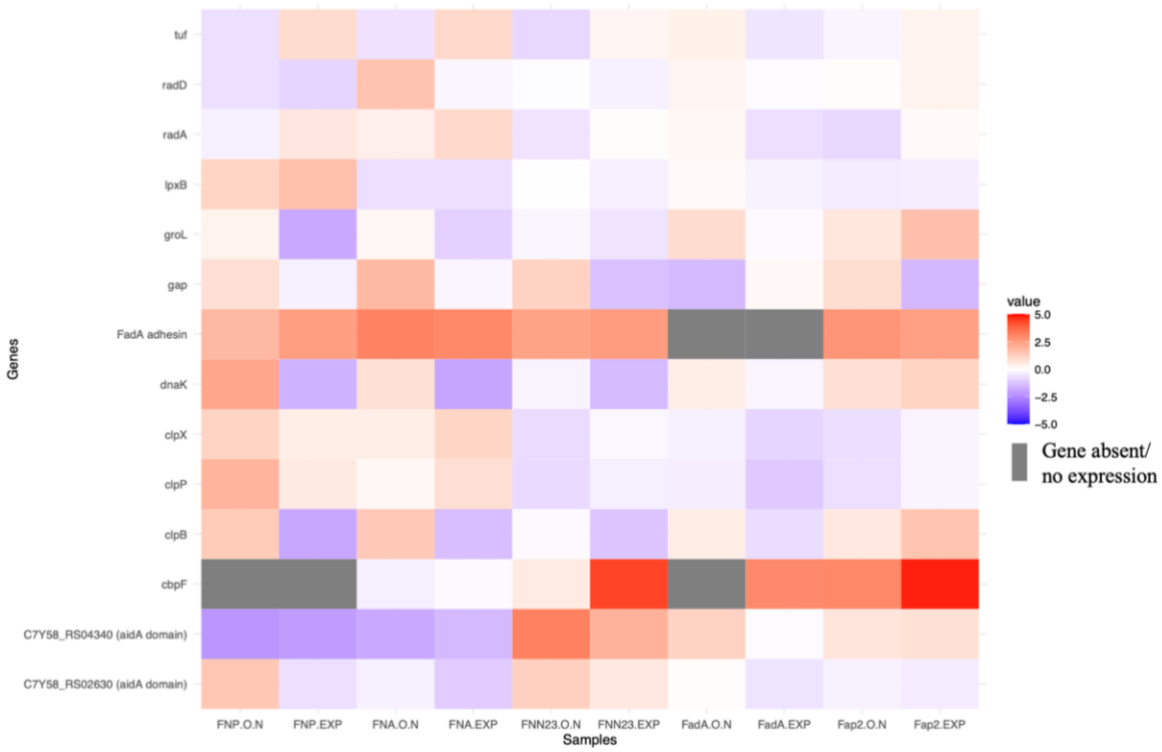


*Figure 5.12:**Heatmap of the log$_2$ foldchanges of the expression across all strains of adhesion-associated genes**.The axis contaim the strain/species name and the y-axis contain the names of the adhesion/invasion associated gene. The expression scale runes from -5 to 5 fold changes. Instances where the gene is missing or there is no expression of the gene registerd are denoted in grey.*

5.14 Expression levels of virulence genes associated with adhesion

Other virulence genes that have been associated with bacterial adherence were also evaluated. These include the chaperone proteins genes *dnak* and *groL* (Figure 5.12). Proteins of these genes, including the *tuf* gene, are considered moonlighting proteins because they have primary roles other than adhesion ; but have been reported to play a role in adhesion in various bacteria (Jeffery, 2018).

For example a *dnaK* mutation reduced the motility and adherence to Caco-2 cells in *Pseudomonas aeruginosa* (Okuda et al., 2017). *DnaK*-mediated reduction in adherence was also reported for *S. aureus* (Singh et al., 2012). In this dataset, the expression of *dnaK* was comparatively lower in the exponential samples of FNN23, FNA, FNP and the *FadA* mutant strain. *GroL* has been associated with attachment to plasminogen, fibronectin and mucin (Hagemann et al., 2017). Tuf has also been reported to bind to several components of the extracellular matrix, including fibronectin and glycosaminoglycans (Harvey et al., 2019). The expression of *groL* and *tuf* also show strain-dependent variation, and *tuf* upregulation is common for exponential phase samples compared to stationary samples. Other factors evaluated were the *clpP* and *clpX* , *clpB* genes, which are part of proteases that are crucial for stress tolerance survival and also form part of the virulence repertoire of bacteria (Aljghami et al., 2022). *ClpP* has been shown to improve bacterial tolerance to stress, promote resistance to phagocytosis and improve adhesion of *Streptococcus suis* to He-p2 cells (Roy et al., 2019). Both *clpP* and *clpX* were relatively overexpressed in FNP and FNA samples compared to the FNN23 wild-type and mutant strains. Overall, there were variations in the expression of adhesion and adhesion-associated virulence factors. This may have consequences for the levels of adhesion of each strain to mammalian cells.

5.15 Discussion

This chapter aimed to determine whether there were differences in the transcriptomic profile of the subspecies FNA, FNP and FNN23. Adhesin mutants in FNN23 strains (*FadA, Fap2*) were also included to evaluate the impact of the mutations on the transcriptomic profile of FNN23. Chapters three and four showed subspecies-specific differences in the adherence of the bacteria to Caco-2 cells. In chapter four, the impact of the bacterial growth phase on the adhesion of the subspecies was also demonstrated. Transcriptomic analysis of the subspecies at different growth phases, evaluated in this chapter, provides insights into bacterial dynamics that may translate to variations in adherence and invasion of the subspecies.

The greatest source of variation between the subspecies was growth phase differences. There was a clear stratification in gene upregulation and downregulation of genes. Genes upregulated at the exponential phase were generally downregulated at the stationary phase. The stationary phase is normally associated with reduced cell growth and an uptick in stress-related genes in accordance with nutrient depletion (Chubukov and Sauer, 2014;Jaishankar and Srivastava, 2017). The exponential phase, on the other hand, is associated with increased cell growth, higher metabolism, and upregulated protein synthesis (Alreshidi et al., 2023). Data from this chapter showed that at the exponential phase, many upregulated genes made up part of the ribosome and were involved in transcription, translation, and protein turnover, suggesting that protein synthesis is upregulated at this growth phase. Ribosomal genes were under-expressed at the stationary phase. Translation

processes are energetically costly, and at the stationary phase, bacteria commonly prioritise nutrient and energy preservation instead of energy expenditure (Feaga and Dworkin, 2021).

Subspecies-specific differences in the transcriptomic profile of the subspecies were observed. Comparison of gene expression differences at the exponential phase and stationary for each strain revealed that some genes are only differentially expressed when comparing stationary to exponential phase for certain strains and not others. FNP had the most genes that were only differentially expressed between its stationary and exponential cultures. This was interesting as in previous work (Chapter 3), FNP showed the highest level of adhesion of all the subspecies. Taken together, this suggests that its gene expression profile may be much more adaptable to changing environments and interactions, contributing to its higher rates of adhesion and invasion compared to the other subspecies. Another noteworthy observation was that FNA had the least number of DEGs when comparing exponential vs stationary. Compared to the other strains, the transcriptomic profile of FNA may be a lot more opposed to changes. FNA could be prioritising the expression of more "necessary" genes for adaptation to the changes in the growth phase. Whether this provides an advantage may depend on the environment and the stressors within the environment, considering that sometimes, "unnecessary " genes may have subtle functions that contribute to fitness in certain situations (Price et al., 2016).

Additionally, the gene expression profiles of the mutants of *FadA* and *Fap2* were compared to the wild-type strain. In line with the deletion created in the *FadA* mutant, the expression of the *FadA*

gene was downregulated compared to the other two strains. This was not the case with the *Fap2* gene, as a hit for this gene was not found in the dataset. This may be explained by annotation errors (common with large genes) resulting in poorly annotated genes (Umaña et al., 2019). Interestingly, only three genes were differentially expressed between the wild type and mutants at the exponential phase, whereas over 40 genes were differentially expressed at the stationary phase. Furthermore, there was an increase in the expression of other adhesin-associated genes, such as the CAECAM 1- binding protein gene *cbpf*, and genes for transporter proteins. This implies that these adhesins could be more beneficial at the stationary phase than at the exponential phase, as these gene expression changes could have compensatory roles for the functions and impacts of the deleted adhesins.

Targeted analysis of known adhesins and adhesion-associated genes showed strain-dependent, growth phase-influenced differences. The *FadA* gene showed consistent upregulation across the strains at different growth states, except for the mutant stain. *FadA* has been reported to be highly conserved in *F. nucleatum*, however others have shown that in some strains it may be absent (Nithianantham et al., 2009a;Muchova et al., 2022). In this data set, *FadA* was expressed in all three subspecies used. A striking contrast in the expression of *cbpf* was observed, with the gene being significantly upregulated in FNN23 and having more than a 5-fold decrease in FNP samples. Interestingly, *cbpf* was also downregulated in the FadA stationary samples but not exponential samples, suggesting growth phase-dependent differences in adhesin expression. Another gene which showed clear growth phase heterogeneity in expression was *dnaK*, which tended to be more upregulated with stationary samples than with exponential phase samples. The *DnaK chaperone*

system is important for stress adaptation, among other functions; therefore, its upregulation in the stationary samples aligns with the increase in stress-related genes at the stationary phase. In the physiological context of *F. nucleatum* colonisation and infection of host tissues, a lot more factors in the environment will influence the bacterial growth phase. The nutrient composition of the environment, competition from other bacteria and other microorganisms and the host immune response are some factors that can influence what growth phase the bacteria will be in (Shi et al., 2021). Some bacteria can remain in the stationary phase or the lag phase for extended periods if there are nutrient deficiencies and other stressors (Gray et al., 2019). It is likely that physiologically, there is heterogeneity in the growth stage of the bacteria, and cells can change or remain in one phase depending on the environment. In addition, the gene expression levels may be regulated differently when the bacteria interact with the host, and synergistic relationships between adhesins and virulence factors may exist but cannot be directly determined from the current experiment. Nonetheless, this data sheds light on the intrinsic differences between the subspecies, which are more likely to translate to observed differences in the interactions of the bacteria with their host. Dual RNA-seq of the bacteria and the host could improve this data by showing interaction-specific transcriptomic changes to the host and the bacteria upon co-culture (Nuss et al., 2017). It is worth noting that the reference strain used was the FNN25 reference genome, which was reflected by better mapping of the FNN23 strain compared to the other two subspecies. Using a common reference genome is beneficial as it simplifies comparative analysis for genes found in all three species; however, if there are significant differences in the genome content, there is a possibility of expression loss (Price and Gibas, 2017). Therefore, some under-expressed genes in FNA and FNP observed in this chapter may be because of some expression loss. Generally, as these are the same species, the differences in the genome content will not be as pronounced the

same way they would be with different bacterial species. In conclusion, using RNA sequences to understand gene expression between subspecies at different growth phases can reveal insights into the adaptation of bacteria to different conditions and provide knowledge of key differences between bacteria that can be explored in subsequent studies.

6 UNCOVERING THE METAGENOMIC AND TRANSCRIPTOMIC PATTERNS OF CRC TUMOURS AND CONTROLS, THE ROLE OF TSPAN6 AND *F. NUCLEATUM*

6.1 Introduction

The use of omics data has been pivotal in establishing the role of bacteria in CRC. Several metagenomic studies have been conducted on faecal, colon and saliva samples to gain insights into the differences between CRC patients/samples and healthy/normal controls (Debesa-Tur et al., 2021;Rezasoltani et al., 2022;Kharofa et al., 2023). As a result, the enrichment of certain bacteria, such as *F. nucleatum* in CRC, has been extensively reported (Kostic et al., 2013;Wu et al., 2022), and the role of bacteria in CRC has gained a growing appreciation. However, as mentioned previously, the exact mechanism of microbial dysbiosis in CRC still needs more research (Sanders et al., 2022). Techniques such as RNA sequencing of CRC samples and controls can provide insights into alterations to the tumour microenvironment resulting from microbial dysbiosis. Transcriptomic analysis has also been critical in improving the typing and classification of CRC based on molecular signatures of samples due to the advent of subtyping methods such as consensus molecular subtype (CMS) prediction, which relies on gene expression data (Guinney et al., 2015). However, few studies have utilised paired omics techniques to unveil the relationship between the tumour metagenome and alterations in the tumour microenvironment.

6.1.1 Tumour staging

Standard CRC diagnosis involves histological examination. There are different CRC staging methods, and the most common is the American Joint Committee on Cancer (AJCC) TNM staging system (Keung and Gershenwald, 2018). Using this method, tumours are staged based on three characteristics: tumour (T) referring to the size and extent of the main tumour, number of nodes (N) affected, and whether the tumour has spread (M). Depending on the TNM combinations, tumours are ranked from stage 0 to stage IV, of which IV is the most advanced stage. Tumour EMVI (Extramural venous invasion) status is also determined in some instances. EMVI refers to a condition whereby the tumour has invaded veins beyond the smooth muscles (muscularis propia) in the colon (Guner and Tumay, 2021). Patients who are positive for EMVI have a poor prognosis and reduced overall survival (Inoue et al., 2021). Tumour staging is important as it influences treatment practices and disease management (De Rosa et al., 2015)

6.1.2 Background

Chapter 3 of this thesis demonstrated the differential interaction of bacteria (*F. nucleatum, Sgg)* with CRC cells. Chapter 4 looked more into pathogen characteristics influencing *F. nucleatum* adherence to Caco-2 cells and associated alteration of immune cell mediators. Chapter 5 then delved into understanding the transcriptomic profile of the *F. nucleatum* subspecies to determine

features contributing to its virulence in CRC. Additionally, the role of Tspan6 in bacteria interactions with CRC cells was explored. To further understand the interaction dynamics between bacteria and the CRC environment, this chapter combines analysis of 112 metagenomic and 80 transcriptomic data from 57 CRC tumours and their paired controls to explore associations between the tumour bacterial metagenome and its corresponding expression changes. Emphasis was placed on uncovering the relationship between *F. nucleatum* and alterations in the CRC environment. Additionally, the dataset was used to determine the relationship between Tspan6 expression and the metagenome of the samples.

6.1.3 Aims

1. To investigate the relationship between the metagenomic profile of tumours and their transcriptomic profile, emphasising the influences of Tspan6 and *F. nucleatum.*
2. To evaluate the impact on the CRC tumour immune microenvironment of the bacterial metagenome.
3. To determine the association of the tumour's clinicopathological features with the metagenomic and transcriptomic characteristics of the samples.

6.2 Descriptive statistics of the datasets (clinicopathological information)

Clinicopathological information was shared for the 57 patients whose samples were utilised for this study. Gender information was only available for 49 patients, of which 38.8 % were female and 61.2 % were male. The median patient age was 71, the youngest was 34, and the oldest was 91. Histological classification was adenocarcinoma for all patients except for one (female, 34) whose histology was adenocarcinoma with neuroendocrine differentiation. All patients underwent surgery, and disease recurrence was 3.5 %, with a mortality rate of 5.26 % at the time of receipt of patient information (05/04/2023). Tumours were detected in one of the following sites: ascending colon (12.3%), descending colon (7.0%), rectum (22.8), transverse colon (right side) (7.0%), sigmoid (24.6%), sigmoid with hepatic flexure (1.8%) and caecum (24.6%).

6.3 Overview of overall metagenomic differences of tumour vs normal samples

Counts for bacterial metagenomic data containing bacterial species and their corresponding number of reads from normal and tumour samples were combined into a single file. Count data was filtered to obtain only bacteria with at least ten counts in at least 10% of the samples leading to a total of 517 out of 5963 bacterial species. The dataset was normalised in DESeq2 using the variance stabilising transformation considering tumour/normal difference. DESeq2 is normally

used to perform differential gene expression analysis on expression data. It was used here with metagenomic data counts, and the output referred to as differentially abundant genes. Log fold changes (LFC) for the abundance of bacteria in tumour vs normal samples for each bacterial species were obtained. LFC shrinkage was performed using apeglm to reduce noise (Zhu et al., 2019). Out of 517 species that had a non-zero total, 81 (61%) were overabundant, and 7 (14%) were under-abundant using log foldchange thresholds of $-2 < \log_2$ fold change > 2 and adjusted p-value of < 0.05. The top 10 over-abundant bacteria when considering tumour vs normal samples are shown (table 6.1). 70% of the top ten over-abundant species were from the phylum Fusobacteria, comprised of bacteria from the Fusobacterium and Leptotrichia genera. The remaining species (30%) belonged to the genus Campylobacter. There was a lot more genus diversity for the ten most under-abundant species. *Ralstonia pickettii, Cloacibacillus porcorum* and *Streptococcus parasanguinis, Alistipes sp. dk3624,* and *Phocaeiocola vulgatus* were some of the most depleted or under-abundant species in tumour vs normal samples.

Table 6.1:Top ten over and under-abundant species for tumour vs normal bacterial metagenome

Bacterial species	Base mean	Log_2 Fold Change	p value	adjusted p value
Over-abundant				
Fusobacterium nucleatum	22585.12	6.35	1.76E-33	9.09E-31
Fusobacterium hwasookii	501.55	5.49	6.09E-26	1.08E-23
Campylobacter gracilis	135.01	7.45	6.27E-26	1.08E-23
Fusobacterium canifelinum	320.41	5.50	2.26E-24	2.93E-22
Fusobacterium pseudoperiodonticum	626.16	4.86	1.00E-22	1.04E-20
Fusobacterium sp. oral taxon 203	199.91	5.37	4.43E-20	3.81E-18
Campylobacter rectus	256.81	6.70	6.89E-20	5.09E-18
Campylobacter showae	93.82	5.71	2.06E-19	1.33E-17
Leptotrichia wadei	24.33	3.50	6.90E-18	3.96E-16
Fusobacterium periodonticum	55.18	4.72	5.78E-16	2.99E-14
Under-abundant				
Ralstonia pickettii	14.67	-2.72	1.57E-05	1.94E-04
Cloacibacillus porcorum	15.80	-1.70	6.12E-04	5.02E-03
Streptococcus parasanguinis	46.98	-1.37	3.44E-03	2.47E-02
Akkermansia muciniphila	2373.75	-1.64	3.92E-03	2.70E-02
Phocaeicola vulgatus	16442.38	-1.08	6.38E-03	4.03E-02
Parabacteroides goldsteinii	345.61	-1.19	7.47E-03	4.59E-02
Alistipes sp. dk3624	34.77	-1.18	8.30E-03	4.88E-02
Sutterella wadsworthensis	786.87	-1.56	1.12E-02	6.25E-02
Priestia megaterium	11.75	-0.48	1.14E-02	6.28E-02
Lachnospira eligens	977.40	-1.08	1.32E-02	7.156433

6.4 Sample grouping based on the metagenomic profile

To assess the impact of the metagenome on the sample profile, combined tumour and normal samples were visualised and grouped using the t-SNE dimensionality reduction algorithm and the Louvain algorithm. T-SNE provides a method to visualise high-dimensional data in a low-dimensional space (Maaten and Hinton, 2008). Louvain, on the other hand, is a machine-learning algorithm that groups highly connected nodes/data points (also known as communities) (Blondel et al., 2008). Using these algorithms, the combined normal and tumour samples (114 samples) were grouped based on the metagenomics profile of the samples such that samples that contained similar bacterial species profiles clustered together. Five sample clusters were obtained (Figure 6.1 left panel), and the cluster composition, showing the number of normal and tumour samples in each group, is shown (Figure 6.1 right panel). Louvain Cluster 1 contained the highest number of samples, mostly normal samples. Louvain Clusters 5 and 4 had more tumour samples than normal samples. Interestingly Louvain cluster 3 had an equal number of tumour and normal samples, while Louvain Cluster 2 contained more normal samples than tumour samples.

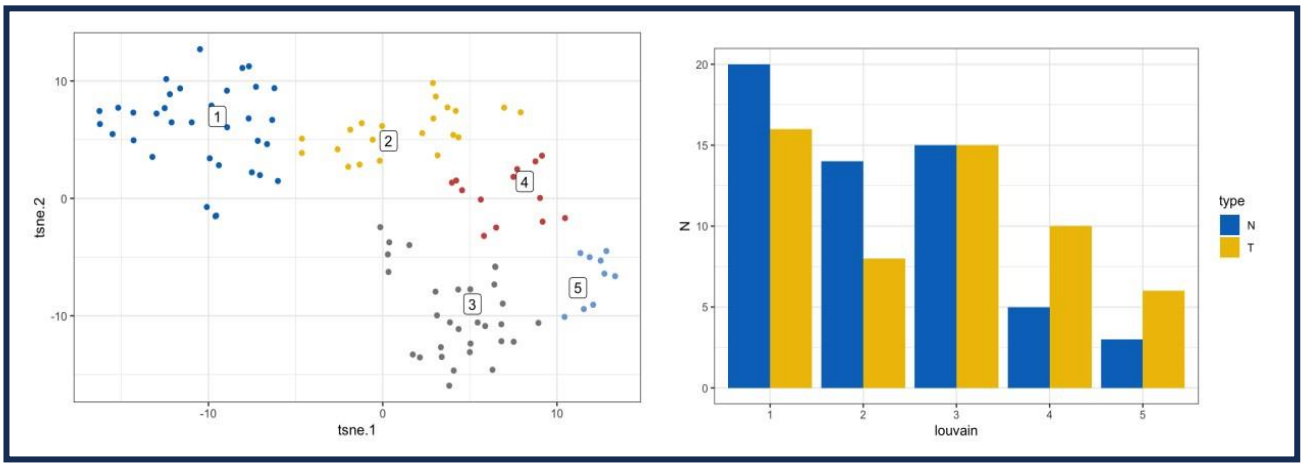


Figure 6.1:Sample grouping based on the metagenomic profile. *A) Samples form five clusters.Dots with same colours represent a cluster. B) Clusters comprise varying numbers of tumour and normal samples, as shown by the bar chart. The number of samples(N) within a cluster is shown on the y-axis, and the sample cluster (Louvain) is on the x-axis. The colour of the bars represents either normal samples(blue) or tumour samples (yellow).*

6.5 Sample clusters have significant differences in some of the bacterial species' composition.

Then, to better understand the differences in the metagenome of the clusters, the number of differentially abundant bacteria between the Louvain groups was determined (Table 6.2). Cluster 3 vs. 1 showed the greatest difference in bacterial species, followed by Cluster 5 vs. 1. Cluster 4 vs. 2 had 71 differentially abundant bacteria.

Table 6.2: Number of differentially expressed species between the clusters

Cluster	**2 vs 1**	**3 vs 1**	**4 vs 1**	**5 vs 1**	**3 vs 2**	**4 vs 2**	**5 vs 2**	**4 vs 3**	**5 vs 3**	**5 vs 4**
Number of differentially abundant bacteria	256	403	297	402	341	71	349	340	223	323

Bacteria that were significantly differentially enriched or depleted in the clusters were visualised on volcano plots using the EnhnacedVolcano package in R to get a general idea of the direction of the differences between the groups (Figure 6.2). A log2 fold change of -2 <log_2fold change >2 and p-value of 0.05 were used to determine significantly differentially abundant bacteria; for most of the plots a great number of bacteria had positive log_2 fold changes except for the comparisons of cluster 4 vs cluster 5. The volcano plot for cluster 4 vs. cluster 2 had the least number of differentially abundant bacteria, as expected.

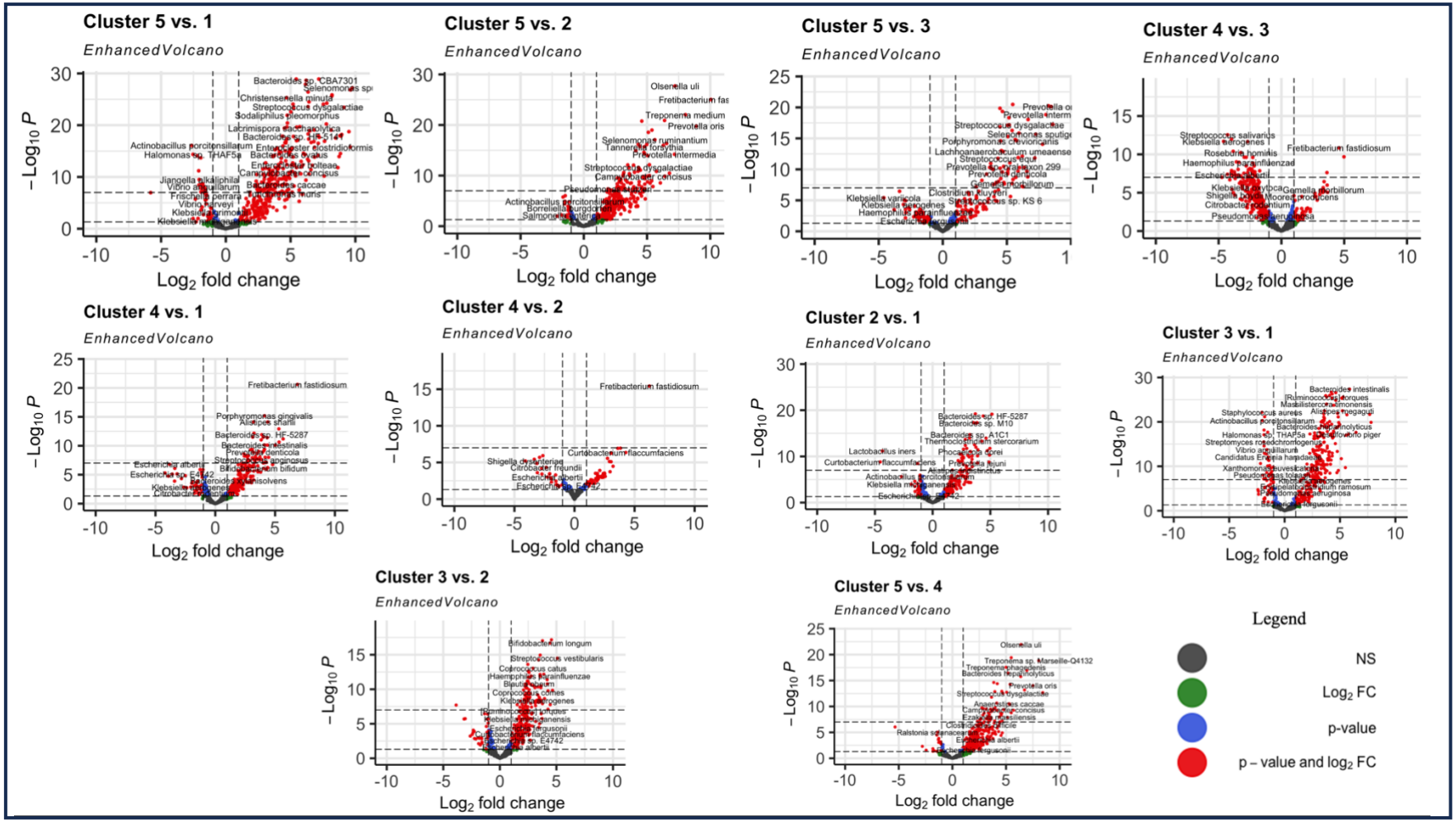


Figure 6.2*:*Volcano plots of differences in the bacterial composition of the 5 Louvain clusters. *The log2fold change threshold was between -1< log2 fold change>1 and a p-value threshold of 0.05. Red dots represent species that are significantly under or overrepresented based on both p-value and log2fold change thresholds, while green shows significant under or over abundance based on log2fold change only. In blue are those that only have a significant p-value. Grey dots show no significant difference based on p-value and in log2 fold change thresholds.Volcano plots were created using the EnhacedVolcano package in R.*

A cluster analysis was performed to determine how the bacterial species grouped. The bacteria species were also grouped using the t-SNE algorithm resulting in 10 bacterial species groups (Appendix 6.1). The heatmap shows the patterns of the different species groups and which metagenomic cluster they belong to (Figure 6.3).

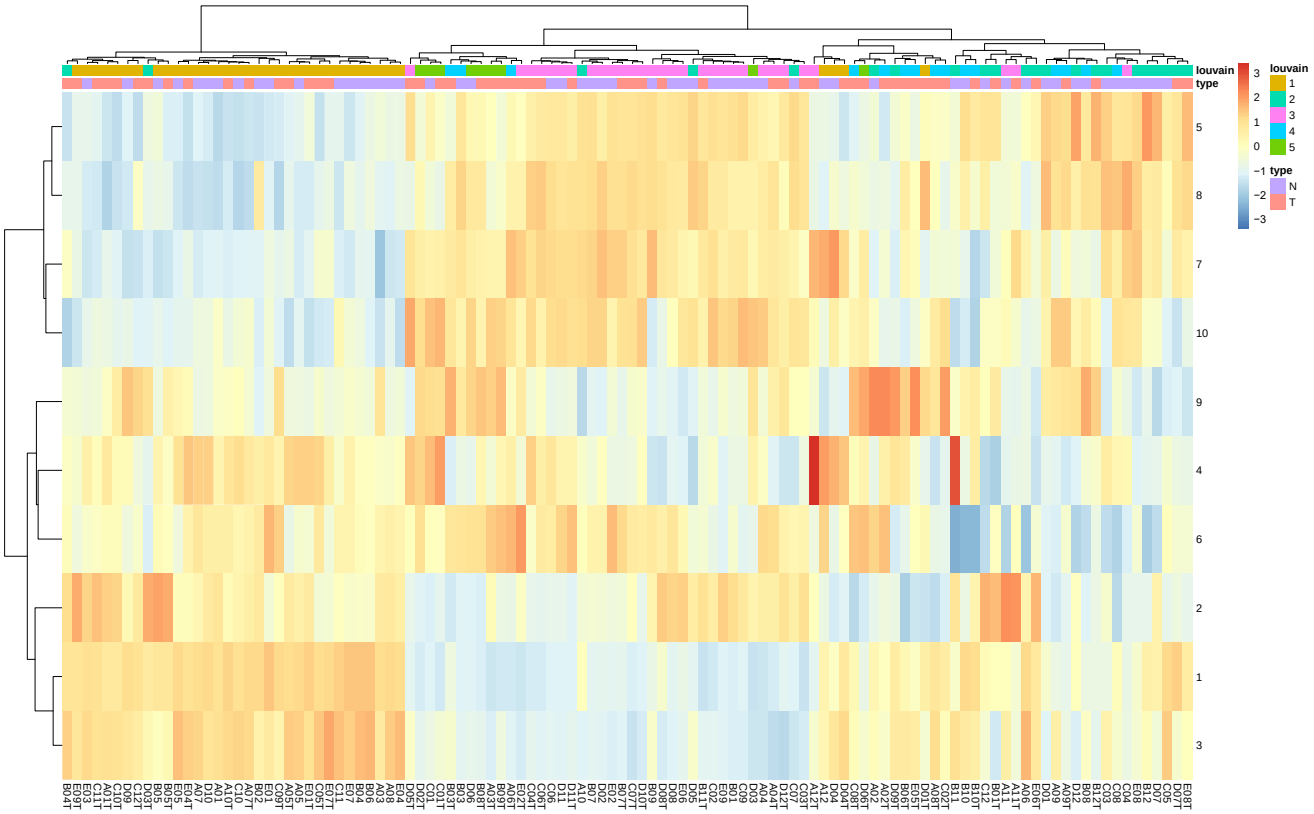


*Figure 6.3:**Heatmap of the bacterial species groups and their distribution across the louvain clusters**. Sample names are on the x-axis showing both normal and tumour samples. The ten groups of species are shown on the right y-axis while the left y-axis shows hierarchical clustering of the species groups. The louvain group and sample type for each sample type is shown at the top of the heatmap and hierarchical clustering of the samples according to the louvain group was applied. The colours of the bars of the heatmap indicate the abundance of a certain species group in a sample ranging from -3 which denotes low abundance to 3 which indicates high abundance.*

Bacterial groups 5, 8, 7 and 10 are common to Louvain cluster 3, while Louvain cluster 1 has an overrepresentation of species from groups 1, 2, 3, 4 and 6. Louvain clusters 2 and 4 overrepresent species of groups 1, 3, 5 and 8. A Fisher's exact test revealed a significant relationship (p-value < 2.2e-16) between the species groups and the Louvain clusters. However, the relationship between the sample type and species groups was not statistically significant (p-value ≥ 0.83).

6.6 Distribution of key pathogens across the Louvain clusters

The presence of bacteria between tumour and normal samples was compared, and normalised expression data was plotted on a t-SNE plot to visualise the distribution of the bacteria across the different metagenomics clusters. The distribution across the clusters of some important pathogens in CRC was evaluated (Figure 6.4). Two CRC-associated bacteria, *F. nucleatum* and *P. micra*, were evaluated. Additionally, the distribution of a commensal *Bifidobacterium breve* (*B. breve*) was also determined. Another bacterium reported to be more enriched in normal samples than control *Blautia producta* (*B. producta)* (Zhang et al., 2023) was also queried to see how its distribution across tissues varies across the Louvain clusters. *B. producta* showed widespread abundance across the Louvain clusters, whereas the other bacteria were enriched mainly in clusters 3 and 5. A wald t-test p-value comparing the abundance of the bacteria in normal compared to tumour samples was performed for each strain. *F. nucleatum* was significantly enriched in tumour samples compared to normal (log_2fold change 22585.1, adjusted p-value 9.0925e-3). The abundance in tumours compared to normals was also significant for *P. micra* (log_2fold change

2411.53 adjusted p values =0.0263902). *B. breve* and *B. producta* were not significantly different between normal and tumours with log_2fold changes of -1.49162 and 0.461066 and adjusted p values of 0.102519 and 0.381836, respectively. For *F. nucleatum* and *P. micra* samples that have low (blue dots. Figure 6.4) abundance of the bacteria are most likely the normal samples and the ones with high abundance (yellow/green dots) are most likely the tumour samples. *B. producta*, a common gut commensal (Yang et al., 2022), appears to have indiscriminate high abundance across all samples, except for a few samples in Louvain cluster 1.

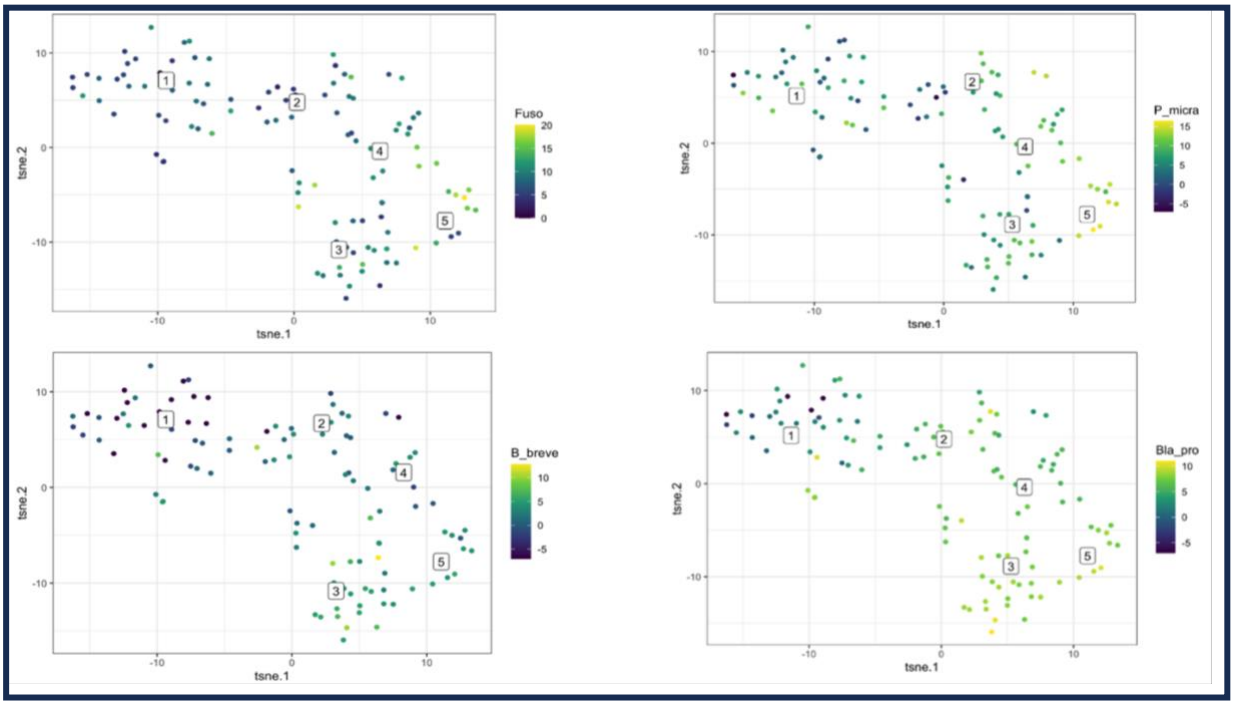


Figure 6.4: Distribution of some CRC-associated bacteria across the samples. *F. nucleatum (Fuso), P. micra (P. micra), B. breve(B_breve) and B. producta (Bla_pro) presence was visualised across the sample in the different metagenomic clusters. The intensity of the colour of the dots ranges from navy for low abundance samples to yellow/green for samples wih high abundance in the metagenomic cluster.*

6.7 Evaluation of the immune profile of the samples reveals significant downregulation of the M2 macrophage populations in normal sample.

Count matrix of combined normal and tumour gene expression data was used for differential gene expression analysis with the Limma package in R. (Ritchie et al., 2015). Then, the differential expression analysis results were used for immune cell expression analysis using the consensus TME package (Jiménez-Sánchez et al., 2019) with gene set variation analysis (gvsa) method to allow for assessment of variation between immune gene sets (Hänzelmann et al., 2013). Hierarchical clustering revealed that samples belonged to one of four immune clusters (Immune clusters 1-4) (Figure 6.5). Immune cluster 1 had an increased expression of NK cells, CD4 T-cells, and dendritic cells but had a downregulation in the expression of macrophages. Conversely, macrophages, including M1 and M2 macrophages, were overexpressed in cluster 3 and immune cluster 2. Immune cluster 2 had a downregulation in the expression of dendritic and NK cells. It also shows an increased expression of T-regulatory cells. Furthermore, immune cluster 4 cells have a particularly high expression of fibroblasts.

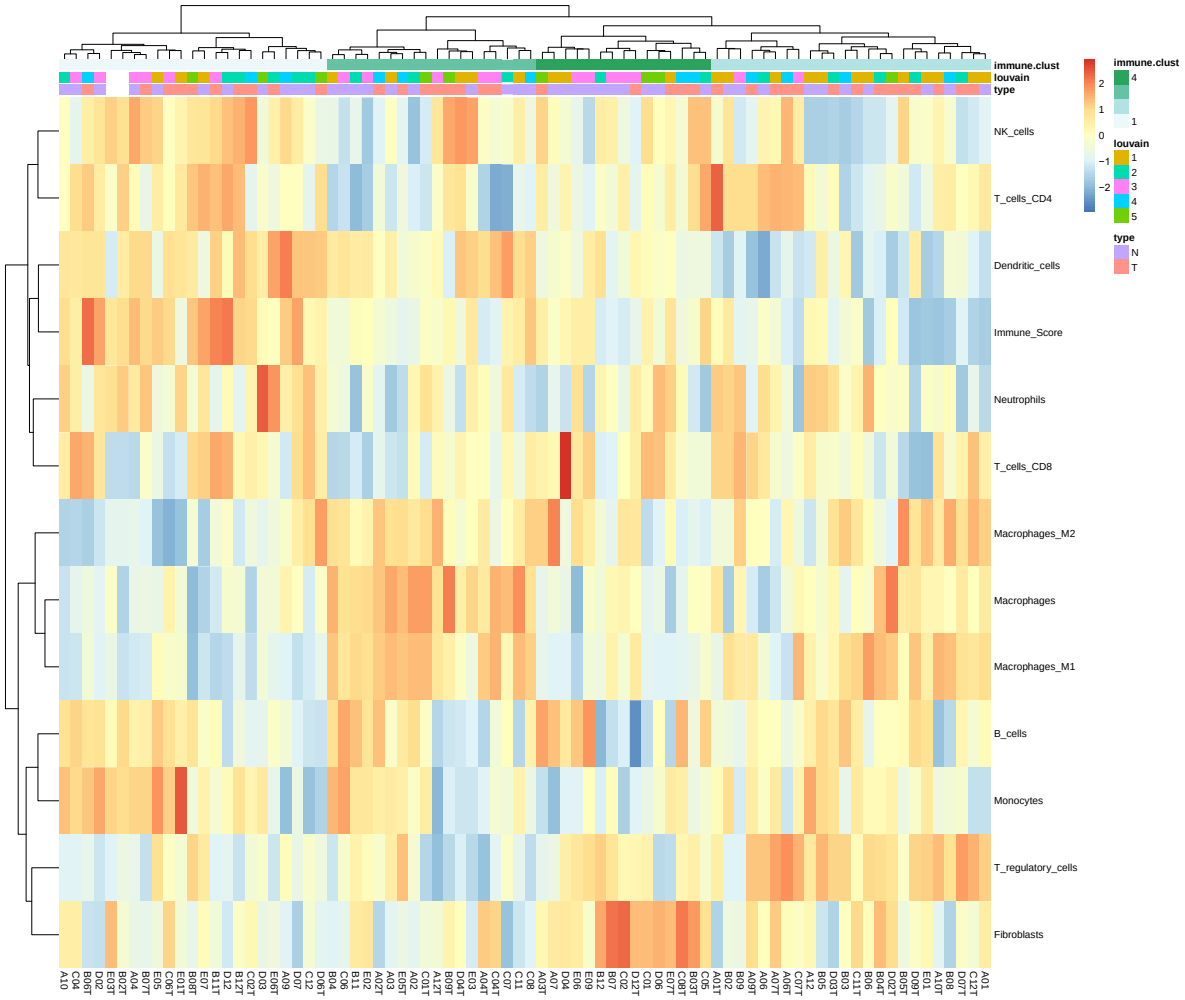


Figure 6.5:Immune profile of the samples (combines normal and tumour samples). *Heatmap shows the expression levels of immune cells in the samples. Sample names, for both normal and tumour samples are shown on the x-axis. The right y-axis show the immune cell types. Samples fall into one of the four immune groups and have been clusteres according to the immune group they belong to as depicted on the top of the heatmap. Louvain group and sample type are shown at the top of the heatmap. The sample type and the louvain group are also indicated on the heatmap.*

The immune cell scores between Louvain metagenomic clusters were compared (Table 6.3). The differences in immune scores between the clusters were not statistically significant (p adjusted<0.05). Based on raw p-values (Table 6.3), the expression of dendritic cells was higher in Clusters 2 and 3 than in Cluster 1. However, dendritic cell expression was decreased in Cluster 5 compared to Cluster 2 and to Cluster 3. Additionally, Cluster 5 had lower immune scores for T

regulatory cell expression than Cluster 1(Log FC=-0.470, p=0.0320). The same analysis was done for normal and tumour samples independently. The immune cell scores between sample clusters showed no statistically significant differences for tumours at an adjusted p-value of <0.05. When considering raw p values, Clusters 4 and 5 tumour samples had lower immune scores for dendritic cells than those in Cluster 3. Furthermore, Cluster 3 had higher dendritic cell immune scores than Cluster 1. Neutrophils were also differentially expressed, with decreased expression in Cluster 4 than in Cluster 1. Interestingly, when this was done for the normal samples only, macrophages and B cells but not dendritic cells, were differentially expressed. A statistically significant downregulation in M2 macrophages was observed (logFC -1.27 adjusted p-value =0.0195) when comparing immune scores of Clusters 5 vs Cluster 4. M1 macrophages also showed differential expression for the same Cluster combination; however, only the raw p-value was significant. Based on p values of <0.05, Cluster 5 showed under expression of Macrophages, M1 macrophages and M2 compared to Cluster 1. Cluster 3 also had higher M2 macrophage immune scores compared to Cluster 5. Additionally, Cluster 3 B cell immune scores surpassed those of Cluster 5. These data suggest that dendritic cells tend to be more differentially expressed in tumours while macrophages especially M2 macrophage differential expression are more associated with normal samples.

Table 6.3: Comparison of immune cell scores between Louvain metagenomic clusters

Louvain clusters (combined tumours and normals)						
Cluster	**Cell type**	**LogFC**	**Ave expr**	**t**	**P value**	**P adjusted**
2 vs 1	DC	0.364	0.0169	2.39	0.0290	0.248
3 vs 1	DC	0.402	0.0169	2.72	0.00778	0.101
5 vs 1	T-reg	-0.470	0.0144	-2.18	0.0320	-0.193
5 vs 2	DC	-0.424	0.0169	-2.17	0.032	0.370
5 vs 3	DC	-0.462	0.0169	-2.41	0.0179	0.23
Tumours only						
3 vs 1	DC	0.639	-0.0418	2.63	0.0115	0.149
4 vs 1	NEUT	-0.533	-0.0599	-2.08	0.0429	0.556
4 vs 3	DC	-0.600	-0.0418	-2.40	0.0203	0.263
5 vs 3	DC	-0.632	-0.0418	-2.34	0.0235	0.306
Normals only						
5 vs 1	M2 Mφ	-0.870	0.0215	-2.75	0.0080	0.0813
	M1 Mφ	-0.816	-0.269	-2.58	0.0125	0.0813
	Mφ	-0.531	-0.0371	-2.08	0.0418	0.168
5 vs 2	M2 Mφ	-0.804	0.0215	2.46	0.0170	0.221
5 vs 3	M2 Mφ	-0.8334	0.0215	-2.55	0.0135	0.176
	B cells	-0.688	0.0292	-2.26	0.0277	0.180
5 vs 4	M2 Mφ	-1.27	0.0215	-3.34	0.00150	0.0195
	M1 Mφ	-0.872	-0.0269	-2.296	0.0254	0.165

DC-dendritic cells, T-reg=T-regulatory cells, NEUT= neutrophils, Mφ = Macrophages, M1 Mφ= Macrophage_M1, M2 Mφ= Macrophages_M2, LogFc= log foldchange Padj = adjusted p value

6.8 Comparison of hallmark activity between the metagenomic (Louvain) clusters

To comprehensively explore the gene expression differences between the Louvain metagenomics clusters, a gene set enrichment analysis (GSEA) approach was used. This used 50 hallmark pathways from the Human Molecular Signatures Database (MSigDB) (https://www.gsea-msigdb.org/gsea/msigdb/human/collections.jsp). The 50 hallmark pathways are comprised of gene sets that represent certain biological processes and functions based on gene expression data (Liberzon et al., 2015). Based on the hallmark pathways, the relative enrichment scores of the 50 hallmark pathways were compared for the Louvain metagenomic clusters. The pathway myc targets v1 and G2M checkpoint were significantly differentially enriched for clusters 2 vs 1 and 4 vs 2 (Table 6.4) These pathways were negatively enriched for cluster 2 vs 1 but positively enriched for cluster 4 vs cluster 2, suggesting that cluster 2 has a downregulation in the candidate genes of these pathways. In contrast, cluster 4 has a significant upregulation in the candidate genes for these pathways compared to cluster 2. The E2F targets pathway was also significantly negatively enriched for cluster 2 vs cluster 1 (NES, -1.35), suggesting a downregulation in some genes of this pathway for cluster 2 relative to cluster 1. Significant differences were also found in comparing cluster 5 vs cluster 3 in the coagulation (NES, -1.39), and TNFA signalling via NFKB (NES, -1.38). Candidate genes involved in the pathway enrichment, normalised enrichment scores (NES (N+T)) and the corresponding p-values(N+T) are shown (Table 6. 1). Comparisons of the gene expression using hallmark pathways were also performed for tumours only to determine if any

tumour-specific pathways were differentially enriched. The same pathways (similar to in the combined tumours +normals) were differentially enriched in the same clusters. NES and p-values for tumours only are denoted with the letter (T) (Table 6.4). These pathways are negatively enriched in cluster 2 compared to cluster 1 with NES values of -1.36 and -1.45 for E2F targets and G2M pathways, respectively. Myc targets v1, which is important for cellular proliferation, is also significantly negatively enriched for cluster 2 compared to cluster 1 (NES = -1.62). In contrast, in cluster 4 vs cluster 2, G2M and myc targets v1 are positively enriched with respective NES values of 1.37 and 1.41, respectively. Cluster 5, on the other hand, had a negative enrichment in genes involved in the coagulation (NES, -1.37) and TNFA signalling via NFKB (NES, -1.37) compared to cluster 3.

Table 6.4:Pathways differentially enriched for combined normal and tumours (N+T) and for tumour(T) samples only

Cluster	Hallmark	Genes	P-value (N+T)	NES (N+T)	P - value(T)	NES (T)
2 vs 1	E2F targets	*PSMC3IP, MXD3, KIF4A, CKS2, PCNA, CDKN3*	0.0411	-1.35	0.0353	-1.36
	Myc targets v1	*RAD54L, E2F1, NOTCH2, CCND1, CCNA2, TGFB1*	0.00966	-1.60	0.000183	-1.62
	G2M checkpoint	*SNRPG, CCNA2, ABCE1, PHB, PCNA, HSPE1*	0.000373	-1.44	0.00702	-1.45
4 vs 2	G2M checkpoint	*RAD54L, E2F1, NOTCH2, CCND1, CCNA2, TGFB1*	0.0312	1.36	0.0360	1.37
	Myc targets v1	*CCNA2, SNRPG, PHB, AIMP2, ABCE1, HSPE1*	0.0170	1.40	0.0184	1.41
5 vs 3	Coagulation	*SERPINB2, MASP2, MMP10, HPN, MMP1, ITGB3*	0.0256	-1.39	0.0254	-1.37
	TNFA signalling via nfκb	*SERPINB2, ICAM1, REL, CXCL6, IL23A, CXCL10*	0.0100	-1.38	0.0261	-1.36

*N+T =combined tumours and normal samples dataset, T =tumours only, NES=Normalised enrichment score.

For the differences in gene expression for normal samples, Cluster 2 vs Cluster 1 had a positive enrichment in the hallmark apical surface (NES, 1.42). On the other hand, the hallmark coagulation was differentially enriched in Cluster 4 vs. Cluster 3 (Table 6.5). Taken together, the combined and independent enrichment analysis shows that more pathways are differentially expressed between tumour samples than between normal samples. Furthermore, similar pathways in the same clusters are differentially enriched for tumours and for tumours and normalised, suggesting that the tumour gene expression profile has more influence in the observed differences than the normal samples.

Table 6.5:Pathways differentially enriched for normal samples only

Cluster	Hallmark	Genes	P - value (Normals)	NES (Normals)
2 vs 1	Apical surface	*RTN4RL1, GHRL, CD160, IL2RG*	0.0459	1.42
4 vs 3	Coagulation	*TCIRG1, MRPL11, PDHA1, UQCRQ, IDH3G, ACO2*	0.0441	1.32

*NES = Normalised enrichement score

6.9 Comparison of hallmark activity between the clusters as a function of Tspan6

Tspan6 was demonstrated to play a strain-specific role in bacterial adhesion to CRC cells (Section 3.9). Therefore, the activity of hallmark pathways as a function of Tspan6 expression was

evaluated. Initially, analysis was performed on combined normal and tumour samples (Figure 6.6). Sample-specific differences were revealed; for example, samples with high Tspan6 expression showed increased activity in the hallmarks mtorc1 signalling, myc targets v1 and hallmark E2F targets. Those with decreased expression, on the other hand, had decreased activity in the same hallmark pathways.

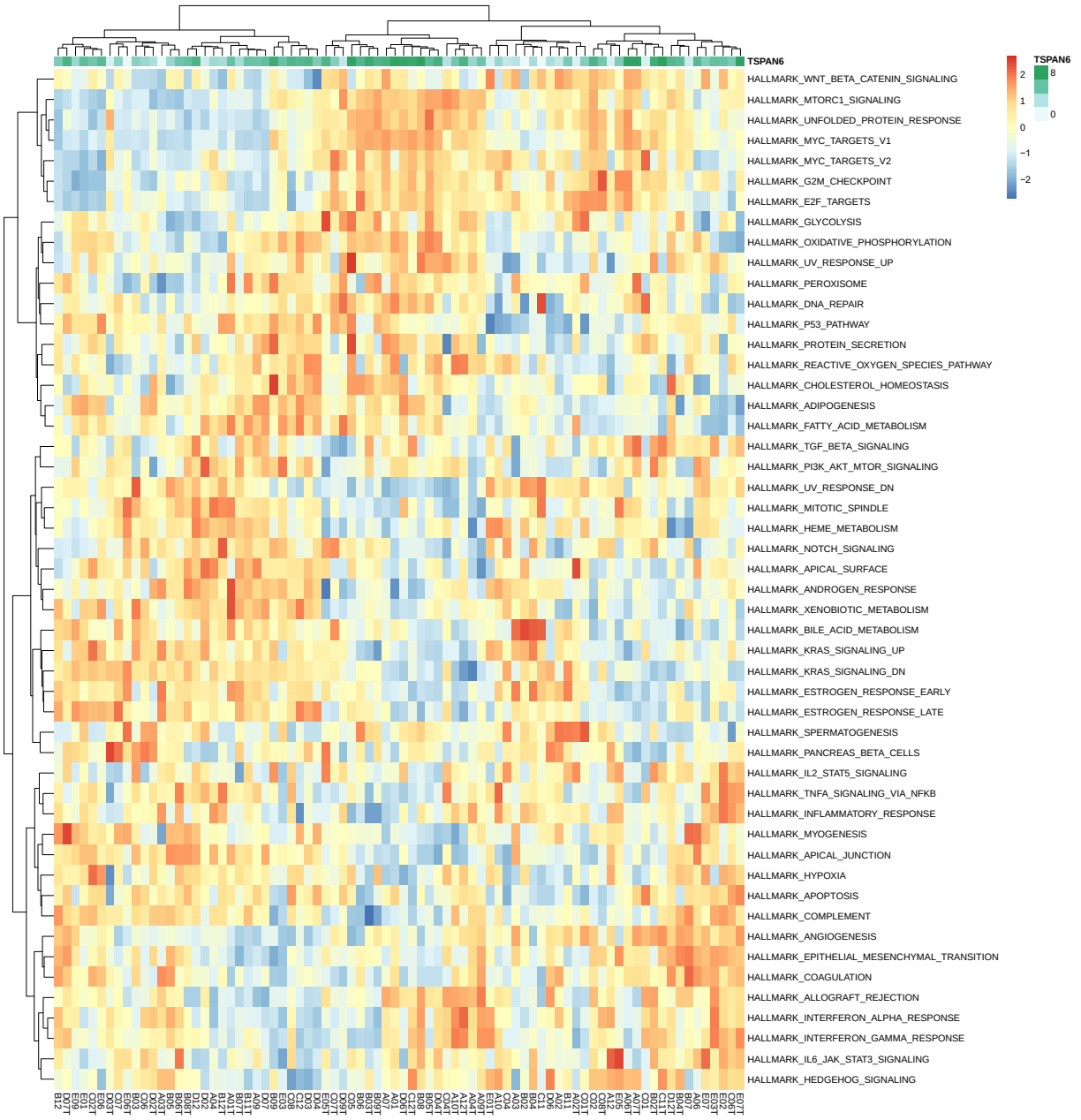


Figure 6.6:Heatmap of hallmark activity as a function of Tspan6 expression level. *Sample names for both tumours and normal samples are on the x-axis. The y-axis on the right shows the hallmark activity investigated. The bars of the heatmap show the levels of activity of the different hallmarks from low(-2) to high (2) and corresponding Tspan6 activity from low (0) to high (6) is shown at the top of the heatmap.*

Correlation analysis was performed to determine the global impact of Tspan6 expression on hallmark activity (Figure 6.7). Tspan6 expression was moderately positively correlated (Cor = ~0.5-0.6) with increased activity of myc targets v1, uv response up, mtorc1 signalling, peroxisome, phosphorylation, and protein secretion.

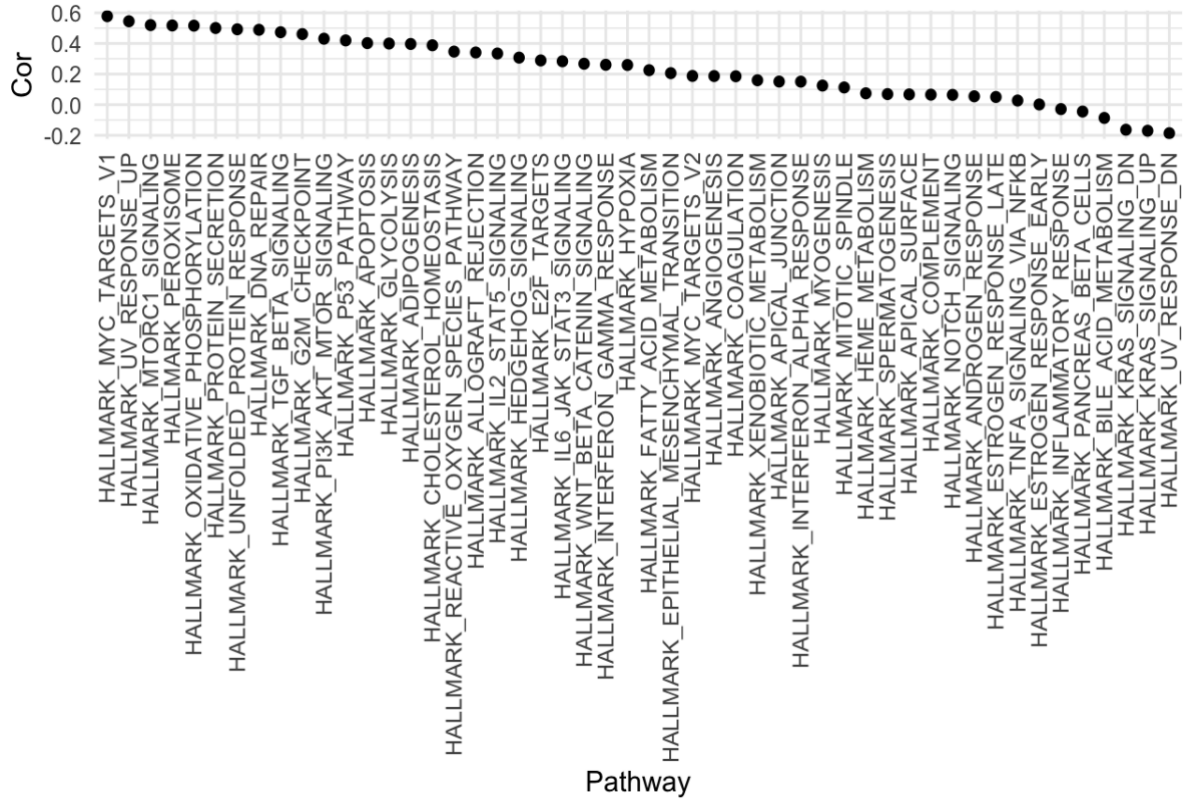


Figure 6.7: Panel showing correlation of tetraspanin expression with hallmark activities for all samples, normal and tumours. *The correlation value is shown on the y-axis and the hallmark pathways are shown on they x-axis.Correlation analysis were performed for the 50 hallmark pathways from the Human Molecular Signatures Database (section 6.8) using a gene set enrichment analysis (GSEA) approach.*

Next, the impact of the expression of Tspan6 on the activity of hallmark pathways between the Louvain clusters was determined. There were no statistically significant differences observed. Tumours and normals were then analysed separately to reduce confounding. A heatmap (Figure 6.9 right panel) shows differences in hallmark activity across tumour samples depending on Tspan6 expression levels. Increased Tspan6 expression in tumours was moderately positively correlated (Cor=0.5-0.6) with hallmark myc targets v1, hallmark peroxisome and hallmark TGFB signalling (Figure 6.8A). In contrast, when similar analyses were done for normal samples only, Tspan6 expression was moderately positively correlated ($0.7 < Cor \geq 0.6$) with hallmark uv response up, hallmark oxidative phosphorylation, and hallmark protein_secretion (Figure 6.9 left panel). The heatmap (Figure 6.8B) for normals only also shows sample-specific differences in hallmark activity across samples depending on Tspan6 expression levels.

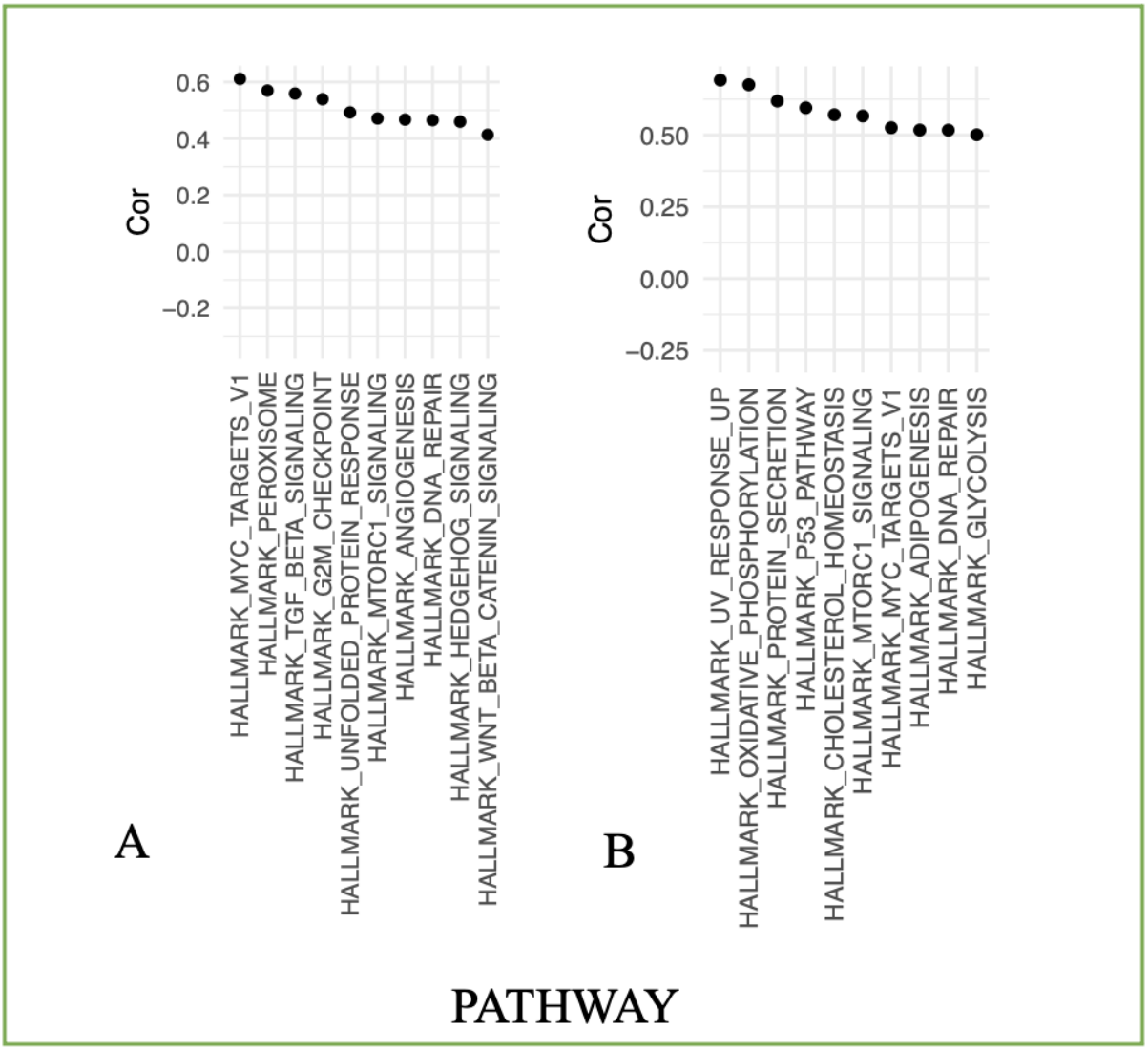


Figure 6.8: Panel showing the top (+ve) correlations of tetraspanin expression with hallmark activities for tumour (A) and normal (B) samples.*Pathway names are shown on the x-axis and the corresponding correlation value is shown on the y-axis. Correlation analysis were performed for the 50 hallmark pathways from the Human Molecular Signatures Database (section 6.8) using a gene set enrichment analysis (GSEA) approach.*

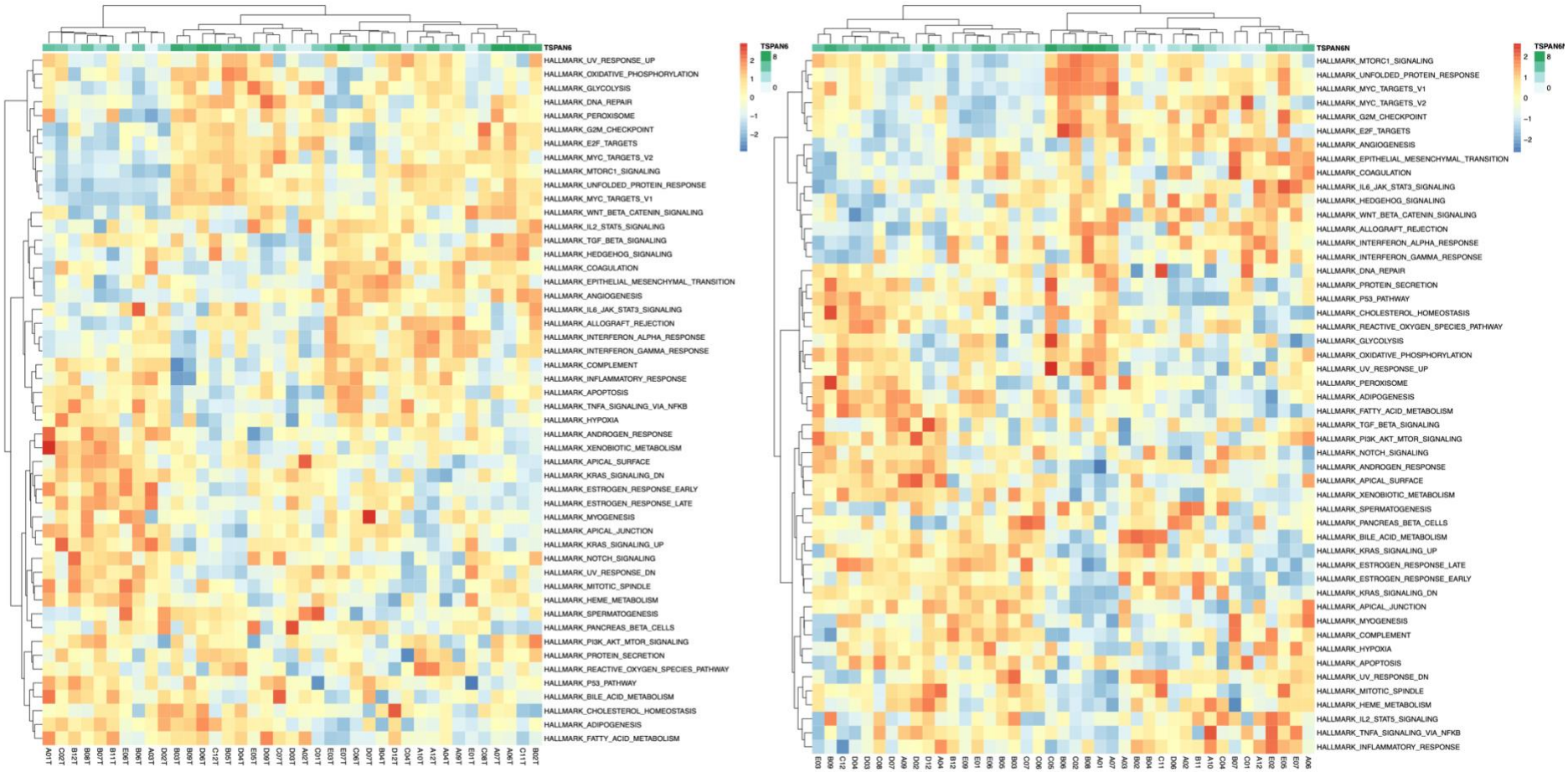


Figure 6.9: Heatmap of pathway activity as a function of Tspan6 expression intumours and normal sample. *Left panel shows activity for tumour samples only and right panel shows activity for Normal samples only. The x-axis contains the sample names, and the y-axis shows the 50 hallmark pathways. Each bar in the heatmap shows the activity of the hallmark pathways and the corresponding Tspan6 expression levels are shown at the top of the heatmap for each sample.*

6.10 Sample grouping based on the gene expression profile of the samples

Expression data was available for 80 samples, 41 normal samples and 39 tumour samples. The t-SNE program assessed how the samples grouped based on their expression clusters. Like with metagenomics data, five clusters were formed for the combined normal and tumour samples. As introduced in Chapter 1, CRC samples can be subtyped into CMS groups based on their gene expression profile (Chapter 1, section 1.4.3). The CMS prediction of every sample in the clusters was determined (Figure 6.10).

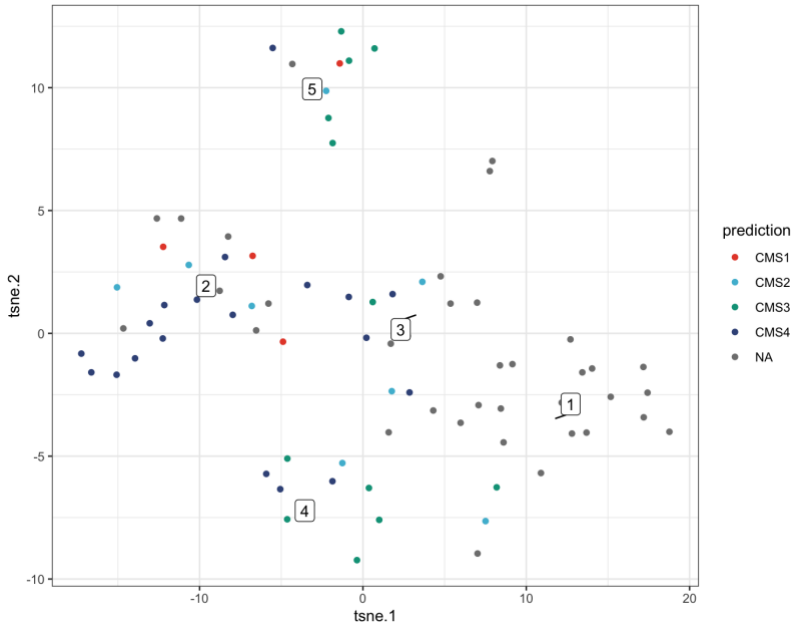


Figure 6.10: T-SNE clusters of samples based on their expression profile. *CMS prediction was performed for the samples and is indicated by colour on the t-SNE plot. Samples without a CMS prediction are indicated with grey dots. The numbers 1-5 are represent the five expression based clusters.*

Cluster 1 had a lot of samples that did not have a CMS prediction, whereas Cluster 2 had a lot of CMS4 samples. Clusters 3 and 4 both had a lot more samples with CMS3 prediction. Using the Limma package in R, gene expression results were filtered for samples with adjusted p-value <0.1. The number of differentially expressed genes between the clusters was determined (Table 6.6). Many differentially expressed genes were observed between Cluster 1 and clusters 2, 3, 4, and 5 suggesting that cluster 1 may have a significantly different expression profile compared to the other clusters. This aligns with CMS prediction data which showed that Cluster 1 had a lot of samples without a CMS prediction. The other major differences were between cluster 2 and clusters 3, 4 and 5. As demonstrated above (Figure 6.10), Cluster 2 had more CMS1 and CMS4 predicted samples than any other cluster suggesting a different expression profile. Only 81 genes were differentially expressed when comparing cluster 5 vs cluster 4. This was also not surprising as samples in these clusters had a similar CMS prediction pattern i.e., most of the samples had a CMS3 classification.

Table 6.6:Number of differentially expressed genes for the expression-based clusters.

Cluster	**2 v 1**	**3 v 1**	**4 v 1**	**5 v 1**	**3 v 2**	**4 v 2**	**5 v 2**	**4 v 3**	**5 v 3**	**5 v 4**
Number of DEG[5]	14218	12100	11103	10529	11006	6726	7019	2829	2732	81

[5] DEG = differentially expressed genes

6.11 Differential pathway analysis of expression-based clusters

Pathway analysis using 50 hallmark pathways from MSigDB was used to improve understanding of the processes in which the differentially expressed genes are involved. Significantly enriched pathways and corresponding p and p-adjusted values are shown (Table 6.8). Most comparisons except three (5 vs 2, 4 vs 3 and 5 vs 4) showed significant enrichment in some hallmark pathways (Table 6.8). The most differentially enriched pathways were myc targets v1 and oxidative phosphorylation pathways. These pathways were positively enriched in clusters 2, 3 and 5 compared to cluster 1. Cluster 4 vs 1 was also positively enriched in the myc targets v1pathway compared to Cluster 1. This suggests that cluster 1 has a significant reduction in the expression of genes involved in these pathways compared to all the other 4 clusters. This is unsurprising, as Cluster 1 has many CMS negative samples compared to the other clusters. Other clusters differed with cluster 1 in other pathways, such as hallmark apoptosis, p53 and hallmark mtorc1 signalling pathways. Cluster 1 was often negatively enriched in these pathways. As these pathways are cancer-associated, Cluster 1 may have samples with a normal-like phenotype. Additionally, hallmark myc targets activity was higher in Cluster 2 compared to Clusters 3 and 4. These comparisons were done on combined normal and tumour samples; therefore, they represent an unsupervised clustering based on the expression of the samples.

Table 6.7: Differentially enriched pathways between expression-based clusters

Cluster	Pathway	Genes	NES[6]	P value	P adjusted
2 v1	mtorc1_signaling	*DDIT3, PNP, NUFIP1, DDIT4, GAPDH, BUB1*	1.54	1.63e-03	2.721e-02
	myc_targets_v1	*AIMP2, PHB, CDK4, CCNA2, C1QBP, RANBP1*	1.71	4.011e-07	2.01e-05
	oxidative phosphorylation	*TCIRG1, ACAA2, AFG3L2, DLAT, BDH2, COX17*	1.57	7.98e-05	1.99e-03
3 v 1	oxidative phosphorylation	*TCIRG1, ACAA2, AFG3L2, BDH2, DLAT, NDUFA3*	1.64	0.000227	0.0113
	myc_targets_v1	*PHB, AIMP2, CDK4, RANBP1, C1QBP, SNRPG,*	1.58	0.000567	0.0142
4 v 1	oxidative phosphorylation	*TCIRG1, ACAA2, AFG3L2, BDH2, DLAT, COX17*	1.54	0.00137	0.0686
5 v1	apoptosis	*DDIT3, F2R, CCND1, GSTM1, ERBB2, BCL2L2*	1.54	3.00e-03	3.74e-02
	mtorc1_signaling	*DDIT3, PNP, GAPDH, BUB1, DDIT4, UCHL5*	1.61	3.58e-04	6.00e-03
	myc_targets_v1	*PHB, CDK4, CCNA2, C1QBP, AIMP2, RANBP1*	1.84	1.41e-06	7.04e-05
	oxidative_ phosphorylation	*TCIRG1, ACAA2, AFG3L2,*	1.74	2.07e-05	5.17e-04
	p53_pathway	*TRAF4, DDIT3, CEBPA, F2R, XPC, TGFB1*	1.45	6.23e-03	6.23e-02
3 v 2	myc_targets_v1	*CCNA2, AIMP2, CDK4, CCT5, HSPE1, C1QBP*	-1.77	4.88e-06	0.000244
4 v 2	E2F_targets	*PSMC3IP, MXD3, DEPDC1, PRPS1, PTTG1, MCM2*	-1.573	0.00112	0.0281
	myc_targets_v1	*SRM, AIMP2, MCM2, SNRPD1, MYC, PCNA*	-1.63	0.000355	0.0177
5 v 3	myc_targets_v1	*CCNA2, CDK4, CCT5,*	1.76	7.81e-05	0.00390

[6] Normalised enrichment score

Next, normal and tumour samples were clustered independently, resulting in 4 clusters for normal samples (Figure 6.11 left panel) and 3 for tumours (Figure 6.11 right panel). This was based on filtering for those with significant expression at $p < 0.05$.

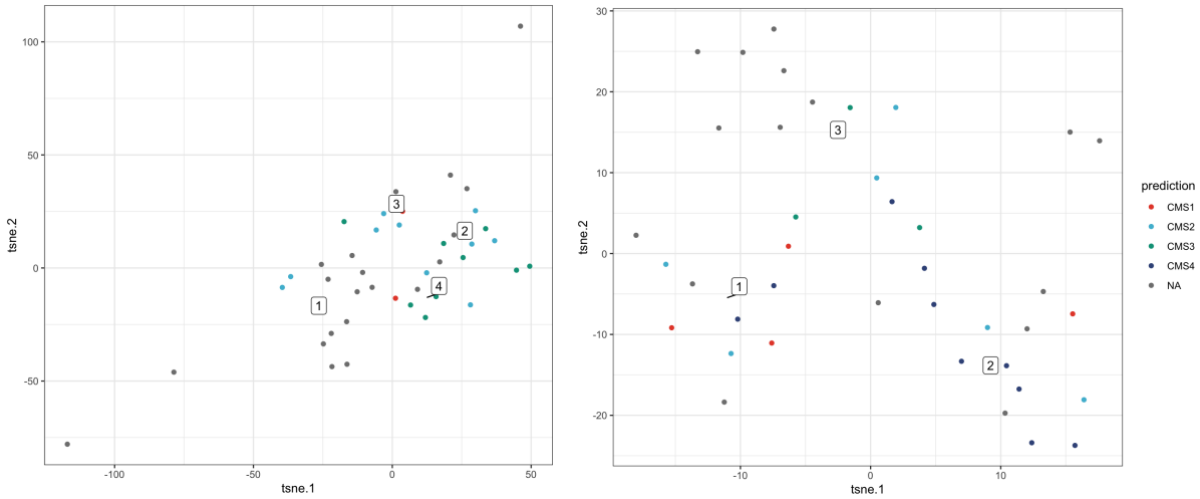


Figure 6.11*:*T-sne clusters of normal samples (left panel) and tumour samples (right panel) based on their expression profile. *Numbers on the graph indicate the different clusters. Four clusters were obtained for the normal samples and 3 clusters for the tumour samples. CMS prediction for each sample was performed for the samples and is indicated by colour on the t-SNE plot.Samples without a CMS prediction are indicated as N/A and have grey dots.*

The normal clusters had CMS-positive samples, mostly in Clusters 2, 3 and 4. Based on the premise that CMS typing is for tumours, this suggests that some normal samples may have an expression profile like tumours or that there has been "contamination" with tumour tissue in the Normals. Additionally, tumour-adjacent "normal" tissues can sometimes have some levels of transformation; therefore, they could return a positive CMS prediction. Tumour samples, on the other hand, formed three expression-based clusters (Figure 6.11 right panel). Interestingly, some of the samples in Cluster 3 were CMS-negative. The expectation was that tumour samples would have a CMS prediction. This could either be because some of the tumours have a normal-like phenotype, and therefore do not have a CMS prediction or that there was potential contamination with normal samples. Some tumours have been shown to maintain a "normal" like phenotype. The number of differentially expressed genes was determined for both the normal and tumour clusters, as shown (table 6.9). For the normal samples, there were a lot of differentially expressed genes between cluster one and the other three clusters .The highest difference was observed for cluster 1 vs cluster 32 with over 9162 differentially expressed genes. However, no differentially expressed genes were observed for cluster 2 vs cluster 4; only one was observed for cluster 2 vs cluster 3 and for cluster 3 vs cluster 4. This suggests that the other 3 clusters (clusters 2, 3, and 4) are more similar to each other than they are to cluster 1. Cluster 2 vs cluster 3 had the highest number of differentially expressed genes for the tumour samples at 12549. There were 9545 differentially expressed for cluster 1 v 2 and 7963 for Cluster 1 vs 3. The metagenomic profiles between the tumour clusters were done for the tumour sample. There are 16 differentially expressed species for the first and second comparisons. Finally, there were 11 for the last comparison.

Table 6.8:Number of differentially expressed genes between the clusters for normals and for tumours

Normals						
Cluster	1 v 2	1 v 3	1 v 3	2 v 3	2 v 4	3 v 4
Number of DEG (Normals)	9162	6741	4726	1	0	1
Tumours						
Cluster	1 v 2	1 v 3	2 v 3			
Number of DEG (Tumours)	9545	7963	12539			

6.12 Differences in gene expression and pathway enrichment between samples with high and low numbers of *F. nucleatum*

Species-specific influences on the activity of hallmark pathways were explored using *F. nucleatum*. Differences in the activity of hallmark pathways between samples with high *F. nucleatum* and those with low *F. nucleatum* scores were determined. To do this, samples with *F. nucleatum* scores higher than the median were given a score of 1 to represent high *F. nucleatum* abundance, and those that were less than or equal to the median were given a score of 0 and were considered low *F. nucleatum* samples. Of 50 hallmark pathways, 37 were differentially enriched in high compared to low *F. nucleatum* samples (table 6.10). All 37 pathways were positively and statistically significantly enriched. A higher abundance of *F. nucleatum* is associated with increased positive enrichment in many hallmark activities. Two pathways, hallmark oxidative phosphorylation and hallmark apoptosis, had two of the highest net enrichment scores (3.43 adjusted p-value 5.80e-23) and (3.03, adjusted p-value 1.80e-15), respectively.

Table 6.10: Differences in pathway activity of samples with high vs low *F. nucleatum* abundance

Pathway	p value	P adjusted	Enrichment score	size
Adipogenesis	2.16276e-12	1.85997e-11	2.723726893	184
hallmark_allograft_rejection	0.000344065	0.000510165	1.782389309	163
hallmark_androgen_response	8.19643e-13	8.81116e-12	2.8058555	95
hallmark_angiogenesis	0.005557826	0.006828186	1.781398315	30
hallmark_apical_junction	0.000205425	0.000327158	1.811699445	174
hallmark_apical_surface	0.021384134	0.024851832	1.568382673	34
hallmark_apoptosis	8.35663e-17	1.79668e-15	3.034767663	148
hallmark_bile_acid_metabolism	0.002340459	0.003049689	1.802844568	87
hallmark_coagulation	0.000120743	0.000211383	2.01569646	105
hallmark_complement	1.01974e-05	2.08805e-05	2.004464891	174
hallmark_dna_repair	6.48954e-05	0.000121326	2.029245883	144
hallmark_epithelial_mesenchymal_transition	7.92873e-09	3.40936e-08	2.444253098	183
hallmark_estrogen_response_early	3.33429e-08	1.19479e-07	2.388918806	183
hallmark_estrogen_response_late	7.97777e-06	1.71522e-05	2.127799888	179
hallmark_fatty_acid_metabolism	5.14723e-07	1.35115e-06	2.273813857	140
hallmark_glycolysis	3.08782e-11	2.21294e-10	2.640506381	181
hallmark_heme_metabolism	5.34174e-07	1.35115e-06	2.135355186	171
hallmark_hypoxia	2.79916e-09	1.33738e-08	2.36121657	171
hallmark_il2_stat5_signaling	1.8861e-09	1.01378e-08	2.426481599	176
hallmark_il6_jak_stat3_signaling	0.000154479	0.000255484	1.99374859	78
hallmark_inflammatory_response	0.003353367	0.004241023	1.563105831	166
hallmark_interferon_alpha_response	0.000892014	0.00123731	1.800624202	94
hallmark_kras_signaling_up	7.63259e-06	1.71522e-05	1.972894225	168
hallmark_myc_targets_v2	0.001384073	0.001859849	1.875767701	58
hallmark_myogenesis	0.000378992	0.000543221	1.780804526	154
hallmark_notch_signaling	0.027273121	0.030861689	1.624182832	30
hallmark_oxidative_phosphorylation	1.34872e-24	5.79949e-23	3.425644504	184
hallmark_peroxisome	1.56103e-05	3.05111e-05	2.114070454	92
hallmark_pi3k_akt_mtor_signaling	3.10678e-08	1.19479e-07	2.390280394	99
hallmark_protein_secretion	2.8793e-14	4.127e-13	2.838012658	93
hallmark_reactive_oxygen_species_pathway	0.000122897	0.000211383	2.067017696	45
hallmark_spermatogenesis	0.007592272	0.009068548	1.694958716	79
hallmark_tgf_beta_signaling	4.12882e-08	1.36569e-07	2.472393959	53
hallmark_tnfa_signaling_via_nfkb	4.14305e-11	2.54501e-10	2.612573025	184
hallmark_unfolded_protein_response	2.153e-06	5.14327e-06	2.301209198	106
hallmark_uv_response_dn	2.70126e-07	8.05419e-07	2.325903809	138
hallmark_uv_response_up	2.8096e-07	8.05419e-07	2.308348706	141

hallmark_xenobiotic_metabolism	0.000254404	0.000390691	1.80165159	164

6.13 Correlation between Louvain clusters with tumour status and CMS prediction

Lastly, the associations between Louvain metagenomic clusters with various clinicopathological features were investigated to better understand the relationship between the clusters and patient characteristics. The association of the clusters with tumour T stage, regional lymph node (N), CMS prediction Tumour T (T) stage was tested. The association of Louvain clusters with N status was significant as determined by a two-sided Fishers Exact test for count data (p- value= 0.0142 (Figure 6.12 A). The same test revealed insignificant associations with tumour T stage (p value = of 0.878) (Figure 6.12 B). T Stage 1 was made up entirely from Louvain clusters 1 and 3 samples. When the association between the Louvain clusters and CMS prediction was investigated, it was revealed that over 50% of the samples in CMS1 belonged to Louvain cluster 3, and no samples from Louvain clusters 4 and 5 had this prediction. Furthermore, 50% of CMS2 samples were from Louvain cluster 4 and none from Louvain 2 or Louvain 5. CMS3 and CMS4 had a lot more diversity of represented Louvain clusters. However, a Fisher exact test revealed insignificant associations between Louvain and CMS prediction (p value= 0.6609). The relationship between the Louvain clusters and tumour EMVI status was also determined. Of the 5 clusters, Cluster 4 had the most EMVI-positive tumours, with over 50 % of the Cluster 4 samples having EMVI yes (y) status. Cluster 1 and 3 have

the least number of EMVI positive samples, with under 0.25 samples having an EMVI positive status.

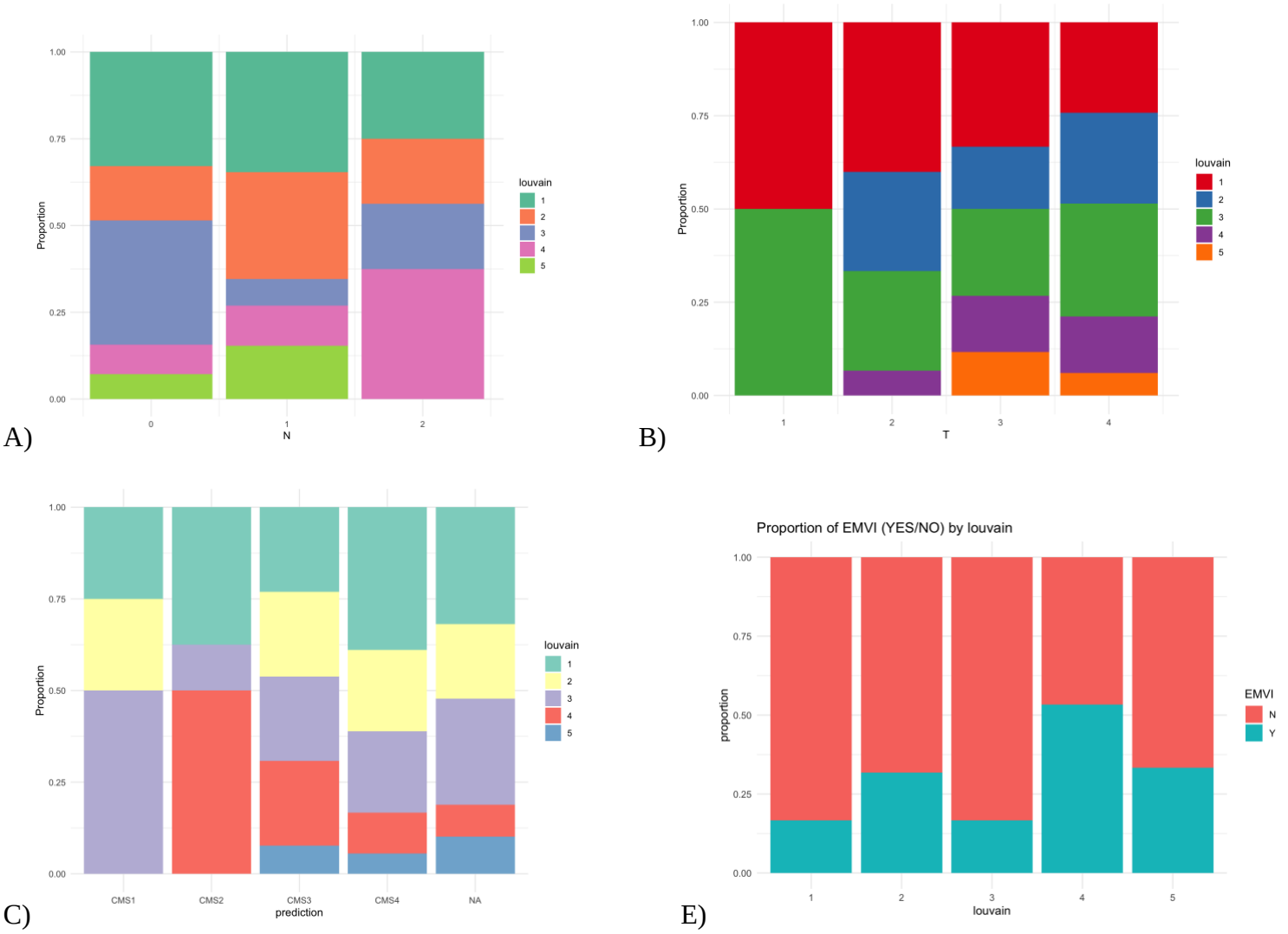


Figure 6.12: Associations of Louvain clusters with clinicopathological features. *Association of Louvain clusters with A) N stage, B) T stage, C) CMS prediction and D) Tumour EMVI status. Proportion of association of the different features is shown on the y-axis of each graph and the queried features are on the x-axis.*

6.14 Discussion

This study used paired metagenomics and transcriptomic data from tumour and normal samples of 57 CRC patients to gain insights into the relationship between bacteria and CRC transcriptomic landscape. Firstly, the differences in the metagenomic profile of tumours compared to normal samples were evaluated. There was significant enrichment of Fusobacteria and Campylobacter species in tumours compared to normal samples. *F. nucleatum* was the most significantly over-abundant species in the dataset. This aligns with several publications that have reported the overrepresentation of *F. nucleatum* in tissues and faecal samples of CRC patients compared to healthy controls (Castellarin et al., 2012;Kostic et al., 2013;Nakatsu et al., 2015) . Three Campylobacter species; *rectus*, *showae* and *gracilis* were among the top 10 significantly enriched species tumours compared to normal samples. Others have also reported the association of Campylobacter species, including *C. showae* and *C. rectus* with CRC (Han and Wang, 2013;Hsu et al., 2019). Another Campylobacter species, *Campylobacter jejuni* has been associated with CRC and is believed to mediate CRC tumorigenesis through the action of its toxin; the cytolethal distending toxin (CDT) (He et al., 2019). Interestingly, Warren et al., reported the co-occurrence of Fusobacterium, Campylobacter and Leptotrichia in CRC, which is in line with findings from this work, where the top ten most significantly enriched bacteria were from these three genera (Warren et al., 2013).

The top ten most significantly depleted bacteria in tumour compared to normal samples consisted of bacteria with diverse roles in disease and the gut. One of the depleted bacteria, *Phocaeicola vulgatus* is normally over-abundant in the healthy human gut and contributes to better outcomes for patients with colitis (Evans et al., 2022;Liu et al., 2022). For one of the depleted bacteria, *Alistipes sp. dk3624,* there is still a lack of consensus on whether its role in tumorigenesis in CRC, if any, is beneficial or detrimental (Parker et al., 2020). *Sutterella wadsworthensis*, another significantly depleted species in CRC tumours compared to controls, was previously isolated from patients with ulcerative colitis. However, it is believed to be more of a bystander in conditions such as colitis and IBD (Mukhopadhya et al., 2011). *Priestia megaterium*, also significantly depleted in the tumour samples of the current dataset, has been shown to produce arginase, which can act as an anti-cancer agent through an arginine depletion mechanism (Jiao et al., 2023). In the same vein, *Parabacteroides goldsteinii*, which is a gut commnesal, has been associated with improved outcomes in patients with metabolic disorders such as obesity and has anti-virulence activity against *Helicobacter pylori* (Wu et al., 2019b;Lai et al., 2022). This suggests that it may have anti-inflammatory and antagonistic effects on oncogenic pathogens. Similarly, *Akkermansia muciniphila* a mucin-degrading gut bacterium, also under-abundant in the CRC sample set, has been linked with benefits for patients with diabetes (Cani et al., 2022). Overall, enriched bacteria in the study were those with reported links to CRC, and the depleted ones are more associated with better health outcomes for certain conditions. This could be because the CRC microenvironment is favourable to bacteria such as *F. nucleatum* and allows them to thrive better than some gut commensal bacteria.

Next, the combined normal and tumour samples were grouped using the Louvain algorithm following dimensionality reduction by t-SNE, forming five Louvain metagenomic clusters. Significant differences were found in the abundance of bacterial species between the Louvain metagenomic clusters. Differences in the immune scores of the Louvain clusters were determined to establish whether variations in the bacterial composition affected the immune profile. There were no statistically significant differences in the immune scores of the candidate immune cells between the clusters. This was true even when comparisons were made for tumour samples only. However, when normal samples in the clusters were compared, there was a statistically significant downregulation in M2 macrophage immune scores of Clusters 5 compared to Cluster 4. M2 macrophages have anti-inflammatory properties and have been implicated in tumour promotion (Boutilier and Elsawa, 2021, Jayasingam et al., 2019). Sickert et al. observed an increase in tumour-associated macrophages in CRC tumours compared to normal cells (Sickert et al., 2005). Therefore, a downregulation in M2 macrophages for some of the normal samples in the clusters is on par with observations by others. Based on raw p values, there was differential expression of dendritic cells, while in Normals most of the differences were in macrophages. Dendritic cells are important in the tumour environment for tumour antigen presentation activation of other immune cells such as T cells necessary for anti-tumour responses (Del Prete et al., 2023). M2-macrophages have tumour-promoting, immune system-dampening effect in cancer and have been associated with increased tumour invasiveness and can promote angiogenesis (Jayasingam et al., 2019;Lin et al., 2019;Ahmed and Ismail, 2020). This data suggests that differences in metagenomic profiles can impact immune cell infiltration to the tumour microenvironment.

Then, to investigate the impact of the metagenomic profiles on the gene expression of the samples, the enrichment scores of 50 hallmark pathways were compared between the clusters. The pathways that were differentially enriched included the cell cycle-related pathways, E2F targets, and hallmark G2M. Furthermore, myc targets v1, TNFA signalling via NF-KB were also differentially enriched. At the G2M checkpoint, cell cycle arrest occurs before cells move from the G2 to M phase DNA to allow for DNA damage repair before cells can proceed through the cell cycle (2011). The E2F targets pathway, on the other hand, has roles in cell cycle regulation, angiogenesis, and apoptosis (Xie et al., 2021). Hallmark coagulation, another differentially enriched pathway, comprises genes that code for the coagulation pathway. In contrast, hallmark TNFA signalling via NF-KB genes comprises of NFKB-regulated genes because of TNF activity. These genes are involved in many functions, including cell survival, immune structure, and activity (Webster and Vucic, 2020). It is therefore likely that the metagenomic profile of the samples can impact the function of key pathways that have consequences on cancer progression and DNA damage repair. In previous chapters, the importance of Tspan6 in CRC adherence was established. The differences in the expression of Tspan6 between the metagenomic Louvain clusters were investigated to determine whether there was an association between Tspan6 expression and the metagenomic profile of the samples. No significant differences were found in Tspan6 expression between the Louvain clusters for combined samples and even when the tumour and normal samples in the clusters were compared independently. This suggests that there might not be a direct relationship between Tspan6 expression and the metagenomic composition of tumours and their normal controls. It is possible that the differences in Tspan6 expression could be patient/sample-specific rather than a community effect. Then, the influence of Tspan6 expression on the activity of hallmark genes was evaluated. High Tspan6 expression was moderately correlated

with increased activity in the hallmark pathways: myc targets v1, uv response up, mtorc1 signalling, peroxisome, oxidative phosphorylation, protein secretion and TGF-B signalling.

The samples were then grouped based on their expression profile to delve deeper into the expression data. Five expression-based clusters were formed, and these clusters had statistically significant differences in the enrichment of some pathways in the 50 hallmark pathway gene set. Differentially enriched pathways included the myc targets v1, oxidative phosphorylation, mtorc1, p53, E2F targets and apoptosis. Interestingly, 3 of these pathways, myc targets v1, and E2F targets, were among some of the differentially enriched pathways based on metagenomic clustering. This consistency in the difference emphasises a potential relationship between the expression of genes in these pathways and the metagenomic profile of the samples. The expression clusters were subjected to CMS prediction, a subtyping method based on gene expression data. All CMS subtypes were represented in the data, including CMS-negative samples. This is in contrast to Purcell et al., who could not detect CMS4 in their tumour samples when they used the CMS molecular prediction method (Purcell et al., 2017). The expression data was further queried by creating clusters of tumours and normal samples independent of each other based on the expression profile of the samples. Tumour samples formed three clusters, while the normal samples only formed four gene expression-based clusters. The groups were also subjected to CMS subtyping. Interestingly, three out of the four clusters of the normal clusters had CMS-positive samples. The three CMS positive “normal” clusters showed no significant differences between the number of differentially expressed genes, but all showed significant differences when compared to cluster 1, which contained CMS negative samples. Adjacent normal cells can sometimes undergo malignant changes that are not morphologically

obvious, a phenomenon known as Field cancerisation first described in 1953 (Slaughter et al., 1953). Others have demonstrated that in line with the Field characterisation theory, the gene expression profile of normal adjacent tissue can even be useful in predicting prognostic outcomes of patients with CRC (Kim et al., 2023). Additionally, some tumour-only clusters had no CMS classification, which could indicate a more "normal" phenotype. Others who have used the CMS classification reported some tumours not having CMS classification and purported that this could result from intratumoral heterogeneity or be due to inherent methodological differences of their study and the original study (Purcell et al., 2017). In their study, Purcell et al. also looked at the association of key bacteria with CMS subtypes, showing that bacteria such as *F. nucleatum* were associated more with CMS1 (Purcell et al., 2017). However, the approach of the current study focused on groups of bacteria rather than individual bacterial species when considering the associations of metagenomics with CMS. Therefore, associations of individual species with CMS subtypes were not determined.

Finally, the association between the Louvain clusters and various patient clinicopathological data was explored. The Louvain metagenomic clusters showed a statistically significant association with the tumour T stage but not with the N stage. Mima and colleagues demonstrated a strong association between *F. nucleatum* in tissues and the T status of tumours (Mima et al., 2016). However, they found the relationship between *F. nucleatum* burden and NM status insignificant. This aligns with the findings of this study, where T status was found to have a significant association with metagenomic clusters while N status was not. Though similar results were observed, it is worth noting that the focus here was on more than one species but on groups of bacterial species. For the dataset used in this study, all the samples had an M

status of 0, implying that no metastasis had been recorded for this cohort. The TNM status of tumours has consequences for patient prognosis and overall survival (Kim et al., 2017). Due to a lack of up-to-date clinicopathological data, survival analysis could not be performed. Therefore, the impact on the survival of the metagenomic profile of samples could not be determined.

6.15 Conclusion

Microbial composition differences between CRC tumours and normal samples point to an important role of bacteria in CRC. This study, consistent with others, has shown an enrichment of bacterial species such as *F. nucleatum* in CRC tissues compared to normal controls. Furthermore, it was demonstrated here that the metagenomic profile of CRC samples can lead to differences in gene expression and the tumour immune microenvironment. The impact and influence on gene expression and the patient clinicopathology of microbial profile of samples were evaluated and revealed an association of the metagenomic profile with the tumour T stage. Understanding the impact of the bacterial metagenome on different aspects of tumorigenesis can help decipher the processes and mechanisms by which bacteria such as *F. nucleatum* contribute to tumorigenesis in CRC.

7 GENERAL DISCUSSION

The main aim of this study was to evaluate the role of bacteria in CRC with a targeted focus on the bacterium *F. nucleatum*. First, the adhesion and invasion dynamics of three *F. nucleatum* subspecies (FNN23, FNP and FNA) were determined using a Caco-2 cell model. This was followed by proteomic profiling of the inflammatory mediators secreted upon *F. nucleatum* subspecies adhesion and invasion to Caco-2 cells. Additionally, transcriptomic analyses were used to determine the gene expression profile of the subspecies used at different growth phases. Finally, metagenomic and transcriptomic analysis of data from 57 CRC samples and their paired normal controls was performed to explore the impact of microbial dysbiosis on aspects of CRC, such as immune cell infiltration, activity of hallmark genes and patient clinicopathology. The aims of this study were met and will provide insights into *F. nucleatum* interaction with CRC cells. The main findings, future directions and limitations of the study are discussed in this chapter.

7.1 Adhesion and invasion dynamics of *F. nucleatum*

The enrichment of *F. nucleatum* in CRC tumours has garnered much interest recently. However, only a few studies describe to the subspecies level, the *F. nucleatum* found in or isolated from tumours (Bi et al., 2021;Queen et al., 2021;Morsi et al., 2022). Additionally, most experimental studies investigating various aspects of the biology and virulence characteristics of *F. nucleatum* in CRC tend to focus on one of the subspecies; subspecies *nucleatum* (Ben Lagha et al., 2017;Wang et al., 2023a;Wu et al., 2023). This is unsurprising, as you would expect subspecies to behave in similar or representative ways. However, reports have found differences in the representation of the subspecies in CRC tumours; for example, Bi and

colleagues reported that ssp. *animalis* was more commonly isolated than the others in their study (Bi et al., 2021). As a result, this thesis investigated the adhesion and invasion dynamics of *F. nucleatum* to Caco-2 cells using three subspecies (FNN23, FNA and FNP). The hypothesis was that the subspecies would adhere to and invade Caco-2 cells at varying levels. Indeed**,** adhesion and invasion assays revealed differences in the adherence of the *F. nucleatum* subspecies tested. FNP was the most highly adherent subspecies, but the differences in invasion were not statistically significant. FNPs different adherence levels compared to the other two subspecies is interesting as FNP has also been shown to differ from other *F. nucleatum* subspecies in its ability to form biofilms (Muchova et al., 2022). Understanding the factors that contribute to the differential interaction of the subspecies with CRC cells could contribute to insights into the mechanism of action of *F. nucleatum* in CRC, which largely remains enigmatic.

Invasion and adhesion assays are often done with either exponential or stationary phase bacteria depending on the experiment design and research question. However, how the growth phase impacts levels of adherence and invasion in *F. nucleatum* has not yet been elucidated. Therefore, Chapter 4 of this thesis looked at the impact of the bacterial growth phase on the adherence and invasion levels of the subspecies and found growth phase-dependent differences in adherence and invasion for some of the subspecies. FNN23 was especially sensitive to growth phase differences, while FNP adherence remained unaffected by the growth phase. Based on the subspecies-specific differences in adherence, the hypothesis was that the subspecies may have differences in the expression of adhesins or other factors necessary for interacting with host tissues. Furthermore, the improvement of adherence for FNN23 and FNA

at the exponential phase suggested that the adhesins may be increased at the exponential phase for the two subspecies. Mutants in the Fap2 and FadA adhesins of FNN23 were then used to explore the role of adhesins in adherence to Caco-2 cells. In alignment with reports from others, there was a significant reduction in the adherence and invasion of the two mutant strains to Caco-2 cells compared to the wild-type strain (Ikegami et al., 2009;Abed et al., 2016b). Interestingly, the exponential phase FNN23 Δ*fadA* had increased adherence, compared to the stationary phase, while the exponential phase FNN23 Δ*fap2* had decreased adherence to Caco-2 cells relative to stationary phase cells. The improved adherence of FNN23 at the exponential phase, combined with the observed increase in adherence for the FadA mutant at the exponential phase, suggests that Fap2 expression levels might be higher at the exponential phase than FadA levels. However, the combined impact of the mutations could not be tested due to the lack of a double mutant. Additionally, complementation studies could not be performed due to time constraints.

Along with uncovering pathogen characteristics, such as growth phase, that may impact adherence, host factors that may do the same were also investigated. Therefore, the role of Tspan6, a tetraspanin protein recently shown to contribute to the early stages of CRC through EGFR-mediated signalling, was determined (Andrijes et al., 2021). Subspecies-specific differences were observed in the interaction of bacteria with Tspan6-expressing cells. Tspan6 expression improved adherence of some strains (FNN23 wild-type strain, FadA mutant strain) and significantly reduced that of others (FNA and FNP). Subspecies-specific differences in invasion to T6 cells were also observed, with the invasion of most of the strains (FNN23, FNA, Fap2 mutant strain) to T6 cells being decreased. When two control strains (*S. epidermidis*, and

Sgg) were used in adherence and invasion, no significant impact in adherence and invasion were found. This suggests that the observed Tspan6 impact on adhesion and invasion may be specific to *F. nucleatum* subspecies. More strains must be tested to see if this suggestion holds for other bacteria. The exact mechanism of action of Tspan6 in interaction with bacteria remains elusive. Studies with another member of the tetraspanin family CD9 demonstrated that CD9 coordinates with syndecan-1 to organise fibronectin into adhesion platforms that allow for *S. aureus* adhesion (Green et al., 2023a). Because Tspan6 is associated with syndecan 4 (Ghossoub et al., 2020), a similar mechanism may exist, where Tspan6 associates with syndecan proteins and impacts the levels of adherence of *F. nucleatum*. Interactions of the subspecies with host factors such as Tspan6 may be partly through adhesins on the bacterial surface. We hypothesised that the observed subspecies-specific differences in adherence and in interactions with host factors (e.g., Tspan6) may be due to variations in the expression of adhesins. One of the adhesins could be Fap2, whose deletion significantly reduced FNN23 adherence to Caco-2 cells. Even though the FadA deletion also reduced adherence and invasion, it was not to the same extent as the Fap2 deletion. FadA in *F. nucleatum* has been shown to have multiple orthologues (FadA2 and FadA3), which may compensate, to a degree, for the deletion of the other ortholog (Umaña et al., 2019). Adhesion and invasion experiments revealed insights into the differences in the *F. nucleatum* subspecies. In the future, other cell lines could be used as controls to establish whether the observed results are cell line specific. Additionally, this work could be made more sophisticated by using 3D organoid cultures or tumour spheroids for adherence and invasion experiments. These would represent a more physiologically relevant model than monolayers due to improved cellular interactions and communication.

7.2 Impact of adherence and invasion on the release of inflammatory mediators

The impact on the release of inflammatory mediators by Caco-2 cells was explored to determine if differences in adherence translated to differential cytokine release by the subspecies. In a more physiological set-up, there are interactions between different cells, impacting cytokine release and regulation in the tumour microenvironment. In this study, cytokine release by epithelial cells, in the absence of immune cells, was investigated using the Olink target 48 panel. A total of 3 cytokines were significantly differentially expressed out of 45 cytokines, of which two (CXCL10 and IL17C) were highly expressed in the *F. nucleatum* subspecies compared to the *Sgg* control. As *F. nucleatum* subspecies had shown subspecies specific differences in adherence, the expectation was that the difference may lead to variations in the levels of secreted cytokines. However, this was not the case as the differences in cytokine secretion between the *F. nucleatum* subspecies were not statistically significant. This could be explained by the fact that Caco-2 cells were infected for 24 hours for cytokine profiling, whereas in the adherence assays, the infection length was either 1 hour or 4 hours. However, it is worth mentioning that only FNP and FNN23 had significantly higher induction of IL17C than the *Sgg* strain, while FNA did not, which may reflect some subspecies-specific effects. Even though both *F. nucleatum* and *Sgg* are associated with CRC, data from adhesion assays and the cytokine secretion experiments show that their interaction with the host cells differs in many aspects. This may be because the bacteria employ different CRC-associated virulence mechanisms. The impact of the bacterial growth phase on cytokine expression revealed

variations in the induction of IL6, with higher expression levels in exponential-phase bacteria relative to stationary-phase bacteria. Some of the tested cytokines showed levels below the limit of detection and this may be because some of the cytokines are secreted more by immune cells and not epithelial cells. In the future, co-culturing Caco-2 cells and bacteria with immune cells such as neutrophils and/or macrophages would help inform understanding of the molecular immune crosstalk between bacteria, epithelial cells, immune mediators, and immune cells, which are all critical aspects of the CRC tumour microenvironment. A major limitation of this part of the study was that only two biological and experimental replicates were used, which may impact the statistical power of the analysis and prevent meaningful conclusions from being drawn. To augment this work, ELISA experiments using CXCL10, IL17C and IL6 could be carried out, using different infection times and different MOIs to gain a better appreciation of their differential induction in *F. nucleatum* and *Sgg.*

7.3 Transcriptional analysis of *F. nucleatum* subspecies reveals differences in the expression of adhesion genes

To test the hypothesis that differences in adherence and invasion were due to variations in the expression of adhesins, we performed RNA sequencing of the *F. nucleatum* subspecies at exponential and stationary phases. As expected, there were gene expression differences between stationary and exponential phase cultures. Additionally, as per our hypothesis, differences in the expression of some adhesin and adhesin-related genes were observed. Though informative, outside of the known adhesin genes, this data is not definitive regarding which differentially expressed genes contribute to the observed differences in adherence and invasion. For example, FNP had the most differentially expressed genes of the exponential vs stationary phase, which was interesting as FNP had not shown sensitivity to changes in the growth phase when adherence assays were performed. How this gene expression profile translates to absence of observable changes in adherence at exponential vs stationary phase, cannot be deduced from the RNA sequencing experiment. This is because sequencing only the bacteria does not account for alterations that occur during the interactions of the bacteria with epithelial cells. Bacterial genes may be expressed and regulated differently when in contact with epithelial cells than in culture without epithelial cells. To address this, others have used dual RNA sequencing to examine infection-related transcriptomic changes (Westermann et al., 2012). Simillar to the cytokine profiling experiments, this part of the study was limited by using only two biological replicates per subspecies, which may impact the statistical power of the analysis and affect meaningful biological interpretations from the data. Furthermore, annotation errors in the reference genome impacted detection of some key adhesins such as Fap2.

7.4 Bioinformatic analytics to understand the impact of microbial dysbiosis on CRC

The final result chapter of this study evaluated the role that different bacterial metagenomic profiles have on the CRC environment. *F. nucleatum* is one of the most enriched bacteria in CRC*;* however, it is not the only CRC-associated bacterium. The possible contributing mechanism to CRC of other bacteria, such as *Sgg*, *Bacteroides fragilis* and *P. micra*, has been reported in the literature (Kordahi et al., 2021;Kamali et al., 2022;Zhao et al., 2022b). These bacteria may act in concert or have synergistic associations in the tumour environment that promote CRC. Indeed, some bacteria have been found to frequently co-occur in cancerous tumours compared to normal tissues (Xu et al., 2023). Clusters containing different species representing different metagenomic profiles of 57 CRC samples and 57 controls were compared to each other to better understand the role of microbial dysbiosis in CRC. Differences in immune cell scores and differences in the gene expression profiles of the samples were observed. Tumour samples in the clusters had alterations in dendritic cells, whereas, with the normal samples, there were alterations in macrophage populations. This illustrated that immune-associated differences are affected by the metagenomic profile of samples, suggesting that bacteria can influence the structure and composition of the immune environment. Gene expression differences were explored using 50 hallmark pathways comprising different gene sets representing various processes and/or pathways. Some hallmark genes that were differentially enhanced, such as the myc pathways, E2F, G2M checkpoints and apoptosis, are pathways generally associated with carcinogenesis. This suggests that certain bacterial profiles

can lead to a dysregulation or differential regulation and activity of pathways necessary for different aspects of CRC. Interestingly, metagenomic clusters did not show differences in their Tspan6 expression. The bioinformatics data was also used to further understand *F. nucleatum* pathogenesis in CRC. The impact on hallmark pathway expression of samples with high compared to those with low *F. nucleatum* abundance was evaluated. Analysis revealed significant enrichment for high *F. nucleatum* samples in 37 pathways, including cancer-related pathways such as apoptosis pathway, myc targets, G2M checkpoints, to name a few. This suggests that *F. nucleatum* carriage may illicit global gene expression alterations and dysregulation of normal processes, contributing to CRC. Furthermore, alterations of over 30 pathways show how complex and multifaceted the role of microbial dysbiosis in CRC is. How microbial dysbiosis relates to various clinicopathological features is still an unexplored area. We found a strong relationship between the bacterial metagenomic profile and the tumour T stage. In CRC, the T stage is correlated with survival, and advanced T stages (T3, T4) are usually associated with a poor prognosis (Chen et al., 2021). This suggests that specific metagenomic profiles may be associated with larger primary tumours while others may not. More analysis based on the clinicopathological data, such as survival analysis, could not be performed due to a lack of up-to-date survival metadata. Additionally, time constraints limited deeper exploration of the gene expression data to look at specific genes and signalling pathways that may be altered.

7.5 *F. nucleatum*, cause or consequence

Whether *F. nucleatum* is involved in the initiation or progression of CRC remains unclear. Additionally, it is still unclear whether bacteria in tumours result from the tumour microenvironment being more conducive for colonisation by these bacteria or if they are present as critical initiators of malignancy. Animal models have shown in $Apc^{Min/+}$ mice that *F. nucleatum* infection increases the CRC tumours frequency and size. However, as these mice are already CRC-prone, they allude more to the role of *F. nucleatum* as a promoter of carcinogenesis rather than an initiator. In this study, we demonstrate that *F. nucleatum* can bind to Caco-2 cells and induce cytokines that have cancer-promoting effects. Furthermore, in alignment with other studies, we show that *F. nucleatum* is one of the most abundant species in CRC tumours compared to the control samples. The gut is undoubtedly not *F. nucleatums* primary site of colonisation; therefore, its presence and enrichment in tumours point to a more substantial and/or specific role of this bacterium in CRC. *F. nucleatum* has been associated with other cancers, such as breast cancer, gastric cancer, and oral cancer, to name a few. It has also been associated with conditions such as IBD, colitis and rheumatoid arthritis (Liu et al., 2020;Hong et al., 2023). One common factor of all these conditions is altered inflammation, so that *F. nucleatum* may be attracted to areas of increased inflammation. For instance, *F. nucleatum* is one of the bacteria that are strongly associated with periodontitis, which is a condition that is hallmarked by chronic inflammation (Hoare et al., 2019). Various aspects of *F. nucleatum* pathogenesis in CRC have been described in the literature, including the role of Fap2 and FadA in CRC. However, more is still to be understood about the relationship of the bacterium with CRC. This project's objectives were met and contribute knowledge into the

general impact of dysbiosis on aspects of CRC. It also contributes to more specific knowledge on the dynamics of adherence and invasion of *F. nucleatum* subspecies to CRC cells.

8 REFERENCES

ABANCENS, M., BUSTOS, V., HARVEY, H., MCBRYAN, J. & HARVEY, B. J. 2020. Sexual Dimorphism in Colon Cancer. *Frontiers in Oncology,* 10.

ABED, J., EMGÅRD, J. E., ZAMIR, G., FAROJA, M., ALMOGY, G., GRENOV, A., SOL, A., NAOR, R., PIKARSKY, E., ATLAN, K. A., MELLUL, A., CHAUSHU, S., MANSON, A. L., EARL, A. M., OU, N., BRENNAN, C. A., GARRETT, W. S. & BACHRACH, G. 2016a. Fap2 Mediates Fusobacterium nucleatum Colorectal Adenocarcinoma Enrichment by Binding to Tumor-Expressed Gal-GalNAc. *Cell Host Microbe,* 20**,** 215-25.

ABED, J., EMGÅRD, JOHANNA E. M., ZAMIR, G., FAROJA, M., ALMOGY, G., GRENOV, A., SOL, A., NAOR, R., PIKARSKY, E., ATLAN, KARINE A., MELLUL, A., CHAUSHU, S., MANSON, ABIGAIL L., EARL, ASHLEE M., OU, N., BRENNAN, CAITLIN A., GARRETT, WENDY S. & BACHRACH, G. 2016b. Fap2 Mediates Fusobacterium nucleatum Colorectal Adenocarcinoma Enrichment by Binding to Tumor-Expressed Gal-GalNAc. *Cell Host & Microbe,* 20**,** 215-225.

ADVANI, S. M., ADVANI, P. S., BROWN, D. W., DESANTIS, S. M., KORPHAISARN, K., VONVILLE, H. M., BRESSLER, J., LOPEZ, D. S., DAVIS, J. S., DANIEL, C. R., SARSHEKEH, A. M., BRAITHWAITE, D., SWARTZ, M. D. & KOPETZ, S. 2019. Global differences in the prevalence of the CpG island methylator phenotype of colorectal cancer. *BMC Cancer,* 19**,** 964.

AFRA, K., LAUPLAND, K., LEAL, J., LLOYD, T. & GREGSON, D. 2013. Incidence, risk factors, and outcomes of Fusobacterium species bacteremia. *BMC Infect Dis,* 13**,** 264.

AHMED, I. & ISMAIL, N. 2020. M1 and M2 Macrophages Polarization via mTORC1 Influences Innate Immunity and Outcome of Ehrlichia Infection. *J Cell Immunol,* 2**,** 108-115.

AITCHISON, A., PEARSON, J. F., PURCELL, R. V., FRIZELLE, F. A. & KEENAN, J. I. 2022. Detection of Fusobacterium nucleatum DNA in primary care patient stool samples does not predict progression of colorectal neoplasia. *PLoS One,* 17**,** e0269541.

ALBERS, T., MANIAK, M., BEITZ, E. & VON BÜLOW, J. 2016. The C Isoform of Dictyostelium Tetraspanins Localizes to the Contractile Vacuole and Contributes to Resistance against Osmotic Stress. *PLOS ONE,* 11**,** e0162065.

ALJGHAMI, M. E., BARGHASH, M. M., MAJAESIC, E., BHANDARI, V. & HOURY, W. A. 2022. Cellular functions of the ClpP protease impacting bacterial virulence. *Front Mol Biosci,* 9**,** 1054408.

ALRESHIDI, M., DUNSTAN, H., MACDONALD, M., SAEED, M., ELKAHOUI, S. & ROBERTS, T. 2023. Significant Changes in Cytoplasmic Amino Acid Composition Occur in the Transition between Mid-Exponential and Stationary Phases of Growth of Staphylococcus aureus: An Example of Adaptive Homeostasis in Response to Nutrient Limitations. *Microorganisms,* 11**,** 147.

ALVES MARTINS, B. A., DE BULHÕES, G. F., CAVALCANTI, I. N., MARTINS, M. M., DE OLIVEIRA, P. G. & MARTINS, A. M. A. 2019. Biomarkers in Colorectal Cancer: The Role of Translational Proteomics Research. *Frontiers in Oncology,* 9.

AMITAY, E. L., CARR, P. R., JANSEN, L., ROTH, W., ALWERS, E., HERPEL, E., KLOOR, M., BLÄKER, H., CHANG-CLAUDE, J., BRENNER, H. & HOFFMEISTER, M. 2020. Smoking, alcohol consumption and colorectal cancer risk by molecular pathological subtypes and pathways. *British Journal of Cancer,* 122**,** 1604-1610.

ANDERSON, N. M. & SIMON, M. C. 2020. The tumor microenvironment. *Curr Biol,* 30**,** R921-r925.

ANDRIJES, R., HEJMADI, R. K., PUGH, M., RAJESH, S., NOVITSKAYA, V., IBRAHIM, M., OVERDUIN, M., TSELEPIS, C., MIDDLETON, G. W., GYŐRFFY, B., BEGGS, A. D. & BERDITCHEVSKI, F. 2021. Tetraspanin 6 is a regulator of carcinogenesis in colorectal cancer. *Proc Natl Acad Sci U S A,* 118.

ARORA, S., LI, X., HILLHOUSE, A., KONGANTI, K., LITTLE, S. V., LAWHON, S. D., THREADGILL, D., SHELBURNE, S. & HOOK, M. 2020. Staphylococcus epidermidis MSCRAMM SesJ Is Encoded in Composite Islands. *mBio,* 11.

ARUNI, A. W., DOU, Y., MISHRA, A. & FLETCHER, H. M. 2015. The Biofilm Community-Rebels with a Cause. *Curr Oral Health Rep,* 2**,** 48-56.

AVRIL, M. & DEPAOLO, R. W. 2021. "Driver-passenger" bacteria and their metabolites in the pathogenesis of colorectal cancer. *Gut Microbes,* 13**,** 1941710.

AZEEM, S., GILLANI, S. W., SIDDIQUI, A., JANDRAJUPALLI, S. B., POH, V. & SYED SULAIMAN, S. A. 2015. Diet and Colorectal Cancer Risk in Asia--a Systematic Review. *Asian Pac J Cancer Prev,* 16**,** 5389-96.

BADDER, L. M., HOLLINS, A. J., HERPERS, B., YAN, K., EWAN, K. B., THOMAS, M., SHONE, J. R., BADDER, D. A., NAVEN, M., ASHELFORD, K. E., HARGEST, R., CLARKE, A. R., ESDAR, C., BUCHSTALLER, H.-P., TREHERNE, J. M., BOJ, S., RAMEZANPOUR, B., WIENKE, D., PRICE, L. S., SHAW, P. H. & DALE, T. C. 2020. 3D imaging of colorectal cancer organoids identifies responses to Tankyrase inhibitors. *PLOS ONE,* 15**,** e0235319.

BAI, M., CHEN, X. & BA, Y. 2016. CXCL10/CXCR3 overexpression as a biomarker of poor prognosis in patients with stage II colorectal cancer. *Mol Clin Oncol,* 4**,** 23-30.

BAIK, J. E., LI, L., SHAH, M. A., FREEDBERG, D. E., JIN, Z., WANG, T. C. & HAN, Y. W. 2022. Circulating IgA Antibodies Against Fusobacterium nucleatum Amyloid Adhesin FadA are a Potential Biomarker for Colorectal Neoplasia. *Cancer Research Communications,* 2**,** 1497-1503.

BARDOU, P., MARIETTE, J., ESCUDIÉ, F., DJEMIEL, C. & KLOPP, C. 2014. jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics,* 15**,** 293.

BEATTY, W. L. 2006. Trafficking from CD63-positive late endocytic multivesicular bodies is essential for intracellular development of Chlamydia trachomatis. *J Cell Sci,* 119**,** 350-9.

BECIC, A., LEIFELD, J., SHAUKAT, J. & HOLLMANN, M. 2021. Tetraspanins as Potential Modulators of Glutamatergic Synaptic Function. *Front Mol Neurosci,* 14**,** 801882.

BEN LAGHA, A., HAAS, B. & GRENIER, D. 2017. Tea polyphenols inhibit the growth and virulence properties of Fusobacterium nucleatum. *Scientific Reports,* 7**,** 44815.

BENNETT, K. W. & ELEY, A. 1993. Fusobacteria: new taxonomy and related diseases. *J Med Microbiol,* 39**,** 246-54.

BI, D., ZHU, Y., GAO, Y., LI, H., ZHU, X., WEI, R., XIE, R., WEI, Q. & QIN, H. 2021. A newly developed PCR-based method revealed distinct Fusobacterium nucleatum subspecies infection patterns in colorectal cancer. *Microbial Biotechnology,* 14**,** 2176-2186.

BILLER, L. H. & SCHRAG, D. 2021. Diagnosis and Treatment of Metastatic Colorectal Cancer: A Review. *JAMA,* 325**,** 669-685.

BLONDEL, V. D., GUILLAUME, J.-L., LAMBIOTTE, R. & LEFEBVRE, E. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment,* 2008**,** P10008.

BONASONI, M. P., COMITINI, G., PATI, M., BARDARO, M., RUSSELLO, G., CARRETTO, E., DALLA DEA, G., PALICELLI, A., BERNARDELLI, G., CHESI, E. & GARGANO, G. 2022. Fulminant Sepsis and Perinatal Death at 23 Weeks Due to Fusobacterium nucleatum. *Fetal Pediatr Pathol***,** 1-8.

BOYLE, T. 2012. Physical Activity and Colon Cancer:Timing, Intensity, and Sedentary Behavior. *American Journal of Lifestyle Medicine,* 6**,** 204-215.

BRADBURY, K. E., MURPHY, N. & KEY, T. J. 2020. Diet and colorectal cancer in UK Biobank: a prospective study. *Int J Epidemiol,* 49**,** 246-258.

BRANGER, J., KNAPP, S., WEIJER, S., LEEMANS, J. C., PATER, J. M., SPEELMAN, P., FLORQUIN, S. & VAN DER POLL, T. 2004. Role of Toll-like receptor 4 in gram-positive and gram-negative pneumonia in mice. *Infect Immun,* 72**,** 788-94.

BRENNAN, C. A. & GARRETT, W. S. 2019. Fusobacterium nucleatum - symbiont, opportunist and oncobacterium. *Nat Rev Microbiol,* 17**,** 156-166.

BRENNER, D. R., RUAN, Y., SHAW, E., DE, P., HEITMAN, S. J. & HILSDEN, R. J. 2017. Increasing colorectal cancer incidence trends among younger adults in Canada. *Preventive Medicine,* 105**,** 345-349.

BRODERICK, P., CARVAJAL-CARMONA, L., PITTMAN, A. M., WEBB, E., HOWARTH, K., ROWAN, A., LUBBE, S., SPAIN, S., SULLIVAN, K., FIELDING, S., JAEGER, E., VIJAYAKRISHNAN, J., KEMP, Z., GORMAN, M., CHANDLER, I., PAPAEMMANUIL, E., PENEGAR, S., WOOD, W., SELLICK, G., QURESHI, M., TEIXEIRA, A., DOMINGO, E., BARCLAY, E., MARTIN, L., SIEBER, O., KERR, D., GRAY, R., PETO, J., CAZIER, J.-B., TOMLINSON, I., HOULSTON, R. S. & MEMBERS OF THE, C. C. 2007. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. *Nature Genetics,* 39**,** 1315-1317.

BUCCAFUSCA, G., PROSERPIO, I., TRALONGO, A. C., RAMETTA GIULIANO, S. & TRALONGO, P. 2019. Early colorectal cancer: diagnosis, treatment and survivorship care. *Critical Reviews in Oncology/Hematology,* 136**,** 20-30.

BUI, F. Q., JOHNSON, L., ROBERTS, J., HUNG, S. C., LEE, J., ATANASOVA, K. R., HUANG, P. R., YILMAZ, Ö. & OJCIUS, D. M. 2016. Fusobacterium nucleatum infection of gingival epithelial cells leads to NLRP3 inflammasome-dependent secretion of IL-1β and the danger signals ASC and HMGB1. *Cell Microbiol,* 18**,** 970-81.

CANI, P. D., DEPOMMIER, C., DERRIEN, M., EVERARD, A. & DE VOS, W. M. 2022. Akkermansia muciniphila: paradigm for next-generation beneficial microorganisms. *Nature Reviews Gastroenterology & Hepatology,* 19**,** 625-637.

CANTALAPIEDRA, C. P., HERNÁNDEZ-PLAZA, A., LETUNIC, I., BORK, P. & HUERTA-CEPAS, J. 2021. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution,* 38**,** 5825-5829.

CASASANTA, M. A., YOO, C. C., UDAYASURYAN, B., SANDERS, B. E., UMAÑA, A., ZHANG, Y., PENG, H., DUNCAN, A. J., WANG, Y., LI, L., VERBRIDGE, S. S. & SLADE, D. J. 2020. Fusobacterium nucleatum host-cell binding and invasion induces IL-8 and CXCL1 secretion that drives colorectal cancer cell migration. *Sci Signal,* 13.

CASTELLARIN, M., WARREN, R. L., FREEMAN, J. D., DREOLINI, L., KRZYWINSKI, M., STRAUSS, J., BARNES, R., WATSON, P., ALLEN-VERCOE, E., MOORE, R. A. & HOLT, R. A. 2012. Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. *Genome Res,* 22**,** 299-306.

CHEN, K., COLLINS, G., WANG, H. & TOH, J. W. T. 2021. Pathological Features and Prognostication in Colorectal Cancer. *Curr Oncol,* 28**,** 5356-5383.

CHEN, L., DENG, H., CUI, H., FANG, J., ZUO, Z., DENG, J., LI, Y., WANG, X. & ZHAO, L. 2018. Inflammatory responses and inflammation-associated diseases in organs. *Oncotarget,* 9**,** 7204-7218.

CHEN, X., LI, H., GUO, F., HOFFMEISTER, M. & BRENNER, H. 2022a. Alcohol consumption, polygenic risk score, and early- and late-onset colorectal cancer risk. *eClinicalMedicine,* 49.

CHEN, Y., CHEN, Y., ZHANG, J., CAO, P., SU, W., DENG, Y., ZHAN, N., FU, X., HUANG, Y. & DONG, W. 2020. Fusobacterium nucleatum Promotes Metastasis in Colorectal Cancer by Activating Autophagy Signaling via the Upregulation of CARD3 Expression. *Theranostics,* 10**,** 323-339.

CHEN, Y., HUANG, Z., TANG, Z., HUANG, Y., HUANG, M., LIU, H., ZIEBOLZ, D., SCHMALZ, G., JIA, B. & ZHAO, J. 2022b. More Than Just a Periodontal Pathogen - the Research Progress on Fusobacterium nucleatum. *Front Cell Infect Microbiol,* 12**,** 815318.

CHEN, Z., GU, S., TROJANOWICZ, B., LIU, N., ZHU, G., DRALLE, H. & HOANG-VU, C. 2011. Down-regulation of TM4SF is associated with the metastatic potential of gastric carcinoma TM4SF members in gastric carcinoma. *World J Surg Oncol,* 9**,** 43.

CHENG, Y., LING, Z. & LI, L. 2020. The Intestinal Microbiota and Colorectal Cancer. *Frontiers in Immunology,* 11.

CHIDIMATSU, H., TSUNEDOMI, R., NAKAGAMI, Y., XU, M., NAKAJIMA, M., NAKASHIMA-NAKASUGA, C., TOMOCHIKA, S., YOSHIDA, S., SUZUKI, N., WATANABE, Y., MATSUI, H., SHINDO, Y., TOKUMITSU, Y., IIDA, M., TAKEDA, S., IOKA, T., UENO, T., TANABE, T., HOSHII, Y., HAZAMA, S. & NAGANO, H. 2023. Serum CCL7 Is a Novel Prognostic Biomarker of Metastatic Colorectal Cancer. *Anticancer Research,* 43**,** 105-114.

CHRISMAN, B., HE, C., JUNG, J.-Y., STOCKHAM, N., PASKOV, K., WASHINGTON, P. & WALL, D. P. 2022. The human "contaminome": bacterial, viral, and computational contamination in whole genome sequences from 1000 families. *Scientific Reports,* 12**,** 9863.

CHUBUKOV, V. & SAUER, U. 2014. Environmental dependence of stationary-phase metabolism in Bacillus subtilis and Escherichia coli. *Appl Environ Microbiol,* 80**,** 2901-9.

CHUN, E., LAVOIE, S., MICHAUD, M., GALLINI, C. A., KIM, J., SOUCY, G., ODZE, R., GLICKMAN, J. N. & GARRETT, W. S. 2015. CCL2 Promotes Colorectal Carcinogenesis by Enhancing Polymorphonuclear Myeloid-Derived Suppressor Cell Population and Function. *Cell Rep,* 12**,** 244-57.

CHUNSONG, H., YULING, H., LI, W., JIE, X., GANG, Z., QIUPING, Z., QINGPING, G., KEJIAN, Z., LI, Q., CHANG, A. E., YOUXIN, J. & JINQUAN, T. 2006. CXC Chemokine Ligand 13 and CC Chemokine Ligand 19 Cooperatively Render Resistance to Apoptosis in B Cell Lineage Acute and Chronic Lymphocytic Leukemia CD23+CD5+ B Cells1. *The Journal of Immunology,* 177**,** 6713-6722.

CI, H., XU, Z., XU, J., WANG, Y. & WU, S. 2018. Expressions of KAI1 and E-cadherin in nonsmall cell lung cancer and their correlation with vasculogenic mimicry. *Medicine (Baltimore),* 97**,** e12293.

CITRON, D. M. 2002. Update on the taxonomy and clinical aspects of the genus fusobacterium. *Clin Infect Dis,* 35**,** S22-7.

COLLIER, J. J., SUOMI, F., OLÁHOVÁ, M., MCWILLIAMS, T. G. & TAYLOR, R. W. 2021. Emerging roles of ATG7 in human health and disease. *EMBO Molecular Medicine,* 13**,** e14824.

COPPENHAGEN-GLAZER, S., SOL, A., ABED, J., NAOR, R., ZHANG, X., HAN, Y. W. & BACHRACH, G. 2015. Fap2 of Fusobacterium nucleatum is a galactose-inhibitable adhesin involved in coaggregation, cell adhesion, and preterm birth. *Infect Immun,* 83**,** 1104-13.

CRABTREE, M., SIEBER, O. M., LIPTON, L., HODGSON, S. V., LAMLUM, H., THOMAS, H. J. W., NEALE, K., PHILLIPS, R. K. S., HEINIMANN, K. & TOMLINSON, I. P. M. 2003. Refining the relation between 'first hits' and 'second hits' at the APC locus: the 'loose fit' model and evidence for differences in somatic mutation spectra among patients. *Oncogene,* 22**,** 4257-4265.

DABIJA-WOLTER, G., CIMPAN, M.-R., COSTEA, D. E., JOHANNESSEN, A. C., SØRNES, S., NEPPELBERG, E., AL-HARONI, M., SKAUG, N. & BAKKEN, V. 2009. Fusobacterium nucleatum Enters Normal Human Oral Fibroblasts In Vitro. *Journal of Periodontology,* 80**,** 1174-1183.

DAKOWICZ, D., ZAJKOWSKA, M. & MROCZKO, B. 2022. Relationship between VEGF Family Members, Their Receptors and Cell Death in the Neoplastic Transformation of Colorectal Cancer. *Int J Mol Sci,* 23.

DE ROSA, M., PACE, U., REGA, D., COSTABILE, V., DURATURO, F., IZZO, P. & DELRIO, P. 2015. Genetics, diagnosis and management of colorectal cancer (Review). *Oncol Rep,* 34**,** 1087-96.

DEBESA-TUR, G., PÉREZ-BROCAL, V., RUIZ-RUIZ, S., CASTILLEJO, A., LATORRE, A., SOTO, J. L. & MOYA, A. 2021. Metagenomic analysis of formalin-fixed paraffin-embedded tumor and normal mucosa reveals differences in the microbiome of colorectal cancer patients. *Scientific Reports,* 11**,** 391.

DEKKER, E., TANIS, P. J., VLEUGELS, J. L. A., KASI, P. M. & WALLACE, M. B. 2019. Colorectal cancer. *Lancet,* 394**,** 1467-1480.

DEL PRETE, A., SALVI, V., SORIANI, A., LAFFRANCHI, M., SOZIO, F., BOSISIO, D. & SOZZANI, S. 2023. Dendritic cell subsets in cancer immunity and tumor antigen sensing. *Cellular & Molecular Immunology,* 20**,** 432-447.

DETCHOKUL, S., WILLIAMS, E. D., PARKER, M. W. & FRAUMAN, A. G. 2014. Tetraspanins as regulators of the tumour microenvironment: implications for metastasis and therapeutic strategies. *Br J Pharmacol,* 171**,** 5462-90.

DEVRIESE, S., VAN DEN BOSSCHE, L., VAN WELDEN, S., HOLVOET, T., PINHEIRO, I., HINDRYCKX, P., DE VOS, M. & LAUKENS, D. 2017. T84 monolayers are superior to Caco-2 as a model system of colonocytes. *Histochem Cell Biol,* 148**,** 85-93.

DEY, S., BASU, S. & RANJAN, A. 2023. Revisiting the Role of CD63 as Pro-Tumorigenic or Anti-Tumorigenic Tetraspanin in Cancers and its Theragnostic Implications. *Advanced Biology,* 7**,** 2300078.

DHALIWAL, D., BHARGAVA, R. & MOVAHED, M. R. 2023. Fusobacterium nucleatum endocarditis: a case report and literature review. *Am J Cardiovasc Dis,* 13**,** 29-31.

DIENSTMANN, R., VERMEULEN, L., GUINNEY, J., KOPETZ, S., TEJPAR, S. & TABERNERO, J. 2017. Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nature Reviews Cancer,* 17**,** 79-92.

DORNIER, E., COUMAILLEAU, F., OTTAVI, J. F., MORETTI, J., BOUCHEIX, C., MAUDUIT, P., SCHWEISGUTH, F. & RUBINSTEIN, E. 2012. TspanC8 tetraspanins regulate ADAM10/Kuzbanian trafficking and promote Notch activation in flies and mammals. *J Cell Biol,* 199**,** 481-96.

DOUGHERTY, M. W. & JOBIN, C. 2023. Intestinal bacteria and colorectal cancer: etiology and treatment. *Gut Microbes,* 15**,** 2185028.

DREWES, J. L., HOUSSEAU, F. & SEARS, C. L. 2016. Sporadic colorectal cancer: microbial contributors to disease prevention, development and therapy. *British Journal of Cancer,* 115**,** 273-280.

DRULINER, B. R., WANG, P., BAE, T., BAHETI, S., SLETTEDAHL, S., MAHONEY, D., VASMATZIS, N., XU, H., KIM, M., BOCKOL, M., O'BRIEN, D., GRILL, D., WARNER, N., MUNOZ-GOMEZ, M., KOSSICK, K., JOHNSON, R., MOUCHLI, M., FELMLEE-DEVINE, D., WASHECHEK-ALETTO, J., SMYRK, T., OBERG, A., WANG, J., CHIA, N., ABYZOV, A., AHLQUIST, D. & BOARDMAN, L. A. 2018. Molecular characterization of colorectal adenomas with and without malignancy reveals distinguishing genome, transcriptome and methylome alterations. *Sci Rep,* 8**,** 3161.

DUAN, C., TANG, X., WANG, W., QIAN, W., FU, X., DENG, X., ZHOU, S., HAN, C. & HOU, X. 2021. Lactobacillus rhamnosus attenuates intestinal inflammation induced by Fusobacterium nucleatum infection by restoring the autophagic flux. *Int J Mol Med,* 47**,** 125-136.

DUNLOP, M. G., DOBBINS, S. E., FARRINGTON, S. M., JONES, A. M., PALLES, C., WHIFFIN, N., TENESA, A., SPAIN, S., BRODERICK, P., OOI, L.-Y., DOMINGO, E., SMILLIE, C., HENRION, M., FRAMPTON, M., MARTIN, L., GRIMES, G., GORMAN, M., SEMPLE, C., MA, Y. P., BARCLAY, E., PRENDERGAST, J., CAZIER, J.-B., OLVER, B., PENEGAR, S., LUBBE, S., CHANDER, I., CARVAJAL-CARMONA, L. G., BALLEREAU, S., LLOYD, A., VIJAYAKRISHNAN, J., ZGAGA, L., RUDAN, I., THEODORATOU, E., THOMAS, H., MAHER, E., EVANS, G., WALKER, L., HALLIDAY, D., LUCASSEN, A., PATERSON, J., HODGSON, S., HOMFRAY, T., SIDE, L., IZATT, L., DONALDSON, A., TOMKINS, S., MORRISON, P., BREWER, C., HENDERSON, A., DAVIDSON, R., MURDAY, V., COOK, J., HAITES, N., BISHOP, T., SHERIDAN, E., GREEN, A., MARKS, C., CARPENTER, S., BROUGHTON, M., GREENHALGE, L., SURI, M., STARR, J. M., DEARY, I., KIRAC, I., KOVACEVIĆ, D., AALTONEN, L. A., RENKONEN-SINISALO, L., MECKLIN, J.-P., MATSUDA, K., NAKAMURA, Y., OKADA, Y., GALLINGER, S., DUGGAN, D. J., CONTI, D., NEWCOMB, P., HOPPER, J., JENKINS, M. A., SCHUMACHER, F., CASEY, G., EASTON, D., SHAH, M., PHAROAH, P., LINDBLOM, A., LIU, T., EDLER, D., LENANDER, C., DALÉN, J., HJERN, F., LUNDQVIST, N., LINDFORSS, U., PÅHLMAN, L., SMEDH, K., TÖRNQVIST, A., HOLM, J., JANSON, M., ANDERSSON, M., EKELUND, S., OLSSON, L., SMITH, C. G., WEST, H., et al. 2012. Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. *Nature Genetics,* 44**,** 770-776.

DZINK, J. L., SHEENAN, M. T. & SOCRANSKY, S. S. 1990. Proposal of Three Subspecies of Fusobacterium nucleatum Knorr 1922: Fusobacterium nucleatum subsp. nucleatum

subsp. nov., comb. nov.; Fusobacterium nucleatum subsp. polymorphum subsp. nov., nom. rev., comb. nov.; and Fusobacterium nucleatum subsp. vincentii subsp. nov., nom. rev., comb. nov. *International Journal of Systematic and Evolutionary Microbiology,* 40**,** 74-78.

EARNEST, J. T., HANTAK, M. P., PARK, J. E. & GALLAGHER, T. 2015. Coronavirus and influenza virus proteolytic priming takes place in tetraspanin-enriched membrane microdomains. *J Virol,* 89**,** 6093-104.

EGESTEN, A., ELIASSON, M., JOHANSSON, H. M., OLIN, A. I., MORGELIN, M., MUELLER, A., PEASE, J. E., FRICK, I. M. & BJORCK, L. 2007. The CXC chemokine MIG/CXCL9 is important in innate immunity against Streptococcus pyogenes. *J Infect Dis,* 195**,** 684-93.

ELGAWIDI, A., MOHSIN, M. I., ALI, F., WATTS, A., MONK, P. N., THOMAS, M. S. & PARTRIDGE, L. J. 2020. A role for tetraspanin proteins in regulating fusion induced by Burkholderia thailandensis. *Med Microbiol Immunol,* 209**,** 473-487.

ELIASSON, M., MÖRGELIN, M., FARBER, J. M., EGESTEN, A. & ALBIGER, B. 2010. Streptococcus pneumoniae induces expression of the antibacterial CXC chemokine MIG/CXCL9 via MyD88-dependent signaling in a murine model of airway infection. *Microbes and Infection,* 12**,** 565-573.

EVANS, J. C., MCENEANY, V. L., COYNE, M. J., CALDWELL, E. P., SHEAHAN, M. L., VON, S. S., COYNE, E. M., TWETEN, R. K. & COMSTOCK, L. E. 2022. A proteolytically activated antimicrobial toxin encoded on a mobile plasmid of Bacteroidales induces a protective response. *Nature Communications,* 13**,** 4258.

EWING, I., HURLEY, J. J., JOSEPHIDES, E. & MILLAR, A. 2014. The molecular genetics of colorectal cancer. *Frontline Gastroenterol,* 5**,** 26-30.

EYKYN, S. J. 1989. Necrobacillosis. *Scand J Infect Dis Suppl,* 62**,** 41-6.

FAGERBERG, L., HALLSTRÖM, B. M., OKSVOLD, P., KAMPF, C., DJUREINOVIC, D., ODEBERG, J., HABUKA, M., TAHMASEBPOOR, S., DANIELSSON, A., EDLUND, K., ASPLUND, A., SJÖSTEDT, E., LUNDBERG, E., SZIGYARTO, C. A., SKOGS, M., TAKANEN, J. O., BERLING, H., TEGEL, H., MULDER, J., NILSSON, P., SCHWENK, J. M., LINDSKOG, C., DANIELSSON, F., MARDINOGLU, A., SIVERTSSON, A., VON FEILITZEN, K., FORSBERG, M., ZWAHLEN, M., OLSSON, I., NAVANI, S., HUSS, M., NIELSEN, J., PONTEN, F. & UHLÉN, M. 2014. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics,* 13**,** 397-406.

FANG, C. Y., CHEN, J. S., HSU, B. M., HUSSAIN, B., RATHOD, J. & LEE, K. H. 2021. Colorectal Cancer Stage-Specific Fecal Bacterial Community Fingerprinting of the Taiwanese Population and Underpinning of Potential Taxonomic Biomarkers. *Microorganisms,* 9.

FARDINI, Y., WANG, X., TÉMOIN, S., NITHIANANTHAM, S., LEE, D., SHOHAM, M. & HAN, Y. W. 2011. Fusobacterium nucleatum adhesin FadA binds vascular endothelial cadherin and alters endothelial integrity. *Mol Microbiol,* 82**,** 1468-80.

FEAGA, H. A. & DWORKIN, J. 2021. Transcription regulates ribosome hibernation. *Mol Microbiol,* 116**,** 663-673.

FEARON, E. R. & VOGELSTEIN, B. 1990. A genetic model for colorectal tumorigenesis. *Cell,* 61**,** 759-767.

FÉNÉANT, L., LEVY, S. & COCQUEREL, L. 2014. CD81 and hepatitis C virus (HCV) infection. *Viruses,* 6**,** 535-72.

FENG, X., ZHANG, Z., SUN, P., SONG, G., WANG, L., SUN, Z., YUAN, N., WANG, Q. & LUN, L. 2020. Interleukin-18 Is a Prognostic Marker and Plays a Tumor Suppressive Role in Colon Cancer. *Dis Markers,* 2020**,** 6439614.

FOGH, J., FOGH, J. M. & ORFEO, T. 1977. One hundred and twenty-seven cultured human tumor cell lines producing tumors in nude mice. *J Natl Cancer Inst,* 59**,** 221-6.

FOSTER, T. J. 2020. Surface Proteins of Staphylococcus epidermidis. *Front Microbiol,* 11**,** 1829.

FRANCESCHI, D., GIULIANI, V., GIUNTINI, V. & PINI PRATO, G. 2020. Brain abscess and periodontal pathogens (Fusobacterium Nucleatum). Report of a case. *Clin Case Rep,* 8**,** 2488-2493.

FRANCO, M., MURATORI, C., CORSO, S., TENAGLIA, E., BERTOTTI, A., CAPPARUCCIA, L., TRUSOLINO, L., COMOGLIO, P. M. & TAMAGNONE, L. 2010. The tetraspanin CD151 is required for Met-dependent signaling and tumor cell growth. *J Biol Chem,* 285**,** 38756-64.

FRIESS, H., GUO, X. Z., BERBERAT, P., GRABER, H. U., ZIMMERMANN, A., KORC, M. & BÜCHLER, M. W. 1998. Reduced KAI1 expression in pancreatic cancer is associated with lymph node and distant metastases. *Int J Cancer,* 79**,** 349-55.

GALEANO NIÑO, J. L., WU, H., LACOURSE, K. D., KEMPCHINSKY, A. G., BARYIAMES, A., BARBER, B., FUTRAN, N., HOULTON, J., SATHER, C., SICINSKA, E., TAYLOR, A., MINOT, S. S., JOHNSTON, C. D. & BULLMAN, S. 2022. Effect of the intratumoral microbiota on spatial and cellular heterogeneity in cancer. *Nature,* 611**,** 810-817.

GAUTHIER, S., TÉTU, A., HIMAYA, E., MORAND, M., CHANDAD, F., RALLU, F. & BUJOLD, E. 2011. The origin of Fusobacterium nucleatum involved in intra-amniotic infection and preterm birth. *J Matern Fetal Neonatal Med,* 24**,** 1329-32.

GE, Y., HUANG, M. & YAO, Y.-M. 2020. Biology of Interleukin-17 and Its Pathophysiological Significance in Sepsis. *Frontiers in Immunology,* 11.

GENG, F., ZHANG, Y., LU, Z., ZHANG, S. & PAN, Y. 2020. Fusobacterium nucleatum Caused DNA Damage and Promoted Cell Proliferation by the Ku70/p53 Pathway in Oral Cancer Cells. *DNA Cell Biol,* 39**,** 144-151.

GHARBIA, S. E. & SHAH, H. N. 1992. Fusobacterium nucleatum subsp. fusiforme subsp. nov. and Fusobacterium nucleatum subsp. animalis subsp. nov. as additional subspecies within Fusobacterium nucleatum. *Int J Syst Bacteriol,* 42**,** 296-8.

GHARBIA, S. E., SHAH, H. N., LAWSON, P. A. & HAAPASALO, M. 1990. Distribution and frequency of Fusobacterium nucleatum subspecies in the human oral cavity. *Oral Microbiol Immunol,* 5**,** 324-7.

GHOSSOUB, R., CHÉRY, M., AUDEBERT, S., LEBLANC, R., EGEA-JIMENEZ, A. L., LEMBO, F., MAMMAR, S., LE DEZ, F., CAMOIN, L., BORG, J. P., RUBINSTEIN, E., DAVID, G. & ZIMMERMANN, P. 2020. Tetraspanin-6 negatively regulates exosome production. *Proc Natl Acad Sci U S A,* 117**,** 5913-5922.

GOLDBERG, E. A., VENKAT-RAMANI, T., HEWIT, M. & BONILLA, H. F. 2013. Epidemiology and clinical outcomes of patients with Fusobacterium bacteraemia. *Epidemiol Infect,* 141**,** 325-9.

GOPINATH, D., MENON, R. K., WIE, C. C., BANERJEE, M., PANDA, S., MANDAL, D., BEHERA, P. K., ROYCHOUDHURY, S., KHEUR, S., BOTELHO, M. G. & JOHNSON, N. W. 2021. Differences in the bacteriome of swab, saliva, and tissue biopsies in oral cancer. *Scientific Reports,* 11**,** 1181.

GRADY, W. M. & CARETHERS, J. M. 2008. Genomic and epigenetic instability in colorectal cancer pathogenesis. *Gastroenterology,* 135**,** 1079-99.

GRÄSSE, L., FAST, L. A., SCHEFFER, K. D., BOUKHALLOUK, F., SPODEN, G. A., TENZER, S., BOLLER, K., BAGO, R., RAJESH, S., OVERDUIN, M., BERDITCHEVSKI, F. & FLORIN, L. 2016. The CD63-Syntenin-1 Complex Controls Post-Endocytic Trafficking of Oncogenic Human Papillomaviruses. *Sci Rep,* 6**,** 32337.

GRASSI, E., CORBELLI, J., PAPIANI, G., BARBERA, M. A., GAZZANEO, F. & TAMBERI, S. 2021. Current Therapeutic Strategies in BRAF-Mutant Metastatic Colorectal Cancer. *Frontiers in Oncology,* 11.

GRAY, D. A., DUGAR, G., GAMBA, P., STRAHL, H., JONKER, M. J. & HAMOEN, L. W. 2019. Extreme slow growth as alternative strategy to survive deep starvation in bacteria. *Nature Communications,* 10**,** 890.

GREEN, L. R., ISSA, R., ALBALDI, F., URWIN, L., THOMPSON, R., KHALID, H., TURNER, C. E., CIANI, B., PARTRIDGE, L. J. & MONK, P. N. 2023a. CD9 co-operation with syndecan-1 is required for a major staphylococcal adhesion pathway. *mBio,* 14**,** e01482-23.

GREEN, L. R., ISSA, R., ALBALDI, F., URWIN, L., THOMPSON, R., KHALID, H., TURNER, C. E., CIANI, B., PARTRIDGE, L. J. & MONK, P. N. 2023b. CD9 co-operation with syndecan-1 is required for a major staphylococcal adhesion pathway. *mBio,* 14**,** e0148223.

GREEN, L. R., MONK, P. N., PARTRIDGE, L. J., MORRIS, P., GORRINGE, A. R. & READ, R. C. 2011. Cooperative role for tetraspanins in adhesin-mediated attachment of bacterial species to human epithelial cells. *Infect Immun,* 79**,** 2241-9.

GREENE, N. P., KAPLAN, E., CROW, A. & KORONAKIS, V. 2018. Corrigendum: Antibiotic Resistance Mediated by the MacB ABC Transporter Family: A Structural and Functional Perspective. *Frontiers in Microbiology,* 9.

GROTHEY, A., FAKIH, M. & TABERNERO, J. 2021. Management of BRAF-mutant metastatic colorectal cancer: a review of treatment options and evidence-based guidelines. *Annals of Oncology,* 32**,** 959-967.

GRYFE, R. 2009. Inherited colorectal cancer syndromes. *Clin Colon Rectal Surg,* 22**,** 198-208.

GUERMONPREZ, P., HELFT, J., CLASER, C., DEROUBAIX, S., KARANJE, H., GAZUMYAN, A., DARASSE-JÈZE, G., TELERMAN, S. B., BRETON, G., SCHREIBER, H. A., FRIAS-STAHELI, N., BILLERBECK, E., DORNER, M., RICE, C. M., PLOSS, A., KLEIN, F., SWIECKI, M., COLONNA, M., KAMPHORST, A. O., MEREDITH, M., NIEC, R., TAKACS, C., MIKHAIL, F., HARI, A., BOSQUE, D., EISENREICH, T., MERAD, M., SHI, Y., GINHOUX, F., RÉNIA, L., URBAN, B. C. & NUSSENZWEIG, M. C. 2013. Inflammatory Flt3l is essential to mobilize dendritic cells and for T cell responses during Plasmodium infection. *Nat Med,* 19**,** 730-8.

GUINNEY, J., DIENSTMANN, R., WANG, X., DE REYNIÈS, A., SCHLICKER, A., SONESON, C., MARISA, L., ROEPMAN, P., NYAMUNDANDA, G., ANGELINO, P., BOT, B. M., MORRIS, J. S., SIMON, I. M., GERSTER, S., FESSLER, E., DE SOUSA E MELO, F., MISSIAGLIA, E., RAMAY, H., BARRAS, D., HOMICSKO, K., MARU, D., MANYAM, G. C., BROOM, B., BOIGE, V., PEREZ-VILLAMIL, B., LADERAS, T., SALAZAR, R., GRAY, J. W., HANAHAN, D., TABERNERO, J., BERNARDS, R., FRIEND, S. H., LAURENT-PUIG, P., MEDEMA, J. P.,

SADANANDAM, A., WESSELS, L., DELORENZI, M., KOPETZ, S., VERMEULEN, L. & TEJPAR, S. 2015. The consensus molecular subtypes of colorectal cancer. *Nature Medicine,* 21**,** 1350-1356.

GUIX, F. X., SANNERUD, R., BERDITCHEVSKI, F., ARRANZ, A. M., HORRÉ, K., SNELLINX, A., THATHIAH, A., SAIDO, T., SAITO, T., RAJESH, S., OVERDUIN, M., KUMAR-SINGH, S., RADAELLI, E., CORTHOUT, N., COLOMBELLI, J., TOSI, S., MUNCK, S., SALAS, I. H., ANNAERT, W. & DE STROOPER, B. 2017. Tetraspanin 6: a pivotal protein of the multiple vesicular body determining exosome release and lysosomal degradation of amyloid precursor protein fragments. *Molecular Neurodegeneration,* 12**,** 25.

GUNER, O. S. & TUMAY, L. V. 2021. Persistent extramural vascular invasion positivity on magnetic resonance imaging after neoadjuvant chemoradiotherapy predicts poor outcome in rectal cancer. *Asian Journal of Surgery,* 44**,** 841-847.

GUO, P., TIAN, Z., KONG, X., YANG, L., SHAN, X., DONG, B., DING, X., JING, X., JIANG, C., JIANG, N. & YU, Y. 2020. FadA promotes DNA damage and progression of Fusobacterium nucleatum-induced colorectal cancer through up-regulation of chk2. *Journal of Experimental & Clinical Cancer Research,* 39**,** 202.

GUO, Q., XIA, B., ZHANG, F., RICHARDSON, M. M., LI, M., ZHANG, J. S., CHEN, F. & ZHANG, X. A. 2012. Tetraspanin CO-029 inhibits colorectal cancer cell movement by deregulating cell-matrix and cell-cell adhesions. *PLoS One,* 7**,** e38464.

GUR, C., IBRAHIM, Y., ISAACSON, B., YAMIN, R., ABED, J., GAMLIEL, M., ENK, J., BAR-ON, Y., STANIETSKY-KAYNAN, N., COPPENHAGEN-GLAZER, S., SHUSSMAN, N., ALMOGY, G., CUAPIO, A., HOFER, E., MEVORACH, D., TABIB, A., ORTENBERG, R., MARKEL, G., MIKLIĆ, K., JONJIC, S., BRENNAN, C. A., GARRETT, W. S., BACHRACH, G. & MANDELBOIM, O. 2015. Binding of the Fap2 protein of Fusobacterium nucleatum to human inhibitory receptor TIGIT protects tumors from immune cell attack. *Immunity,* 42**,** 344-355.

GURSOY, U. K., KÖNÖNEN, E. & UITTO, V.-J. 2008. Intracellular replication of fusobacteria requires new actin filament formation of epithelial cells *APMIS,* 116**,** 1063-1070.

HAGEMANN, L., GRÜNDEL, A., JACOBS, E. & DUMKE, R. 2017. The surface-displayed chaperones GroEL and DnaK of Mycoplasma pneumoniae interact with human plasminogen and components of the extracellular matrix. *Pathog Dis,* 75.

HAJJAR, R., RICHARD, C. S. & SANTOS, M. M. 2021. The role of butyrate in surgical and oncological outcomes in colorectal cancer. *American Journal of Physiology-Gastrointestinal and Liver Physiology,* 320**,** G601-G608.

HAN, X. Y., WEINBERG, J. S., PRABHU, S. S., HASSENBUSCH, S. J., FULLER, G. N., TARRAND, J. J. & KONTOYIANNIS, D. P. 2003. Fusobacterial brain abscess: a review of five cases and an analysis of possible pathogenesis. *Journal of Neurosurgery,* 99**,** 693-700.

HAN, Y. W., REDLINE, R. W., LI, M., YIN, L., HILL, G. B. & MCCORMICK, T. S. 2004. Fusobacterium nucleatum induces premature and term stillbirths in pregnant mice: implication of oral bacteria in preterm birth. *Infect Immun,* 72**,** 2272-9.

HAN, Y. W. & WANG, X. 2013. Mobile Microbiome:Oral Bacteria in Extra-oral Infections and Inflammation. *Journal of Dental Research,* 92**,** 485-491.

HANKEY, W., FRANKEL, W. L. & GRODEN, J. 2018. Functions of the APC tumor suppressor protein dependent and independent of canonical WNT signaling: implications for therapeutic targeting. *Cancer Metastasis Rev,* 37**,** 159-172.

HANTAK, M. P., QING, E., EARNEST, J. T. & GALLAGHER, T. 2019. Tetraspanins: Architects of Viral Entry and Exit Platforms. *Journal of Virology,* 93**,** 10.1128/jvi.01429-17.

HÄNZELMANN, S., CASTELO, R. & GUINNEY, J. 2013. GSVA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics,* 14**,** 7.

HARVEY, K. L., JAROCKI, V. M., CHARLES, I. G. & DJORDJEVIC, S. P. 2019. The Diverse Functional Roles of Elongation Factor Tu (EF-Tu) in Microbial Pathogenesis. *Frontiers in Microbiology,* 10.

HASSIN, O., NATARAJ, N. B., SHREBERK-SHAKED, M., AYLON, Y., YAEGER, R., FONTEMAGGI, G., MUKHERJEE, S., MADDALENA, M., AVIOZ, A., IANCU, O., MALLEL, G., GERSHONI, A., GROSHEVA, I., FELDMESSER, E., BEN-DOR, S., GOLANI, O., HENDEL, A., BLANDINO, G., KELSEN, D., YARDEN, Y. & OREN, M. 2022. Different hotspot p53 mutants exert distinct phenotypes and predict outcome of colorectal cancer patients. *Nat Commun,* 13**,** 2800.

HASSUNA, N. A., MONK, P. N., ALI, F., READ, R. C. & PARTRIDGE, L. J. 2017. A role for the tetraspanin proteins in Salmonella infection of human macrophages. *Journal of Infection,* 75**,** 115-124.

HASUWA, H., SHISHIDO, Y., YAMAZAKI, A., KOBAYASHI, T., YU, X. & MEKADA, E. 2001. CD9 Amino Acids Critical for Upregulation of Diphtheria Toxin Binding. *Biochemical and Biophysical Research Communications,* 289**,** 782-790.

HAYWARD, S., GACHEHILADZE, M., BADR, N., ANDRIJES, R., MOLOSTVOV, G., PANIUSHKINA, L., SOPIKOVA, B., SLOBODOVÁ, Z., MGEBRISHVILI, G., SHARMA, N., HORIMOTO, Y., BURG, D., ROBERTSON, G., HANBY, A., HOAR, F., REA, D., ECKHARDT, B. L., UENO, N. T., NAZARENKO, I., LONG, H. M., VAN LAERE, S., SHAABAN, A. M. & BERDITCHEVSKI, F. 2020. The CD151-midkine pathway regulates the immune microenvironment in inflammatory breast cancer. *The Journal of Pathology,* 251**,** 63-73.

HE, Z., GHARAIBEH, R. Z., NEWSOME, R. C., POPE, J. L., DOUGHERTY, M. W., TOMKOVICH, S., PONS, B., MIREY, G., VIGNARD, J., HENDRIXSON, D. R. & JOBIN, C. 2019. Campylobacter jejuni promotes colorectal tumorigenesis through the action of cytolethal distending toxin. *Gut,* 68**,** 289-300.

HE, Z., TIAN, W., WEI, Q. & XU, J. 2022. Involvement of Fusobacterium nucleatum in malignancies except for colorectal cancer: A literature review. *Frontiers in Immunology,* 13.

HECKMANN, J. G., LANG, C. J. G., HARTL, H. & TOMANDL, B. 2003. Multiple Brain Abscesses Caused by Fusobacterium nucleatum Treated Conservatively. *Canadian Journal of Neurological Sciences,* 30**,** 266-268.

HEKMATI, M., BEN-SHAUL, Y. & POLAK-CHARCON, S. 1990. A morphological study of a human adenocarcinoma cell line (HT29) differentiating in culture. Similarities to intestinal embryonic development. *Cell Differentiation and Development,* 31**,** 207-218.

HEMLER, M. E. 2008. Targeting of tetraspanin proteins--potential benefits and strategies. *Nat Rev Drug Discov,* 7**,** 747-58.

HEMLER, M. E. 2014. Tetraspanin proteins promote multiple cancer stages. *Nature Reviews Cancer,* 14**,** 49-60.

HEO, K. & LEE, S. 2020. TSPAN8 as a Novel Emerging Therapeutic Target in Cancer for Monoclonal Antibody Therapy. *Biomolecules,* 10.

HERZIG, D. S., LUAN, L., BOHANNON, J. K., TOLIVER-KINSKY, T. E., GUO, Y. & SHERWOOD, E. R. 2014. The role of CXCL10 in the pathogenesis of experimental septic shock. *Crit Care,* 18**,** R113.

HIPPE, A., BRAUN, S. A., OLÁH, P., GERBER, P. A., SCHORR, A., SEELIGER, S., HOLTZ, S., JANNASCH, K., PIVARCSI, A., BUHREN, B., SCHRUMPF, H., KISLAT, A., BÜNEMANN, E., STEINHOFF, M., FISCHER, J., LIRA, S. A., BOUKAMP, P., HEVEZI, P., STOECKLEIN, N. H., HOFFMANN, T., ALVES, F., SLEEMAN, J., BAUER, T., KLUFA, J., AMBERG, N., SIBILIA, M., ZLOTNIK, A., MÜLLER-HOMEY, A. & HOMEY, B. 2020. EGFR/Ras-induced CCL20 production modulates the tumour microenvironment. *British Journal of Cancer,* 123**,** 942-954.

HISCHEBETH, G. T. R., KEIL, V. C., GENTIL, K., BOSTRÖM, A., KUCHELMEISTER, K. & BEKEREDJIAN-DING, I. 2014. Rapid brain death caused by a cerebellar abscess with Fusobacterium nucleatum in a young man with drug abuse: a case report. *BMC Research Notes,* 7**,** 353.

HOARE, A., SOTO, C., ROJAS-CELIS, V. & BRAVO, D. 2019. Chronic Inflammation as a Link between Periodontitis and Carcinogenesis. *Mediators Inflamm,* 2019**,** 1029857.

HOFSTAD, T. 2006. The Genus Fusobacterium. *In:* DWORKIN, M., FALKOW, S., ROSENBERG, E., SCHLEIFER, K.-H. & STACKEBRANDT, E. (eds.) *The Prokaryotes: Volume 7: Proteobacteria: Delta, Epsilon Subclass.* New York, NY: Springer New York.

HONG, M., LI, Z., LIU, H., ZHENG, S., ZHANG, F., ZHU, J., SHI, H., YE, H., CHOU, Z., GAO, L., DIAO, J., ZHANG, Y., ZHANG, D., CHEN, S., ZHOU, H. & LI, J. 2023. Fusobacterium nucleatum aggravates rheumatoid arthritis through FadA-containing outer membrane vesicles. *Cell Host & Microbe,* 31**,** 798-810.e7.

HORI, H., YANO, S., KOUFUJI, K., TAKEDA, J. & SHIROUZU, K. 2004. CD9 expression in gastric cancer and its significance. *J Surg Res,* 117**,** 208-15.

HSU, T., GEMMELL, M. R., FRANZOSA, E. A., BERRY, S., MUKHOPADHYA, I., HANSEN, R., MICHAUD, M., NIELSEN, H., MILLER, W. G., NIELSEN, H., BAJAJ-ELLIOTT, M., HUTTENHOWER, C., GARRETT, W. S. & HOLD, G. L. 2019. Comparative genomics and genome biology of Campylobacter showae. *Emerging Microbes & Infections,* 8**,** 827-840.

HUANG, R., SUN, H., LIN, R., ZHANG, J., YIN, H., XIAN, S., LI, M., WANG, S., LI, Z., QIAO, Y., JIANG, M., YAN, P., MENG, T. & HUANG, Z. 2022. The role of tetraspanins pan-cancer. *iScience,* 25**,** 104777.

HUANG, S., YUAN, S., DONG, M., SU, J., YU, C., SHEN, Y., XIE, X., YU, Y., YU, X., CHEN, S., ZHANG, S., PONTAROTTI, P. & XU, A. 2005. The phylogenetic analysis of tetraspanins projects the evolution of cell–cell interactions from unicellular to multicellular organisms. *Genomics,* 86**,** 674-684.

HUH, J. W. & ROH, T. Y. 2020. Opportunistic detection of Fusobacterium nucleatum as a marker for the early gut microbial dysbiosis. *BMC Microbiol,* 20**,** 208.

HUMMEL, M., HEGEWISCH-BECKER, S., NEUMANN, J. H. L. & VOGEL, A. 2021. BRAF testing in metastatic colorectal carcinoma and novel, chemotherapy-free therapeutic options. *Pathologe,* 42**,** 98-109.

HUR, J., SMITH-WARNER, S. A., RIMM, E. B., WILLETT, W. C., WU, K., CAO, Y. & GIOVANNUCCI, E. 2021. Alcohol intake in early adulthood and risk of colorectal cancer: three large prospective cohort studies of men and women in the United States. *Eur J Epidemiol,* 36**,** 325-333.

IKEGAMI, A., CHUNG, P. & HAN, Y. W. 2009. Complementation of the fadA mutation in Fusobacterium nucleatum demonstrates that the surface-exposed adhesin promotes cellular invasion and placental colonization. *Infect Immun,* 77**,** 3075-9.

IMAI, T., KAKIZAKI, M., NISHIMURA, M. & YOSHIE, O. 1995. Molecular analyses of the association of CD4 with two members of the transmembrane 4 superfamily, CD81 and CD82. *J Immunol,* 155**,** 1229-39.

INOUE, A., SHEEDY, S. P., HEIKEN, J. P., MOHAMMADINEJAD, P., GRAHAM, R. P., LEE, H. E., KELLEY, S. R., HANSEL, S. L., BRUINING, D. H., FIDLER, J. L. & FLETCHER, J. G. 2021. MRI-detected extramural venous invasion of rectal cancer: Multimodality performance and implications at baseline imaging and after neoadjuvant therapy. *Insights into Imaging,* 12**,** 110.

IWABUCHI, E., MIKI, Y., KANAI, A., ISHIDA, T. & SASANO, H. 2021. Abstract 2903: The exosome marker CD63 and Caveolin-1 in breast cancer patients. *Cancer Research,* 81**,** 2903-2903.

JADI, P. K., DAVE, A., ISSA, R., TABBASUM, K., OKUROWSKA, K., SAMARTH, A., URWIN, L., GREEN, L. R., PARTRIDGE, L. J., MACNEIL, S., GARG, P., MONK, P. N. & ROY, S. 2023. Tetraspanin CD9-derived peptides inhibit Pseudomonas aeruginosa corneal infection and aid in wound healing of corneal epithelial cells. *The Ocular Surface.*

JAEGER, E., WEBB, E., HOWARTH, K., CARVAJAL-CARMONA, L., ROWAN, A., BRODERICK, P., WALTHER, A., SPAIN, S., PITTMAN, A., KEMP, Z., SULLIVAN, K., HEINIMANN, K., LUBBE, S., DOMINGO, E., BARCLAY, E., MARTIN, L., GORMAN, M., CHANDLER, I., VIJAYAKRISHNAN, J., WOOD, W., PAPAEMMANUIL, E., PENEGAR, S., QURESHI, M., FARRINGTON, S., TENESA, A., CAZIER, J.-B., KERR, D., GRAY, R., PETO, J., DUNLOP, M., CAMPBELL, H., THOMAS, H., HOULSTON, R., TOMLINSON, I. & MEMBERS OF THE, C. C. 2008. Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. *Nature Genetics,* 40**,** 26-28.

JAISHANKAR, J. & SRIVASTAVA, P. 2017. Molecular Basis of Stationary Phase Survival and Applications. *Front Microbiol,* 8**,** 2000.

JASPERSON, K. W., TUOHY, T. M., NEKLASON, D. W. & BURT, R. W. 2010. Hereditary and familial colon cancer. *Gastroenterology,* 138**,** 2044-58.

JAYASIMHAN, D., WU, L. & HUGGAN, P. 2017. Fusobacterial liver abscess: a case report and review of the literature. *BMC Infect Dis,* 17**,** 440.

JAYASINGAM, S. D., CITARTAN, M., THANG, T. H., MAT ZIN, A. A., ANG, K. C. & CH'NG, E. S. 2019. Evaluating the Polarization of Tumor-Associated Macrophages Into M1 and M2 Phenotypes in Human Cancer Tissue: Technicalities and Challenges in Routine Clinical Practice. *Front Oncol,* 9**,** 1512.

JEFFERY, C. 2018. Intracellular proteins moonlighting as bacterial adhesion factors. *AIMS Microbiol,* 4**,** 362-376.

JI, N., YANG, J. & JI, Y. 2020. Determining Impact of Growth Phases on Capacity of Staphylococcus aureus to Adhere to and Invade Host Cells. *In:* JI, Y. (ed.) *Methicillin-Resistant Staphylococcus Aureus (MRSA) Protocols: Cutting-Edge Technologies and Advancements.* New York, NY: Springer US.

JIAO, Y., SHEN, P., WANG, S., CHEN, J., ZHOU, X. & MA, G. 2023. Arginase from Priestia megaterium and the Effects of CMCS Conjugation on Its Enzymological Properties. *Current Microbiology,* 80.

JIMÉNEZ-SÁNCHEZ, A., CAST, O. & MILLER, M. L. 2019. Comprehensive Benchmarking and Integration of Tumor Microenvironment Cell Estimation Methods. *Cancer Research,* 79**,** 6238-6246.

JOHDI, N. A. & SUKOR, N. F. 2020. Colorectal Cancer Immunotherapy: Options and Strategies. *Front Immunol,* 11**,** 1624.

JONKER, D. J., O'CALLAGHAN, C. J., KARAPETIS, C. S., ZALCBERG, J. R., TU, D., AU, H.-J., BERRY, S. R., KRAHN, M., PRICE, T., SIMES, R. J., TEBBUTT, N. C., VAN HAZEL, G., WIERZBICKI, R., LANGER, C. & MOORE, M. J. 2007. Cetuximab for the Treatment of Colorectal Cancer. *New England Journal of Medicine,* 357**,** 2040-2048.

KAMALI, N., TALEBI BEZMIN ABADI, A., ABADI, B., RAHIMI, F. & FOROOTAN, M. 2022. Identification of Streptococcus gallolyticus in tumor samples of Iranian patients diagnosed with colorectal cancer. *BMC Research Notes,* 15**,** 316.

KAPLAN, C. W., LUX, R., HAAKE, S. K. & SHI, W. 2009. The Fusobacterium nucleatum outer membrane protein RadD is an arginine-inhibitable adhesin required for inter-species adherence and the structured architecture of multispecies biofilm. *Mol Microbiol,* 71**,** 35-47.

KAPLAN, C. W., MA, X., PARANJPE, A., JEWETT, A., LUX, R., KINDER-HAAKE, S. & SHI, W. 2010. Fusobacterium nucleatum outer membrane proteins Fap2 and RadD induce cell death in human lymphocytes. *Infect Immun,* 78**,** 4773-8.

KAPRIO, T., HAGSTRÖM, J., ANDERSSON, L. C. & HAGLUND, C. 2020. Tetraspanin CD63 independently predicts poor prognosis in colorectal cancer. *Histol Histopathol,* 35**,** 887-892.

KARAM, J., MÉRESSE, S., KREMER, L. & DAHER, W. 2020. The roles of tetraspanins in bacterial infections. *Cellular Microbiology,* 22**,** e13260.

KASPER, S. H., MORELL-PEREZ, C., WYCHE, T. P., SANA, T. R., LIEBERMAN, L. A. & HETT, E. C. 2020. Colorectal cancer-associated anaerobic bacteria proliferate in tumor spheroids and alter the microenvironment. *Scientific Reports,* 10**,** 5321.

KAUFFMAN, A. L., GYURDIEVA, A. V., MABUS, J. R., FERGUSON, C., YAN, Z. & HORNBY, P. J. 2013. Alternative functional in vitro models of human intestinal epithelia. *Front Pharmacol,* 4**,** 79.

KEUNG, E. Z. & GERSHENWALD, J. E. 2018. The eighth edition American Joint Committee on Cancer (AJCC) melanoma staging system: implications for melanoma treatment and care. *Expert Rev Anticancer Ther,* 18**,** 775-784.

KHAN, S. Z. & LENGYEL, C. G. 2023. Challenges in the management of colorectal cancer in low- and middle-income countries. *Cancer Treatment and Research Communications,* 35**,** 100705.

KHAROFA, J., APEWOKIN, S., ALENGHAT, T. & OLLBERDING, N. J. 2023. Metagenomic analysis of the fecal microbiome in colorectal cancer patients compared to healthy controls as a function of age. *Cancer Medicine,* 12**,** 2945-2957.

KIM, H.-S., LEE, D.-S., CHANG, Y.-H., KIM, M. J., KOH, S., KIM, J., SEONG, J.-H., SONG, S. K., SHIN, H. S., SON, J.-B., JUNG, M. Y., PARK, S.-N., YOO, S. Y., CHO, K. W., KIM, D.-K., MOON, S., KIM, D., CHOI, Y., KIM, B.-O., JANG, H.-S., KIM, C. S., KIM, C., CHOE, S.-J. & KOOK, J.-K. 2010. Application of rpoB and Zinc Protease Gene for Use in Molecular Discrimination of Fusobacterium nucleatum Subspecies. *Journal of Clinical Microbiology,* 48**,** 545-553.

KIM, J., KIM, H., LEE, M.-S., LEE, H., KIM, Y. J., LEE, W. Y., YUN, S. H., KIM, H. C., HONG, H. K., HANNENHALLI, S., CHO, Y. B., PARK, D. & CHOI, S. S. 2023.

Transcriptomes of the tumor-adjacent normal tissues are more informative than tumors in predicting recurrence in colorectal cancer patients. *Journal of Translational Medicine,* 21**,** 209.

KIM, S. G., SEO, H. S., LEE, H. H., SONG, K. Y. & PARK, C. H. 2017. Comparison of the Differences in Survival Rates between the 7th and 8th Editions of the AJCC TNM Staging System for Gastric Adenocarcinoma: a Single-Institution Study of 5,507 Patients in Korea. *J Gastric Cancer,* 17**,** 212-219.

KOMATSU, M., WANG, Q. J., HOLSTEIN, G. R., FRIEDRICH, V. L., IWATA, J.-I., KOMINAMI, E., CHAIT, B. T., TANAKA, K. & YUE, Z. 2007. Essential role for autophagy protein Atg7 in the maintenance of axonal homeostasis and the prevention of axonal degeneration. *Proceedings of the National Academy of Sciences,* 104**,** 14489-14494.

KONDOH, N. & MIZUNO-KAMIYA, M. 2022. The Role of Immune Modulatory Cytokines in the Tumor Microenvironments of Head and Neck Squamous Cell Carcinomas. *Cancers (Basel),* 14.

KOOK, J.-K., PARK, S.-N., LIM, Y. K., CHO, E., JO, E., ROH, H., SHIN, Y., PAEK, J., KIM, H.-S., KIM, H., SHIN, J. H. & CHANG, Y.-H. 2017. Genome-Based Reclassification of Fusobacterium nucleatum Subspecies at the Species Level. *Current Microbiology,* 74**,** 1137-1147.

KORDAHI, M. C., STANAWAY, I. B., AVRIL, M., CHAC, D., BLANC, M. P., ROSS, B., DIENER, C., JAIN, S., MCCLEARY, P., PARKER, A., FRIEDMAN, V., HUANG, J., BURKE, W., GIBBONS, S. M., WILLIS, A. D., DARVEAU, R. P., GRADY, W. M., KO, C. W. & DEPAOLO, R. W. 2021. Genomic and functional characterization of a mucosal symbiont involved in early-stage colorectal cancer. *Cell Host Microbe,* 29**,** 1589-1598.e6.

KOSTIC, A. D., CHUN, E., ROBERTSON, L., GLICKMAN, J. N., GALLINI, C. A., MICHAUD, M., CLANCY, T. E., CHUNG, D. C., LOCHHEAD, P., HOLD, G. L., EL-OMAR, E. M., BRENNER, D., FUCHS, C. S., MEYERSON, M. & GARRETT, W. S. 2013. Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe,* 14**,** 207-15.

KUCKI, M., DIENER, L., BOHMER, N., HIRSCH, C., KRUG, H. F., PALERMO, V. & WICK, P. 2017. Uptake of label-free graphene oxide by Caco-2 cells is dependent on the cell differentiation status. *Journal of Nanobiotechnology,* 15**,** 46.

KUMAR, R., HEROLD, J. L., TAYLOR, J., XU, J. & XU, Y. 2018. Variations among Streptococcus gallolyticus subsp. gallolyticus strains in connection with colorectal cancer. *Scientific Reports,* 8**,** 1514.

KUMAR, R., TAYLOR, J. C., JAIN, A., JUNG, S. Y., GARZA, V. & XU, Y. 2022. Modulation of the extracellular matrix by Streptococcus gallolyticus subsp. gallolyticus and importance in cell proliferation. *PLoS Pathog,* 18**,** e1010894.

KUMARI, S., DEVI, G. T., BADANA, A., DASARI, V. R. & MALLA, R. R. 2015. CD151-A Striking Marker for Cancer Therapy. *Biomark Cancer,* 7**,** 7-11.

KURGAN, Ş., KANSAL, S., NGUYEN, D., STEPHENS, D., KORONEOS, Y., HASTURK, H., VAN DYKE, T. E. & KANTARCI, A. 2017. Strain-Specific Impact of Fusobacterium nucleatum on Neutrophil Function. *J Periodontol,* 88**,** 380-389.

KWON, T., LAMSTER, I. B. & LEVIN, L. 2021. Current Concepts in the Management of Periodontitis. *Int Dent J,* 71**,** 462-476.

KWONG, L. N. & DOVE, W. F. 2009. APC and its modifiers in colon cancer. *Adv Exp Med Biol,* 656**,** 85-106.

LAI, C.-H., LIN, T.-L., HUANG, M.-Z., LI, S.-W., WU, H.-Y., CHIU, Y.-F., YANG, C.-Y., CHIU, C.-H. & LAI, H.-C. 2022. Gut Commensal Parabacteroides goldsteinii MTS01 Alters Gut Microbiota Composition and Reduces Cholesterol to Mitigate Helicobacter pylori-Induced Pathogenesis. *Frontiers in Immunology,* 13.

LAMMERS, A., NUIJTEN, P. J. M. & SMITH, H. E. 1999. The fibronectin binding proteins of Staphylococcus aureus are required for adhesion to and invasion of bovine mammary gland cells. *FEMS Microbiology Letters,* 180**,** 103-109.

LAW, P. J., TIMOFEEVA, M., FERNANDEZ-ROZADILLA, C., BRODERICK, P., STUDD, J., FERNANDEZ-TAJES, J., FARRINGTON, S., SVINTI, V., PALLES, C., ORLANDO, G., SUD, A., HOLROYD, A., PENEGAR, S., THEODORATOU, E., VAUGHAN-SHAW, P., CAMPBELL, H., ZGAGA, L., HAYWARD, C., CAMPBELL, A., HARRIS, S., DEARY, I. J., STARR, J., GATCOMBE, L., PINNA, M., BRIGGS, S., MARTIN, L., JAEGER, E., SHARMA-OATES, A., EAST, J., LEEDHAM, S., ARNOLD, R., JOHNSTONE, E., WANG, H., KERR, D., KERR, R., MAUGHAN, T., KAPLAN, R., AL-TASSAN, N., PALIN, K., HÄNNINEN, U. A., CAJUSO, T., TANSKANEN, T., KONDELIN, J., KAASINEN, E., SARIN, A.-P., ERIKSSON, J. G., RISSANEN, H., KNEKT, P., PUKKALA, E., JOUSILAHTI, P., SALOMAA, V., RIPATTI, S., PALOTIE, A., RENKONEN-SINISALO, L., LEPISTÖ, A., BÖHM, J., MECKLIN, J.-P., BUCHANAN, D. D., WIN, A.-K., HOPPER, J., JENKINS, M. E., LINDOR, N. M., NEWCOMB, P. A., GALLINGER, S., DUGGAN, D., CASEY, G., HOFFMANN, P., NÖTHEN, M. M., JÖCKEL, K.-H., EASTON, D. F., PHAROAH, P. D. P., PETO, J., CANZIAN, F., SWERDLOW, A., EELES, R. A., KOTE-JARAI, Z., MUIR, K., PASHAYAN, N., HENDERSON, B. E., HAIMAN, C. A., SCHUMACHER, F. R., AL OLAMA, A. A., BENLLOCH, S., BERNDT, S. I., CONTI, D. V., WIKLUND, F., CHANOCK, S., GAPSTUR, S., STEVENS, V. L., TANGEN, C. M., BATRA, J., CLEMENTS, J., GRONBERG, H., SCHLEUTKER, J., ALBANES, D., WOLK, A., WEST, C., MUCCI, L., CANCEL-TASSIN, G., KOUTROS, S., et al. 2019. Association analyses identify 31 new risk loci for colorectal cancer susceptibility. *Nature Communications,* 10**,** 2154.

LEE, H.-R., JUN, H.-K., KIM, H.-D., LEE, S.-H. & CHOI, B.-K. 2012. Fusobacterium nucleatum GroEL induces risk factors of atherosclerosis in human microvascular endothelial cells and ApoE−/− mice. *Molecular Oral Microbiology,* 27**,** 109-123.

LEE, J., BYUN, H. J., LEE, M. S., JIN, Y. J., JEOUNG, D., KIM, Y. M. & LEE, H. 2017. The metastasis suppressor CD82/KAI1 inhibits fibronectin adhesion-induced epithelial-to-mesenchymal transition in prostate cancer cells by repressing the associated integrin signaling. *Oncotarget,* 8**,** 1641-1654.

LEUNG, P.-S., ARONSON, W. J., NGO, T. H., GOLDING, L. A. & BARNARD, R. J. 2004. Exercise alters the IGF axis in vivo and increases p53 protein in prostate tumor cells in vitro. *Journal of Applied Physiology,* 96**,** 450-454.

LEWANDOWSKA, A., RUDZKI, G., LEWANDOWSKI, T., STRYJKOWSKA-GÓRA, A. & RUDZKI, S. 2022. Title: Risk Factors for the Diagnosis of Colorectal Cancer. *Cancer Control,* 29**,** 10732748211056692.

LI, H., SUN, Y. & SUN, L. 2022. A Teleost CXCL10 Is Both an Immunoregulator and an Antimicrobial. *Frontiers in Immunology,* 13.

LI, J., BUTCHER, J., MACK, D. & STINTZI, A. 2014. Functional Impacts of the Intestinal Microbiome in the Pathogenesis of Inflammatory Bowel Disease. *Inflammatory Bowel Diseases,* 21**,** 139-153.

LI, J., CHEN, X., ZHU, L., LAO, Z., ZHOU, T., ZANG, L., GE, W., JIANG, M., XU, J., CAO, Y., DU, S., YU, Y., FAN, G. & WANG, H. 2021a. SOX9 is a critical regulator of TSPAN8-mediated metastasis in pancreatic cancer. *Oncogene,* 40**,** 4884-4893.
LI, K., LUO, H., HUANG, L., LUO, H. & ZHU, X. 2020. Microsatellite instability: a review of what the oncologist should know. *Cancer Cell International,* 20**,** 16.
LI, S., LIU, J., ZHENG, X., REN, L., YANG, Y., LI, W., FU, W., WANG, J. & DU, G. 2021b. Tumorigenic bacteria in colorectal cancer: mechanisms and treatments. *Cancer Biol Med,* 19**,** 147-62.
LI, X., ZHONG, Q., LUO, D., DU, Q. & LIU, W. 2019. The prognostic value of CXC subfamily ligands in stage I-III patients with colorectal cancer. *PLOS ONE,* 14**,** e0214611.
LI, Y. Y., GE, Q. X., CAO, J., ZHOU, Y. J., DU, Y. L., SHEN, B., WAN, Y. J. & NIE, Y. Q. 2016. Association of Fusobacterium nucleatum infection with colorectal cancer in Chinese patients. *World J Gastroenterol,* 22**,** 3227-33.
LIAO, S., LIN, Y., LIU, L., YANG, S., LIN, Y., HE, J. & SHAO, Y. 2023. ADAM10-a "multitasker" in sepsis: focus on its posttranslational target. *Inflammation Research,* 72**,** 395-423.
LIBERZON, A., BIRGER, C., THORVALDSDÓTTIR, H., GHANDI, M., MESIROV, JILL P. & TAMAYO, P. 2015. The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Systems,* 1**,** 417-425.
LIM, S. M., MOHAMAD HANIF, E. A. & CHIN, S.-F. 2021. Is targeting autophagy mechanism in cancer a good approach? The possible double-edge sword effect. *Cell & Bioscience,* 11**,** 56.
LIMA, B. P., SHI, W. & LUX, R. 2017a. Identification and characterization of a novel Fusobacterium nucleatum adhesin involved in physical interaction and biofilm formation with Streptococcus gordonii. *MicrobiologyOpen,* 6**,** e00444.
LIMA, B. P., SHI, W. & LUX, R. 2017b. Identification and characterization of a novel Fusobacterium nucleatum adhesin involved in physical interaction and biofilm formation with Streptococcus gordonii. *Microbiologyopen,* 6.
LIN, M. G. & HURLEY, J. H. 2016. Structure and function of the ULK1 complex in autophagy. *Curr Opin Cell Biol,* 39**,** 61-8.
LIN, Y., HE, Z., YE, J., LIU, Z., SHE, X., GAO, X. & LIANG, R. 2020. Progress in Understanding the IL-6/STAT3 Pathway in Colorectal Cancer. *Onco Targets Ther,* 13**,** 13023-13032.
LIN, Y., XU, J. & LAN, H. 2019. Tumor-associated macrophages in tumor metastasis: biological roles and clinical therapeutic applications. *Journal of Hematology & Oncology,* 12**,** 76.
LIN, Y.-W., WANG, L.-C., LEE, C.-K. & CHEN, S.-H. 2018. Flt3 ligand treatment reduces enterovirus A71 lethality in mice with enhanced B cell responses. *Scientific Reports,* 8**,** 12184.
LIU, H., HONG, X. L., SUN, T. T., HUANG, X. W., WANG, J. L. & XIONG, H. 2020. Fusobacterium nucleatum exacerbates colitis by damaging epithelial barriers and inducing aberrant inflammation. *Journal of Digestive Diseases,* 21**,** 385-398.
LIU, L., XU, M., LAN, R., HU, D., LI, X., QIAO, L., ZHANG, S., LIN, X., YANG, J., REN, Z. & XU, J. 2022. Bacteroides vulgatus attenuates experimental mice colitis through modulating gut microbiota and immune responses. *Frontiers in Immunology,* 13.
LIU, M., LIU, S., HUANG, M., WANG, Y., WANG, M., TIAN, X., LI, L., YANG, Z., WANG, M., ZHU, D., JIA, R., CHEN, S., ZHAO, X., YANG, Q., WU, Y., ZHANG,

S., HUANG, J., OU, X., MAO, S., GAO, Q., SUN, D., YU, Y. L. & CHENG, A. 2021. An Exposed Outer Membrane Hemin-Binding Protein Facilitates Hemin Transport by a TonB-Dependent Receptor in Riemerella anatipestifer. *Applied and Environmental Microbiology,* 87**,** e00367-21.

LORICO, A., LORICO-RAPPA, M., KARBANOVÁ, J., CORBEIL, D. & PIZZORNO, G. 2021. CD9, a tetraspanin target for cancer therapy? *Exp Biol Med (Maywood),* 246**,** 1121-1138.

LOURENÇO, T. G., HELLER, D., SILVA-BOGHOSSIAN, C. M., COTTON, S. L., PASTER, B. J. & COLOMBO, A. P. 2014. Microbial signature profiles of periodontally healthy and diseased patients. *J Clin Periodontol,* 41**,** 1027-36.

LUPIA, A., PEPPICELLI, S., WITORT, E., BIANCHINI, F., CARLONI, V., PIMPINELLI, N., URSO, C., BORGOGNONI, L., CAPACCIOLI, S., CALORINI, L. & LULLI, M. 2014. CD63 Tetraspanin Is a Negative Driver of Epithelial-to-Mesenchymal Transition in Human Melanoma Cells. *Journal of Investigative Dermatology,* 134**,** 2947-2956.

MA, T. F., FAN, Y. R., ZHAO, Y. H. & LIU, B. 2023a. Emerging role of autophagy in colorectal cancer: Progress and prospects for clinical intervention. *World J Gastrointest Oncol,* 15**,** 979-987.

MA, X., SUN, T., ZHOU, J., ZHI, M., SHEN, S., WANG, Y., GU, X., LI, Z., GAO, H., WANG, P. & FENG, Q. 2023b. Pangenomic Study of Fusobacterium nucleatum Reveals the Distribution of Pathogenic Genes and Functional Clusters at the Subspecies and Strain Levels. *Microbiology Spectrum,* 11**,** e05184-22.

MAATEN, L. V. D. & HINTON, G. E. 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research,* 9**,** 2579-2605.

MAIER, R. M. 2009. Chapter 3 - Bacterial Growth. *In:* MAIER, R. M., PEPPER, I. L. & GERBA, C. P. (eds.) *Environmental Microbiology (Second Edition).* San Diego: Academic Press.

MAJEK, O., GONDOS, A., JANSEN, L., EMRICH, K., HOLLECZEK, B., KATALINIC, A., NENNECKE, A., EBERLE, A., BRENNER, H. & THE, G. C. S. W. G. 2013. Sex Differences in Colorectal Cancer Survival: Population-Based Analysis of 164,996 Colorectal Cancer Patients in Germany. *PLOS ONE,* 8**,** e68077.

MANDAL, D. P., MOHANTY, N., BEHERA, P. K., GOPINATH, D., PANDA, S., AL-KHERAIF, A. A., DIVAKAR, D. D., ANIL, S. & PANDA, S. 2021. A Plausible Proposition of CCL20-Related Mechanism in Fusobacterium nucleatum-Associated Oral Carcinogenesis. *Life (Basel),* 11.

MANZONI, G., MARINACH, C., TOPÇU, S., BRIQUET, S., GRAND, M., TOLLE, M., GRANSAGNE, M., LESCAR, J., ANDOLINA, C., FRANETICH, J.-F., ZEISEL, M. B., HUBY, T., RUBINSTEIN, E., SNOUNOU, G., MAZIER, D., NOSTEN, F., BAUMERT, T. F. & SILVIE, O. 2017. Plasmodium P36 determines host cell receptor usage during sporozoite invasion. *eLife,* 6**,** e25903.

MARCOTTE, H. & LAVOIE, M. C. 1998. Oral microbial ecology and the role of salivary immunoglobulin A. *Microbiol Mol Biol Rev,* 62**,** 71-109.

MARIAPPAN, V., VELLASAMY, K. M., HASHIM, O. H. & VADIVELU, J. 2011. Profiling of Burkholderia cepacia secretome at mid-logarithmic and early-stationary phases of growth. *PLoS One,* 6**,** e26518.

MARONGIU, L. & ALLGAYER, H. 2022. Viruses in colorectal cancer. *Mol Oncol,* 16**,** 1423-1450.

MARTIN-GALLAUSIAUX, C., MALABIRADE, A., HABIER, J. & WILMES, P. 2020. Fusobacterium nucleatum Extracellular Vesicles Modulate Gut Epithelial Cell Innate Immunity via FomA and TLR2. *Front Immunol,* 11**,** 583644.

MATTHEWS, A. L., SZYROKA, J., COLLIER, R., NOY, P. J. & TOMLINSON, M. G. 2017. Scissor sisters: regulation of ADAM10 by the TspanC8 tetraspanins. *Biochem Soc Trans,* 45**,** 719-730.

MAY, F. P., GLENN, B. A., CRESPI, C. M., PONCE, N., SPIEGEL, B. M. R. & BASTANI, R. 2017. Decreasing Black-White Disparities in Colorectal Cancer Incidence and Stage at Presentation in the United States. *Cancer Epidemiology, Biomarkers & Prevention,* 26**,** 762-768.

MAZUROV, D., HEIDECKER, G. & DERSE, D. 2007. The Inner Loop of Tetraspanins CD82 and CD81 Mediates Interactions with Human T Cell Lymphotrophic Virus Type 1 Gag Protein*. *Journal of Biological Chemistry,* 282**,** 3896-3903.

MCCOOL, D. J., MARCON, M. A., FORSTNER, J. F. & FORSTNER, G. G. 1990. The T84 human colonic adenocarcinoma cell line produces mucin in culture and releases it in response to various secretagogues. *Biochem J,* 267**,** 491-500.

MCGEACHY, M. J., CUA, D. J. & GAFFEN, S. L. 2019. The IL-17 Family of Cytokines in Health and Disease. *Immunity,* 50**,** 892-906.

MCLAUGHLIN, K. A., TOMBS, M. A. & CHRISTIE, M. R. 2020. Autoimmunity to tetraspanin-7 in type 1 diabetes. *Med Microbiol Immunol,* 209**,** 437-445.

MENG, M., ZHONG, K., JIANG, T., LIU, Z., KWAN, H. Y. & SU, T. 2021. The current understanding on the impact of KRAS on colorectal cancer. *Biomedicine & Pharmacotherapy,* 140**,** 111717.

MEYERHARDT, J. A., GIOVANNUCCI, E. L., OGINO, S., KIRKNER, G. J., CHAN, A. T., WILLETT, W. & FUCHS, C. S. 2009. Physical activity and male colorectal cancer survival. *Arch Intern Med,* 169**,** 2102-8.

MILES, A. A., MISRA, S. S. & IRWIN, J. O. 1938. The estimation of the bactericidal power of the blood. *J Hyg (Lond),* 38**,** 732-49.

MILWARD, M. R., CHAPPLE, I. L., WRIGHT, H. J., MILLARD, J. L., MATTHEWS, J. B. & COOPER, P. R. 2007. Differential activation of NF-kappaB and gene expression in oral epithelial cells by periodontal pathogens. *Clin Exp Immunol,* 148**,** 307-24.

MIMA, K., NISHIHARA, R., QIAN, Z. R., CAO, Y., SUKAWA, Y., NOWAK, J. A., YANG, J., DOU, R., MASUGI, Y., SONG, M., KOSTIC, A. D., GIANNAKIS, M., BULLMAN, S., MILNER, D. A., BABA, H., GIOVANNUCCI, E. L., GARRAWAY, L. A., FREEMAN, G. J., DRANOFF, G., GARRETT, W. S., HUTTENHOWER, C., MEYERSON, M., MEYERHARDT, J. A., CHAN, A. T., FUCHS, C. S. & OGINO, S. 2016. Fusobacterium nucleatum in colorectal carcinoma tissue and patient prognosis. *Gut,* 65**,** 1973-1980.

MOLOSTVOV, G., GACHECHILADZE, M., SHAABAN, A. M., HAYWARD, S., DEAN, I., DIAS, I. H. K., BADR, N., DANIAL, I., MOHAMMED, F., NOVITSKAYA, V., PANIUSHKINA, L., SPEIRS, V., HANBY, A., NAZARENKO, I., WITHERS, D. R., VAN LAERE, S., LONG, H. M. & BERDITCHEVSKI, F. 2023. Tspan6 stimulates the chemoattractive potential of breast cancer cells for B cells in an EV- and LXR-dependent manner. *Cell Reports,* 42**,** 112207.

MORGAN, E., ARNOLD, M., GINI, A., LORENZONI, V., CABASAG, C. J., LAVERSANNE, M., VIGNAT, J., FERLAY, J., MURPHY, N. & BRAY, F. 2023. Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. *Gut,* 72**,** 338-344.

MORKEL, M., RIEMER, P., BLÄKER, H. & SERS, C. 2015. Similar but different: distinct roles for KRAS and BRAF oncogenes in colorectal cancer development and therapy resistance. *Oncotarget,* 6**,** 20785-800.

MORRIS, K. T., KHAN, H., AHMAD, A., WESTON, L. L., NOFCHISSEY, R. A., PINCHUK, I. V. & BESWICK, E. J. 2014. G-CSF and G-CSFR are highly expressed in human gastric and colon cancers and promote carcinoma cell proliferation and migration. *Br J Cancer,* 110**,** 1211-20.

MORSI, H., GOLIZEH, M., BROSSEAU, N., JANATI, A. I., EMAMI, E., NDAO, M. & TRAN, S. D. 2022. Detection of Fusobacterium nucleatum subspecies in the saliva of pre-colorectal cancer patients, using tandem mass spectrometry. *Archives of Oral Biology,* 134**,** 105337.

MOVSISYAN, A. S., MAGUIRE, F. B., PARIKH-PATEL, A., GHOLAMI, S. & KEEGAN, T. H. M. 2021. Increasing Rates of Colorectal Cancer Among Young People in California, 1988-2017. *J Registry Manag,* 48**,** 152-160.

MUCHOVA, M., BALACCO, D. L., GRANT, M. M., CHAPPLE, I. L. C., KUEHNE, S. A. & HIRSCHFELD, J. 2022. Fusobacterium nucleatum Subspecies Differ in Biofilm Forming Ability in vitro. *Front Oral Health,* 3**,** 853618.

MUKHOPADHYA, I., HANSEN, R., NICHOLL, C. E., ALHAIDAN, Y. A., THOMSON, J. M., BERRY, S. H., PATTINSON, C., STEAD, D. A., RUSSELL, R. K., EL-OMAR, E. M. & HOLD, G. L. 2011. A comprehensive evaluation of colonic mucosal isolates of Sutterella wadsworthensis from inflammatory bowel disease. *PLoS One,* 6**,** e27076.

MURAOKA, A., SUZUKI, M., HAMAGUCHI, T., WATANABE, S., IIJIMA, K., MUROFUSHI, Y., SHINJO, K., OSUKA, S., HARIYAMA, Y., ITO, M., OHNO, K., KIYONO, T., KYO, S., IWASE, A., KIKKAWA, F., KAJIYAMA, H. & KONDO, Y. 2023. Fusobacterium infection facilitates the development of endometriosis through the phenotypic transition of endometrial fibroblasts. *Science Translational Medicine,* 15**,** eadd1531.

MURDOCCA, M., MANGO, R., PUCCI, S., BIOCCA, S., TESTA, B., CAPUANO, R., PAOLESSE, R., SANCHEZ, M., ORLANDI, A., NATALE, C. D., NOVELLI, G. & SANGIUOLO, F. 2016. The lectin-like oxidized LDL receptor-1: a new potential molecular target in colorectal cancer. *Oncotarget,* 7.

NAGPAL, S. J., MUKHIJA, D. & PATEL, P. 2015. Fusobacterium nucleatum: a rare cause of pyogenic liver abscess. *Springerplus,* 4**,** 283.

NAGY, K. N., SONKODI, I., SZÖKE, I., NAGY, E. & NEWMAN, H. N. 1998. The microflora associated with human oral carcinomas. *Oral Oncology,* 34**,** 304-308.

NAKATSU, G., LI, X., ZHOU, H., SHENG, J., WONG, S. H., WU, W. K. K., NG, S. C., TSOI, H., DONG, Y., ZHANG, N., HE, Y., KANG, Q., CAO, L., WANG, K., ZHANG, J., LIANG, Q., YU, J. & SUNG, J. J. Y. 2015. Gut mucosal microbiome across stages of colorectal carcinogenesis. *Nature Communications,* 6**,** 8727.

NAVARRO LLORENS, J. M., TORMO, A. & MARTÍNEZ-GARCÍA, E. 2010. Stationary phase in gram-negative bacteria. *FEMS Microbiology Reviews,* 34**,** 476-495.

NEW, C., LEE, Z. Y., TAN, K. S., WONG, A. H., WANG, Y. & TRAN, T. 2021. Tetraspanins: Host Factors in Viral Infections. *Int J Mol Sci,* 22.

NGUYEN, L. H., GOEL, A. & CHUNG, D. C. 2020. Pathways of Colorectal Carcinogenesis. *Gastroenterology,* 158**,** 291-302.

NIES, J. F. & PANZER, U. 2020. IL-17C/IL-17RE: Emergence of a Unique Axis in TH17 Biology. *Frontiers in Immunology,* 11.

NITHIANANTHAM, S., XU, M., YAMADA, M., IKEGAMI, A., SHOHAM, M. & HAN, Y. W. 2009a. Crystal structure of FadA adhesin from Fusobacterium nucleatum reveals a novel oligomerization motif, the leucine chain. *J Biol Chem,* 284**,** 3865-72.

NITHIANANTHAM, S., XU, M., YAMADA, M., IKEGAMI, A., SHOHAM, M. & HAN, Y. W. 2009b. Crystal Structure of FadA Adhesin from Fusobacterium nucleatum Reveals a Novel Oligomerization Motif, the Leucine Chain*. *Journal of Biological Chemistry,* 284**,** 3865-3872.

NOJADEH, J. N., BEHROUZ SHARIF, S. & SAKHINIA, E. 2018. Microsatellite instability in colorectal cancer. *Excli j,* 17**,** 159-168.

NUSS, A. M., BECKSTETTE, M., PIMENOVA, M., SCHMÜHL, C., OPITZ, W., PISANO, F., HEROVEN, A. K. & DERSCH, P. 2017. Tissue dual RNA-seq allows fast discovery of infection-specific functions and riboregulators shaping host–pathogen transcriptomes. *Proceedings of the National Academy of Sciences,* 114**,** E791-E800.

O'BRIEN, M. J., YANG, S., HUANG, C. S., SHEPHERD, C., CERDA, S. & FARRAYE, F. A. 2008. The serrated polyp pathway to colorectal carcinoma. *Diagnostic Histopathology,* 14**,** 78-93.

OKUDA, J., YAMANE, S., NAGATA, S., KUNIKATA, C., SUEZAWA, C. & YASUDA, M. 2017. The Pseudomonas aeruginosa dnaK gene is involved in bacterial translocation across the intestinal epithelial cell barrier. *Microbiology (Reading),* 163**,** 1208-1216.

ORDAS, L., COSTA, L., LOZANO, A., CHEVILLARD, C., CALOVOULOS, A., KANTAR, D., FERNANDEZ, L., CHAUVIN, L., DOSSET, P., DOUCET, C., HERON-MILHAVET, L., ODINTSOVA, E., BERDITCHEVSKI, F., MILHIET, P. E. & BÉNISTANT, C. 2021. Mechanical Control of Cell Migration by the Metastasis Suppressor Tetraspanin CD82/KAI1. *Cells,* 10.

OSMAN, M. A., NEOH, H. M., AB MUTALIB, N. S., CHIN, S. F. & JAMAL, R. 2018. 16S rRNA Gene Sequencing for Deciphering the Colorectal Cancer Gut Microbiome: Current Protocols and Workflows. *Front Microbiol,* 9**,** 767.

PARHI, L., ALON-MAIMON, T., SOL, A., NEJMAN, D., SHHADEH, A., FAINSOD-LEVI, T., YAJUK, O., ISAACSON, B., ABED, J., MAALOUF, N., NISSAN, A., SANDBANK, J., YEHUDA-SHNAIDMAN, E., PONATH, F., VOGEL, J., MANDELBOIM, O., GRANOT, Z., STRAUSSMAN, R. & BACHRACH, G. 2020. Breast cancer colonization by Fusobacterium nucleatum accelerates tumor growth and metastatic progression. *Nat Commun,* 11**,** 3259.

PARKER, B. J., WEARSCH, P. A., VELOO, A. C. M. & RODRIGUEZ-PALACIOS, A. 2020. The Genus Alistipes: Gut Bacteria With Emerging Implications to Inflammation, Cancer, and Mental Health. *Front Immunol,* 11**,** 906.

PARZYCH, K. R. & KLIONSKY, D. J. 2014. An overview of autophagy: morphology, mechanism, and regulation. *Antioxid Redox Signal,* 20**,** 460-73.

PASQUEREAU-KOTULA, E., MARTINS, M., AYMERIC, L. & DRAMSI, S. 2018. Significance of Streptococcus gallolyticus subsp. gallolyticus Association With Colorectal Cancer. *Front Microbiol,* 9**,** 614.

PATEL, S. G., KARLITZ, J. J., YEN, T., LIEU, C. H. & BOLAND, C. R. 2022. The rising tide of early-onset colorectal cancer: a comprehensive review of epidemiology, clinical features, biology, risk factors, prevention, and early detection. *The Lancet Gastroenterology & Hepatology,* 7**,** 262-274.

PENNEL, K. A., QUINN, J. A., NIXON, C., INTHAGARD, J., VAN WYK, H. C., CHANG, D., REBUS, S., GROUP, G., HAY, J., MAKA, N. N., ROXBURGH, C. S.,

HORGAN, P. G., MCMILLAN, D. C., PARK, J. H., ROSEWEIR, A. K., STEELE, C. W. & EDWARDS, J. 2022. CXCL8 expression is associated with advanced stage, right sidedness, and distinct histological features of colorectal cancer. *The Journal of Pathology: Clinical Research,* 8**,** 509-520.

PIHLSTROM, B. L., MICHALOWICZ, B. S. & JOHNSON, N. W. 2005. Periodontal diseases. *The Lancet,* 366**,** 1809-1820.

PINO, M. S. & CHUNG, D. C. 2010. The chromosomal instability pathway in colon cancer. *Gastroenterology,* 138**,** 2059-72.

POL, J. G., LE NAOUR, J. & KROEMER, G. 2020. FLT3LG - a biomarker reflecting clinical responses to the immunogenic cell death inducer oxaliplatin. *Oncoimmunology,* 9**,** 1755214.

POPAT, S., STONE, J. & HOULSTON, R. S. 2003. Allelic imbalance in colorectal cancer at the CRAC1 locus in early-onset colorectal cancer. *Cancer Genet Cytogenet,* 145**,** 70-3.

POWELL, D. 2019.

PRICE, A. & GIBAS, C. 2017. The quantitative impact of read mapping to non-native reference genomes in comparative RNA-Seq studies. *PLoS One,* 12**,** e0180904.

PRICE, M. N., WETMORE, K. M., DEUTSCHBAUER, A. M. & ARKIN, A. P. 2016. A Comparison of the Costs and Benefits of Bacterial Gene Expression. *PLOS ONE,* 11**,** e0164314.

PROENÇA, M. A., BISELLI, J. M., SUCCI, M., SEVERINO, F. E., BERARDINELLI, G. N., CAETANO, A., REIS, R. M., HUGHES, D. J. & SILVA, A. E. 2018. Relationship between Fusobacterium nucleatum, inflammatory mediators and microRNAs in colorectal carcinogenesis. *World J Gastroenterol,* 24**,** 5351-5365.

PURCELL, R. V., VISNOVSKA, M., BIGGS, P. J., SCHMEIER, S. & FRIZELLE, F. A. 2017. Distinct gut microbiome patterns associate with consensus molecular subtypes of colorectal cancer. *Scientific Reports,* 7**,** 11590.

QUEEN, J., DOMINGUE, J. C., WHITE, J. R., STEVENS, C., UDAYASURYAN, B., NGUYEN, T. T. D., WU, S., DING, H., FAN, H., MCMANN, M., CORONA, A., LARMAN, T. C., VERBRIDGE, S. S., HOUSSEAU, F., SLADE, D. J., DREWES, J. L. & SEARS, C. L. 2021. Comparative Analysis of Colon Cancer-Derived Fusobacterium nucleatum Subspecies: Inflammation and Colon Tumorigenesis in Murine Models. *mBio,* 13**,** e0299121.

RAHMAN, R., SCHMALTZ, C., JACKSON, C. S., SIMOES, E. J., JACKSON-THOMPSON, J. & IBDAH, J. A. 2015. Increased risk for colorectal cancer under age 50 in racial and ethnic minorities living in the United States. *Cancer Medicine,* 4**,** 1863-1870.

RAMACHANDRAN, R. P., VENCES-CATALÁN, F., WISEMAN, D., ZLOTKIN-RIVKIN, E., SHTEYER, E., MELAMED-BOOK, N., ROSENSHINE, I., LEVY, S. & AROETI, B. 2018. EspH Suppresses Erk by Spatial Segregation from CD81 Tetraspanin Microdomains. *Infect Immun,* 86.

RAWLA, P., SUNKARA, T. & BARSOUK, A. 2019. Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors. *Prz Gastroenterol,* 14**,** 89-103.

REID-YU, S. A., TUINEMA, B. R., SMALL, C. N., XING, L. & COOMBES, B. K. 2015. CXCL9 Contributes to Antimicrobial Protection of the Gut during Citrobacter rodentium Infection Independent of Chemokine-Receptor Signaling. *PLOS Pathogens,* 11**,** e1004648.

REZASOLTANI, S., AGHDAEI, H. A., JASEMI, S., GAZOULI, M., DOVROLIS, N., SADEGHI, A., SCHLÜTER, H., ZALI, M. R., SECHI, L. A. & FEIZABADI, M. M. 2022. Oral Microbiota as Novel Biomarkers for Colorectal Cancer Screening. *Cancers (Basel),* 15.

RITCHIE, M. E., PHIPSON, B., WU, D., HU, Y., LAW, C. W., SHI, W. & SMYTH, G. K. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research,* 43**,** e47-e47.

ROLFE, M. D., RICE, C. J., LUCCHINI, S., PIN, C., THOMPSON, A., CAMERON, A. D., ALSTON, M., STRINGER, M. F., BETTS, R. P., BARANYI, J., PECK, M. W. & HINTON, J. C. 2012. Lag phase is a distinct growth phase that prepares bacteria for exponential growth and involves transient metal accumulation. *J Bacteriol,* 194**,** 686-701.

ROMANSKA, H. M. & BERDITCHEVSKI, F. 2011. Tetraspanins in human epithelial malignancies. *The Journal of Pathology,* 223**,** 4-14.

ROY, S., ZHU, Y., MA, J., ROY, A. C., ZHANG, Y., ZHONG, X., PAN, Z. & YAO, H. 2019. Role of ClpX and ClpP in Streptococcus suis serotype 2 stress tolerance and virulence. *Microbiological Research,* 223-225**,** 99-109.

RUBINSTEIN, E. 2011. The complexity of tetraspanins. *Biochem Soc Trans,* 39**,** 501-5.

RUBINSTEIN, M. R., BAIK, J. E., LAGANA, S. M., HAN, R. P., RAAB, W. J., SAHOO, D., DALERBA, P., WANG, T. C. & HAN, Y. W. 2019a. Fusobacterium nucleatum promotes colorectal cancer by inducing Wnt/β-catenin modulator Annexin A1. *EMBO reports,* 20**,** e47638.

RUBINSTEIN, M. R., BAIK, J. E., LAGANA, S. M., HAN, R. P., RAAB, W. J., SAHOO, D., DALERBA, P., WANG, T. C. & HAN, Y. W. 2019b. Fusobacterium nucleatum promotes colorectal cancer by inducing Wnt/β-catenin modulator Annexin A1. *EMBO Rep,* 20.

RUBINSTEIN, MARA R., WANG, X., LIU, W., HAO, Y., CAI, G. & HAN, YIPING W. 2013a. Fusobacterium nucleatum Promotes Colorectal Carcinogenesis by Modulating E-Cadherin/β-Catenin Signaling via its FadA Adhesin. *Cell Host & Microbe,* 14**,** 195-206.

RUBINSTEIN, M. R., WANG, X., LIU, W., HAO, Y., CAI, G. & HAN, Y. W. 2013b. Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating E-cadherin/β-catenin signaling via its FadA adhesin. *Cell Host Microbe,* 14**,** 195-206.

RUSTGI, A. K. 2007. The genetics of hereditary colon cancer. *Genes Dev,* 21**,** 2525-38.

RUTTER, C. M., MAY, F. P., CORONADO, G. D., PUJOL, T. A., THOMAS, E. G. & CABREROS, I. 2022. Racism Is a Modifiable Risk Factor: Relationships Among Race, Ethnicity, and Colorectal Cancer Outcomes. *Gastroenterology,* 162**,** 1053-1055.

SAHOO, M., CEBALLOS-OLVERA, I., DEL BARRIO, L. & RE, F. 2011. Role of the inflammasome, IL-1β, and IL-18 in bacterial infections. *ScientificWorldJournal,* 11**,** 2037-50.

SALAS, I. H., CALLAERTS-VEGH, Z., ARRANZ, A. M., GUIX, F. X., D'HOOGE, R., ESTEBAN, J. A., DE STROOPER, B. & DOTTI, C. G. 2017. Tetraspanin 6: A novel regulator of hippocampal synaptic transmission and long term plasticity. *PLOS ONE,* 12**,** e0171968.

SALVUCCI, M., CRAWFORD, N., STOTT, K., BULLMAN, S., LONGLEY, D. B. & PREHN, J. H. M. 2022. Patients with mesenchymal tumours and high Fusobacteriales prevalence have worse prognosis in colorectal cancer (CRC). *Gut,* 71**,** 1600-1612.

SAMEER, A. S. 2013. Colorectal cancer: molecular mutations and polymorphisms. *Front Oncol,* 3**,** 114.

SANDERS, B. E., QIAO, G., HIRSCHMAN, J. & MEYERSON, M. 2022. Fusobacterium and Colorectal Cancer: Important Association or Random Coincidence? *Cancer Research,* 82**,** 3671-3672.

SAUNDERS, A. S., BENDER, D. E., RAY, A. L., WU, X. & MORRIS, K. T. 2021. Colony-stimulating factor 3 signaling in colon and rectal cancers: Immune response and CMS classification in TCGA data. *PLOS ONE,* 16**,** e0247233.

SCHAEFFER, C. R., WOODS, K. M., LONGO, G. M., KIEDROWSKI, M. R., PAHARIK, A. E., BÜTTNER, H., CHRISTNER, M., BOISSY, R. J., HORSWILL, A. R., ROHDE, H. & FEY, P. D. 2015. Accumulation-associated protein enhances Staphylococcus epidermidis biofilm formation under dynamic conditions and is required for infection in a rat catheter model. *Infect Immun,* 83**,** 214-26.

SCHELLER, J., CHALARIS, A., SCHMIDT-ARRAS, D. & ROSE-JOHN, S. 2011. The pro- and anti-inflammatory properties of the cytokine interleukin-6. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research,* 1813**,** 878-888.

SCHOTT, B. H., ANTONIA, A. L., WANG, L., PITTMAN, K. J., SIXT, B. S., BARNES, A. B., VALDIVIA, R. H. & KO, D. C. 2020. Modeling of variables in cellular infection reveals CXCL10 levels are regulated by human genetic variation and the Chlamydia-encoded CPAF protease. *Scientific Reports,* 10**,** 18269.

SCHUMACHER, F. R., SCHMIT, S. L., JIAO, S., EDLUND, C. K., WANG, H., ZHANG, B., HSU, L., HUANG, S.-C., FISCHER, C. P., HARJU, J. F., IDOS, G. E., LEJBKOWICZ, F., MANION, F. J., MCDONNELL, K., MCNEIL, C. E., MELAS, M., RENNERT, H. S., SHI, W., THOMAS, D. C., VAN DEN BERG, D. J., HUTTER, C. M., ARAGAKI, A. K., BUTTERBACH, K., CAAN, B. J., CARLSON, C. S., CHANOCK, S. J., CURTIS, K. R., FUCHS, C. S., GALA, M., GIOVANNUCCI, E. L., GOGARTEN, S. M., HAYES, R. B., HENDERSON, B., HUNTER, D. J., JACKSON, R. D., KOLONEL, L. N., KOOPERBERG, C., KÜRY, S., LACROIX, A., LAURIE, C. C., LAURIE, C. A., LEMIRE, M., LEVINE, D., MA, J., MAKAR, K. W., QU, C., TAVERNA, D., ULRICH, C. M., WU, K., KONO, S., WEST, D. W., BERNDT, S. I., BEZIEAU, S., BRENNER, H., CAMPBELL, P. T., CHAN, A. T., CHANG-CLAUDE, J., COETZEE, G. A., CONTI, D. V., DUGGAN, D., FIGUEIREDO, J. C., FORTINI, B. K., GALLINGER, S. J., GAUDERMAN, W. J., GILES, G., GREEN, R., HAILE, R., HARRISON, T. A., HOFFMEISTER, M., HOPPER, J. L., HUDSON, T. J., JACOBS, E., IWASAKI, M., JEE, S. H., JENKINS, M., JIA, W.-H., JOSHI, A., LI, L., LINDOR, N. M., MATSUO, K., MORENO, V., MUKHERJEE, B., NEWCOMB, P. A., POTTER, J. D., RASKIN, L., RENNERT, G., ROSSE, S., SEVERI, G., SCHOEN, R. E., SEMINARA, D., SHU, X.-O., SLATTERY, M. L., TSUGANE, S., WHITE, E., XIANG, Y.-B., ZANKE, B. W., ZHENG, W., LE MARCHAND, L., CASEY, G., GRUBER, S. B., et al. 2015. Genome-wide association study of colorectal cancer identifies six new susceptibility loci. *Nature Communications,* 6**,** 7138.

SEARS, C. L. & PARDOLL, D. M. 2011. Perspective: alpha-bugs, their microbial partners, and the link to colon cancer. *J Infect Dis,* 203**,** 306-11.

SETO, S., MATSUMOTO, S., TSUJIMURA, K. & KOIDE, Y. 2010. Differential recruitment of CD63 and Rab7-interacting-lysosomal-protein to phagosomes containing Mycobacterium tuberculosis in macrophages. *Microbiol Immunol,* 54**,** 170-4.

SEVERN, M. M. & HORSWILL, A. R. 2023. Staphylococcus epidermidis and its dual lifestyle in skin health and infection. *Nature Reviews Microbiology,* 21**,** 97-111.

SHAO, D., VOGTMANN, E., LIU, A., QIN, J., CHEN, W., ABNET, C. C. & WEI, W. 2019. Microbial characterization of esophageal squamous cell carcinoma and gastric cardia adenocarcinoma from a high-risk region of China. *Cancer,* 125**,** 3993-4002.

SHEN, T., YANG, Z., CHENG, X., XIAO, Y., YU, K., CAI, X., XIA, C. & LI, Y. 2017. CXCL8 induces epithelial-mesenchymal transition in colon cancer cells via the PI3K/Akt/NF-κB signaling pathway. *Oncol Rep,* 37**,** 2095-2100.

SHI, H., WESTFALL, C. S., KAO, J., ODERMATT, P. D., ANDERSON, S. E., CESAR, S., SIEVERT, M., MOORE, J., GONZALEZ, C. G., ZHANG, L., ELIAS, J. E., CHANG, F., HUANG, K. C. & LEVIN, P. A. 2021. Starvation induces shrinkage of the bacterial cytoplasm. *Proceedings of the National Academy of Sciences,* 118**,** e2104686118.

SHI, M., LUO, F., SHAO, T., ZHANG, H., YANG, T., WEI, Y., CHEN, R. & GUO, R. 2022. Positive Correlation Between LTA Expression and Overall Immune Activity Suggests an Increased Probability of Survival in Uterine Corpus Endometrial Carcinoma. *Frontiers in Cell and Developmental Biology,* 9.

SHIBUYA, M. 2011. Vascular Endothelial Growth Factor (VEGF) and Its Receptor (VEGFR) Signaling in Angiogenesis: A Crucial Target for Anti- and Pro-Angiogenic Therapies. *Genes Cancer,* 2**,** 1097-105.

SICKERT, D., AUST, D. E., LANGER, S., HAUPT, I., BARETTON, G. B. & DIETER, P. 2005. Characterization of macrophage subpopulations in colon cancer using tissue microarrays. *Histopathology,* 46**,** 515-521.

SIEGEL, R., DESANTIS, C. & JEMAL, A. 2014. Colorectal cancer statistics, 2014. *CA: A Cancer Journal for Clinicians,* 64**,** 104-117.

SIEGEL, R. L., FEDEWA, S. A., ANDERSON, W. F., MILLER, K. D., MA, J., ROSENBERG, P. S. & JEMAL, A. 2017. Colorectal cancer incidence patterns in the United States, 1974–2013. *JNCI: Journal of the National Cancer Institute,* 109**,** djw322.

SIEGEL, R. L., MILLER, K. D., GODING SAUER, A., FEDEWA, S. A., BUTTERLY, L. F., ANDERSON, J. C., CERCEK, A., SMITH, R. A. & JEMAL, A. 2020. Colorectal cancer statistics, 2020. *CA: A Cancer Journal for Clinicians,* 70**,** 145-164.

SINGH, V. K., SYRING, M., SINGH, A., SINGHAL, K., DALECKI, A. & JOHANSSON, T. 2012. An insight into the significance of the DnaK heat shock system in Staphylococcus aureus. *International Journal of Medical Microbiology,* 302**,** 242-252.

SLAUGHTER, D. P., SOUTHWICK, H. W. & SMEJKAL, W. 1953. "Field cancerization" in oral stratified squamous epithelium. Clinical implications of multicentric origin. *Cancer,* 6**,** 963-968.

SONG, X., GAO, H., LIN, Y., YAO, Y., ZHU, S., WANG, J., LIU, Y., YAO, X., MENG, G., SHEN, N., SHI, Y., IWAKURA, Y. & QIAN, Y. 2014. Alterations in the Microbiota Drive Interleukin-17C Production from Intestinal Epithelial Cells to Promote Tumorigenesis. *Immunity,* 40**,** 140-152.

SONG, X., ZHU, S., SHI, P., LIU, Y., SHI, Y., LEVIN, S. D. & QIAN, Y. 2011. IL-17RE is the functional receptor for IL-17C and mediates mucosal immunity to infection with intestinal pathogens. *Nat Immunol,* 12**,** 1151-8.

STIPP, C. S., KOLESNIKOVA, T. V. & HEMLER, M. E. 2003. Functional domains in tetraspanin proteins. *Trends in Biochemical Sciences,* 28**,** 106-112.

STONES, D. H. & KRACHLER, A. M. 2016. Against the tide: the role of bacterial adhesion in host colonization. *Biochem Soc Trans,* 44**,** 1571-1580.

STRICKLER, J. H., SATAKE, H., GEORGE, T. J., YAEGER, R., HOLLEBECQUE, A., GARRIDO-LAGUNA, I., SCHULER, M., BURNS, T. F., COVELER, A. L., FALCHOOK, G. S., VINCENT, M., SUNAKAWA, Y., DAHAN, L., BAJOR, D., RHA, S.-Y., LEMECH, C., JURIC, D., REHN, M., NGARMCHAMNANRITH, G., JAFARINASABIAN, P., TRAN, Q. & HONG, D. S. 2022. Sotorasib in KRAS p.G12C–Mutated Advanced Pancreatic Cancer. *New England Journal of Medicine,* 388**,** 33-43.

SU, W., CHEN, Y., CAO, P., CHEN, Y., GUO, Y., WANG, S. & DONG, W. 2020. Fusobacterium nucleatum Promotes the Development of Ulcerative Colitis by Inducing the Autophagic Cell Death of Intestinal Epithelial. *Front Cell Infect Microbiol,* 10**,** 594806.

SUÁREZ, H., ROCHA-PERUGINI, V., ÁLVAREZ, S. & YÁÑEZ-MÓ, M. 2018. Tetraspanins, Another Piece in the HIV-1 Replication Puzzle. *Front Immunol,* 9**,** 1811.

SUN, C.-H., LI, B.-B., WANG, B., ZHAO, J., ZHANG, X.-Y., LI, T.-T., LI, W.-B., TANG, D., QIU, M.-J., WANG, X.-C., ZHU, C.-M. & QIAN, Z.-R. 2019. The role of Fusobacterium nucleatum in colorectal cancer: from carcinogenesis to clinical management. *Chronic Diseases and Translational Medicine,* 5**,** 178-187.

SUNG, H., FERLAY, J., SIEGEL, R. L., LAVERSANNE, M., SOERJOMATARAM, I., JEMAL, A. & BRAY, F. 2021. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin,* 71**,** 209-249.

SUSA, K. J., KRUSE, A. C. & BLACKLOW, S. C. 2024. Tetraspanins: structure, dynamics, and principles of partner-protein recognition. *Trends in Cell Biology,* 34**,** 509-522.

SVEEN, A., KOPETZ, S. & LOTHE, R. A. 2020. Biomarker-guided therapy for colorectal cancer: strength in complexity. *Nature Reviews Clinical Oncology,* 17**,** 11-32.

SZKLARCZYK, D., GABLE, A. L., LYON, D., JUNGE, A., WYDER, S., HUERTA-CEPAS, J., SIMONOVIC, M., DONCHEVA, N. T., MORRIS, J. H., BORK, P., JENSEN, L. J. & MERING, C. V. 2019. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res,* 47**,** D607-d613.

TANG, B., WANG, K., JIA, Y. P., ZHU, P., FANG, Y., ZHANG, Z. J., MAO, X. H., LI, Q. & ZENG, D. Z. 2016. Fusobacterium nucleatum-Induced Impairment of Autophagic Flux Enhances the Expression of Proinflammatory Cytokines via ROS in Caco-2 Cells. *PLoS One,* 11**,** e0165701.

TANG, Y., CHENG, Y., MARTINKA, M., ONG, C. J. & LI, G. 2014. Prognostic significance of KAI1/CD82 in human melanoma and its role in cell migration and invasion through the regulation of ING4. *Carcinogenesis,* 35**,** 86-95.

TAPIAL, S., OLMEDILLAS-LÓPEZ, S., RUEDA, D., ARRIBA, M., GARCÍA, J. L., VIVAS, A., PÉREZ, J., PENA-COUSO, L., OLIVERA, R., RODRÍGUEZ, Y., GARCÍA-ARRANZ, M., GARCÍA-OLMO, D., GONZÁLEZ-SARMIENTO, R., URIOSTE, M., GOEL, A. & PEREA, J. 2019. Cimp-Positive Status is More Representative in Multiple Colorectal Cancers than in Unique Primary Colorectal Cancers. *Scientific Reports,* 9**,** 10516.

TATUSOV, R. L., KOONIN, E. V. & LIPMAN, D. J. 1997. A genomic perspective on protein families. *Science,* 278**,** 631-7.

TAYLOR, J. C., GAO, X., XU, J., HOLDER, M., PETROSINO, J., KUMAR, R., LIU, W., HÖÖK, M., MACKENZIE, C., HILLHOUSE, A., BRASHEAR, W., NUNEZ, M. P. & XU, Y. 2021. A type VII secretion system of Streptococcus gallolyticus subsp. gallolyticus contributes to gut colonization and the development of colon tumors. *PLoS Pathog,* 17**,** e1009182.

TAYLOR, J. C., KUMAR, R., XU, J. & XU, Y. 2023. A pathogenicity locus of Streptococcus gallolyticus subspecies gallolyticus. *Scientific Reports,* 13**,** 6291.

THAM, T. N., GOUIN, E., RUBINSTEIN, E., BOUCHEIX, C., COSSART, P. & PIZARRO-CERDA, J. 2010. Tetraspanin CD81 is required for Listeria monocytogenes invasion. *Infect Immun,* 78**,** 204-9.

THURNHEER, T., KARYGIANNI, L., FLURY, M. & BELIBASAKIS, G. N. 2019. Fusobacterium Species and Subspecies Differentially Affect the Composition and Architecture of Supra- and Subgingival Biofilms Models. *Front Microbiol,* 10**,** 1716.

TIJHUIS, A. E., JOHNSON, S. C. & MCCLELLAND, S. E. 2019. The emerging links between chromosomal instability (CIN), metastasis, inflammation and tumour immunity. *Molecular Cytogenetics,* 12**,** 17.

TODD, S. C., DOCTOR, V. S. & LEVY, S. 1998. Sequences and expression of six new members of the tetraspanin/TM4SF family. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression,* 1399**,** 101-104.

TOUSSI, D. N., LIU, X. & MASSARI, P. 2012. The FomA porin from Fusobacterium nucleatum is a Toll-like receptor 2 agonist with immune adjuvant activity. *Clin Vaccine Immunol,* 19**,** 1093-101.

TSOI, K. K. F., PAU, C. Y. Y., WU, W. K. K., CHAN, F. K. L., GRIFFITHS, S. & SUNG, J. J. Y. 2009. Cigarette Smoking and the Risk of Colorectal Cancer: A Meta-analysis of Prospective Cohort Studies. *Clinical Gastroenterology and Hepatology,* 7**,** 682-688.e5.

TURNER, M. D., NEDJAI, B., HURST, T. & PENNINGTON, D. J. 2014. Cytokines and chemokines: At the crossroads of cell signalling and inflammatory disease. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research,* 1843**,** 2563-2582.

TZENG, J., LU, H. H.-S. & LI, W.-H. 2008. Multidimensional scaling for large genomic data sets. *BMC Bioinformatics,* 9**,** 179.

UMAÑA, A., SANDERS, B. E., YOO, C. C., CASASANTA, M. A., UDAYASURYAN, B., VERBRIDGE, S. S. & SLADE, D. J. 2019. Utilizing Whole Fusobacterium Genomes To Identify, Correct, and Characterize Potential Virulence Protein Families. *J Bacteriol,* 201.

UMEDA, R., SATOUH, Y., TAKEMOTO, M., NAKADA-NAKURA, Y., LIU, K., YOKOYAMA, T., SHIROUZU, M., IWATA, S., NOMURA, N., SATO, K., IKAWA, M., NISHIZAWA, T. & NUREKI, O. 2020. Structural insights into tetraspanin CD9 function. *Nature Communications,* 11**,** 1606.

USUI, M., ONIZUKA, S., SATO, T., KOKABU, S., ARIYOSHI, W. & NAKASHIMA, K. 2021. Mechanism of alveolar bone destruction in periodontitis — Periodontal bacteria and inflammation. *Japanese Dental Science Review,* 57**,** 201-208.

VAN SPRIEL, A. B. & FIGDOR, C. G. 2010. The role of tetraspanins in the pathogenesis of infectious diseases. *Microbes and Infection,* 12**,** 106-112.

VENCES-CATALAN, F. & LEVY, S. 2022. Tetraspanins in cell stemness and cancer initiation: markers or active players? *Trends in Cell Biology,* 32**,** 377-379.

VERNIA, F., LONGO, S., STEFANELLI, G., VISCIDO, A. & LATELLA, G. 2021. Dietary Factors Modulating Colorectal Carcinogenesis. *Nutrients,* 13.

VESELOVSKY, V. A., DYACHKOVA, M. S., BESPIATYKH, D. A., YUNES, R. A., SHITIKOV, E. A., POLYAEVA, P. S., DANILENKO, V. N., OLEKHNOVICH, E. I. & KLIMINA, K. M. 2022. The Gene Expression Profile Differs in Growth Phases of the Bifidobacterium Longum Culture. *Microorganisms,* 10.

VIERA, M., YIP, G. W. C., SHEN, H. M., BAEG, G. H. & BAY, B. H. 2021. Targeting CD82/KAI1 for Precision Therapeutics in Surmounting Metastatic Potential in Breast Cancer. *Cancers (Basel),* 13.

WALKER, S. L., HILL, J. E., REDMAN, J. A. & ELIMELECH, M. 2005. Influence of growth phase on adhesion kinetics of Escherichia coli D21g. *Appl Environ Microbiol,* 71**,** 3093-9.

WANG, K., ZHENG, J., YU, J., WU, Y., GUO, J., XU, Z. & SUN, X. 2020. Knockdown of MMP-1 inhibits the progression of colorectal cancer by suppressing the PI3K/Akt/c-myc signaling pathway and EMT. *Oncol Rep,* 43**,** 1103-1112.

WANG, Q., ZHAO, L., XU, C., ZHOU, J. & WU, Y. 2019a. Fusobacterium nucleatum stimulates monocyte adhesion to and transmigration through endothelial cells. *Archives of Oral Biology,* 100**,** 86-92.

WANG, T., SONG, P., ZHONG, T., WANG, X., XIANG, X., LIU, Q., CHEN, H., XIA, T., LIU, H., NIU, Y., HU, Y., XU, L., SHAO, Y., ZHU, L., QI, H., SHEN, J., HOU, T., FODDE, R. & SHAO, J. 2019b. The inflammatory cytokine IL-6 induces FRA1 deacetylation promoting colorectal cancer stem-like properties. *Oncogene,* 38**,** 4932-4947.

WANG, X., O'CONNELL, K., JEON, J., SONG, M., HUNTER, D., HOFFMEISTER, M., LIN, Y., BERNDT, S., BRENNER, H., CHAN, A. T., CHANG-CLAUDE, J., GONG, J., GUNTER, M. J., HARRISON, T. A., HAYES, R. B., JOSHI, A., NEWCOMB, P., SCHOEN, R., SLATTERY, M. L., VARGAS, A., POTTER, J. D., LE MARCHAND, L., GIOVANNUCCI, E., WHITE, E., HSU, L., PETERS, U. & DU, M. 2019c. Combined effect of modifiable and non-modifiable risk factors for colorectal cancer risk in a pooled analysis of 11 population-based studies. *BMJ Open Gastroenterol,* 6**,** e000339.

WANG, Y., REN, Y., HUANG, Y., YU, X., YANG, Y., WANG, D., SHI, L., TAO, K., WANG, G. & WU, K. 2021. Fungal dysbiosis of the gut microbiota is associated with colorectal cancer in Chinese patients. *Am J Transl Res,* 13**,** 11287-11301.

WANG, Y., TONG, X., OMOREGIE, E. S., LIU, W., MENG, S. & YE, X. 2012. Tetraspanin 6 (TSPAN6) negatively regulates retinoic acid-inducible gene I-like receptor-mediated immune signaling in a ubiquitination-dependent manner. *J Biol Chem,* 287**,** 34626-34.

WANG, Y., WANG, L., SUN, T., SHEN, S., LI, Z., MA, X., GU, X., ZHANG, X., PENG, A., XU, X. & FENG, Q. 2023a. Study of the inflammatory activating process in the early stage of Fusobacterium nucleatum infected PDLSCs. *International Journal of Oral Science,* 15**,** 8.

WANG, Z., PETERS, B. A., BRYANT, M., HANNA, D. B., SCHWARTZ, T., WANG, T., SOLLECITO, C. C., USYK, M., GRASSI, E., WIEK, F., PETER, L. S., POST, W. S., LANDAY, A. L., HODIS, H. N., WEBER, K. M., FRENCH, A., GOLUB, E. T., LAZAR, J., GUSTAFSON, D., SHARMA, A., ANASTOS, K., CLISH, C. B., BURK, R. D., KAPLAN, R. C., KNIGHT, R. & QI, Q. 2023b. Gut microbiota, circulating inflammatory markers and metabolites, and carotid artery atherosclerosis in HIV infection. *Microbiome,* 11**,** 119.

WARREN, R. L., FREEMAN, D. J., PLEASANCE, S., WATSON, P., MOORE, R. A., COCHRANE, K., ALLEN-VERCOE, E. & HOLT, R. A. 2013. Co-occurrence of anaerobic bacteria in colorectal carcinomas. *Microbiome,* 1**,** 16.

WATANABE, S., ZENKE, K. & MUROI, M. 2023. Lipoteichoic Acid Inhibits Lipopolysaccharide-Induced TLR4 Signaling by Forming an Inactive TLR4/MD-2 Complex Dimer. *The Journal of Immunology,* 210**,** 1386-1395.

WEBSTER, J. D. & VUCIC, D. 2020. The Balance of TNF Mediated Pathways Regulates Inflammatory Cell Death Signaling in Healthy and Diseased Tissues. *Frontiers in Cell and Developmental Biology,* 8.

WEITZ, J., KOCH, M., DEBUS, J., HÖHLER, T., GALLE, P. R. & BÜCHLER, M. W. 2005. Colorectal cancer. *The Lancet,* 365**,** 153-165.

WESTERMANN, A. J., GORSKI, S. A. & VOGEL, J. 2012. Dual RNA-seq of pathogen and host. *Nature Reviews Microbiology,* 10**,** 618-630.

WHITE, A., IRONMONGER, L., STEELE, R. J. C., ORMISTON-SMITH, N., CRAWFORD, C. & SEIMS, A. 2018. A review of sex-related differences in colorectal cancer incidence, screening uptake, routes to diagnosis, cancer stage and survival in the UK. *BMC Cancer,* 18**,** 906.

WOLANIN, P. M., WEBRE, D. J. & STOCK, J. B. 2003. Mechanism of Phosphatase Activity in the Chemotaxis Response Regulator CheY. *Biochemistry,* 42**,** 14075-14082.

WOLIN, K. Y., YAN, Y., COLDITZ, G. A. & LEE, I. M. 2009. Physical activity and colon cancer prevention: a meta-analysis. *British Journal of Cancer,* 100**,** 611-616.

WRIGHT, M. D., MOSELEY, G. W. & VAN SPRIEL, A. B. 2004. Tetraspanin microdomains in immune cell signalling and malignant disease. *Tissue Antigens,* 64**,** 533-542.

WROBLEWSKI, L. E., PEEK, R. M., JR. & WILSON, K. T. 2010. Helicobacter pylori and gastric cancer: factors that modulate disease risk. *Clin Microbiol Rev,* 23**,** 713-39.

WU, C., CHEN, Y.-W., SCHEIBLE, M., CHANG, C., WITTCHEN, M., LEE, J. H., LUONG, T. T., TINER, B. L., TAUCH, A., DAS, A. & TON-THAT, H. 2021. Genetic and molecular determinants of polymicrobial interactions in Fusobacterium nucleatum. *Proceedings of the National Academy of Sciences,* 118**,** e2006482118.

WU, J., LI, Q. & FU, X. 2019a. Fusobacterium nucleatum Contributes to the Carcinogenesis of Colorectal Cancer by Inducing Inflammation and Suppressing Host Immunity. *Translational Oncology,* 12**,** 846-851.

WU, N., FENG, Y. Q., LYU, N., WANG, D., YU, W. D. & HU, Y. F. 2022. Fusobacterium nucleatum promotes colon cancer progression by changing the mucosal microbiota and colon transcriptome in a mouse model. *World J Gastroenterol,* 28**,** 1981-1995.

WU, T. R., LIN, C. S., CHANG, C. J., LIN, T. L., MARTEL, J., KO, Y. F., OJCIUS, D. M., LU, C. C., YOUNG, J. D. & LAI, H. C. 2019b. Gut commensal Parabacteroides goldsteinii plays a predominant role in the anti-obesity effects of polysaccharides isolated from Hirsutella sinensis. *Gut,* 68**,** 248-262.

WU, Y., GUO, S., CHEN, F., LI, Y., HUANG, Y., LIU, W. & ZHANG, G. 2023. Fn-Dps, a novel virulence factor of Fusobacterium nucleatum, disrupts erythrocytes and promotes metastasis in colorectal cancer. *PLoS Pathog,* 19**,** e1011096.

WU, Y., WANG, M., ZHU, Y. & LIN, S. 2016a. Serum interleukin-6 in the diagnosis of bacterial infection in cirrhotic patients: A meta-analysis. *Medicine (Baltimore),* 95**,** e5127.

WU, Z., HUANG, X., HAN, X., LI, Z., ZHU, Q., YAN, J., YU, S., JIN, Z., WANG, Z., ZHENG, Q. & WANG, Y. 2016b. The chemokine CXCL9 expression is associated with better prognosis for colorectal carcinoma patients. *Biomedicine & Pharmacotherapy,* 78**,** 8-13.

XIE, D., PEI, Q., LI, J., WAN, X. & YE, T. 2021. Emerging Role of E2F Family in Cancer Stem Cells. *Front Oncol,* 11**,** 723137.

XIE, H., GIBBONS, R. J. & HAY, D. I. 1991. Adhesive properties of strains of Fusobacterium nucleatum of the subspecies nucleatum, vincentii and polymorphum. *Oral Microbiology and Immunology,* 6**,** 257-263.

XU, J., YE, Y., ZHANG, H., SZMITKOWSKI, M., MÄKINEN, M. J., LI, P., XIA, D., YANG, J., WU, Y. & WU, H. 2016. Diagnostic and Prognostic Value of Serum Interleukin-6 in Colorectal Cancer. *Medicine (Baltimore),* 95**,** e2502.

XU, M., YAMADA, M., LI, M., LIU, H., CHEN, S. G. & HAN, Y. W. 2007. FadA from Fusobacterium nucleatum Utilizes both Secreted and Nonsecreted Forms for Functional Oligomerization for Attachment and Invasion of Host Cells*. *Journal of Biological Chemistry,* 282**,** 25000-25009.

XU, Y., ZHAO, J., MA, Y., LIU, J., CUI, Y., YUAN, Y., XIANG, C., MA, D. & LIU, H. 2023. The microbiome types of colorectal tissue are potentially associated with the prognosis of patients with colorectal cancer. *Frontiers in Microbiology,* 14.

XU, Z., ZHANG, Y., XU, M., ZHENG, X., LIN, M., PAN, J., YE, C., DENG, Y., JIANG, C., LIN, Y., LU, X. & CHI, P. 2019. Demethylation and Overexpression of CSF2 are Involved in Immune Response, Chemotherapy Resistance, and Poor Prognosis in Colorectal Cancer. *Onco Targets Ther,* 12**,** 11255-11269.

YAMAGISHI, H., KURODA, H., IMAI, Y. & HIRAISHI, H. 2016. Molecular pathogenesis of sporadic colorectal cancers. *Chin J Cancer,* 35**,** 4.

YAMAMURA, K., BABA, Y., NAKAGAWA, S., MIMA, K., MIYAKE, K., NAKAMURA, K., SAWAYAMA, H., KINOSHITA, K., ISHIMOTO, T., IWATSUKI, M., SAKAMOTO, Y., YAMASHITA, Y., YOSHIDA, N., WATANABE, M. & BABA, H. 2016. Human Microbiome Fusobacterium Nucleatum in Esophageal Cancer Tissue Is Associated with Prognosis. *Clinical Cancer Research,* 22**,** 5574-5581.

YAMANE, L., SCAPULATEMPO-NETO, C., REIS, R. M. & GUIMARÃES, D. P. 2014. Serrated pathway in colorectal carcinogenesis. *World J Gastroenterol,* 20**,** 2634-40.

YAMAUCHI, M., LOCHHEAD, P., MORIKAWA, T., HUTTENHOWER, C., CHAN, A. T., GIOVANNUCCI, E., FUCHS, C. & OGINO, S. 2012. Colorectal cancer: a tale of two sides or a continuum? *Gut,* 61**,** 794-7.

YANG, X. H., RICHARDSON, A. L., TORRES-ARZAYUS, M. I., ZHOU, P., SHARMA, C., KAZAROV, A. R., ANDZELM, M. M., STROMINGER, J. L., BROWN, M. & HEMLER, M. E. 2008. CD151 Accelerates Breast Cancer by Regulating α6 Integrin Function, Signaling, and Molecular Organization. *Cancer Research,* 68**,** 3204-3213.

YANG, Y., WENG, W., PENG, J., HONG, L., YANG, L., TOIYAMA, Y., GAO, R., LIU, M., YIN, M., PAN, C., LI, H., GUO, B., ZHU, Q., WEI, Q., MOYER, M. P., WANG, P., CAI, S., GOEL, A., QIN, H. & MA, Y. 2017. Fusobacterium nucleatum Increases Proliferation of Colorectal Cancer Cells and Tumor Development in Mice by Activating Toll-Like Receptor 4 Signaling to Nuclear Factor-κB, and Up-regulating Expression of MicroRNA-21. *Gastroenterology,* 152**,** 851-866.e24.

YANG, Y.-N., WANG, Q.-C., XU, W., YU, J., ZHANG, H. & WU, C. 2022. The berberine-enriched gut commensal Blautia producta ameliorates high-fat diet (HFD)-induced

hyperlipidemia and stimulates liver LDLR expression. *Biomedicine & Pharmacotherapy,* 155**,** 113749.

YORK, W. S., MAZUMDER, R., RANZINGER, R., EDWARDS, N., KAHSAY, R., AOKI-KINOSHITA, K. F., CAMPBELL, M. P., CUMMINGS, R. D., FEIZI, T., MARTIN, M., NATALE, D. A., PACKER, N. H., WOODS, R. J., AGARWAL, G., ARPINAR, S., BHAT, S., BLAKE, J., CASTRO, L. J. G., FOCHTMAN, B., GILDERSLEEVE, J., GOLDMAN, R., HOLMES, X., JAIN, V., KULKARNI, S., MAHADIK, R., MEHTA, A., MOUSAVI, R., NAKARAKOMMULA, S., NAVELKAR, R., PATTABIRAMAN, N., PIERCE, M. J., ROSS, K., VASUDEV, P., VORA, J., WILLIAMSON, T. & ZHANG, W. 2020. GlyGen: Computational and Informatics Resources for Glycoscience. *Glycobiology,* 30**,** 72-73.

YU, T., GUO, F., YU, Y., SUN, T., MA, D., HAN, J., QIAN, Y., KRYCZEK, I., SUN, D., NAGARSHETH, N., CHEN, Y., CHEN, H., HONG, J., ZOU, W. & FANG, J.-Y. 2017a. Fusobacterium nucleatum Promotes Chemoresistance to Colorectal Cancer by Modulating Autophagy. *Cell,* 170**,** 548-563.e16.

YU, T., GUO, F., YU, Y., SUN, T., MA, D., HAN, J., QIAN, Y., KRYCZEK, I., SUN, D., NAGARSHETH, N., CHEN, Y., CHEN, H., HONG, J., ZOU, W. & FANG, J. Y. 2017b. Fusobacterium nucleatum Promotes Chemoresistance to Colorectal Cancer by Modulating Autophagy. *Cell,* 170**,** 548-563.e16.

ZEPEDA-RIVERA, M., MINOT, S. S., BOUZEK, H., WU, H., BLANCO-MÍGUEZ, A., MANGHI, P., JONES, D. S., LACOURSE, K. D., WU, Y., MCMAHON, E. F., PARK, S.-N., LIM, Y. K., KEMPCHINSKY, A. G., WILLIS, A. D., COTTON, S. L., YOST, S. C., SICINSKA, E., KOOK, J.-K., DEWHIRST, F. E., SEGATA, N., BULLMAN, S. & JOHNSTON, C. D. 2024. A distinct Fusobacterium nucleatum clade dominates the colorectal cancer niche. *Nature,* 628**,** 424-432.

ZHANG, L., LIU, Y., ZHENG, H. J. & ZHANG, C. P. 2019. The Oral Microbiota May Have Influence on Oral Cancer. *Front Cell Infect Microbiol,* 9**,** 476.

ZHANG, S., LI, C., LIU, J., GENG, F., SHI, X., LI, Q., LU, Z. & PAN, Y. 2020. Fusobacterium nucleatum promotes epithelial-mesenchymal transiton through regulation of the lncRNA MIR4435-2HG/miR-296-5p/Akt2/SNAI1 signaling pathway. *Febs j,* 287**,** 4032-4047.

ZHANG, X., YU, D., WU, D., GAO, X., SHAO, F., ZHAO, M., WANG, J., MA, J., WANG, W., QIN, X., CHEN, Y., XIA, P. & WANG, S. 2023. Tissue-resident Lachnospiraceae family bacteria protect against colorectal carcinogenesis by promoting tumor immune surveillance. *Cell Host & Microbe,* 31**,** 418-432.e8.

ZHANG, Y., ZHANG, L., ZHENG, S., LI, M., XU, C., JIA, D., QI, Y., HOU, T., WANG, L., WANG, B., LI, A., CHEN, S., SI, J. & ZHUO, W. 2022. Fusobacterium nucleatum promotes colorectal cancer cells adhesion to endothelial cells and facilitates extravasation and metastasis by inducing ALPK1/NF-κB/ICAM1 axis. *Gut Microbes,* 14**,** 2038852.

ZHANG, Y. J., LI, S., GAN, R. Y., ZHOU, T., XU, D. P. & LI, H. B. 2015. Impacts of gut bacteria on human health and diseases. *Int J Mol Sci,* 16**,** 7493-519.

ZHAO, H., CHU, M., HUANG, Z., YANG, X., RAN, S., HU, B., ZHANG, C. & LIANG, J. 2017. Variations in oral microbiota associated with oral cancer. *Scientific Reports,* 7**,** 11773.

ZHAO, H., MING, T., TANG, S., REN, S., YANG, H., LIU, M., TAO, Q. & XU, H. 2022a. Wnt signaling in colorectal cancer: pathogenic role and therapeutic target. *Molecular Cancer,* 21**,** 144.

ZHAO, H., WU, L., YAN, G., CHEN, Y., ZHOU, M., WU, Y. & LI, Y. 2021a. Inflammation and tumor progression: signaling pathways and targeted intervention. *Signal Transduction and Targeted Therapy,* 6**,** 263.

ZHAO, L., ZHANG, X., ZHOU, Y., FU, K., LAU, H. C.-H., CHUN, T. W.-Y., CHEUNG, A. H.-K., COKER, O. O., WEI, H., WU, W. K.-K., WONG, S. H., SUNG, J. J.-Y., TO, K. F. & YU, J. 2022b. Parvimonas micra promotes colorectal tumorigenesis and is associated with prognosis of colorectal cancer patients. *Oncogene,* 41**,** 4200-4210.

ZHAO, T., LI, Y., SHEN, K., WANG, Q. & ZHANG, J. 2021b. Knockdown of OLR1 weakens glycolytic metabolism to repress colon cancer cell proliferation and chemoresistance by downregulating SULT2B1 via c-MYC. *Cell Death Dis,* 13**,** 4.

ZHOU, G., MO, W. J., SEBBEL, P., MIN, G., NEUBERT, T. A., GLOCKSHUBER, R., WU, X. R., SUN, T. T. & KONG, X. P. 2001. Uroplakin Ia is the urothelial receptor for uropathogenic Escherichia coli: evidence from in vitro FimH binding. *J Cell Sci,* 114**,** 4095-103.

ZHOU, J., LIU, L., WU, P., ZHAO, L. & WU, Y. 2022. Fusobacterium nucleatum Accelerates Atherosclerosis via Macrophage-Driven Aberrant Proinflammatory Response and Lipid Metabolism. *Frontiers in Microbiology,* 13.

ZHOU, L.-J., LIN, W.-Z., MENG, X.-Q., ZHU, H., LIU, T., DU, L.-J., BAI, X.-B., CHEN, B.-Y., LIU, Y., XU, Y., XIE, Y., SHU, R., CHEN, F.-M., ZHU, Y.-Q. & DUAN, S.-Z. 2023a. Periodontitis exacerbates atherosclerosis through Fusobacterium nucleatum-promoted hepatic glycolysis and lipogenesis. *Cardiovascular Research*.

ZHOU, Z., YANG, Z., ZHOU, L., YANG, M. & HE, S. 2023b. The versatile roles of testrapanins in cancer from intracellular signaling to cell–cell communication: cell membrane proteins without ligands. *Cell & Bioscience,* 13**,** 59.

ZHU, A., IBRAHIM, J. G. & LOVE, M. I. 2019. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics,* 35**,** 2084-2092.

ZHU, G., PEI, L., XIA, H., TANG, Q. & BI, F. 2021a. Role of oncogenic KRAS in the prognosis, diagnosis and treatment of colorectal cancer. *Molecular Cancer,* 20**,** 143.

ZHU, J., CAI, T., ZHOU, J., DU, W., ZENG, Y., LIU, T., FU, Y., LI, Y., QIAN, Q., YANG, X. H., LI, Q., HUANG, J.-A. & LIU, Z. 2021b. CD151 drives cancer progression depending on integrin α3β1 through EGFR signaling in non-small cell lung cancer. *Journal of Experimental & Clinical Cancer Research,* 40**,** 192.

ZHU, L., CHEN, S. & CHEN, Y. 2011. Unraveling the biological functions of Smad7 with mouse models. *Cell & Bioscience,* 1**,** 44.

ZHU, L., CHEN, Y., DU, H., CONG, Y., YAN, W., MA, K. & HUANG, X. 2022. N-glycosylation of CD82 at Asn157 is required for suppressing migration and invasion by reversing EMT via Wnt/β-catenin pathway in colon cancer. *Biochemical and Biophysical Research Communications,* 629**,** 121-127.

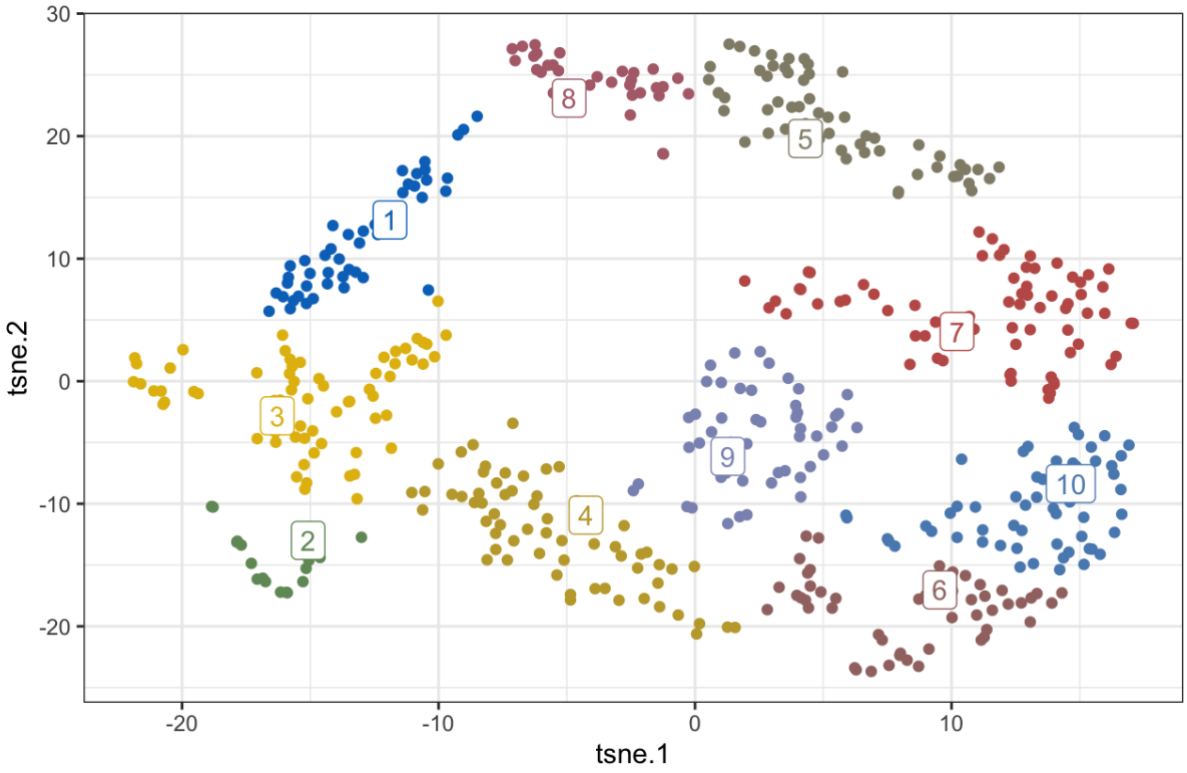
ZIPIN-ROITMAN, A., MESHEL, T., SAGI-ASSIF, O., SHALMON, B., AVIVI, C., PFEFFER, R. M., WITZ, I. P. & BEN-BARUCH, A. 2007. CXCL10 Promotes Invasion-Related Properties in Human Colorectal Carcinoma Cells. *Cancer Research,* 67**,** 3396-3405.

ZOU, F., WANG, X., HAN, X., ROTHSCHILD, G., ZHENG, S. G., BASU, U. & SUN, J. 2018. Expression and Function of Tetraspanins and Their Interacting Partners in B Cells. *Frontiers in Immunology,* 9.

2011. G2/M Checkpoint. *In:* SCHWAB, M. (ed.) *Encyclopedia of Cancer.* Berlin, Heidelberg: Springer Berlin Heidelberg.

9 APPENDICES

9.1 Appendix 6.1



*Supplementary figure 1**: T-SNE clusters of bacterial strains**. 10 groups were formed based on the species profile of the samples.*