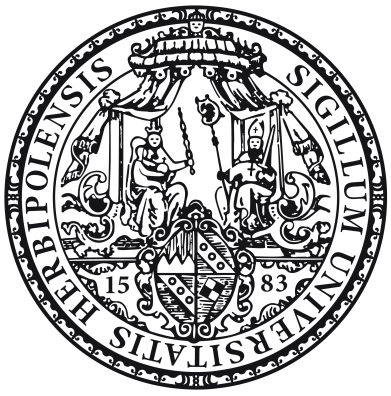

Investigating the molecular biology of *Fusobacterium nucleatum*

Untersuchung der Molekularbiologie von *Fusobacterium nucleatum*

Doctoral thesis for a doctoral degree
at the Graduate School of Life Sciences,
Julius-Maximilians-Universität Würzburg,
Section Infection and Immunity

Falk Fred Finn Ponath

from

Husum, Germany

Würzburg 2022

Submitted on: 29.09.2022

Members of the *Promotionskomitee*:

Chairperson: Prof. Dr. Georg Gasteiger

Primary Supervisor: Prof. Dr. Jörg Vogel

Supervisor (Second): Prof. Dr. Chase Beisel

Supervisor (Third): Dr. Ana Rita Brochado

Supervisor (Fourth): Prof. Dr. Till Strowig

Date of Public Defense:

Date of Receipt of Certificates:

When forgiveness is our torch and imagination our sword
Well I'll tie the ropes of hate and slash open the minds of the bored
And we'll start a world so equal and free
Every inch of this earth is yours all the land and all the sea
Imagine no restrictions but the climate and the weather
Then we can explore space together
Forever
∼Rou Reynolds, Enter Shikari∼

Summary

The anaerobe *Fusobacterium nucleatum* (*F. nucleatum*) is an important member of the oral microbiome but can also colonize different tissues of the human body. In particular, its association with multiple human cancers has drawn much attention. This association has prompted growing interest into the interaction of *F. nucleatum* with cancer, with studies focusing primarily on the host cells. At the same time, *F. nucleatum* itself remains poorly understood, which includes its transcriptomic architecture but also gene regulation such as global stress responses that typically enable survival of bacteria in new environments. An important aspect of such regulatory networks is the post-transcriptional regulation, which is entirely unknown in *F. nucleatum*. This paucity extents to any knowledge on small regulatory RNAs (sRNAs), despite their important role as post-transcriptional regulators of the bacterial physiology.

Investigating the above stated aspects is further complicated by the fact that *F. nucleatum* is phylogenetically distant from all other bacteria, displays very limited genetic tractability and lacks genetic tools for dissecting gene function.

This leaves many open questions on basic gene regulation in *F. nucleatum*, such as if the bacterium combines transcriptional and post-transcriptional regulation in its adaptation to a changing environment.

To begin answering this question, this works elucidated the transcriptomic landscape of *F. nucleatum* by performing differential RNA-seq (dRNA-seq). Conducted for five representative strains of all *F. nucleatum* subspecies and the closely related *F. periodonticum*, the analysis globally uncovered transcriptional start sites (TSS), 5'untranslated regions (UTRs) and improved the existing annotation. Importantly, the dRNA-seq analysis also identified a conserved suite of sRNAs specific to *Fusobacterium*.

The development of five genetic tools enabled further investigations of gene functions in *F. nucleatum*. These include vectors that enable the expression of different fluorescent proteins, inducible gene expression and scarless gene deletion in addition to transcriptional and translational reporter systems.

These tools enabled the dissection of a σ^E response and uncovered several commonalities with its counterpart in the phylogenetically distant Proteobacteria. The

similarities include the upregulation of genes involved in membrane homeostasis but also a σ^E-dependent regulatory sRNA. Surprisingly, oxygen was found to activated σ^E in *F. nucleatum* contrasting the typical role of the σ factor in envelope stress.

The non-coding σ^E-dependent sRNA, named FoxI, was shown to repress the translation of several envelope proteins which represented yet another parallel to the envelope stress response in Proteobacteria.

Overall, this work sheds light on the RNA landscape of the cancer-associated bacterium leading to the discovery of a conserved global stress response consisting of a coding and a non-coding arm. The development of new genetic tools not only aided the latter discovery but also provides the means for further dissecting the molecular and infection biology of this enigmatic bacterium.

Zusammenfassung

Das anaerobe Bakterium *Fusobacterium nucleatum* (*F. nucleatum*) ist ein wichtiger Bestandteil des oralen Mikrobioms, kann aber auch verschiedene Gewebe des menschlichen Körpers besiedeln. Insbesondere seine Verbindung mit mehreren menschlichen Krebsarten hat viel Aufmerksamkeit auf sich gezogen. Diese Assoziation hat zu einem wachsenden Interesse an der Interaktion von *F. nucleatum* mit Krebs geführt, wobei sich die Untersuchungen in erster Linie auf die Wirtszellen konzentrieren. Gleichzeitig ist *F. nucleatum* selbst nach wie vor schlecht verstanden, einschließlich seiner transkriptomischen Architektur, als auch der Genregulation, wie z. B. globale Stressreaktionen, die typischerweise das Überleben von Bakterien in neuen Umgebungen ermöglichen. Ein wichtiger Aspekt solcher regulatorischer Netzwerke ist die post-transkriptionelle Regulation, die bei *F. nucleatum* völlig unbekannt ist. Diese Unkenntnis erstreckt sich auch auf das Wissen über kleine regulatorische RNAs, trotz ihrer wichtigen Rolle als post-transkriptionelle Regulatoren der bakteriellen Physiologie.

Die Untersuchung der oben genannten Aspekte wird zusätzlich durch die Tatsache erschwert, dass *F. nucleatum* phylogenetisch von allen anderen Bakterien weit entfernt ist, eine sehr begrenzte genetische Traktabilität aufweist und keine genetischen Werkzeuge zur Untersuchung der Genfunktion vorliegen.

Dies führt zu vielen offenen Fragen bezüglich grundlegendener Genregulation in *F. nucleatum*, z. B. ob das Bakterium transkriptionelle und post-transkriptionelle Regulation kombiniert, um sich an eine sich verändernde Umwelt anzupassen.

Als erster Schritt zur Beantwortung dieser Frage wurde in dieser Arbeit die transkriptomische Landschaft von *F. nucleatum* durch differential RNA-seq (dRNA-seq) aufgeklärt. Anhand von fünf repräsentativen Stämmen aller Unterarten von *F. nucleatum* und dem eng verwandten *F. periodonticum* wurden durch die Analyse global transkriptionelle Startstellen (TSS) und 5'untranslatierte Regionen (5'UTRs) aufgedeckt als auch die bestehende Annotation verbessert. Weiterhin konnte die dRNA-seq-Analyse auch eine konservierte Anzahl von *Fusobacterium*-spezifischen sRNAs identifizieren.

Die Entwicklung von fünf genetischen Werkzeugen ermöglichte weitere Untersuchungen der Genfunktionen in *F. nucleatum*. Dazu gehören Vektoren, welche die

Expression verschiedener fluoreszierender Proteine ermöglichen als auch Systeme für die induzierbare Genexpression, narbenlose Gendeletion sowie transkriptionelle und translationale Reportersysteme.

Mit diesen Werkzeugen konnte die σ^E Antwort entschlüsselt werden, welche mehrere Gemeinsamkeiten mit ihrem Gegenstück in den phylogenetisch entfernten Proteobakterien aufweist. Zu diesen Gemeinsamkeiten gehört die Hochregulierung von Genen, die an der Membranhomöostase beteiligt sind, aber auch eine σ^E-abhängige regulatorische sRNA. Überraschenderweise wurde festgestellt, dass Sauerstoff σ^E in *F. nucleatum* aktiviert, was im Gegensatz zu der typischen Rolle des σ-Faktors bei Membranstress steht.

Die nicht-kodierende sRNA mit dem Namen FoxI, die von σ^E abhängt, unterdrückt nachweislich die Translation verschiedener Membranproteine, was eine weitere Parallele zur Membranstressreaktion in Proteobakterien darstellt.

Insgesamt wirft diese Arbeit Licht auf die RNA-Landschaft des krebsassoziierten Bakteriums und führt zur Entdeckung einer konservierten globalen Stressantwort, die aus einem kodierenden und einem nicht-kodierenden Arm besteht. Die Entwicklung neuer genetischer Werkzeuge hat nicht nur zu dieser Entdeckung beigetragen, sondern bietet auch die Möglichkeit, die Molekular- und Infektionsbiologie dieses rätselhaften Bakteriums weiter zu entschlüsseln.

Contents

List of Figures

List of Tables

1 Introduction

Recent advances in technology show how important the human microbiome is for overall health (Fan and Pedersen, 2021). However, they also show that the microbiome consists of over 1000 bacterial species with the majority being unidentified or understudied (Gilbert et al., 2018). Concomitantly, we lack knowledge on how these bacteria adapt to their surroundings which can often drastically change as exemplified by pathogens that travel from the environment into the host. Identifying the underlying processes of this adaption for microbiota members can thus deepen our understanding of their general biology but also help discern potential therapeutic targets to combat disease-associated bacteria. One such medically relevant bacterium is *Fusobacterium nucleatum* (*F. nucleatum*) (Brennan and Garrett, 2019) which together with the investigation of a stress response will be the focus of this work.

1.1 *Fusobacterium nucleatum*: An early-branching oral bacterium

The Gram-negative, obligate anaerobe, non-motile and rod-shaped *F. nucleatum* belongs to the distinct phylum of Fusobacteriota (Fig. 1.1). Recent work showed that the entire Fusobacteriota likely diverged early on from the last bacterial common ancestor (LBCA) compared to all known bacterial phyla (Coleman et al., 2021).

The genus of *Fusobacterium* further displays an adaptive radiation with *F. nucleatum* belonging to the lineage capable of actively invading host cells (Manson McGuire et al., 2014). Furthermore, the species of *F. nucleatum* consists out of the four subspecies (subsp.): *F. nucleatum* subsp. *animalis*, *nucleatum*, *polymorphum* and *vincentii* displaying a high degree of heterogeneity commonly found between individual species (Kook et al., 2017; Bi et al., 2022).

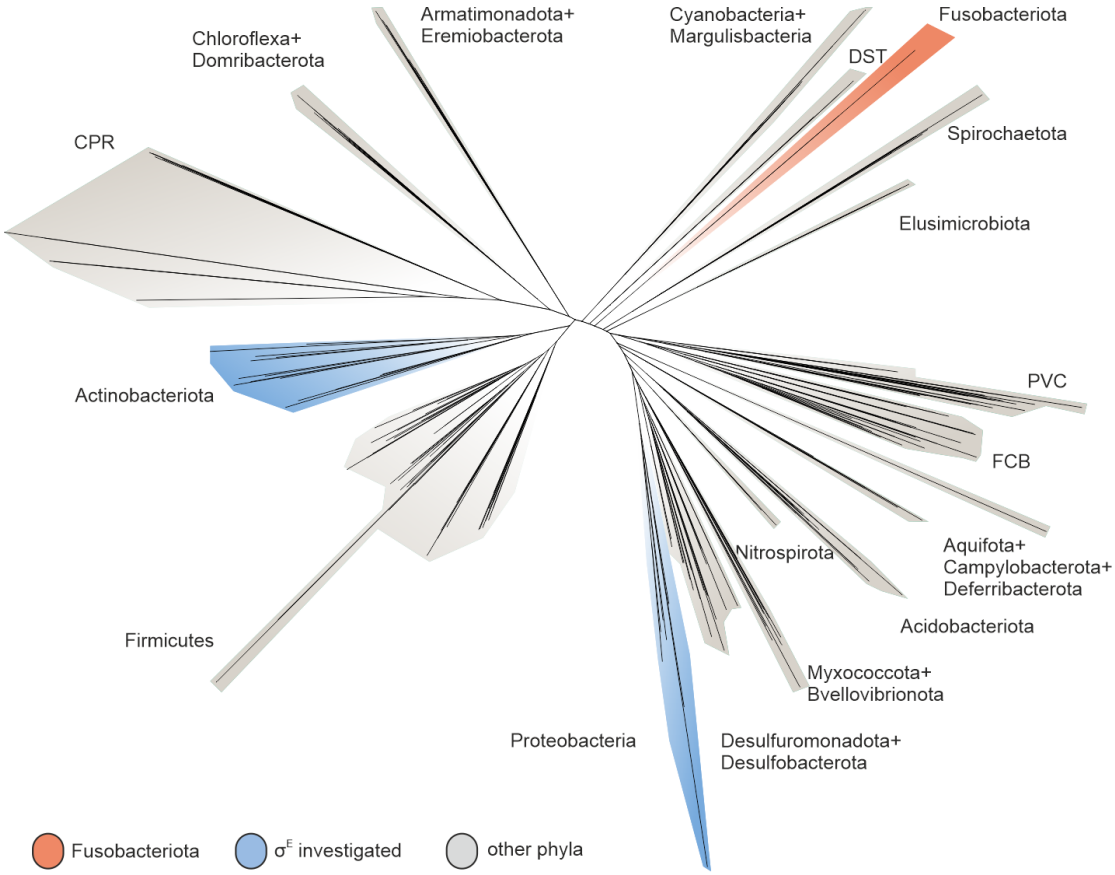

Figure 1.1 | The phylogenetic distance of Fusobacteriota from other phyla. A phylogenetic tree for 265 bacterial species. The the underlying data were generated in Coleman et al. (2021). The figure was modified from Ponath et al. (2022b).

Despite this diversity, all thus far isolated *F. nucleatum* strains share a small genome size of 1.8 to 2.5 megabases (Mbs) with a low GC-content (∼27 %) (Ang et al., 2016).

1.1.1 The microbiome member *F. nucleatum*

F. nucleatum is a host-associated bacterium and an abundant core member of the human oral microbiome (Brennan and Garrett, 2019; Chen and Jiang, 2014; Moore and Moore, 1994). In the oral cavity, *F. nucleatum* is part of a multi-species biofilm, termed the dental plaque. Therein, *F. nucleatum* plays an integral role in forming and maintaining this multi-species community as it serves as a bridging bacterium between the early colonizers (e.g. *Streptococcus* subsps, *Veilonella* subsp. or *Actinomyces* subsp.) and the late colonizers (e.g. *Porphyromonas gingivalis* or *Treponema denticola*) of the dental plaque (Kolenbrander et al., 2002).

This multitude of interactions is brought on by various adhesins expressed on the surface of *F. nucleatum* and serves purposes beyond attachment to the neighboring microbes (see 1.1.1.2 and 1.1.2). For example, the *F. nucleatum*-dependent biofilm formation facilitates cross-feeding and metabolic synergies between the different members of the community (Welch et al., 2016; Kolenbrander et al., 2010; Sakanaka et al., 2022). Further, this multi-species structure allows the generation of an anoxic or micro-aerobic environment supporting survival of *F. nucleatum* itself and other oxygen-sensitive bacteria (Diaz et al., 2002; Kolenbrander et al., 2010).

1.1.1.1 Role in disease

While *F. nucleatum* is generally considered a mutualist, the anaerobic bacterium is also involved in periodontal diseases including endodontic infections, gingivitis and periodontitis (Griffen et al., 2012; Moore and Moore, 1994; Didilescu et al., 2012). These inflammatory diseases can be caused by *F. nucleatum* alone but are more pronounced in multi-microbial infections with other oral species (e.g. *Tannerella forsythia*, *P. gingivalis* or different *Streptococci*) (Settem et al., 2012; Polak et al., 2009; Kesavalu et al., 2007; Han, 2015).

F. nucleatum can also disseminate from the oral cavity reaching distal body sites. For instance, *F. nucleatum* can colonize placental or fetal tissue causing a number of adverse pregnancy outcomes (APO) such as preterm- or still-birth, neonatal sepsis, chorioamnionitis or preclampsia (Vander Haar et al., 2018; Fardini et al., 2010).

This anaerobe bacterium can further be found in a variety of other infections or abscesses throughout the body including: the head and neck area, brain, lungs, abdomen,pelvis, bones,joints and blood (Brook, 1994; Han, 2015). Also caused through dissemination, *F. nucleatum* is a common cause of the inflammatory disease Lemierre syndrome second only to the related species *F. necrophorum* (Williams et al., 2003; Lee et al., 2020).

Furthermore, *F. nucleatum* is also able to colonize the gastrointestinal (GI) tract where it is associated with inflammatory bowl disease and appendicitis (Tahara et al., 2015; Strauss et al., 2011; Swidsinski et al., 2011). Importantly, two landmark publication in 2012 showed that *F. nucleatum* can also colonize the

tissue of colorectal cancer (CRC) (Kostic et al., 2012; Castellarin et al., 2012) adding another major medical problem to the list of diseases associated with this bacterium.

1.1.1.2 Role in cancer

Since 2012, several studies (McCoy, 2013; Feng, 2015; Li, 2016; Mima et al., 2016; Wirbel et al., 2019; Amitay et al., 2017; Nejman et al., 2020; Brennan and Garrett, 2019) could validate the initial observations and prove a clear association of *F. nucleatum* with CRC. *F. nucleatum* has been found in the cancer tissue ranging from the early to late stages of CRC (Ito et al., 2015). Additionally, an increased *F. nucleatum*-burden correlates with overall decreased patient survival (Mima et al., 2016; Yu et al., 2017). This negative outcome is likely caused by a myriad of factors as the colonization by *F. nucleatum* can lead to increased tumor growth, metastasis formation, resistance towards chemotherapeutic agents or mutations through micro-satellite instability (Brennan and Garrett, 2019; Bullman et al., 2017; Kostic et al., 2013; Casasanta et al., 2020; Yu et al., 2017; Mima et al., 2016).

Strikingly, *F. nucleatum* is not restricted in its cancer tropism to CRC as it can also colonize breast, pancreatic and esophageal cancer (Nejman et al., 2020; Parhi et al., 2020; Yamamura et al., 2016; Li et al., 2021; Mitsuhashi et al., 2015; Gaiser et al., 2019). Similar as in CRC, the current evidence strongly indicates that the presence of *F. nucleatum* in these cancers exacerbates the disease (Parhi et al., 2020; Mitsuhashi et al., 2015; Bronzato et al., 2020).

The interaction of *F. nucleatum* with CRC has garnered a lot of attention, but we have uncovered only a few molecular determinants involved in this process such as the OMP Fap2. Functioning as lectin, Fap2 binds to the sugar-moiety Gal-GalNAc which is highly expressed on CRC cells and through this plays a crucial role in the colonization of the cancerous tissue (Gur et al., 2015; Abed et al., 2016). Fap2 can also interact with the immune receptor TIGIT expressed on natural killer cells and T-cells to inhibit their cytotoxic activity (Gur et al., 2015). This immune suppressive effect is further enhanced by binding of the fusobacterial autotransporter CbpF to the T-cell inhibitory receptor CEACAM1 (Galaski et al., 2021; Gur et al., 2019). In contrast, the lipopolysaccharide (LPS) or abundant porin FomA of *F. nucleatum* triggers a pro-inflammatory response via Toll-like-receptors (TLRs) expressed on immune and epithelial cells (Toussi et al., 2012;

Yang et al., 2017).

An additional important finding was the identification of the autotransporter FadA which binds to E-cadherin on cancer cells, thereby leading to the activation of the β-catenin and Wnt-signaling cascade. Consequently, FadA causes the up-regulation of pro-oncogenic and pro-inflammatory genes (Rubinstein et al., 2013).

More recent work demonstrated that *F. nucleatum* also modulates the metabolic milieu of the tumor niche. When co-cultured with a CRC model, *F. nucleatum* displayed increased formate production and lead to an increased glutamine metabolism of the cancer cells. Further, this metabolic shift led to an increased metastasis formation and promoted the expansion of pro-inflammatory Th-17 cells (Ternes et al., 2022).

1.1.2 State of the art for the fusobacterial molecular biology

Even though the medical relevance of *F. nucleatum* has been thoroughly established, we lack a general understanding of its fundamental molecular biology (Brennan and Garrett, 2019). One reason for this is the phylogenetic distance to other model bacteria which hinders direct knowledge transfer (Coleman et al., 2021). Furthermore, poor genetic tractability and a lack of tools for this bacterium represent additional hurdles in the field. This is despite the fact that already in the early 2000 two native fusobacterial plasmids were isolated (Haake et al., 2000; Bachrach et al., 2004a) which could have served as a basis for replicative vectors in *F. nucleatum*. This was followed by the first targeted mutagenesis in *F. nucleatum* whereby suicide vectors were used to generate disruptions in genes encoding exoribonuclease RNase R, the autotransporter Aim1 and the adhesin FadA (Kinder Haake et al., 2006; Han et al., 2005; Kaplan et al., 2005). Yet, it took additional ten years before the first forward-genetic screen was carried out in this bacterium, that used a transposon library and through it identified Fap2 as a galactose-inhibitable adhesin (Coppenhagen-Glazer et al., 2015).

Despite these limitations, the function of several fusobacterial OMPs could be partially elucidated. For example, it was shown that the adhesins Fap2 and FadA play an important role in the interaction with cancer cells (see Chapter 1.1.1.2). However, FadA is not limited to binding the host receptor E-cadherin, as this OMP also forms amyloid-like filaments which plays a role in biofilm formation, adhesion to cancer cells and accelerate periodontal bone loss in an *in vivo* mouse model (Meng et al., 2021).

The surface proteins of *F. nucleatum* also play an important role in the interactions with other microorganism in the dental plaque. For instance, the type 5 autotransporter RadD is an important adhesin that interact with members of the oral microbiome including, *Streptococcus mutants* and *Candida albicans* (Guo et al., 2017; Wu et al., 2015). Like Fap2, RadD can interact with immune cells and cause their death (Kaplan et al., 2010). Less well understood are the autotransporter proteins Aim1, FplA and Fusolisin. In comparison to the examples above, Fusolisin is an autotransporter but also harbors enzymatic function as a serine protease (Doron et al., 2014; Bachrach et al., 2004b). While Fusolisin is able to digest collagen *in vitro* its biological function in *F. nucleatum* is yet to be determined (Doron et al., 2014).

Similarly, FplA displays a phospholipase activity recognizing human phosphoinositides *in vitro* (Casasanta et al., 2017). Even though the role of FplA has not been investigated in context of the host-bacteria interaction, the phospholipase activity indicates its involvement in the manipulation of host cells (Casasanta et al., 2017; Trunk et al., 2019). Our knowledge on Aim1 is even more limited, since it has only been suggested to cause apoptosis in lymphocytes (Kaplan et al., 2005).

Two recent advancements developed a deletion system to yield markerless gene deletions (Wu et al., 2018; Casasanta et al., 2020). Both systems rely on a metabolic deletion background missing the *galK* gene involved in galactose utilization which enables counter-selection using 2-deoxy-galactose otherwise toxic to the wild-type (WT) *F. nucleatum*. This approach helped to shed light on the function of CbpF showing that this autotransporter can directly activate receptor CEACAM1 and through it modulate immune cell activity and potentially cancer growth (Galaski et al., 2021; Gur et al., 2019).

The deletion system was further used to identify a role of the FtsX-EnvC complex in biofilm formation and cell division (Wu et al., 2018) as found for *E. coli*

(Cook et al., 2020; Schmidt et al., 2004). And just recently, Chen et al. (2022) found the L-methionine γ-lyase to contribute to the fusobacterial H_2S production and virulence.

Despite our growing understanding of fusobacterial genes and their function, knowledge about general transcriptome architecture and RNA output in *F. nucleatum* is limited. Exemplifying this is the fact that only a single TSS is known in *Fusobacterium* (Sasaki-Imamura et al., 2010). Apart from a pioneering microarray-based study (Merritt et al., 2009), global transcriptomic approaches have only recently been applied to this bacterium (Mutha et al., 2018; Cochrane et al., 2020; Wu et al., 2021; Scheible et al., 2022). These studies also used automatically annotated genome sequences missing UTRs and non-coding RNAs, including sRNAs. As a result, little is known about transcriptional and post-transcriptional regulation in *F. nucleatum*. Consequently, we lack an understanding of general regulatory networks governing basic processes in *F. nucleatum* but also the ones important for the adaption to changing environments.

1.2 Gene regulation in bacteria

Bacteria must monitor their environment and respond to changes in it for their survival. Such environmental stimuli include fluctuations in nutrients, metals, metabolites or envelope stress. To react appropriately, prokaryotes have developed several mechanisms for controlling transcription and translation of specific regulatory networks. As outlined below, such mechanisms can include a single protein but also different protein complexes and regulatory RNAs working in synergy.

1.2.1 Transcriptional regulation

The RNA polymerase (RNAP) is the key complex for RNA synthesis and consist of five conserved subunits that facilitate the transcriptional activity of the enzyme (Borukhov and Nudler, 2008). However, only the binding of an additional sigma factor (σ) to the RNAP allows the binding of a specific promoter sequence (Travers and Burgess, 1969; Feklístov et al., 2014; Gross et al., 1998). Universally conserved among bacteria, RpoD, or σ^{70}, represents the housekeeping σ factor important for maintaining general transcription (Lonetto et al., 1992). However, during stress conditions alternative σ factors can replace σ^{70} in the RNAP complex

and thus rewire its specificity. These alternate σ factors are part of either the σ^{70}-family or that of the σ^{54}-family which share no sequence homology between each other. The two σ factor classes are further differing in sequence motif recognition, whereby σ^{70} members bind to -35 and -10 promoter region and σ^{54} does so in the -24 and -12 position (Feklístov et al., 2014).

1.2.2 Alternative σ factors

The σ^{54} class is currently limited to RpoN and commonly regulates nitrogen metabolism but can also modulate the expression of virulence genes (Kazmierczak et al., 2005). Beyond RpoD, alternate σ^{70} family members are more diverse in their biological role: σ^{S} functions as a general stress σ factor (Gottesman, 2019), σ^{H} is involved in the response to heat-shock (Arsène et al., 2000), and members of the extracytoplasmic function sigma factors (ECF) such as σ^{E} are frequently associated with envelope stress (Hews et al., 2019). Additional functions of alternate σ factors include regulation of the different stages of spore formation (Fimlaid and Shen, 2015), toxin production (Martin-Verstraete et al., 2016) or flagellar synthesis (Soutourina and Bertin, 2003) However, they can also be involved in very specific processes, such as σ^{V} which responds only lysozyme exposure in Firmicutes (Ho and Ellermeier, 2019). Common among the σ factors is their function as a transcriptional activator in contrast to the varying role of transcription factors and two-component systems.

1.2.3 Transcription factors

Transcription factors (TFs) represent another important component in regulatory networks as they orchestrate the access of the RNAP to promoter sites. For this the TFs bind and recognize specific sequence motifs in the promoter region and either exclude or stabilize RNAP association with the DNA, thereby causing the inhibition or activation of gene expression, respectively (Browning and Busby, 2016). The activity of TFs is regulated by sensing internal or external cues that can either lead to binding of its DNA recognition motif or release thereof (Babu and Teichmann, 2003b; Balleza et al., 2009).

Such is the case for the iron-response regulator Fur that inhibits transcription by binding the operator sites upstream of genes encoding components of the iron acquisition systems and is inactivated with decreasing iron levels. This leads to the release of Fur allowing the transcription of previously repressed genes (Braun,

2005; Escolar et al., 1999; Hantke, 1984). Bacteria such as *E. coli* can harbor ∼300 TFs (Pérez-Rueda and Collado-Vides, 2000; Babu and Teichmann, 2003a). The TFs can exert gene expression control of a larger regulon such as the above-mentioned Fur or the catabolite repressor protein (CRP) which regulates ∼500 genes in *E. coli* (Santos-Zavaleta et al., 2019). Combined, the regulatory of activity of different TFs can overlap or interfere with one another which results in complex gene regulation networks (Browning et al., 2019).

F. nucleatum harbors ∼65 predicted TFs which are fully unexplored. Interestingly, the bacterium encodes two putative TFs related to the Fur-family suggesting that regulation of the metal-uptake might be similarly regulated as in *E. coli*.

1.2.4 Two-component systems

In the case of two-component systems (TCS) the function of receiving and relaying a signal is divided into two factors: a membrane bound histidine kinase as the sensory protein and a cognate response regulator (Mascher et al., 2006). Under the activating conditions, the sensory component phosphorylates the response regulator leading to its activation (Zschiedrich et al., 2016). Similar to TFs, the response regulator controls transcription of its target gene(s) by binding a DNA recognition motifs (Makino et al., 1986; Zschiedrich et al., 2016).

TCSs sense a variety of stimuli, such as levels of Mg^{2+} and Ca^{2+} but also envelope damage by the wide-spread PhoPQ TCS (García Véscovi et al., 1996; Groisman, 2001; Dalebroux and Miller, 2014; Bader et al., 2005) or inner membrane (IM) stress by the Cpx-system (Hunke et al., 2012). The Bae- and Rcs-systems present additional TCS (Raffa and Raivio, 2002; Meng et al., 2021) responding to membrane damage highlighting a common function of TCS in the envelope stress response (Rowley et al., 2006).

Interestingly the two recently characterized TCS in *F. nucleatum* have not been linked to envelope maintenance: the CarRS system is regulating lysine metabolism as well as the adhesin RadD. Apart from biofilm formation and co-aggregation, the CarRS system is also important for virulence as a $\Delta carS$ strain is attenuated in a pre-term birth mouse model (Wu et al., 2021).

The second TCS, termed ModRS, is involved in the response to oxidative stress (Scheible et al., 2022). Upon sensing H_2O_2, ModRS activates oxidative defense genes (*msrAB*, *ccdA*, *trx*) to neutralize the reactive oxygen species. In line with this is the observation that ModRS and its regulon are required for survival in

macrophages as well as successful colonization of CRC cell lines and placental tissue in a pre-term birth mouse model (Scheible et al., 2022). The latter might be linked to the role of MetL in virulence since this metabolic gene is modulated by ModRS and is required for bacterial fitness in a pre-term mouse model (Chen et al., 2022).

1.2.4.1 Regulation by the 6S RNA

Regulatory networks do not solely rely on protein factors. The 6S RNA, encoded by the *ssrS* gene, is a non-coding RNA (ncRNA) found widespread throughout the bacterial kingdom (Wehner et al., 2014). It was discovered more than 50 years ago in *E. coli* (Hindley, 1967) where its function is also best understood (Wassarman, 2018). Remarkably, its secondary structure with a central bulge inside a long stem loop is conserved among all identified 6S RNAs despite a lack of primary sequence conservation (Wassarman, 2018). Work in *E. coli* could show that this structural feature is essential for the function of the ncRNA as it mimics an open dsDNA complex and sequesters the RNAP into a stable complex. The RNA-RNAP complex formation coincides with increased levels of the 6S RNA towards the stationary growth phase and causes global gene expression changes (Wassarman and Storz, 2000). Upon outgrowth, the RNAP uses the 6S RNA as a template leading to the synthesis of a product RNA (pRNA) and destabilization of the RNA-RNAP complex (Wassarman and Saecker, 2006). This process becomes important when subjected to adverse conditions, such as during the infection of the host, where the 6S RNA is upregulated in various pathogens (Wassarman, 2018). Additionally, the 6S RNA can be integrated into larger global pathways such as the stringent response (Cavanagh et al., 2010). Despite the early discovery of the 6S RNA we lack a full understanding of the transcriptional regulation by this ncRNA (Wassarman, 2018).

1.2.5 Post-transcriptional regulation by small regulatory RNAs

An additional important component or regulatory system are small regulatory RNAs (sRNAs) which in contrast to above mentioned examples act as post-transcriptional regulators. The sRNA-based regulation is commonly carried out on the RNA level and usually leading to modulation of protein synthesis. RNA-binding proteins (RBPs) such as Hfq, ProQ or CsrA often play an important role

in the sRNA-activities by affecting their stability and supporting the annealing of the sRNA to an mRNA target (Holmqvist and Vogel, 2018; Storz et al., 2011).

Initially discovered sRNAs originated from intergenic regions but by now it is clear that they can stem from both the 5' and 3'UTR of an mRNA or from inside an ORF (Adams et al., 2021). sRNAs are defined as short 50 to 500 nt transcripts (Gottesman and Storz, 2011) which do not encode for proteins. However, both traditional definitions have been challenged by examples such as the ∼1 kB long sRNA SSR42 in *Staphylococcus aureus* (Morrison et al., 2012) and multiple dual-function sRNAs encoding small proteins such as RNAIII or SpoT42 (Morfeldt et al., 1995; Aoyama et al., 2022; Adams et al., 2021; Gimpel and Brantl, 2017).

Canonically, sRNAs act as translational repressors by directly base-pairing with their mRNA target through the seed region. During this, the sRNAs usually bind in the vicinity of the ribosome-binding site (RBS) and/or the start codon of an mRNA and thus inhibit translation initiation (Storz et al., 2011). The reduced level of translation often promotes increased RNA degradation while this is not strictly necessary (Møller et al., 2002; Storz et al., 2011). Alternatively, sRNAs can also promote translation by changing the secondary structure of its RNA target and thus unmasking the RBS (Prévost et al., 2007) or increasing mRNA stability (Fröhlich et al., 2013). Another less frequent mode-of-action for sRNAs is the sponging of other sRNAs as shown for the SroC sequestering the sRNAs GcvB and MgrR (Miyakoshi et al., 2015; Acuña et al., 2016).

Through the above mentioned mechanisms sRNAs function as the fine-tuning components in regulatory networks. Their importance is reflected in the fact that sRNAs are involved in every facet of the bacterial physiology ranging from response to metal or nutrient limitation to virulence and biofilm formation (Nitzan et al., 2017; Ponath et al., 2022a; Hör et al., 2020b). Thus, insights into these important players have endowed us with a deeper understanding of the larger biology of bacteria.

1.3 The envelope stress response

Regulation of the cellular envelope is vital to bacteria as it forms the first line of defense in Gram-negative bacteria (Rowley et al., 2006). As mentioned

above, its integrity is monitored by several different stress responses (Mitchell and Silhavy, 2019), such as the Cpx-system which modifies the IM when facing envelope stress (Raivio, 2014) (see 1.2). Additionally, the phage shock (PSP) system senses perturbations at the IM and supports the maintenance of the proton motive force and ensuring proper localization of secretins (Flores-Kim and Darwin, 2016).

Yet, probably the best understood envelope stress response concerns the integrity of the outer membrane (OM), regulated by the ECF σ^E.

1.3.1 The σ^E stress response

The σ^E response has been extensively characterized since its discovery more than 30 years ago (Erickson and Gross, 1989), especially in *E. coli* but has also been studied in members of the Proteobacteria and Actinobacteriota (Paget et al., 1999; Hews et al., 2019). Encoded in *E. coli* by *rpoE*, σ^E activity is tightly regulated to avoid unwanted toxicity (Nicoloff et al., 2017). For this, an anti-sigma factor binds σ^E and inhibits it from associating with the RNAP (De Las Peñas et al., 1997; Missiakas et al., 1997). In *E. coli*, σ^E is released by regulated intramembrane proteolysis (RIP) from its IM-bound anti-sigma factor RseA. This is the result of sequential cleavage steps initiated by the protease DegS and followed by the proteolytic activity of RseP (Ades et al., 1999; Alba et al., 2002; Kanehara et al., 2002). In the final step, the ClpXP protease is recruited via the adapter protein SspB to yield active σ^E (Flynn et al., 2004).

Across different phyla, σ^E and its homologs can be activated by various broad stresses such as hyperosmolarity, heat shock, oxidative stress, nutrient shifts or OMP overexpression (Rowley et al., 2006). Conversely, the ECF can also be attuned to a specific stressor as exemplified in *Rhodobacter sphaeroides* where the ECF responds to singlet oxygen generated during photosynthesis (Anthony et al., 2005).

Upon activation, σ^E associates with the RNAP and initiates transcription of its regulon by recognizing a conserved promoter motif in the -10 and -35 region of the transcriptional start site (TSS) (Rhodius et al., 2006; Rhodius and Mutalik, 2010; Todor et al., 2020). This regulon shares functional similarities between different bacterial species by including genes involved in DNA damage repair, LPS biogenesis and OM homeostasis (Dartigalongue et al., 2001; Missiakas et al., 1997; Rhodius et al., 2006; Skovierova et al., 2006; Rowley et al., 2006; Ades,

2008; Mutalik et al., 2009). Strikingly, previous works also identified sRNAs as a conserved components of the σ^E regulon which are considered the "non-coding arm" of the response (Fröhlich and Gottesman, 2018).

1.3.1.1 σ^E and small RNAs

In *E. coli* and in *Salmonella typhimurium*, σ^E upregulates the expression of the three Hfq-dependent sRNAs: MicA (Udekwu et al., 2005; Figueroa-Bossi et al., 2006), RybB (Papenfort et al., 2006; Thompson et al., 2007; Papenfort et al., 2010) and MicL (Guo et al., 2014). In this, MicL targets only the mRNA of the lipo-protein *lpp* to downregulate the most abundant protein in these bacteria (Guo et al., 2014). In contrast to the former, MicA and RybB affect a large targetome to inhibit translation of several major OMPs, such as OmpA (Gogol et al., 2011; Vogel and Papenfort, 2006; Johansen et al., 2006; Udekwu et al., 2005; Rasmussen et al., 2005). Taken together, the three sRNAs act as σ^E negative regulatory arm to achieve envelope homeostasis when facing adverse conditions (Fröhlich and Gottesman, 2018). Furthermore, σ^E-dependent sRNAs with a similar function to the above are found in other Proteobacteria, which underpins their important regulatory function (Table 1.1)

Table 1.1 | Overview of studied σ^E-dependent sRNAs.

sRNA	species	reference
MicA, RybB, MicL	*E.coli/ Salmonella*	(Udekwu et al., 2005; Johansen et al., 2006; Papenfort et al., 2006; Pfeiffer et al., 2007; Thompson et al., 2007; Papenfort et al., 2010; Guo et al., 2014)
VrrA, MicV	*V. cholerae*	(Song et al., 2008; Peschek et al., 2019)
ErsA	*P. aeruginosa*	(Ferrara et al., 2015)
Pos19	*R. sphaeroides*	(Berghoff et al., 2009; Müller et al., 2016)

The importance of these sRNAs in maintaining membrane homeostasis is supported by the fact that an *hfq* deletion strain exhibits aberrant activation of the σ^E response (Ding et al., 2004; Figueroa-Bossi et al., 2006; Guisbert et al., 2007; Thompson et al., 2007). The lack of the sRNA-mediated control of OMP synthesis and subsequent accumulation thereof is thought to trigger the unchecked activation of σ^E (Guisbert et al., 2007; Thompson et al., 2007; Bossi et al., 2008).

1.4 Aim of this study

Recent works on the fusobacterial TCS provided us with a first view on gene regulation in *F. nucleatum*. Yet, their exact regulon or other larger regulatory networks remain unknown in this bacterium. Moreover, there is an absence of any knowledge about its post-transcriptional regulation which plays a major role in all facets of the bacterial lifestyle including the adaption to changing environments. Since *F. nucleatum* drastically changes its niches in the body - from tooth to cancer - knowledge on the regulatory networks supporting this adaption could be of great importance. Different environments that *F. nucleatum* faces raise the question if this phylogenetically early-branching bacterium employs similar regulatory networks as bacteria such as *E. coli* to maintain its envelope homeostasis. However, answering such questions has been hampered by the absence of genetic tools for *F. nucleatum* and a detailed genome annotation that would allow one to fully investigate gene regulation in this bacterium.

Thus, this doctoral thesis focused on three points to build an important foundation for further investigation into the molecular biology of *F. nucleatum*:

- Using RNA-seq to generate updated annotations including 5'UTRs and regulatory RNA elements
- Generating genetic tools that allow studying gene functions or the impact of regulatory events in the bacterium
- Investigating the role of the ECF σ^E as the first global stress response in *F. nucleatum*

2 The primary transcriptome of *Fusobacterium nucleatum*

RNA-seq is a powerful tool and has revolutionized our understanding of bacteria by capturing global gene expression changes that reflect their physiological state (Creecy and Conway, 2015). However, the technique is unable to distinguish primary transcripts from processed transcripts and consequently cannot directly detect transcriptional start sites (TSS).

The pioneering of differential RNA-seq (dRNA-seq) overcame this limitation in 2010 leading to elucidation of the primary transcriptome for the pathogen *Helicobacter pylori* (Sharma et al., 2010). Not only did it enable the precise annotation of transcriptional start sites (TSS) and untranslated regions of genes (UTRs), but it also allowed the global identification of sRNAs.

To achieve this, the dRNA-seq protocol takes advantage of the fact that bacterial primary transcripts harbor a tri-phosphate group (5'PPP) at their 5'end whereas processed RNA species including ribosomal RNA (rRNA) or transfer RNAs (tRNA) have a mono-phosphate group (5'P). Treatment with the 5'P-dependent terminator exonuclease (TEX) can specifically degrade processed RNA and thus enriches for the primary transcripts. A comparative sequencing approach of TEX-treated and untreated samples can thus yield clear enrichment patterns for the TSS of genes on a global scale (Sharma et al., 2010; Sharma and Vogel, 2014).

To similarly advance our understanding on the RNA landscape of *Fusobacterium*, dRNA-seq was applied to the five fusobacterial strains representing all subspecies and a closely related species:

- *Fusobacterium nucleatum* subspecies *nucleatum* ATCC 25586 (*Fnn*)
- *Fusobacterium nucleatum* subspecies *animalis* 7_1(*Fna*)
- *Fusobacterium nucleatum* subspecies *polymorphum* ATCC 10953 (*Fnp*)
- *Fusobacterium nucleatum* subspecies *vincentii* 3_1_36A2 (*Fnv*)
- *Fusobacterium periodonticum* 2_1_31 (*FuP*)

In this, the detailed analysis focused on *Fnn* as it represents the reference strain of *F. nucleatum*.

A large part of this chapter has been published before:

- Ponath, F., Tawk, C., Zhu, Y., Barquist, L., Faber, F., & Vogel, J. (2021). RNA landscape of the emerging cancer-associated microbe *Fusobacterium nucleatum*. *Nature Microbiology*, 6(8), 1007–1020. https://doi.org/10.1038/s41564-021-00927-7

2.1 Characterization of *F. nucleatum* growth

Different growth phases of a bacterium entail strong differences in their gene expression patterns. Thus the growth behavior of the used fusobacterial strains was determined as a first step. Due to the bacteria's anaerobic lifestyle all experiments were performed in an anaerobic environment, using rich media (see Chapter 7 for details).

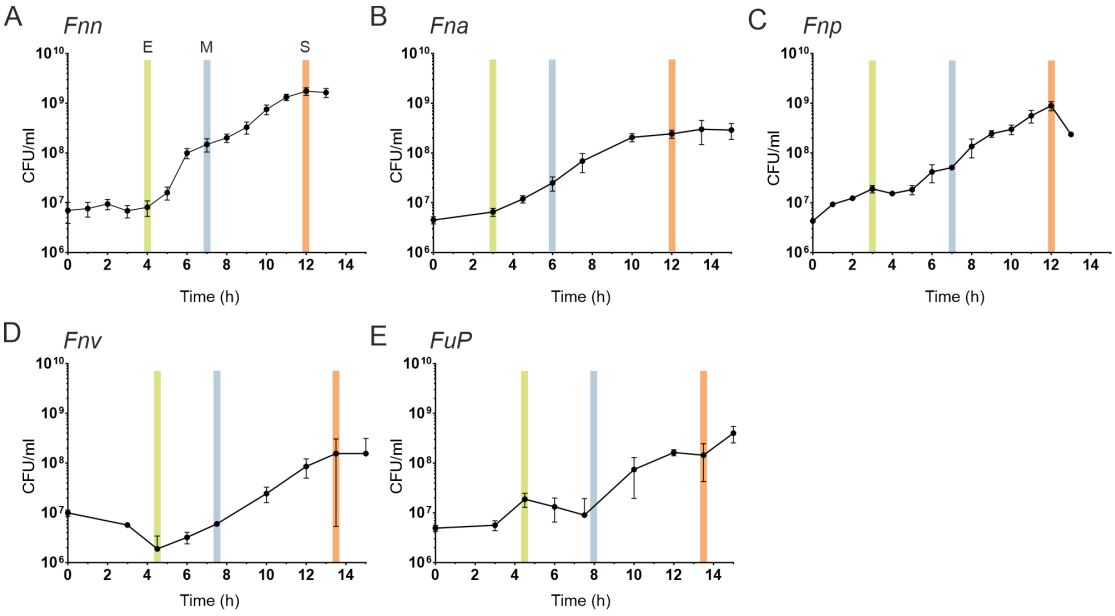

Figure 2.1 | CFU growth of different fusobacterial strains
Colony forming units (CFU) were quantified at the indicated time points for the strain **A.** *Fnn*; **B.** *Fna*; **C.** *Fnp*; **D.** *Fnv*; **E.** *FuP*. The time points analyzed in the RNA-seq analysis are marked as follows: early exponential (E; green), mid-exponential (M; blue) and early stationary phase (S; orange). The data are presented as the mean and standard deviation (SD) of three biological replicates. Panel A of this figure has been adapted from Ponath et al. (2021).

Bacterial growth was assessed by quantifying the colony forming units (CFU) over the time course of 12 to 15 hours (h). *Fnn* reached the early exponential phase

(E) after ∼4 h, the mid-exponential (M) phase at ∼7 h and the early stationary phase (S) at ∼12 h (Fig. 2.1A). The other strains followed an overall similar growth pattern (Fig. 2.1B-E).

2.2 dRNA-seq of *Fusobacterium nucleatum*

To capture the diverse transcriptomic states, total RNA samples were collected from three time points reflecting the different growth stages (Fig 2.1). These samples were subjected to the dRNA-seq cDNA preparation library protocol and sequenced (see Chapter 7 for details). As mentioned above, the in-depth analysis focused on the reference strain (*Fnn*) and the following results will refer to this strain unless otherwise stated (see Chapter 2.4).

2.2.1 Transcriptional start sites

The resulting reads were mapped to the genome of *Fnn* and the resulting coverage files were used as input for the bioinformatic tool ANNOgesic (Yu et al., 2018b) (see Chapter 7). This modular tool enables the prediction of TSS by accounting for the differences at the TSS between coverage files generated from untreated and TEX-treated RNA samples (Fig. 2.2A). For this, ANNOgesic employs a machine-learning algorithm taking following variables into account: height and height reduction, factor and factor reduction, an enrichment factor, a processing factor and finally the base height (Yu et al., 2018b).

The detected TSS are assigned to different classes (Fig. 2.2A): primary TSS (pTSS) are the main starting point of a transcript; secondary TSS (sTSS) are additional starting points of a transcript but expressed weaker than a pTSS; internal TSS (iTSS) start within the ORF of a gene; antisense TSS (aTSS) signal transcription antisense to a gene or within a 100 nt distance of one; orphan TSS (oTSS) are not associated with any annotated gene.

This computational analysis followed by manual curation revealed an total number of 930 TSS detected across all growth stages for *Fnn* (Fig. 2.2B; Supplementary data 8.1). The majority of the TSS (706/930) are pTSS which, combined with the operon prediction(see 2.2.7), account for nearly ∼75% of all genes.

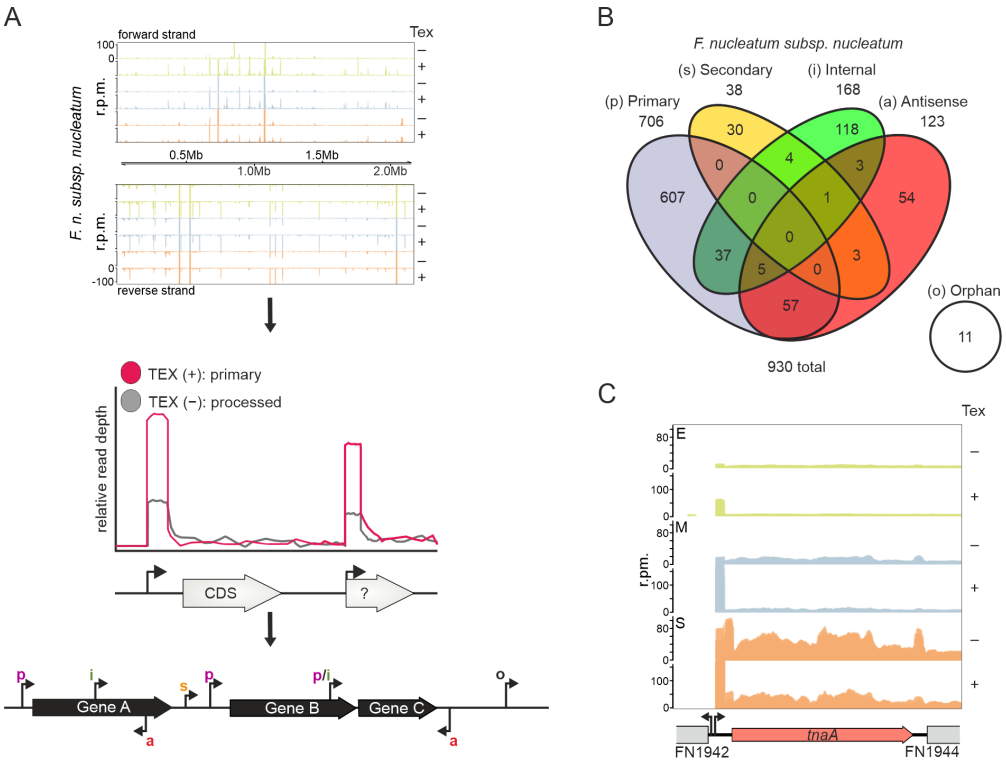

Figure 2.2 | dRNA-seq analysis of *F. nucleatum* subsp. *nucleatum*.
A. Overview of the workflow for the transcriptome analysis via dRNA-seq. Genome-wide read distribution for *Fnn* is shown, followed by schematic representation of the read enrichment upon TEX-treatment and the classification of the different TSS classes. **B.** Venndiagram showing the distribution of the different TSS classes detected for *Fnn*. **C.** Verification of TSS detection for the *tnaAB* operon as shown in Sasaki-Imamura et al. (2010). r.p.m., reads per million. The figure was modified from Ponath et al. (2021).

This included a prediction for the major TSS of the *tnaAB* tryptophanase operon which is in agreement with the previously reported pTSS for this operon (Fig. 2.2C) (Sasaki-Imamura et al., 2010). However, this TSS represents the only example for benchmarking the global analysis as no other TSS has been previously reported for *F. nucleatum*.

The dRNA-seq data also showed growth phase-dependent detection of TSS as in the case of *tnaAB* operon for which the expression increased in the stationary phase (Fig. 2.2C). Transcription of a fructose uptake operon, FN1438-FN1441, displayed the opposite pattern with diminishing RNA levels in the stationary phase (Fig. 2.3A). Overall, 171 genes were differentially expressed (-2$\leq$ $\log_2$ fold change $\geq$2; false discovery rate (FDR) $\leq$0.05) when comparing the expression for the untreated samples between the early exponential and stationary phase (Supplementary data 8.2). Among them, the fructose uptake operon showed the strongest downregulated in the stationary phase while the genes of a putative fructoselysine

metabolism operon (FN0632-FN0627) displayed drastically increased expression levels indicating a metabolic re-wiring with the entry into the stationary phase.

Approximately 12% of the overall TSS belong to multiple of the above introduced TSS classes. The largest group of the multi-class TSS is assigned to the class consisting of primary and antisense TSS (p/aTSS) with a total number of 57 out of 107. The p/aTSS often represent divergently transcribed genes as found for the genes FN1418 and *metL*, encoding a transcriptional regulator and a methionine gamma-lyase, respectively (Fig. 2.3). The short intergenic regions between the genes indicates that both share an overlapping promoter motif and the high constitutive expression of *metL* might inadvertently repress transcription of FN1418 (Warman et al., 2021).

The second largest group of multi-class TSS is represented by primary and internal TSS (p/iTSS) (37/107). These TSS mark the start of suboperons which will be discussed further below in chapter 2.2.7.

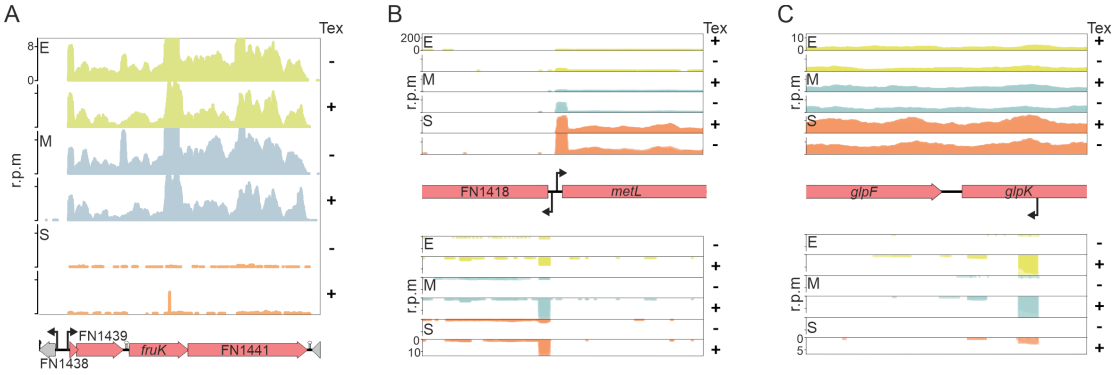

Figure 2.3 | Examples for predicted regulation via TSS prediction.
A. Growth-phase-dependent transcription of the FN1438-FN1441 operon. Normalized read distribution for the FN1438-FN1441 operon including the *fruK* gene likely involved in fructose uptake. **B.** Coverage of the divergently transcribed genes FN1418 and *metL*. **C.** Coverage of the *glpkF-glpK* operon including the detected aTSS in *trans* to *glpK*. Panel **A** of this figure has been adapted from Ponath et al. (2021).

Pervasive antisense transcription, originating from aTSS, has been reported as widespread occurrence in bacteria and can contribute to transcriptional regulation (Georg and Hess, 2018). However, aTSS were only a minor group of detected TSS (54/930) and primarily detected in the E and M growth phase (Supplementary data 8.1). In case of the glycerol kinase *glpK* (FN1839), the antisense transcription originated close to the start codon of the gene. Interestingly, the levels of the glpK mRNA increased during stationary phase when antisense transcription

decreased (Fig. 2.3C). This suggests a possible growth-phase dependent regulation of glycerol metabolism partially controlled by transcriptional interference or a possible post-transcriptional mechanism (Pelechano and Steinmetz, 2013).

2.2.2 5'UTRs

The 5'UTR of the bacterial mRNA contains the Shine-Dalgarno sequence as part of the ribosome binding site (RBS) and thus enables recognition by the ribosome and subsequent translation of the mRNA (Shine and Dalgarno, 1974; Calogero et al., 1988). Additionally, the 5'UTR is also a hot-spot for post-transcriptional regulation in bacteria and thus globally identifying them is vital to fully understand regulatory networks (Geissmann et al., 2009). The annotation of pTSS and sTSS allow the identification of 5'UTRs for *F. nucleatum* which display a median length of ∼36 nt (Fig. 2.4). This distribution of the 5'UTR length is similar to that found in other bacteria such as *H. pylori* or *C. difficile* (Sharma et al., 2010; Fuchs et al., 2021). Conversely, 113 genes also show substantially longer 5'UTRs (>100 nt), as for instance the 324 nt long 5'UTR of the mRNA encoding a putative alanine or glycine:cation symporter FN0328 (Supplementary data 8.3).

A motif analysis of all 5'UTRs showed in 76% of the sequences a shared perfect 5'-AGGAGG-3' motif and over 90 % are associated with this identified sequence (Fig. 2.4), which resembles the Shine-Dalgarno sequence found in other bacteria (Shine and Dalgarno, 1974). This indicates that mRNA recognition by the ribosome is similarly conserved in the evolutionary distant *F. nucleatum*.

Leaderless mRNAs, with a 5'UTR shorter than 10 nt, are prevalent in archaea and some bacterial species. In these organisms, their alternative translation initiation mechanism is suggested to support them in the response to stress (Beck and Moll, 2018). However, *F. nucleatum* shows only a rare occasion of leaderless mRNAs with a 5'UTR length ≤10 nt as the dRNA-seq analysis only identified four such cases: transcripts encoding an integrase/recombinase, a pseudouridine synthetase, a putative transcriptional regulator and a predicted drug efflux pump (Supplementary data 8.3). The low occurrence of leaderless mRNAs (Beck and Moll, 2018; Zheng et al., 2011) suggest that this mRNA species might not play an important role in *F. nucleatum* when facing adverse conditions.

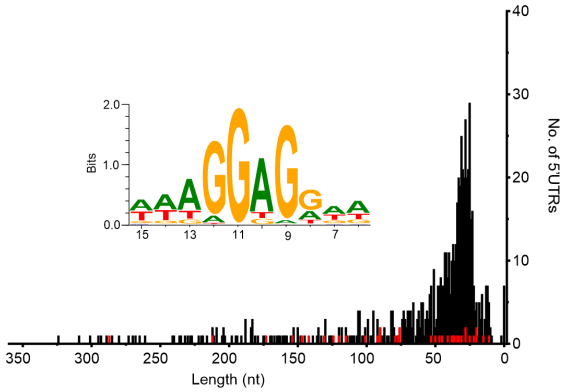

Figure 2.4 | Analysis of 5'UTRs in *F. nucleatum*.
Displayed is the length distribution and counts of all the 5'UTR associated with pTSS (black) and sTSS (red). A consensus motif for the Shine-Dalgarno sequence was predicted using these 5'UTRs. The average distance from the translational start site is indicated. This figure was adapted from Ponath et al. (2021).

2.2.3 Promoter

Bacterial transcription is initiated by DNA binding of the RNA polymerase (RNAP) which forms a complex with a sigma (σ) factor to allow for guided transcription from a specific promoter (Gross et al., 1998). All *F. nucleatum* subspecies encode four σ factors: the housekeeping σ^{70} (*rpoD*); downstream of σ^{70} an unclassified σ factor; a putative homolog to the envelope stress regulator σ^{E} and one sigma factor that shows conservation to the sporulation σ factor SigH from the phylogenetically distant *Clostridiaceae* (Fig. 2.5A). Of note, all subspecies except for *F. nucleatum* subsp. *nucleatum* additionally harbor a member of the σ^{54} family.

In addition, *F. nucleatum* encodes an additional ∼65 putative TFs and five predicted two-component systems (TCS), out of which two have recently been investigated (see Chapter 1.1.2). As no known promoter motifs for *F. nucleatum* were available, the -50 to +1 region of each pTSS was used for a motif analysis using MEME (Bailey et al., 2015). Analysis of this region identified a common extended -10 and -35 box in the upstream region for ∼93% of all pTSS (Fig. 2.5B). This is recognized by the housekeeping σ^{70} as found e.g. in γ-Proteobacteria such *E. coli* (Hawley and Mcclure, 1983; Mitchell et al., 2003). However, both boxes are connected through an A+T-rich spacer more commonly found among members of the ϵ-Proteobacteria (Dugar et al., 2013; Sharma et al., 2010) or Bacteroides (Ryan et al., 2020).

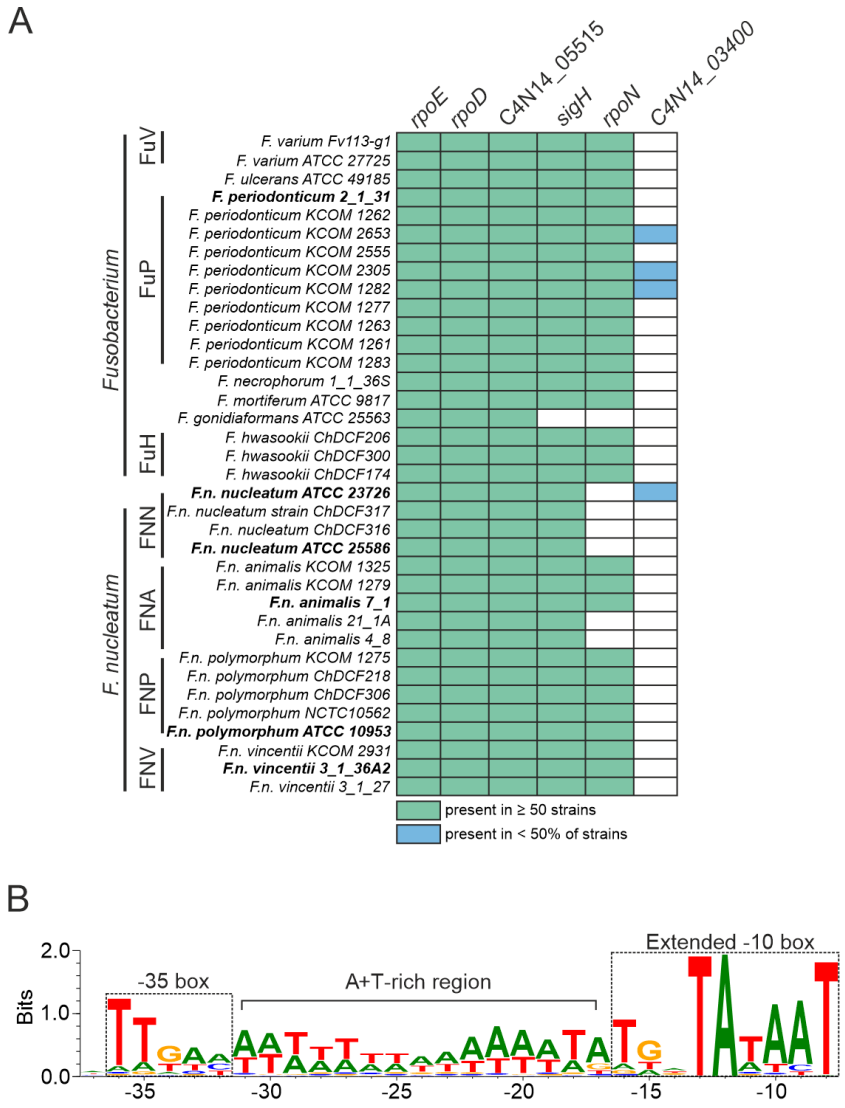

Figure 2.5 | σ factors across *Fusobacterium* and the identified σ^{70} motif.
A. Displayed is the presence (green/blue) or absence (white) of the predicted σ factors based on genome annotations and a homolog search. Strains used in this study are marked in bold. **B.** The identified promoter motif identified via the MEME suite (Bailey et al., 2015) found upstream of ∼93% of all pTSS. The extended -10 box and -35 box are indicated including the A+T-rich spacer separating them. Panel **A** has been modified from Ponath et al. (2022b) and **B** from Ponath et al. (2021).

The non-targeted search did not identify any further promoter motifs indicating that transcription is primarily driven by σ^{70} under the tested growth conditions. This result was unchanged when only analyzing genes transcribed in one or two of the growth stages. This could be partially explained by the observation that only 29 out of the 65 transcription factors were sufficiently expressed and only two showed a strong regulation between early exponential and stationary phase ($-2 \leq \log_2$ fold change (FC) ≥ 2; Supplementary data 8.2). Similarly, none of the three detected TCS were significantly regulated (Supplementary data 8.2).

Overall, these data suggest that σ^{70} is the primary driver of transcription under the studied growth conditions.

2.2.4 Re-annotation

An exact gene annotation is required to accurately perform transcript quantification for RNA-seq studies or when analyzing gene function with molecular tools. High resolution transcriptomic maps, such as those generated by dRNA-seq, are an excellent tool to verify automated ORF annotations for bacterial genomes (Sharma et al., 2010). The manual curation of the TSS displayed several instances of extended ORFs for genes in the annotation for *F. nucleatum* supplied by the NCBI database (accession: AE009951.2; (Kapatral et al., 2002)).

To remedy potential mis-annotations, all CDS were surveyed for a lack of canonical start codon, the lack of an Shine-Dalgarno sequence and the presence of a potential upstream-extended ORF. This led to the identification of 463 CDS exhibiting at least one of these criteria. Using the generated annotation of pTSS, sTSS and the identified Shine-Dalgarno sequence (Fig. 2.4), a total of 460 CDS could be re-annotated and 438/460 out of those were found to carry an 'AUG' start codon and harbor an Shine-Dalgarno sequence in the -20 nt region of the start codon (Supplementary data 8.5).

Further, a comparison to a recent sequence update by Sanders et al. (2018) displays an excellent agreement between both annotations.

2.2.5 Annotation of small ORFs

Small ORFs (sORFs), encode small protein, which are defined as any protein with ≤ 50 amino acids (aa) (Hemm et al., 2020). The sORFs are often overlooked by automated annotation pipelines as these usually consider only ORFs encoding for proteins consisting of ≥ 100 aa (Storz et al., 2014). However, it is becoming more evident that these small proteins can play important roles in different aspects of bacterial physiology (Hemm et al., 2020).

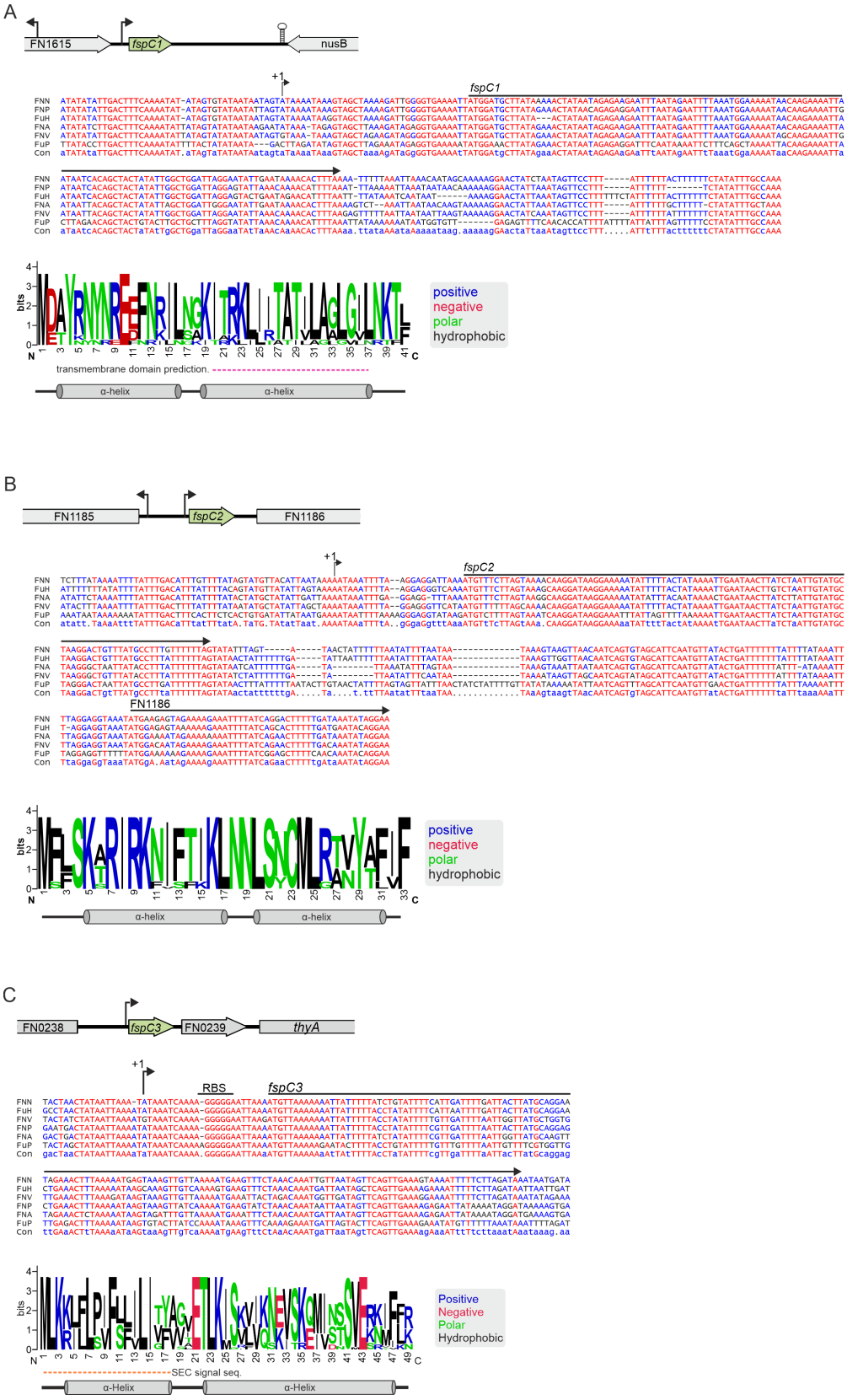

Figure 2.6 | Discovery of sORFs in *F. nucleatum*.
Shown are the genomic locations, sequence alignments and consensus motif of the ORFs for **A.** *fspC1*, **B.** *fspC2* and **C.** *fspC3*. The figure has been modified from (Ponath et al., 2021).

A total of 22 sORFs are annotated in *F. nucleatum*. However, most of them are associated with transposases or are ambiguous candidates as they at lack an AUG start codon or an Shine-Dalgarno sequence. The analysis of long 5'UTRs and orphan TSS identified three high confidence candidates. Now referred to as *fspC1* to *fspC3*, all three sORFs are conserved in their genomic synteny (Fig. 2.6). Additionally, the ORF and several amino acids remain strongly conserved despite varying nucleotide sequences between the different fusobacterial species.

A structure prediction for all three sORF candidates indicates the presence of two α helices and a potential membrane association, which is seems to be a common characteristic of small proteins (Fontaine et al., 2011). In more detail, the 41 aa FspC1 harbors a predicted transmembrane domain and lies downstream of the gene encoding for the anti-terminator NusB (Fig. 2.6A). The gene for the 33 aa long FspC2 is located upstream of the predicted glutamate carboxypeptidase (FN1186) (Fig. 2.6B). Lastly, FspC3 is a $\sim$48 aa ORF and the mRNA is co-transcribed with an operon potentially involved in nucleotide metabolism and transfer RNA (tRNA) maturation or repair function (Fig. 2.6C). Interestingly, FspC3 likely functions in the periplasm as it carries a SEC translocation signal.

The strong conservation in terms of their sequence and genomic synteny across *Fusobacterium* genus, suggest that, FspC1, FspC2 and FspC3 play an important role in the physiology of these bacteria.

2.2.6 Improving the annotation of virulence factors

F. nucleatum lacks all canonical secretion systems for the secretion of effector proteins that can modulate the host (Citron, 2002). Instead, *F. nucleatum* uses a variety of OMPs to interact with the host cells including genes encoding for type 5 autotransporters (Brennan and Garrett, 2019; Kapatral et al., 2002; Desvaux et al., 2005; Umaña et al., 2019). The dRNA-seq analysis revealed that all of these genes are constitutively expressed while also enabling the definition of their operon structure (Fig. 2.7; Supplementary data 8.1).

For example, the β-catenin signaling triggering FadA (Rubinstein et al., 2013) displays a monocistronic expression. In comparison, the gene encoding for the Gal-GalNAc binding Fap2 is transcribed as a dicistron with the first gene encoding for a putative transport protein. The *fadA* paralog, *rapA*, together with *fad-I* and terminally *radD* are part of a multicistronic functional unit important for interspecies co-aggregation (Kaplan et al., 2009; Shokeen et al., 2020). The recently discovered negative regulator of T-cell activity, CbpF, is transcribed as a single

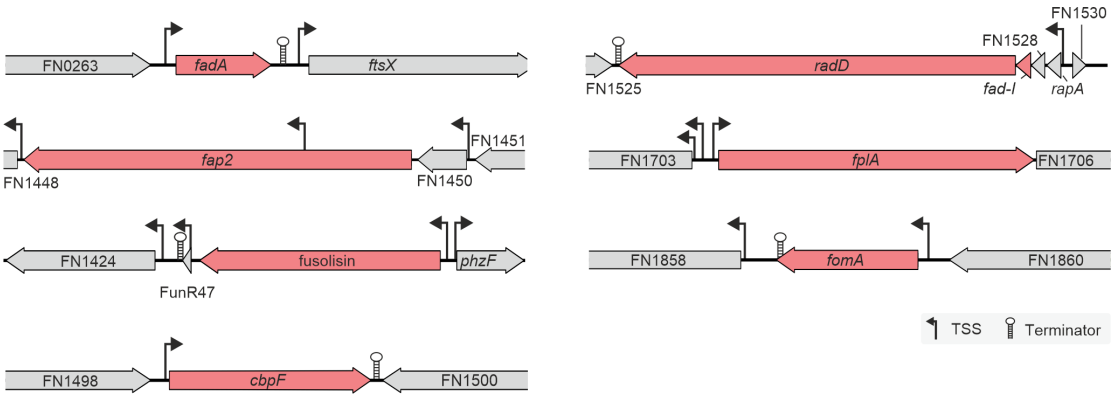

Figure 2.7 | Improving the annotation of virulene factors.
Schematic overview of known fusobacterial virulence factors (red) including associated TSS and terminators. This figure was modified from (Ponath et al., 2021).

gene (Galaski et al., 2021; Brewer et al., 2019). This also the case for the genes of the abundant porin FomA and the phospholipase FplA, whose role in virulence remains to be defined. The serine protease fusolisin (Doron et al., 2014) also shows a monocistronic expression. However, the 3'-region of this gene contains the independently transcribed sRNA FunR47. Additionally, the transcription of 65/208 predicted virulence factors (Kumar et al., 2016) is initiated from a pTSS (Supplementary data 8.4). The improved annotation of fusobacterial virulence factors also defines their 5'UTRs which can serve as a starting point to investigate their post-transcriptional regulation.

2.2.7 Improving the operon annotation

A previous operon prediction from the DOOR database (Cao et al., 2019) for *F. nucleatum* was entirely generated by computational inference from other bacteria. However, this prediction can further be refined by using additional data and predictions, such as TSS, terminators and transcriptomic data. All of these were combined to refine the operon prediction for *F. nucleatum* resulting in a total of 428 operons (Supplementary data 8.6). This analysis further corrected 48 operons, such for the one including for the virulence-associated type Va autotransporter FplA (FN1704; Fig. 2.8A) (Casasanta et al., 2017) prior considered to be monocistronic but is in fact co-transcribed with FN1705 and FN1706, encoding for a putative transporter and an aldose-epimerase, respectively.

The operon annotation was further expanded by eight new operons (Supplementary data 8.6). This includes the operon FN1312-FN1310, in which the first

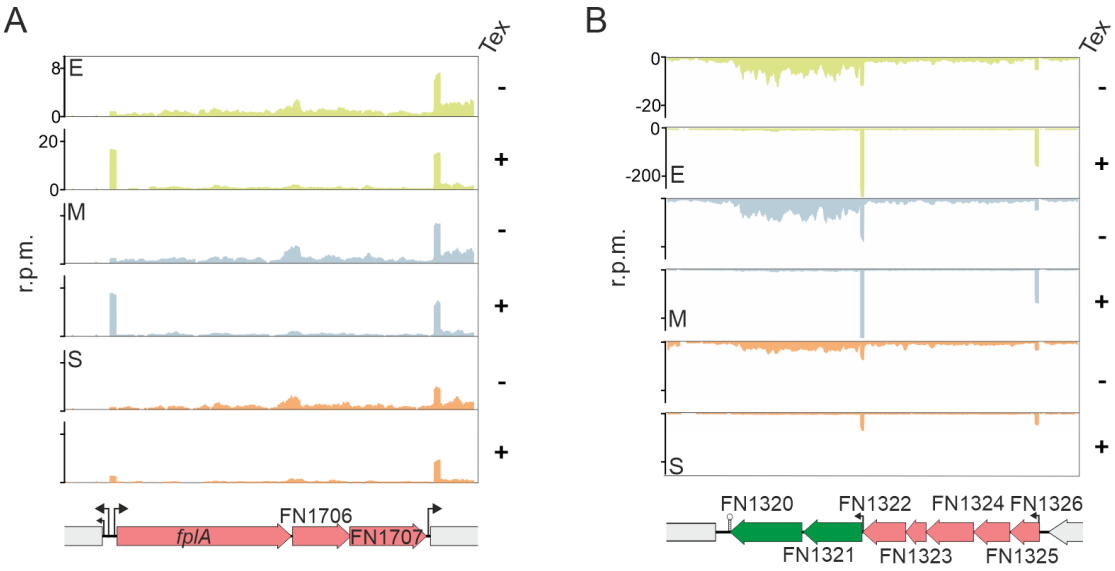

Figure 2.8 | Examples for the updated operon annotation.
A. Normalized coverage and schematic representation of the updated operon annotation for *fplA*-FN1706-FN1707. **B.** Normalized coverage and schematic representation of the updated operon annotation for FN1326-FN1320. The proposed suboperon FN1321-FN1320 (green) is indicated. This figures has been adapted from (Ponath et al., 2021).

gene harbors an ExbB-domain, the second and ExbD domain and the last one a potential a C-terminal TonB-domain. This operon structure suggests that this gene cluster could form a complex and transport larger molecules across the bacterial OM such as iron-siderophores or vitamin B_{12} (Braun, 1995).

Start of transcription inside an operon, marked by internal TSS, can indicate the presence of suboperons. The TSS and operon annotation led to the prediction of 53 suboperons (Supplementary data 8.6). Interestingly, ~40% (21/53) of the suboperons start at the last or the last two genes of the operon and thus uncouple the transcriptional regulation of these genes from the remaining operon. An example for this is a secondary metabolite and pyrimidine biosynthesis operon FN1326-FN1320, in which the orphan response regulator FN1321 and downstream rotamase FN1320 were found to be transcribed independently from the remaining operon (Fig. 2.8B).

2.2.8 CRISPR-Cas system

CRISPR-Cas systems are a widespread anti-phage defense system found in many bacteria and archaea (Makarova et al., 2020). The dRNA-seq analysis identified a type I-B system in *F. nucleatum* together with a downstream encoded

CRISPR RNA (crRNA) locus (Fig. 2.9). The crRNA consists of a unique spacer which contains sequence fragments of foreign DNA acquired e.g. through phage infection. The spacers are gapped by repeat regions which facilitate the processing of mature repeat-spacer pairs from the pre-crRNA. The mature crRNA together with the Cas-complex can then cleave invading nucleic acids in a sequence specific manner thereby endowing bacteria with an adaptive immune system (McGinn and Marraffini, 2019; Barrangou and Horvath, 2017; Jackson et al., 2017).

To check if the identified fusobacterial crRNA was part of such an anti-phage defense, the 16 spacers were searched against the Blast database. This revealed only a single match against the genome of the fusobacterial phage ϕFunu2 (Cochrane et al., 2016), supporting the anti-phage defense function of this system. The function of the CRISPR-Cas system requires regulated transcription and subsequent processing of the crRNA. The latter is achieved by Cas6 in the type I-B system (Carte et al., 2008). Active processing of the crRNA in *F. nucleatum* is evident by the detection of a complex pattern with stable intermediates of double or multiple repeat-spacer pairs via northern blot (Fig. 2.9). This reflects the processing patterns also found for the type I-B systems in *Clostridium thermocellum* and *Methanococcus maripaludis* (Richter et al., 2012).

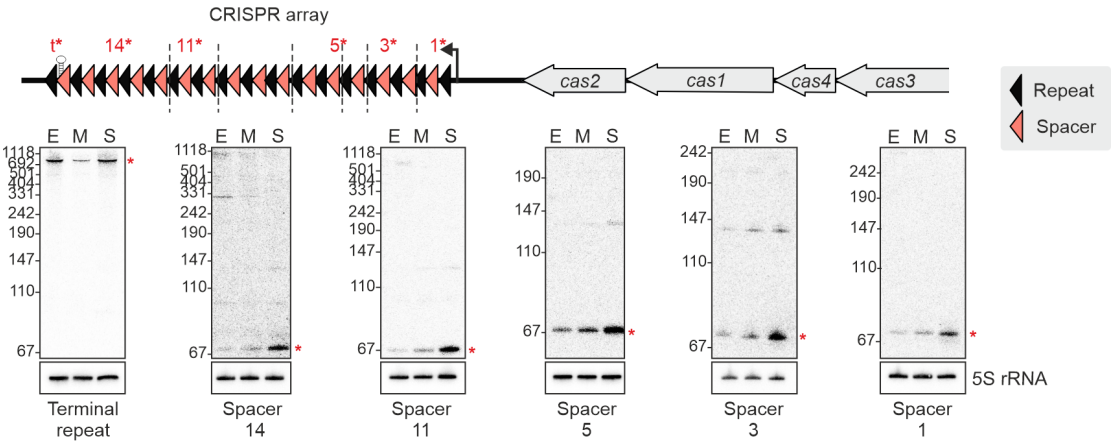

Figure 2.9 | An active CRISPR-Cas system in *F. nucleatum*.
The top shows the schematic representation of the CRISPR-Cas locus in *F. nucleatum* with the marked start and processing sites of the crRNA. The bottom displays the northern blot detection of different processing spacers of the crRNA.

The northern blot validation of the crRNAs indicates increased levels towards the stationary phase (Fig. 2.9) and a similar trend was found for the expression of the Cas operon (Supplementary data 8.2). Recent studies have shown that

the CRISPR-Cas systems can also exhibit regulatory function or be involved in virulence (Mohanraju et al., 2022). While it is unclear if the CRISPR-Cas system of *F. nucleatum* is limited to the anti-phage defense, the expression levels suggest that the bacterium needs to balance an active defense system with its growth optimization.

2.2.9 RNA regulatory elements

The 5'UTR of genes can harbor *cis*-regulatory RNA elements known as riboswitches or RNA thermometers. Both of these usually form specific secondary structure that allow post-transcriptional regulation of the downstream gene (McCown et al., 2017; Loh et al., 2018). Through this, the *cis*-regulatory elements are involved in the post-transcriptional regulation of important biological processes such as metal-uptake, metabolite biosynthesis (Epshtein et al., 2003; Cromie et al., 2006) or location-dependent control of virulence genes (Loh et al., 2018).

To predict these regulatory elements in *F. nucleatum*, the here identified 5'UTRs were used for a computational analysis using the RFAM database as reference (Kalvari et al., 2018).

Riboswitches

The analysis predicted a total of 12 high-confidence riboswitch candidates (Fig. 2.10; Supplementary data 8.7). Three candidates are classified as the cobalamin riboswitches suggesting that the associated genes are linked with uptake of cobalt or vitamin B_{12} (Vitreschak et al., 2003). Indeed, two of these putative riboswitches are found upstream of operons are predicted to be involved in iron- and hemin-acquisition (FN0300-FN0307; FN1971-FN1967). This suggests that *F. nucleatum* utilizes these operons to increase internal iron levels with rising levels of cobalt to avoid cobalt toxicity (Ranquet et al., 2007). The remaining one putative cobalamin riboswitch is found in the 5'UTR of the transcript encoding for a type 5a autotransporter of unknown function (Umaña et al., 2019).

Two additional putative riboswitches belong to the class of flavin mononucleotide (FMN) riboswitches (Supplementary data 8.7). One of them is located in the 5'UTR of the *ribH-ribD* operon (FN1505-1506), involved in the metabolism of the FMN precursor riboflavin itself, while the second one is located upstream of a putative metabolite transporter (FN1498).

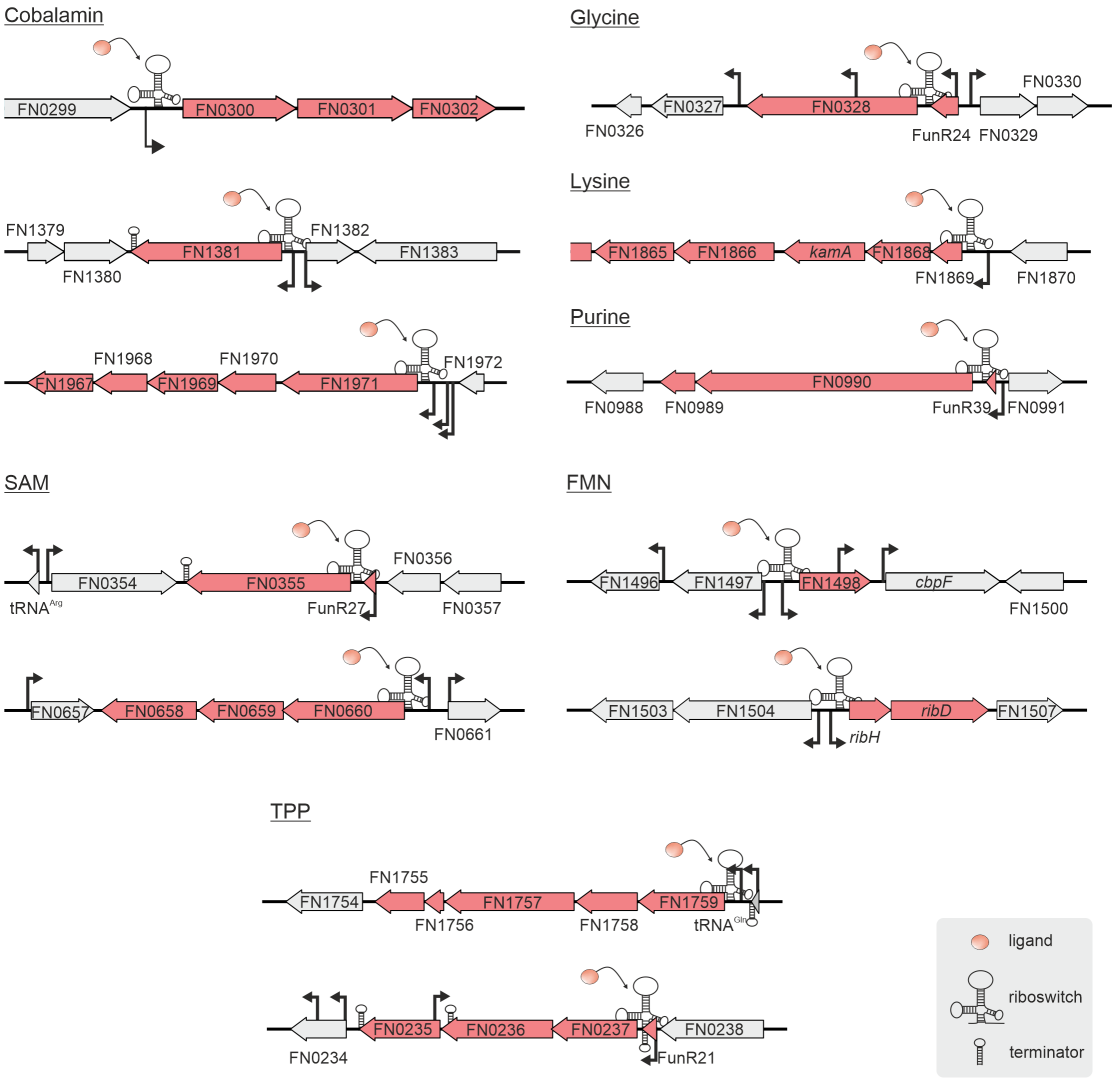

Figure 2.10 | Overview of predicted riboswitches in *F. nucleatum*.
Schematic overview of predicted riboswitches and their associated genes (red) in *F. nucleatum*. This figure was adapted from Ponath et al. (2021).

An identified glycine riboswitch has been previously been shown to bind the amino acid cooperatively *in vitro* and its crystal structure is available (Kwon and Strobel, 2008; Butler et al., 2011). This riboswitch is found in the 5'UTR of the gene for a putative glycine permease (FN0328) and thus likely functions to adjusts levels of the permease to the glycine levels.

Recent work found that the TCS CarRS regulates lysine metabolism and polymicrobial interactions in *F. nucleatum* (see Chapter 1.1.2; (Wu et al., 2021)). Important for this process are the genes *kamA* and *kamD* which belong to lysine metabolic operon and dependent on CarRS for their expression. Interestingly, the analysis identified a putative lysine riboswitch in the 5'UTR of the operon con-

taining *kamA* and *kamD* (FN1869-FN1861). This suggest that lysine metabolism likely controlled on a post-transcriptional level through the availability of lysine and transcriptionally by the activity of the TCS CarRS.

In addition, *cis*-regulatory elements belonging to the families of SAM, purine and thiamine pyrophosphate (TPP) riboswitches were found. Generally, these were located in the 5'UTR of genes or operons related the respective ligand of the riboswitch.

Additional regulatory elements

Synthesis of glucosamine-6-phosphate (GlcN6P) is important for the cell wall biogenesis in bacteria such as *Bacillus subtilis* and is carried out by the amidotransferase GlmS (Milewski, 2002). In *B. subtilis*, the levels of the *glmS* mRNA are post-transcriptionally controlled by the *glmS* ribozyme, which senses GlcN6P and can subsequently cleave the associated mRNA (Winkler et al., 2004). Interestingly, the *glmS* ribozyme is widespread among mostly Gram-positive Firmicutes, but is also found upstream of a *glmS* homologue in the Gram-negative *F. nucleatum*. However, the importance of GlcN6P and its potential relevance in envelope biogenesis has been not explored in *F. nucleatum*.

Two other *cis* elements have been identified in the 5'UTRs, including an ribosomal protein leader commonly associated with the ribosomal L10 protein. As shown for *B. subtilis* or *E. coli*, the RNA elements allow the auto-regulatory control of the ribosomal protein synthesis (Yakhnin et al., 2015; Johnsen et al., 1982). Its function is likely similar in *F. nucleatum* as the here identified ribosomal protein leader is located also upstream of the L10 gene.

The second *cis* element represents a PyrR-binding site in the 5'UTR of the transcriptional regulator *pyrR* itself and the *pyr*-operon involved in pyrimidine biosynthesis. This element can be bound by PyrR and cause transcription termination and subsequent repression of the downstream genes as shown in *B. subtilis* (Bonner et al., 2001). The conserved genomic location of the RNA element suggests that also *F. nucleatum* utilizes this mechanism of transcriptional attenuation to control its pyrimidine biosynthesis.

RNA thermometers

Another class of *cis*-regulatory elements are the RNA thermometers which are located in the 5'UTR of an mRNA's and control its translation efficiency in a temperature-sensitive manner (Kortmann and Narberhaus, 2012). However, none of the five predicted RNA thermometers are located within a 5'UTR (Supplementary data 8.7). This suggests that *F. nucleatum* might not rely on such temperature-sensitive regulatory elements to post-transcriptionally control its gene expression. Supporting this is the strong host-association of the bacterium which likely does not experience strong temperature shifts in contrast to such as pathogen *Listeria monocytogenes* that rely on an RNA thermometer to control the translation of a major virulence regulator (Johansson et al., 2002).

Interestingly, all these *cis*-regulatory elements were found in long 5'UTRs. This raises the question if the remaining longer 5'-regions identified here might carry yet undiscovered classes of RNA elements.

2.3 Non-coding RNAs

The previous annotation for *F. nucleatum* contained only entries for mRNAs, rRNAs and tRNAs, but lacked any non-coding RNAs (ncRNAs). However, ncRNAs play an important role in the physiology of bacteria (see Chapter 1.2.5). To identify potential sRNA candidates the dRNA-seq data were used as input for ANNOgesic (Yu et al., 2018b). This was achieved by comparing the coverage files with the predictions for the TSS, processing sites, transcripts and terminators. In addition, putative candidates needed to exhibit a minimal stable secondary structure (-0.05 kcal/mol) based on the length-normalized minimum free energy prediction by RNAfold of the Vienna RNA Package (Lorenz et al., 2011). This analysis yielded a list of 47 candidate ncRNAs. All candidates were used as input for a BLASTN analysis (Agarwala et al., 2016) against the bacterial small regulatory RNA despository (BSRD) database (Li et al., 2013) which did not yield any hit. This suggest that the ncRNA candidates are likely specific to *F. nucleatum*.

2.3.1 Housekeeping RNAs

The housekeeping RNAs tmRNA, 4.5S RNA, RNaseP-associated M1 RNA and the 6S RNA are highly conserved among bacteria. The former three further exhibit conserved secondary structure features that allow their reliable prediction.

Comparing the ncRNA candidates to the covariance models in the RFAM database (Kalvari et al., 2018) identified all three of them. Indicative of their importance, the primary sequence of each of the RNAs was found to be strongly conserved among Fusobacteria. The expression of each of these RNAs was validated by northern blot detection (Fig. 2.11).

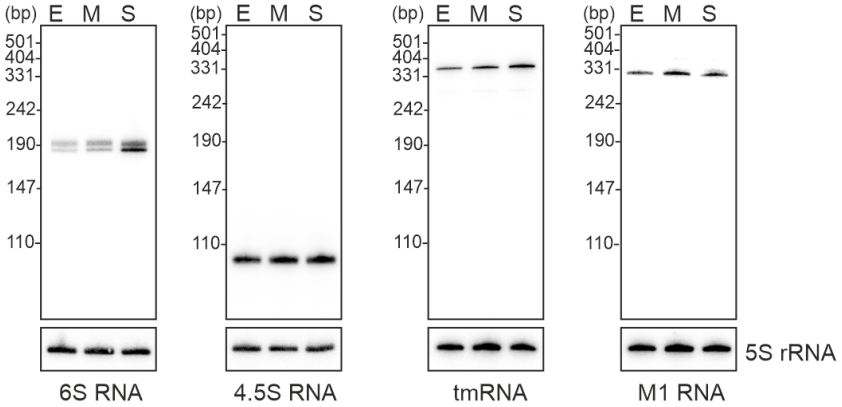

Figure 2.11 | Identification of conserved housekeeping ncRNAs.
Northern blot detection across the different growth stages for the core ncRNAs: 6S RNA, 4.5S RNA, tmRNA and RNase P associated M1 RNA. The 5S rRNA served as loading control. This figure has been adapted from Ponath et al. (2021).

2.3.1.1 M1 RNA

The above-mentioned analysis showed the RNase P RNA (M1 RNA) to be a ∼330 nt long transcript processed off the 3'end of the transcript for the hypothetical protein (FN1315). As found for many other bacteria, the RNase P RNA is expressed independently of its accessory protein counterpart, RnpA (FN0002), with whom it cooperates to generate the mature 5'ends of tRNAs (Hartmann and Hartmann, 2003).

2.3.1.2 4.5S RNA

The 4.5S RNA was detected as a 105 nt long RNA transcribed from its own promoter (Fig. 2.12A). The prediction of its secondary structure shows the conserved apical 'GGAA' tetraloop, which is essential for its interaction with the signal recognition particle protein Ffh to form the signal recognition particle (SRP) (Fig. 2.12) (Akopian et al., 2013). This ribonucleoprotein complex (RNP) is integral to the recognition of the signal peptides on emerging peptide chains and their targeting to the inner membrane for co-translation in or across the inner

membrane (Saraogi and ou Shan, 2014).

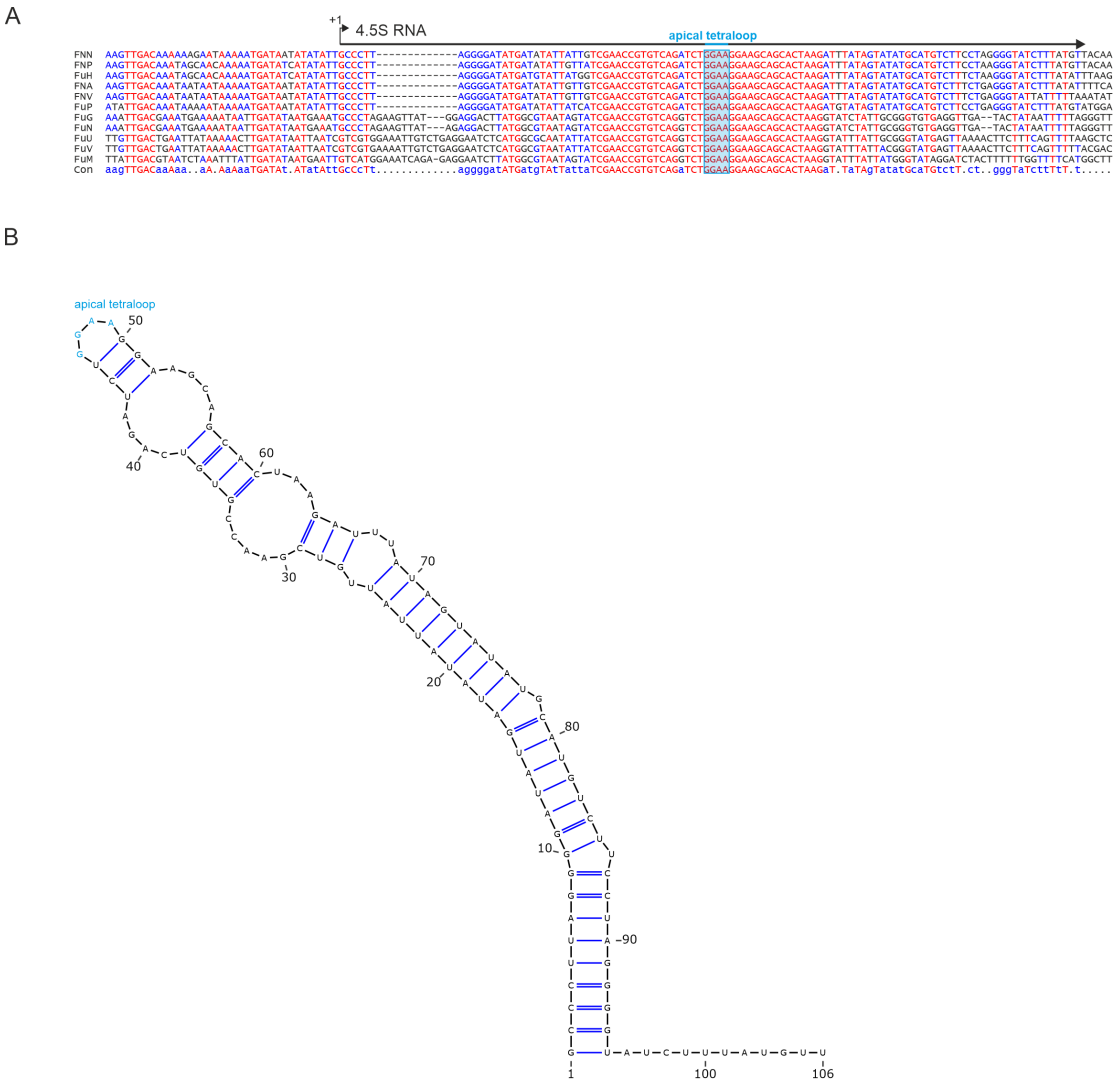

Figure 2.12 | The 4.5S RNA in *F. nucleatum*.
A. Sequence alignment of the 4.5S RNA. **B.** Secondary structure prediction of the 4.5S RNA highlighting the apical GGAA tetraloop. *Fnn*: *F. nucleatum* subspecies *nucleatum*; *Fna*: *F. nucleatum* subspecies *animalis*; *Fnp*: *F. nucleatum* subspecies *polymorphum*; *Fnv*: *F. nucleatum* subspecies *vincentii*; *FuG*: *F. gonidiaiformans*; *FuH*: *F. hwasookii*; *FuM*: *F. mortiferum*; *FuN*: *F. necrophorum*; *FuP*: *F. periodonticum*; *FuU*: *F. ulcerans*; *FuV*: *F. varium*.This figure has been modified from Ponath et al. (2021).

2.3.1.3 tmRNA

The tmRNA was verified as a 363 nt long transcript located in the intergenic region between two genes of unknown function. Previous works have shown that the tmRNA plays an important role in translation. Specifically, it forms an RNP together with the Small Protein B (SmpB) and binds ribosomes that are stalled

on a damaged mRNA without stop codon or when the required tRNA is lacking (Moore and Sauer, 2007). These stalled ribosomes are then released through a trans-translational mechanism of a small ORF on the tmRNA with a characteristic 'YALAA' C-terminus (Flynn et al., 2001). Similarly, the fusobacterial tmRNA contains a conserved 14 aa ORF including the characteristic C-terminal sequence (Fig. 2.13A). Interestingly, a sequence dichotomy between oral and non-oral fusobacterial isolates was observed in which the non-oral isolates harbored an elongated peptide (25 aa) (Fig. 2.13B). Nonetheless, these results argue that the important process of rescuing stalled ribosome in *F. nucleatum* is mediated in a conserved manner by the tmRNA-SmpB (FN0609) RNP.

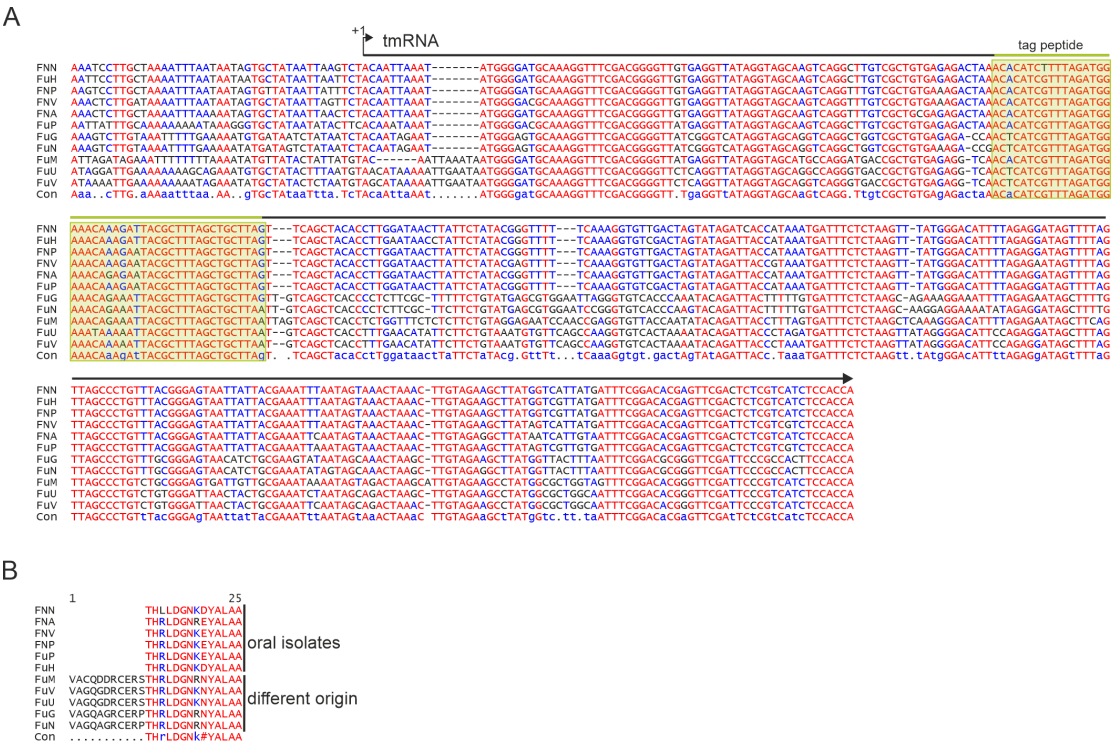

Figure 2.13 | The tmRNA in *F. nucleatum*.
A. Sequence alignment of the tmRNA with ORF of the tag-peptide indicated (green). **B.** Conservation of the tag-peptide displaying differences of the aa sequences between oral isolates and other strains. *Fnn*: *F. nucleatum* subspecies *nucleatum*; *Fna*: *F. nucleatum* subspecies *animalis*; *Fnp*: *F. nucleatum* subspecies *polymorphum*; *Fnv*: *F. nucleatum* subspecies *vincentii*; *FuG*: *F. gonidiaiformans*; *FuH*: *F. hwasookii*; *FuM*: *F. mortiferum*; *FuN*: *F. necrophorum*; *FuP*: *F. periodonticum*; *FuU*: *F. ulcerans*; *FuV*: *F. varium*. This figure has been modified from Ponath et al. (2021).

2.3.1.4 6S RNA

In comparison to the three cases above the 6S RNA cannot be easily predicted in part due to their lack of a conserved primary sequence. Still, the wide-spread riboregulator, together with the associated pRNA, plays an important role by modulating gene expression in *B. subtilis* and *E. coli* (see Chapter 1.2.1; (Wassarman, 2018)).

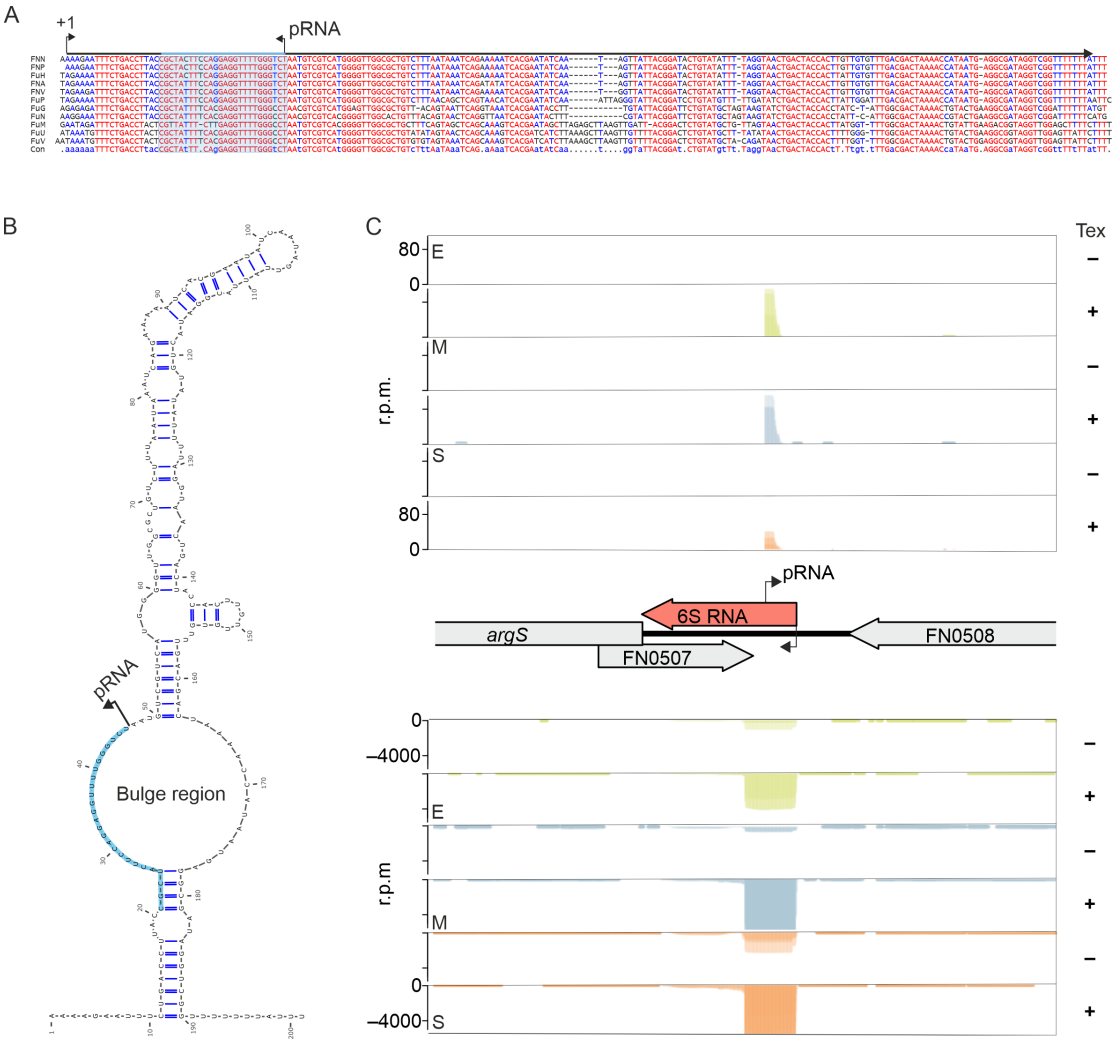

Figure 2.14 | The 6S RNA in *F. nucleatum*.
A. Sequence aligned of the 6S RNA with the highlighted pRNA (blue). **B.** Secondary structure prediction of the 6S RNA. The central bulge region and pRNA transcript (blue) are indicated. **C.** Normalized coverage and schematic overview of the 6S RNA in *F. nucleatum*. The TSSs of the 6S RNA and the pRNA are indicated. *Fnn*: *F. nucleatum* subspecies *nucleatum*; *Fna*: *F. nucleatum* subspecies *animalis*; *Fnp*: *F. nucleatum* subspecies *polymorphum*; *Fnv*: *F. nucleatum* subspecies *vincentii*; *FuG*: *F. gonidiaiformans*; *FuH*: *F. hwasookii*; *FuM*: *F. mortiferum*; *FuN*: *F. necrophorum*; *FuP*: *F. periodonticum*; *FuU*: *F. ulcerans*; *FuV*: *F. varium*. This figure has been modified from Ponath et al. (2021).

Out of the 47 predicted ncRNA candidates a 190 nt long candidate displayed the conserved secondary structure of a 6S RNA with a central bulge inside a long hair-pin (Fig. 2.14A, B). Additionally, its genomic synteny was strongly conserved between the genes encoding for of the serine protease FN0508 and the arginyl-tRNA synthetase ArgS (FN0506). The expression pattern of the 6S RNA also reflects that found in other bacteria as the fusobacterial ncRNA accumulates in the stationary phase (Fig. 2.11 and 2.14C). Strikingly, the transcriptomic data further revealed the transcription of a 26 nt long pRNA originating in the central bulge in *cis* to the 6S RNA as found for *E. coli* or *H. pylori* (Fig. 2.14B, C) (Wassarman, 2018; Sharma et al., 2010).

Combined, the detection of these three conserved features show that this RNA represents the fusobacterial 6S RNA.

2.3.2 Identification of regulatory sRNAs

The remaining ncRNA candidates were distributed across the entire genome of *F. nucleatum* and no did not show sequence similarities to known ncRNAs (Fig. 2.15A). Thus, it was unclear if these candidates represent stable RNA species with a putative sRNA function (see Chapter 1.2.5).

As the first step, the presence of each individual candidate sRNA was evaluated via northern blot analysis across the different growth stages. This extensive validation detected 24 out of 43 stably expressed bona fide sRNAs (Fig. 2.15B). Further, the experimentally observed sizes were in agreement with the computational prediction ranging from 56 nt to 345 nt (Supplementary data 8.8).

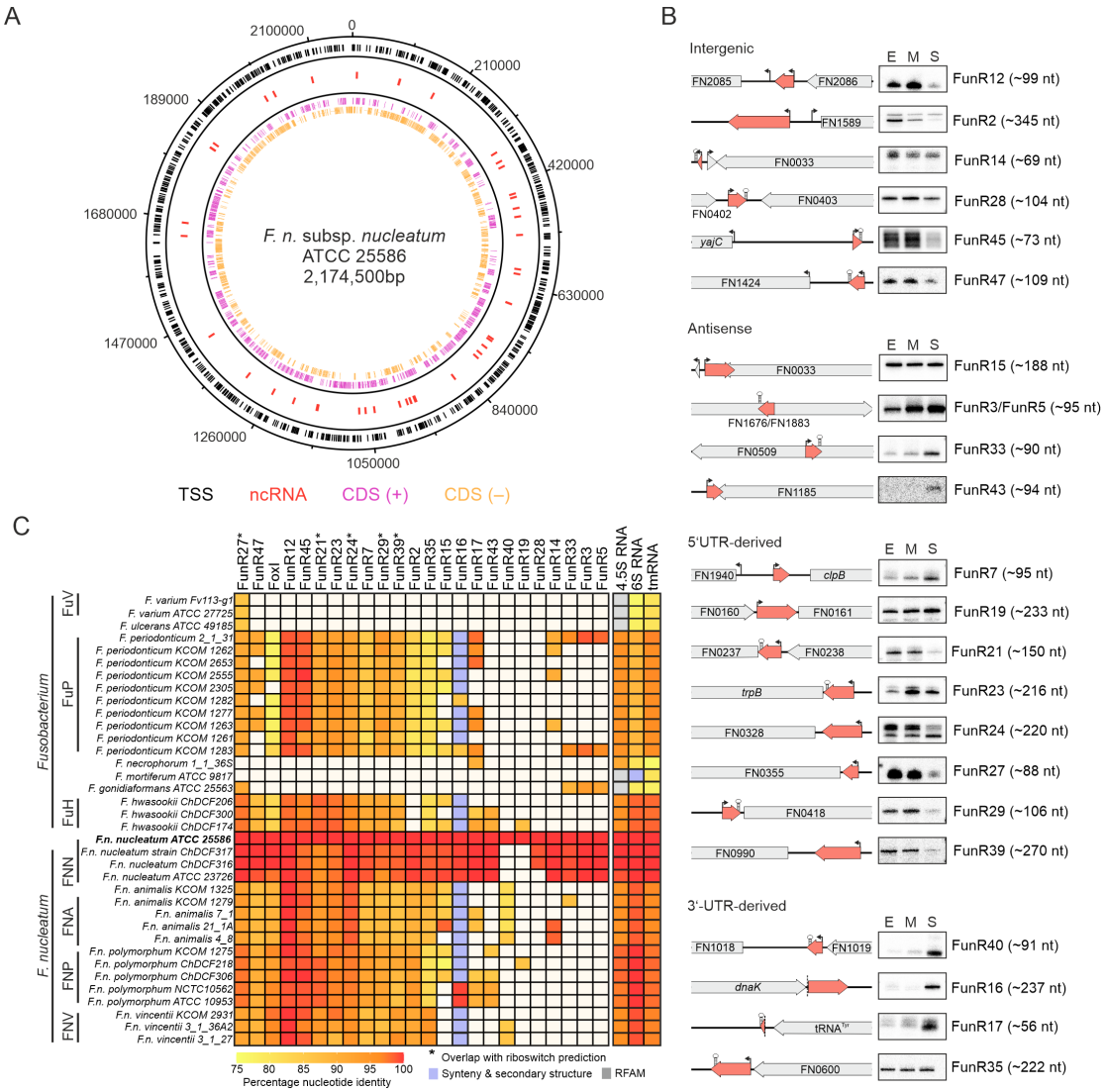

Figure 2.15 | A conserved suite of fusobacterial sRNAs.
A. Genomic location of all predicted TSS and sRNAs across the genome of *F. nucleatum*. **B.** Schematic overview of sRNAs (red) divided into different sRNA classes combined with the northern blot validation across the three different growth stages. **C.** Heatmap depicting the conservational analysis for all validated sRNA in 36 strains of *F. nucleatum* and related species with a complete genome. An sRNA was classified as conserved in a strain when it showed an ≥75% nucleotide identity and length compared to the sRNA in the reference strain. Grey boxes were only identified via the RFAM prediction and purple boxes indicate conservation only when comparing the genomic synteny and secondary structure. An asterisk indicates the overlapping prediction with a riboswitch. This figure was adapted from Ponath et al. (2021).

Based upon the origin of transcription, the sRNAs were categorized in the four common sRNA classes: intergenic, antisense, 5'UTR derived or 3'UTR derived. Intergenic sRNAs are transcribed as an independent gene from an intergenic region (IGR), antisense sRNAs are encoded in *cis* to another gene and the UTR-derived sRNAs are generated from the respective untranslated regions of an mRNA via processing or termination events. In total nine intergenic, five antisense, six 5'UTR

and four 3'UTR-drived sRNAs could be validated in *F. nucleatum* (Fig. 2.15A; Supplementary data 8.8).

Findings in Proteobacteria showed that sRNAs can often be quite conserved on a sequence level among different species, such as the sRNA CpxQ of *E. coli* which is also found in *Salmonella* Typhimurium or *Yersinia pestis* (Chao and Vogel, 2016). Due to the evolutionary distance of *Fusobacterium* from most bacteria (Coleman et al., 2021) it is not surprising that none of the fusobacterial sRNAs are conserved on a sequence level compared to sRNAs in the BSRD database containing over 8000 sRNA sequences (Li et al., 2013).

2.3.2.1 The fusobacterial suite of sRNAs

To investigate if the sRNAs display conservation inside the *Fusobacterium* genus, all validated transcripts were used as input for a BlastN analysis against all complete genome assemblies found on NCBI for *Fusobacterium* (Fig. 2.15C). This search across 36 fusobacterial strains suggests a broad conservation for 15 of these sRNAs (cut-off: 75%-identity and minimum length of sRNA in *F. nucleatum*). This group could present fusobacterial core sRNAs as they are found in members of *F. nucleatum* and the related species *F. hwasookii* and *F. periodonticum*. Such is the case for abundant intergenic sRNA FunR12 which displays a remarkable level of nucleotide conservation of $\geq$99% across all strains. This indicates that FunR12 likely plays a important role during exponential growth when the sRNA is highly abundant (Fig. 2.15B).

2.3.2.2 FunR23 - a conserved riboregulator?

Similarly, the 5'UTR-drived FunR23 also shows a strong conservation in combination with genomic position upstream of gene for the tryptophan synthase subunit beta (*trpB*) (Fig. 2.15B). The location of FunR23 suggesta that this sRNA could be a novel class of riboregulator. Interestingly, RNA leaders of the tryptophan operons are wide-spread as they are present in e.g. Proteobacteria or Firmicutes. Regarding the latter, the tryptophan leader (*trp* leader) is bound by the *trp* RNA-binding attenuation protein (TRAP) that causes premature termination of the tryptophan operon in *B. subtilis* (Potter et al., 2011). However, several other mechanisms are known for the *trp* leader, *trpL*, responding to tryptophan levels in Proteobacteria: causing transcriptional attenuation in a small ORF-dependent manner, as a small peptide generated from the ORF or acting as

an sRNA to repress mRNA targets (Evguenieva-Hackenberg, 2022).

In contrast, FunR23 encodes a small ORF containing consecutive threonine codons instead of tryptophan that is usually found with the *trpL* leader (Yanofsky, 1981; Evguenieva-Hackenberg, 2022). Thus, it is not clear how FunR23 acts in *F. nucleatum* but represents an interesting candidate for future studies, given the diverse mode of actions and the regulatory impact of *trp* leaders in other bacteria (Evguenieva-Hackenberg, 2022). Similarly to *trpL* responding to tryptophan, FunR23 might regulate the expression of *trpB* expression depending on threonine levels, since it contains consecutive codons for this amino acid in its ORF.

2.3.2.3 FunR19 - part of a putative anti-phage defense system

In contrast to the two above mentioned examples, FunR19 can only be found in two additional strains: *F. nucleatum* subsp. *polymorphum* ChDCF218 and *F. nucleatum hwasookii* ChDCF174 (Fig. 2.15C). Interestingly, the sRNA is always found upstream in the 5'UTR of gene encoding for a putative reverse transcriptase. The combination of a reverse transcriptase with an associated ncRNA is termed retron and previous bioinformatic analysis identified these in a variety of microbial genomes including *F. nucleatum* (Rest and Mindell, 2003; Simon et al., 2019). Recent studies highlighted the involvement of retrons in anti-phage defense (Millman et al., 2020; Bobonis et al., 2022) and thus the constitutively transcribed FunR19 together with the downstream reverse transcriptase might play a similar role in protecting *F. nucleatum* from bacteriophages.

2.3.2.4 FunR16 and FunR7 - conserved beyond *Fusobacterium*?

Standing out is the 3'UTR-derived sRNA FunR16 located downstream of gene encoding for the broadly conserved stress chaperone DnaK (Fig. 2.16). The 237 nt long primary sequence is only conserved in the members of the *F. nucleatum* subspecies *nucleatum* in addition to two *F. hwasookii* strains. However, comparing the untranslated region downstream of *dnaK* revealed a conserved secondary structure resembling a long stem-loop structure (Fig. 2.16). The sRNA is likely the result of a processing event from the 3'-end of the *hrcA-grpE-dnaK* mRNA as no enrichment in the TEX-treated libraries at the start site of the sRNA could be detected. As 3'UTR dervied sRNAs often act in conjunction with their parental mRNA (Ponath et al., 2022a), FunR16 might be involved in a general stress response of *F. nucleatum* as it accumulates towards the stationary phase together with the *dnaK* mRNA and other stress chaperones (Fig. 2.15A; Supplementary

data 8.2).

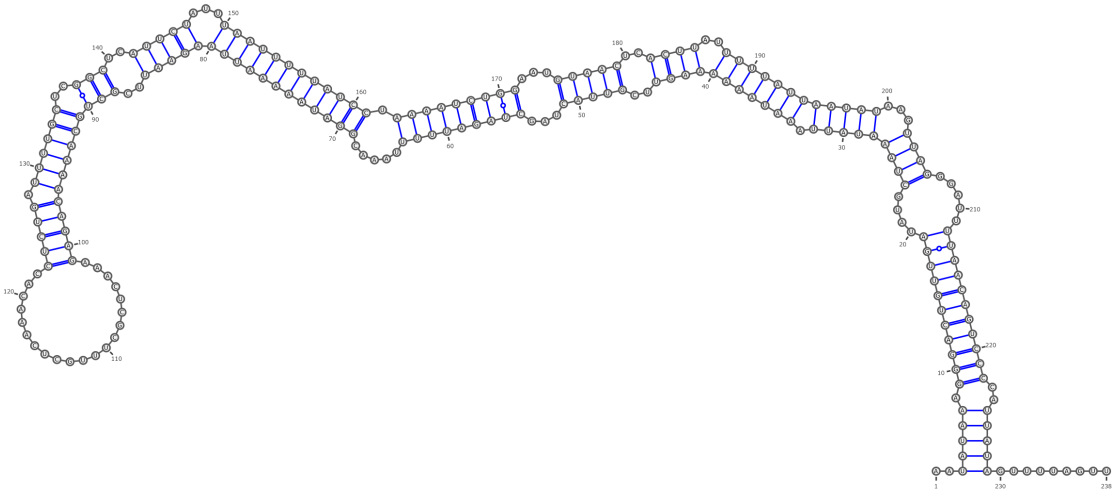

Figure 2.16 | Secondary structure of FunR16
The prediction of FunR16 secondary structure suggest the sRNA to form a stable stem loop with an apical loop.

Interestingly, FunR16 is a positional ortholog to the sRNA Tpke11 found in *E. coli* or *Salmonella enterica* serovar Typhimurium (Rivas et al., 2001; Barnhill et al., 2019). However, Tpke11 is significantly shorter with only 83 nt and no conserved sequences or secondary structure elements could be found comparing it to FunR16.

At the same time, FunR7 is encoded in the 5'UTR of the stress chaperone-encoding *clpB* (Fig. 2.15B) sharing its location with the abundant ProQ-binding sRNA RyfD (Smirnov et al., 2016; Melamed et al., 2020). However, there is no similarity between either of the sRNAs in terms of length, primary sequence, or secondary structure.

Future functional analysis for FunR16 and FunR7 might uncover a shared function with their potential counterparts in *E. coli*. This might involve a role in biofilm formation for FunR7, a characteristic trait of *F. nucleatum*, as RyfD has been implicated in regulating this process (Bak et al., 2015).

2.3.2.5 Potential dual-function sRNAs

Five of the identified 5'UTR-derived sRNAs overlap with the riboswitch prediction (Fig. 2.15C; see Chapter 2.2.9). This includes FunR27 which encompasses part of the 5'UTR of the *metK* mRNA, encoding the S-methyladenosylmethionine

synthetase, which is also predicted to be SAM family riboswitch. A recent study in *Listeria monocytogenes* uncovered the potential dual-function role of riboswitches to not only act in *cis*, regulating the downstream mRNA, but also in *trans* for a SAM family riboswitch (Loh et al., 2009). As all five fusobacterial riboswitches were also detected as stable transcripts (Fig. 2.15B), it is possible that they have a similar dual-function in *F. nucleatum*.

2.3.2.6 *Cis*-regulatory sRNAs

The northern blot analysis further validated the expression of five sRNAs encoded antisense to a coding gene (Fig. 2.15C). Three of them (FunR3, FunR5 and FunR33) are associated with genes encoding transposases. The different patterns detected by northern blot analysis of these sRNAs suggest that they likely undergo different processing steps in the maturation of the transcripts. Ellis et al. (2015) showed that such antisense sRNAs can be involved in the translational repression of the cognate transposase by limiting the overall transposition of a mobile genetic element. Thus, it seems likely that the three sRNAs play a similar role in *F. nucleatum* by minimizing unwarranted transposase activity.

Another interesting example of a *cis*-encoded antisense sRNA is FunR43 which overlaps with the 3'-end of the gene for a SIR2-domain containing protein (FN1185) (Fig. 2.15B). Interestingly, the genomic location on an 'defense island', between the CRISPR-Cas and an abortive infection system (Abi), suggest that FN1185 could play a role in the defense against phages or foreign DNA (Doron et al., 2018; Payne et al., 2021). When encountering such a situation, FunR43 might play an important role in regulating FN1185.

2.3.3 Transcriptional regulation of sRNAs

The transcription of bacterial sRNAs is often tightly regulated by the activity of σ factors or transcription factors (see 1.2 and 1.2.5). In order to identify potential promoter motifs associated with the sRNAs, a similar MEME analysis (see 2.2.3) was performed for the -50 nt upstream region of all sRNAs. This analysis revealed that twelve of these sRNAs displayed a similar σ^{70} motif as found for the majority of coding genes (Fig. 2.17). This indicates that these sRNAs and their regulatory activity are important during the growth of *F. nucleatum*. However, the transcription of the sRNAs harboring this promoter motif is more nuanced as shown by closer a comparison of FunR47 to FunR15. FunR15 is highly abundant

throughout all three analyzed growth phases while the levels of FunR47 decrease upon entry into the stationary phase (Fig. 2.15B). Thus, the expression of at least one of these sRNAs is likely controlled by factors other than σ^{70}. Further, the remaining sRNAs not displaying the canonical promoter motif are likely similarly subject to transcriptional regulation by yet unidentified factors and might thus represent potential starting points for the investigate of condition-specific sRNA-related regulatory networks.

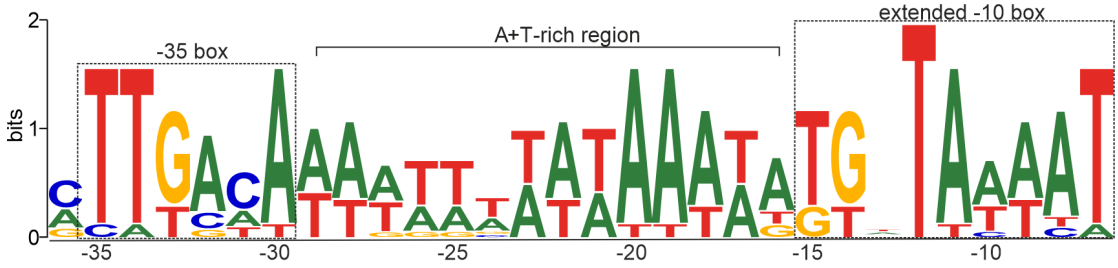

Figure 2.17 | The identified promoter of validated ncRNAs.
In 12 cases, the MEME (Bailey et al., 2015) analysis identified a promoter motif displaying an extended -10 box and -35 box separated by an A+T-rich spacer. This figure has been adapted from Ponath et al. (2021).

2.4 The primary transcriptome of the additional fusobacterial strains

To gain a broader overview of the transcriptome architecture of *Fusobacterium* the dRNA-seq analysis was carried out for representative strains of the remaining *F. nucleatum* subspecies *animalis*, *polymorphum* and *vincentii* as well as the closely related species *F. periodonticum*. As stated above (see 2.2), this included samples for the three different growth phases for all strains, which were analyzed in the same manner as *Fnn* by using ANNOgesic (Yu et al., 2018b).

2.4.1 TSSs in the additional fusobacterial strains

The analysis of the dRNA-seq data by ANNOgesic and manual curation resulted in total of 818 TSS for *Fna*, 946 TSS for *Fnp*, 632 TSS for *Fnv* and 724 TSS for *FuP* (Fig. 2.18; Supplementary data 8.1). In all cases, the primary TSS made up the majority of all detected TSS (between ~85% and ~91%). Just as in *Fnn*, sTSS only made up a small fraction of the total TSS (~2% to ~5%). This

supports the notion, that under these growth conditions, transcription is primarily driven from a single start site.

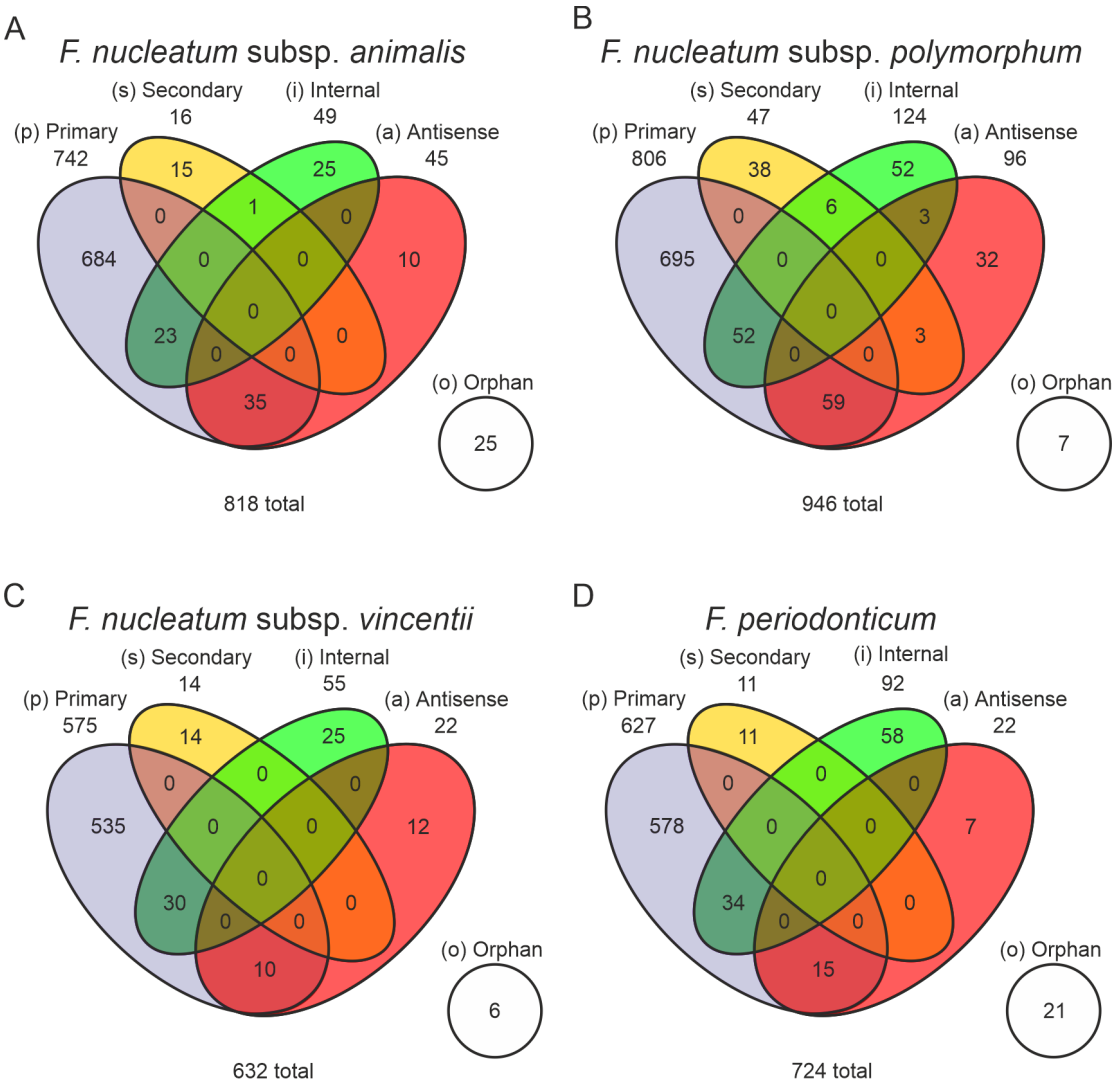

Figure 2.18 | TSS detection for the additional strains.
Venn-diagrams displaying the identified TSS by dRNA-seq for **A.** *F. nucleatum* subsp. *animalis*, **B.** *F. nucleatum* subsp. *polymorphum*, **C.** *F. nucleatum* subsp. *vincentii* and **D.** *F. periodonticum*. This figure has been modified from Ponath et al. (2021).

The second largest group of TSS belonged to the iTSS (∼6% to ∼13%) indicating a transcriptional uncoupling of suboperons also for these strains. Additionally, only little antisense transcription was observed as shown by the low fraction of aTSS (∼3% to ∼10%), with *Fnp* displaying the highest occurrence of aTSSs among the additional fusobacterial strains. This further supports the observation for aTSS transcription in *Fnn* and suggest that pervasive antisense transcription is not commonly found in *Fusobacteria* as in other bacteria (Georg and Hess, 2018).

2.4.2 5'UTRs for the additional fusobacterial strains

The 5'UTRs were predicted for the four additional strains based upon the annotation of the pTSS and sTSS. Overall, the 5'UTRs display a median length of 35 nt to 36 nt as found for *F. nucleatum* (Fig. 2.19; see 2.2.2). This analysis further found 71-144 genes with substantially longer 5'UTRs (>100 nt) among the different fusobacterial strains (Supplementary data 8.3). The longest 5'UTR identified in *Fna* and *Fnv* lies upstream of an operon encoding for an ABC transporter system and also contains a putative TPP riboswitch. Interestingly, *Fnn*, *Fnp* and *FuP* harbor a homolog of the operon together with the riboswitch, but the 5'UTR is significantly shorter (∼170 nt) indicating differences in regulation of the operon between the different species.

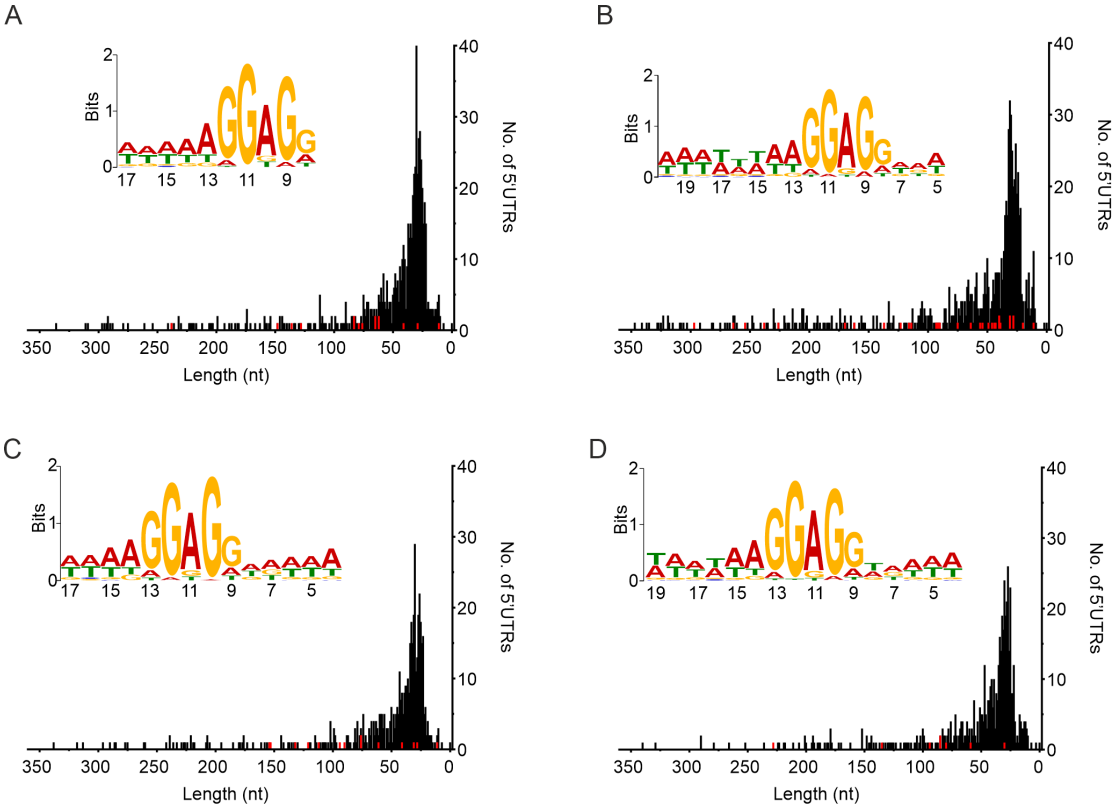

Figure 2.19 | Distribution of 5'UTRs for the additional strains.
Bar chart displaying the length and count for 5'UTRs annotated based on pTSS (black) and sTSS (red) for **A.** *F. nucleatum* subsp. *animalis*, **B.** *F. nucleatum* subsp. *polymorphum*, **C.** *F. nucleatum* subsp. *vincentii* and **D.** *F. periodonticum*. A consensus motif for the individual Shine-Dalgarno sequences was predicted using all 5'UTRs for each strain. The average distance to the translational start site is indicated. This figure has been modified from Ponath et al. (2021).

The longest 5'UTR for *Fnp* and *FuP*, with ∼340 nt, encodes for putative lysine-sensing riboswitch and is found upstream of a CarRS-controlled lysine metabolic operon (Wu et al., 2021) and shows similar length in the other strains.

In addition, the 5'UTRs for all analyzed strains contain a similar Shine-Dalgarno sequence, 5'-AGGAGG-3', pinpointing this motif as the common fusobacterial ribosome recognition site (Fig. 2.19).

Regarding short 5'UTRs, each of the additional strains harbored only 4-5 genes considered leaderless (≤10 nt). Similar to *Fnn*, these are often found associated with genes encoding for transposases or a pseudouridine synthetase. A comparison between the leaderless genes of all five fusobacterial strains uncovered a common homolog always located upstream of the gene *ffh*, which encodes for the signal recognition particle subunit involved in targeting proteins to translocation through the IM (Akopian et al., 2013; Saraogi and ou Shan, 2014). This gene is termed *ylxM* itself and its localization upstream of *ffh* is broadly conserved among Gram-positive bacteria (Williams et al., 2014). While its domain structure suggests a role in transcriptional control, recent results for *Streptococcus mutants* suggest that YlxM rather modulates the GTPase activity of Ffh required for the function of the SRP protein (Williams et al., 2014). Given the apparent role of YlxM in regulating an important biological process, it is unclear why this gene is leaderless in different Fusobacteria and it could be a prime candidate to investigate the role of alternative translation initiation in this genus.

2.4.3 Promoter analysis for the additional fusobacterial strains

As for *Fnn*, the promoter regions of all pTSS were used for a motif search using MEME (Bailey et al., 2015) (see 2.2.3). This analysis revealed a motif similar to that found for *Fnn* (Fig. 2.20). Only in *Fna* and *FuP* the majority of genes (88%-92%) is associated with both the extended -10 and -35 box (Fig. 2.20A, D). In case of *Fnp* and *Fnv* the analysis showed that most genes are (56%-64%) found only with an extended -10 box. The promoter region of the remaining pTSS for both species contains a pronounced -35 box with a varying association of the -10 box (Fig. 2.20).

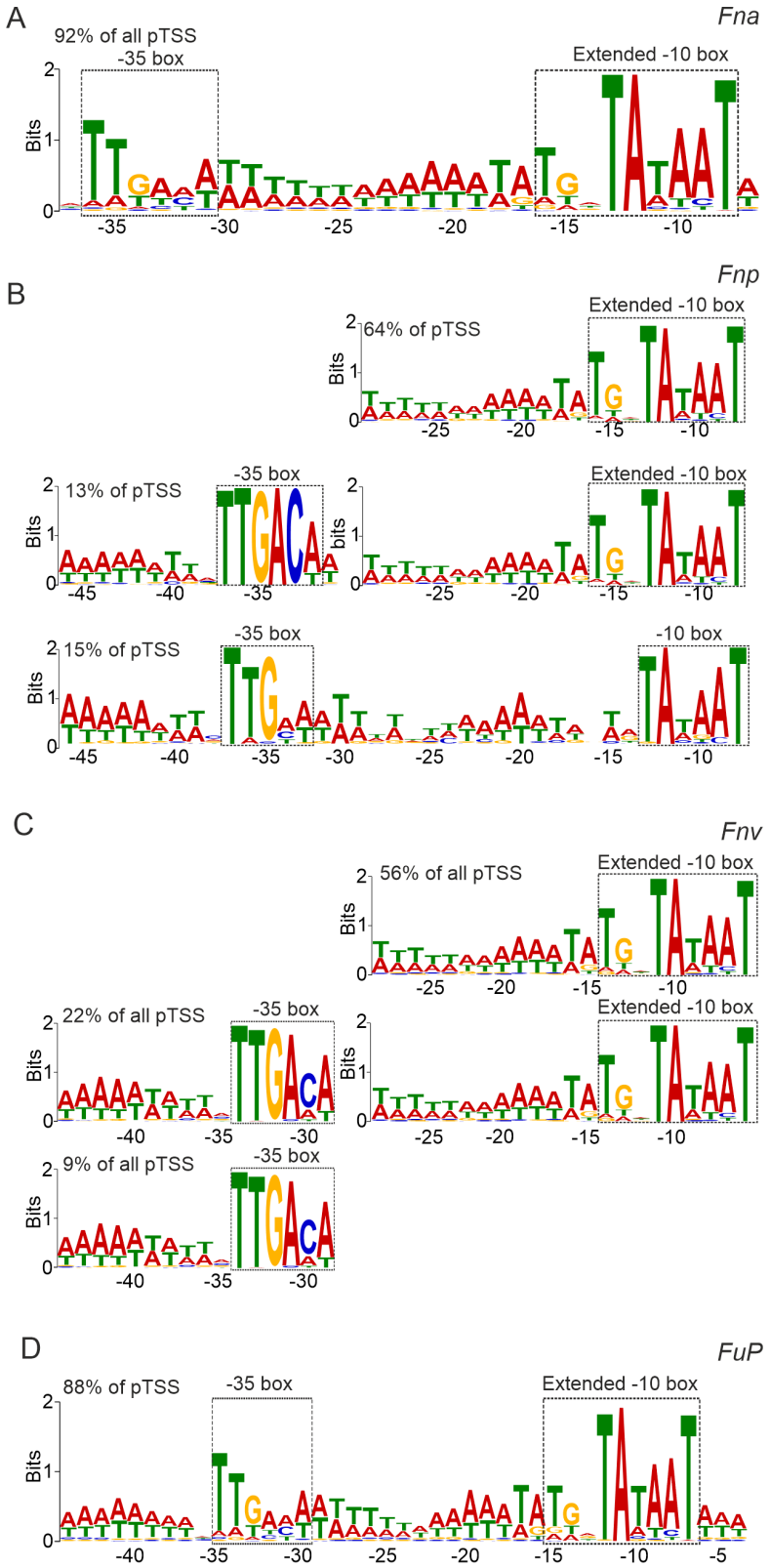

Figure 2.20 | Promoter identification for the additional strains.
Promoter motif identified by a MEME analysis for **A.** *F. nucleatum* subsp. *animalis*, **B.** *F. nucleatum* subsp. *polymorphum*, **C.** *F. nucleatum* subsp. *vincentii* and **D.** *F. periodonticum*. The percentage of pTSS for which the individual motif was detected are indicated. This figure has been modified from Ponath et al. (2021).

Overall, these results suggest that most of the transcription in rich media in *F. nucleatum* and *F. periodonticum* is driven by σ^{70} as in *Fnn*. Furthermore, no major differential expression changes of alternative transcriptional factors were observed (Supplementary data 8.2).

2.4.4 Comparing the 5'UTRs of the additional strains with *Fnn*

The bacterial 5'UTRs can harbor regulatory elements for the transcriptional and post-transcriptional control of gene expression. A comparative approach between different strains or species can identify such sequence regions as these are often more conserved than the remainder of the 5'UTR. Such conservation was investigated for *Fusobacterium* by comparing the 5'UTR of homologous genes present in all five analyzed strains with a detected pTSS. The latter is important as to allow an accurate determination of the 5'end for each gene in the individual strains. This left a total of 205 genes for the comparison.

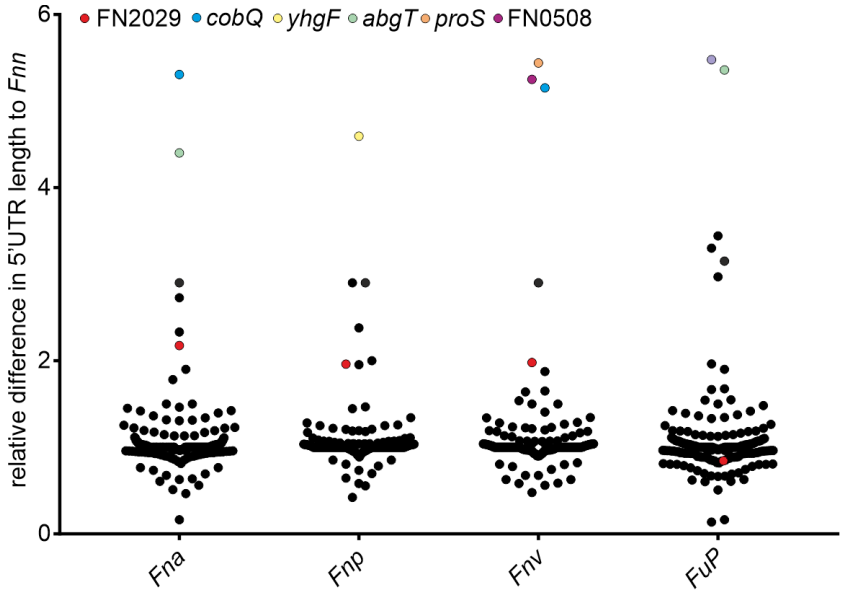

Figure 2.21 | Length comparison of shared 5'UTRs among all strains.
Distribution of the 5'UTR-length relative to *Fnn* for the 205 homologous genes.

A comparison of the 5'UTR-lengths to those of *Fnn* showed that the majority of 5'UTRs displayed a similar length across all strains (Fig. 2.21). This indicates the importance of maintaining the length of the untranslated region even in long 5'UTRs such as those found for transcripts of ribosomal proteins like the large ribosomal subunit protein L33P with an average 5'UTR length of ∼216 nt. Conservation in this example is less surprising due to the general importance of

ribosomal proteins. Yet, an interesting case is portrayed by a small cold shock domain containing protein (Csp) that contains a highly conserved ∼167 nt long and structured 5'UTR (Fig. 2.22). This is reminiscent of the RNA thermometer found upstream of *cspA* in *E. coli* where the RNA element allows increased translation of the mRNA at lower temperatures (Giuliodori et al., 2010). Notwithstanding this possible parallel, the prediction for RNA regulatory elements failed to detect such an element in the 5'UTR in the fusobacterial cold shock protein but might indicate this untranslated regions as a novel cold-sensing RNA element.

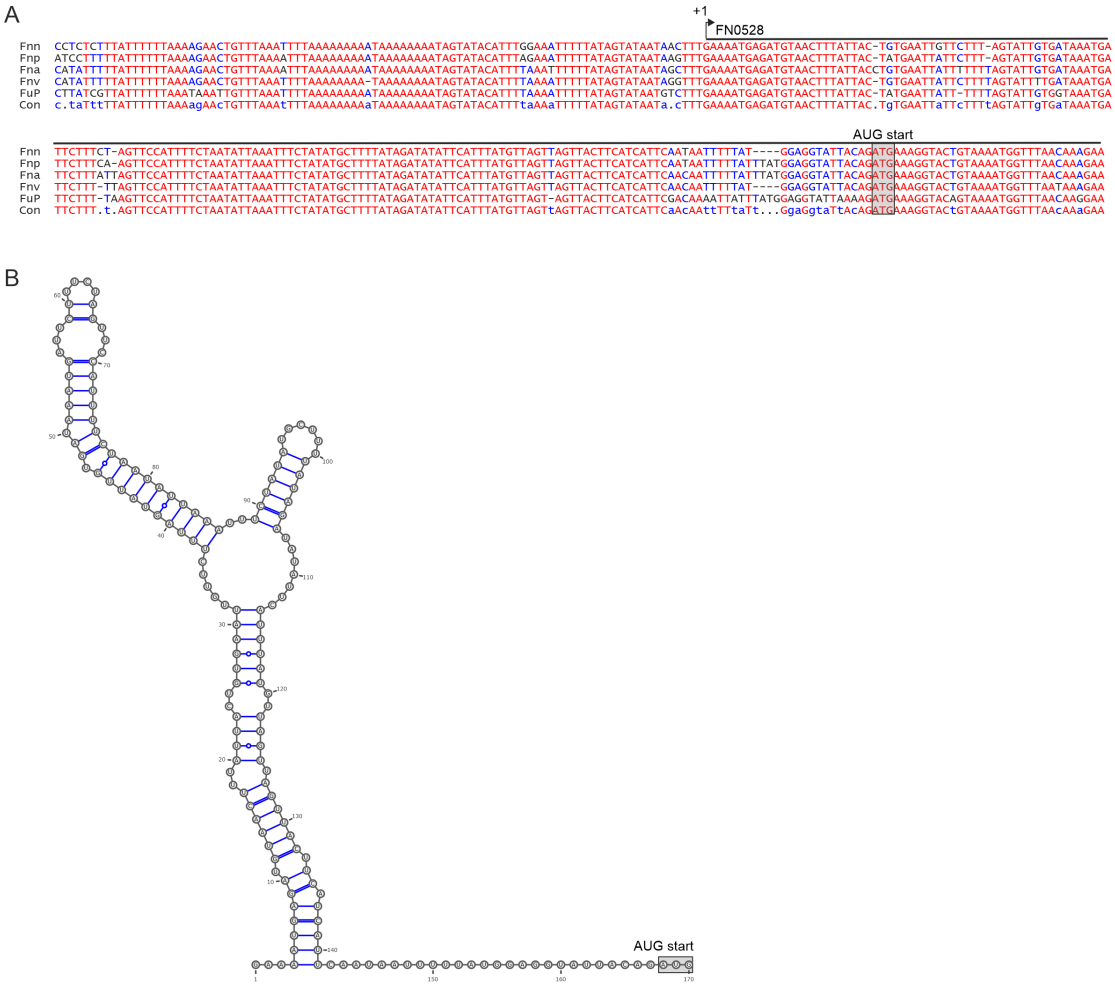

Figure 2.22 | A conserved 5'UTR structure of the cold shock family protein FN0528.
A. Sequence alignment of the FN0528 and its homologs with the conserved TSS and start codon indicated. **B.** Secondary structure prediction for the 5'UTR of *Fnn*.

The 5'UTR of FN2029, encoding a putative acetyltransferase, possesses twice as long untranslated region in all other strains but *FuP* (Fig. 2.23A). An alignment of the 5'UTR further reflects this difference as only the region around the

RBS is conserved among the subspecies. Yet, the promoter region of this gene displayed a strong conservation. Overall, this indicates that post-transcriptional regulation of FN2029 via its 5'UTR could be different between the subspecies while transcriptional control is evolutionary stably maintained.

While the function of FN2029 is unclear, the fusobacterial HutH performs the first step in histidine utilization and is highly conserved among bacteria and mammalians (Bender, 2012; Taylor et al., 1990). Thus it was unexpected to find the 5'UTR length to be shorter in *Fna* and *FuP* (Fig. 2.23B). An alignment for all sequences showed that the start of the 5'UTR is lacking conservation while the RBS and the upstream sequences stretch display high conservation (Fig. 2.23B). Interestingly, the latter resembles an imperfect inverted repeat sequence which could be a conserved binding site for a regulatory RNA or more likely for a transcriptional repressor (Browning and Busby, 2016). Supporting the latter hypothesis is the conserved location of a transcriptional repressor transcribed divergently to HutH, which represents a common regulatory pairing in gene regulatory networks (Hershberg et al., 2005).

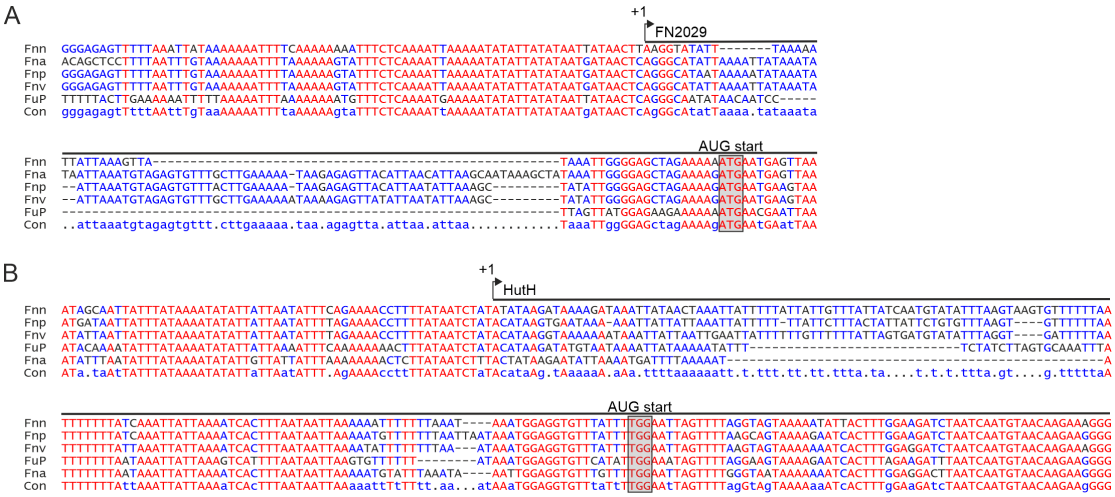

Figure 2.23 | Differences in 5'UTRs of conserved genes.
A. Sequence alignment for the genes encoding the putative O-antigen acetyltransferase FN2029. **B.** Sequence alignment for the histidine kinase HutH. TSS and translational start sites are indicated.

2.4.5 ncRNAs in the additional fusobacterial strains

The sRNA prediction for the four additional strains yielded 26 to 53 putative sRNA candidates. As for *Fnn*, all putative sRNAs were used in search for homologs among the 36 fusobacterial strains with completed genomes. This showed

that only a few additional sRNAs could be predicted that were strongly conserved among the subspecies and related *F. periodonticum* and *F. hwasookii* (Fig. 2.24), when excluding the ones already identified in *Fnn* (Fig. 2.15). Overall, this suggests that the validated sRNAs for *Fnn* already account for the majority of broadly conserved regulatory RNAs expressed under the tested conditions.

Figure 2.24 | Conservation of putative sRNA candidates predicted in the additional strains.
Heatmap depicting the conservational analysis for predicted sRNAs in related species with a complete genome. A putative sRNA was classified as conserved in a strain when it showed a ≥75% nucleotide identity and length compared to the sRNA in the indicated strain. The broadly conserved sRNA candidates FunR50, FunR51 and FunR52 are indicated in bold when they were predicted as sRNA.

A comparison between the different predicted sRNAs candidates showed that three sRNAs were predicted in at least three out of four of the additional strains and further showed a high sequence conservation among at least the subspecies (Fig. 2.24). Interestingly, FunR50 and FunR51 were both associated with genes

involved in the ethanolamine metabolism which is an abundant carbon and nitrogen source in the mammalian gut (Garsin, 2010). FunR50 was predicted to derive from an iTSS within the gene of the sensor kinase EutW as part of the EutVW two-component system (Fig. 2.25) which extents into the IGR between the gene *eutA* encoding for the ethanolamine utilization protein A. In *E. faecalis*, the ethanolamine-dependent regulator EutV acts as a transcriptional anti-terminator by binding the 3'-end between *eutW-eutA*. This 3'-region contains a terminator loop preceded by a conserved sequence motif ("AGCAANGNNGCU") also found in *C. difficile* or *L. monocytogenes* (Fox et al., 2009). In fact, FunR50 also contains this motif (Fig. 2.25) suggesting that the EutV-dependent regulation of the ethanolamine metabolism operon could be partially conserved in *Fusobacterium*.

Figure 2.25 | Conservation of a putative *eutW*-associated sRNA.
Sequence alignment for FunR50 with the indicated transcription start at the 3'end of *eutW*. The stop codon of *eutW* and start codon of *eutA* are indicated.

No such motif could be identified for the EutJ-associated FunR51 (Fig. 2.26) but it is possible that FunR51 function is related to EutJ although the a role of the DnaK-like protein is currently unknown. However, the genomic location is only conserved in *F. nucleatum* while the sRNA is located elsewhere in *F. periodonticum* not supporting the EutJ-associated role at least in the latter.

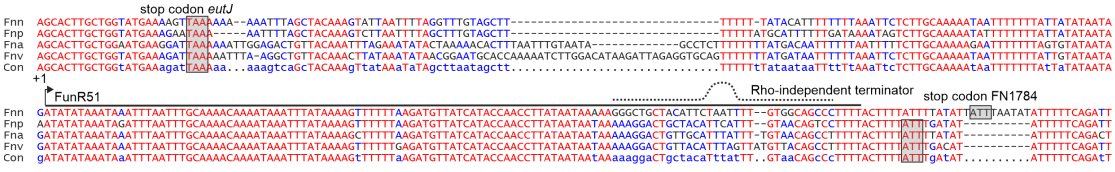

Figure 2.26 | An *eutJ*-associated sRNA candidate in *F. nucleatum*.
Sequence alignment of FunR51 transcribed downstream of *eutJ* in different subspecies of *F. nucleatum*. The stop codons of the adjacent genes are indicated.

Figure 2.27 | The conserved sRNA candidate FunR52.
Sequence alignment of FunR52 displaying the conservation between the genes encoding the ribosomal GTPase Der and the putative RBP YmlH. The respective start and stop codons are indicated.

FunR52 is conserved downstream of the gene for ribosome-associated GTPase Der and upstream of *ylmH* which encodes for a putative RNA-binding protein with a homolog of unknown function in *B. subtilis* (White and Eswara, 2021) (Fig. 2.27). A sequence alignment showed FunR52 and its promoter region to be highly conserved which indicates an important function among *F. nucleatum* and related species. This sRNA had been annotated as a 5'UTR for *ylmH* in *Fnn* but likely represents an independent transcript also in this strain when considering the conservation,a putative terminator loop at the 3'end and an alternative TSS found for the downstream gene.

3 New genetic tools for *Fusobacterium nucleatum*

Our knowledge on gene function and regulatory networks in bacteria is largely derived from model organisms such as *E. coli* or *B. subtilis*, members of the Proteobacteria and Firmicutes, respectively. The phylogenetic distance of *F. nucleatum* to these phyla thus represents a substantial hurdle for direct knowledge transfer through sequence comparisons. Additionally, functional genetics in *F. nucleatum* has been limited by the lack of genetic tools further hampering the decoding of gene function and molecular principles in this remote phylum (see 1.1.2) (Brennan and Garrett, 2019).

To overcoming these limitations, several genetic tools for *F. nucleatum* were developed as described below.

Parts of this chapter have been published before:

- Ponath, F., Tawk, C., Zhu, Y., Barquist, L., Faber, F., & Vogel, J. (2021). RNA landscape of the emerging cancer-associated microbe *Fusobacterium nucleatum*. *Nature Microbiology*, 6(8), 1007–1020. https://doi.org/10.1038/s41564-021-00927-7
- Ponath, F., Zhu, Y., Cosi, V., & Vogel, J. (2022). Expanding the genetic toolkit helps dissect a global stress response in the early-branching species *Fusobacterium nucleatum*. *Proceedings of the National Academy of Sciences of the United States of America*, 119(40), e2201460119. https://doi.org/10.1073/pnas.2201460119

3.1 Development of an *E.coli* - *F. nucleatum* shuttle vector

Plasmids that can be used for *E. coli* cannot always be directly used in other bacterial species as these might employ different mechanisms to stably propagate a vector. Yet, such vectors represent a useful foundation for genetic tools as replication in *E. coli* allows easy manipulation of DNA encoded on such plasmids.

This problem can be circumvented by developing shuttle vectors that enable the plasmid propagation in two different host species (Kreft et al., 1978).

The foundation of a shuttle vector for *F. nucleatum* was the isolation of native plasmids carrying an origin of replication identified (ORI_{FN}) that allows plasmid propagation (Bachrach et al., 2004b; Haake et al., 2000). From this, a basic *E. coli-F.nucleatum* shuttle vector was constructed by transferring the ORI_{FN} into an *E. coli* vector (Haake et al., 2000). However, this plasmid is optimized for *E. coli* and members of the Proteobacteria which differ from *F. nucleatum* in more than just their phylogenetic relationship. This becomes especially apparent when comparing the G+C-content and codon usage of *F. nucleatum* and *E. coli* (Fig. 3.1; Table 3.1).

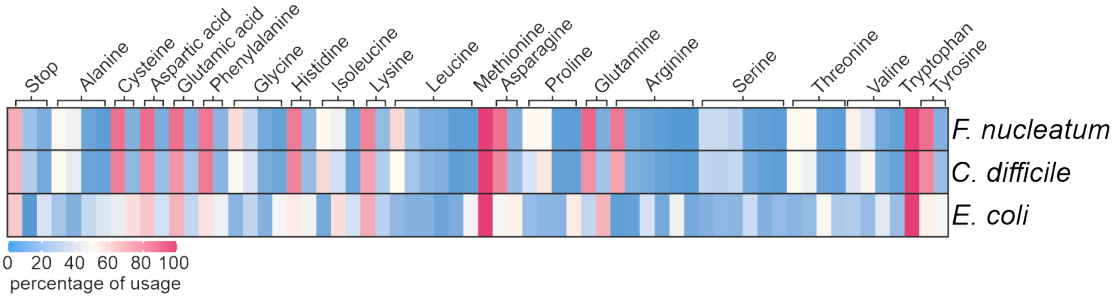

Figure 3.1 | Comparison of codon usage between *F. nucleatum*, *E. coli* and *C. difficile*.
Heatmap showing the percentage of usage for the different amino acids and stop codon according to the individual triplets. The data were derived from Nakamura et al. (2000) for the strains: *F. nucleatum* subsp. *nucleatum* ATCC 25586, *C. difficile* 630 and *E. coli* K12.

Table 3.1 | G+C content for *F. nucleatum*, *C. difficile* and *E. coli*.

F. nucleatum	*C. difficile*	*E. coli*
~27%	~28%	~51%

In contrast, *C. difficile* displays a similar G+C-content and codon usage as *F. nucleatum* (Fig. 3.1; Table 3.1). Taking advantage of these similarities, the ORI_{FN} was cloned into an *E. coli-C. difficile* shuttle vector, pRPF185 (Fagan and Fairweather, 2011) thereby replacing the origin of replication for *C. difficile*. This plasmid was successfully transformed via electroporation into *F. nucleatum* using the antibiotic thiamphenicol as a selection marker and represented a basis to generate genetic tools for *F. nucleatum*.

3.2 A gene overexpression tool

Constitutive overexpression of a gene has been successfully used to identify gene function(s) or the targets of sRNAs. For this, knowledge of a strong and constitutive promoter in the target organism is needed. The 4.5S RNA showed high expression levels (Fig. 2.11) and thus the 100 nt long promoter region of the housekeeping RNA was selected to be tested for an overexpression vector. To validate the stable overexpression of a target, the five sRNAs were cloned at the +1 position of the promoter region in the pEcoFus-backbone (Fig. 3.2A). The empty vector, containing only the promoter region of the 4.5S RNA (control) and the sRNA-expression constructs were transformed into *F. nucleatum*. Northern blot analysis for individual transformants showed an drastic expression increase for each of the sRNAs over the control (Fig. 3.2B). This result demonstrated that this vector, termed pEcoFus, facilitates overexpression of a gene of interest and can be used as a genetic tool to study gene function in *F. nucleatum*.

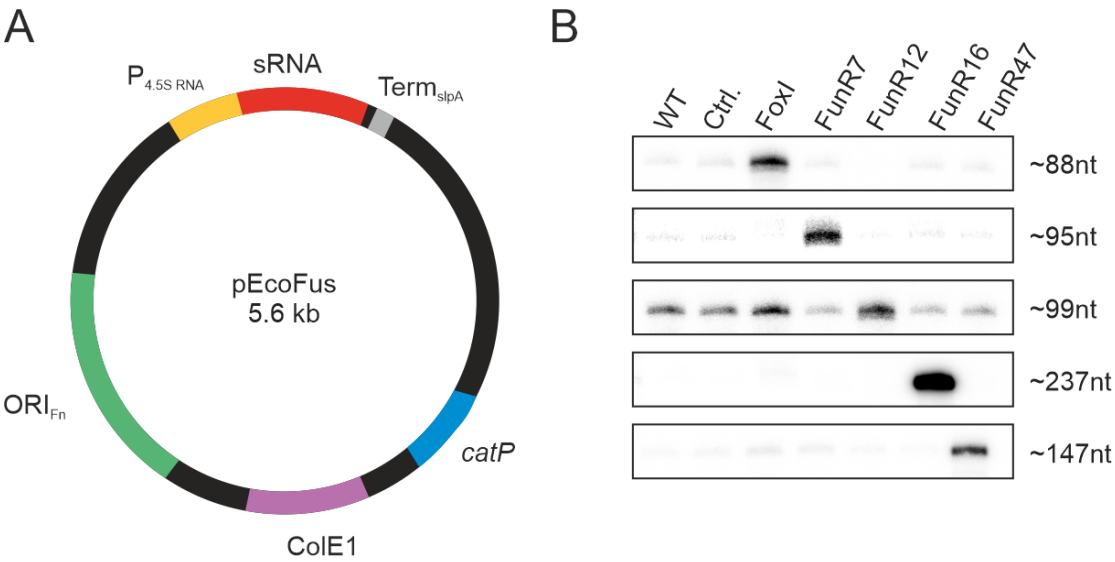

Figure 3.2 | pEcoFus - a 1^{st} generation overexpression vector.
A. Schematic overview of pEcoFus with different regions of the vector indicated. $P_{4.5SRNA}$: promoter region of the 4.5S RNA. This figure has been modified from Ponath et al. (2021).

3.3 Improving pEcoFus - Generation of pVoPo

pEcoFus was generated by using the *E. coli-C. difficile* shuttle vector as a chassis to insert the ORI_{FN} required for plasmid replication in *F. nucleatum*. Consequently, pEcoFus contained several elements such as the *traJ* or *oriT* which

unnecessarily inflate the vectors size to ~5.6 kB. Creating an improved vector, only the ORI_{FN}, the thiamphenicol resistance cassette *catP* and the ColE1 were assembled into a new construct. This process included the addition of several restriction sites for the simple insertion of DNA fragments. Further, the promoter and the 5'UTR of the catP gene was replaced with that of the fusobacterial flavodoxin-encoding *fldA* to facilitate improved expression of the selection marker in *F. nucleatum*. As with pEcoFus, this vector could be maintained in *E. coli* and *F. nucleatum*. The improved backbone, referred to as pVoPo-00, is almost half the size of pEcoFus (5.6 kB vs. 3.6 kB) and includes custom restriction sites. Thus, the resulting plasmid was used as the basis for the development of the following genetic tools (Fig. 3.3).

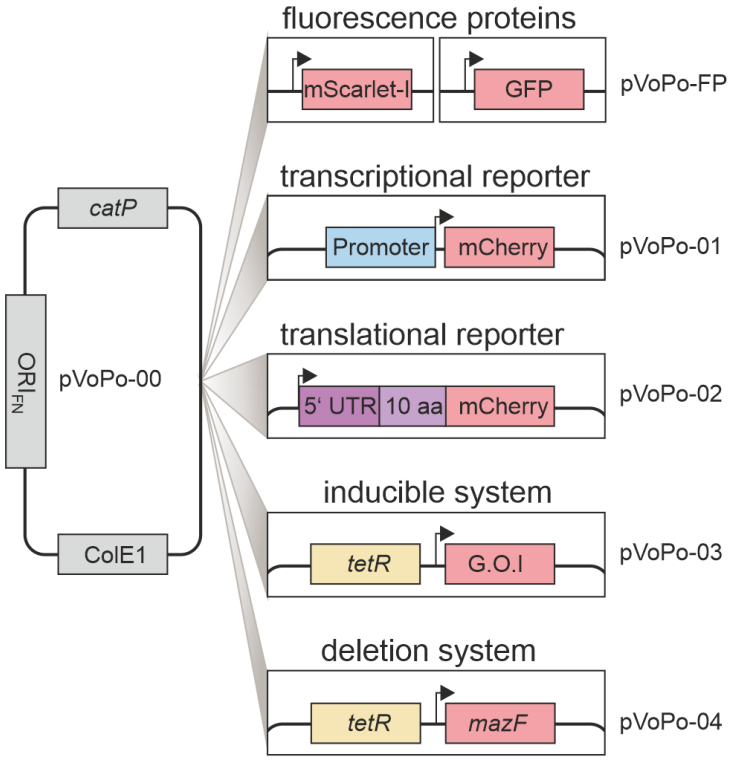

Figure 3.3 | Overview of the five generated genetic tools.
Schematic representation of pVoPo-00 and the individual insert resulting in the different genetic tools. pVoPo-FP (FP: placeholder for different fluorescent proteins); G.O.I: gene of interest. This figures has been adapted from Ponath et al. (2022b).

3.4 A new inducible vector for *F. nucleatum*

While overexpression of a target gene can help elucidate its function, its constitutive expression can also cause secondary effects including cytotoxicity (Prelich, 2012). Consequently, a number of inducible gene expression systems have been developed to limit these effects. These systems often utilize a transcription factor, or more recently riboregulators (Cui et al., 2016; Glasscock et al., 2021), that respond to an exogenous factor and subsequently modulate transcription. One commonly used tool is the LacI inducible system which uses the lac repressor (LacI) to only allow transcription upon its inhibition through the addition of lactose or the synthetic analog IPTG (Isopropyl β-d-1-thiogalactopyranoside) (Amann et al., 1988). However, the LacI system exhibits leaky expression (Rosano and Ceccarelli, 2014).

To circumvent this issue, the tetracycline-inducible TetR-system has been introduced into pVoPo-00. The TetR-system makes use of the antibiotic tetracycline or its derivatives such as the non-bacteriostatic anhydrotetracycline (ATc) as inducers that inhibit the negative regulation by the Tet-repressor protein. Furthermore, this inducible system displays no leaky expression and a strong dose-dependence allowing the fine-tuning of the gene expression levels (Bertram and Hillen, 2008; Lutz and Bujard, 1997).

For proof of concept, mCherry with a short synthetic 5'UTR (Xiao et al., 2020) was placed under control of the TetR-system and transformed into *F. nucleatum*. The addition of different concentrations of ATc to cultures of *F. nucleatum* led to a dose-dependent expression of mCherry after 30 min of induction as shown via western blot analysis (Fig. 3.4). Importantly, no signal for the protein could be detected in the absence of the inducer, which further highlights the degree of control using the TetR-system for gene expression in *F. nucleatum*. The inducible system is further referred to as pVoPo-03.

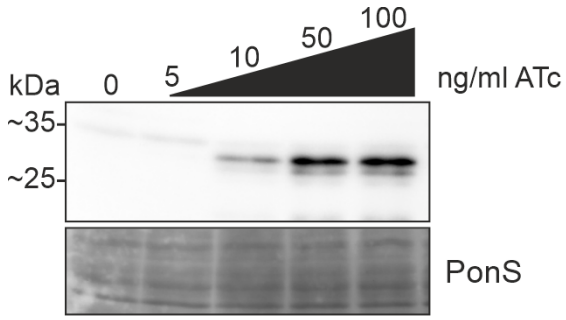

Figure 3.4 | A dose-dependent inducible system for *F. nucleatum*.
Western blot detection of mCherry after 30 min of ATc-induction at the indicated concentrations. Ponceau S (PonS) served as a loading control. This figure has been adapted from Ponath et al. (2022b).

3.5 A new deletion vector for *F.nucleatum*

In addition to the two above introduced gene expression systems, gene inactivation represents another key technique to study gene function in an organism of interest. As described in chapter 1.1.2, this has been previously achieved either by single-cross integration (Haake et al., 2000; Kaplan et al., 2005) or by using a metabolic deletion background (Wu et al., 2018; Casasanta et al., 2020) each with their respective drawbacks. This makes the ability to achieve marker-and background-less gene deletion in *F. nucleatum* highly desirable, especially for future applications in the *in vivo* settings. However, no alternative counter-selection methods to remove the antibiotic selection marker important for the initial integration were tested in *F. nucleatum*. Various methods are used for gene deletion in other bacteria such as the expression of the *sacB* gene conferring sensitivity to sucrose (Gay et al., 1985). Recently, the endonuclease MazF, as part of the toxi-antitoxin *mazEF* system, was applied to a variety of unrelated bacteria such as *B. subtilis* (Zhang et al., 2006), different lactic acid bacteria (Van Zyl et al., 2019) or *C. acetobutylicum* (Al-Hinai et al., 2012). The establishment of an inducible gene expression system for *F. nucleatum* would now allow such an application of toxin expression for counter-selection of plasmid loss.

3.5.1 Testing the toxin MazF as a counter-selection method

Given the proven function of MazF for counter-selection in different bacteria and the fact that *F. nucleatum* lacks a *mazEF* homolog, the MazF was evaluated as potential a counter-selection marker in *F. nucleatum*. The *mazF* gene from *C. difficile* (*mazF*) as chosen due to the similar codon usage between both bacteria.

First, the *mazF* gene was cloned into pVoPo-03 and transformed into *F. nucleatum*. Next, MazF activity was assessed upon induction with ATc by monitoring the growth of either *F. nucleatum* carrying the empty vector control (p.empty) or *mazF* (p.*mazF*) in the absence of antibiotics. All untreated samples and the ATc-induced control samples showed similar growth (Fig. 3.5A). In contrast, expression of *mazF* caused a strong growth delay for ∼16 h (Fig. 3.5A). Importantly, the *mazF*-expressing bacteria could not grow in this time frame when maintaining the selection pressure was maintained by including thiamphenicol in the media (Fig. 3.5B). This suggest the observed re-growth after 16 h was caused by plasmid lost rather than mutations of the toxin itself.

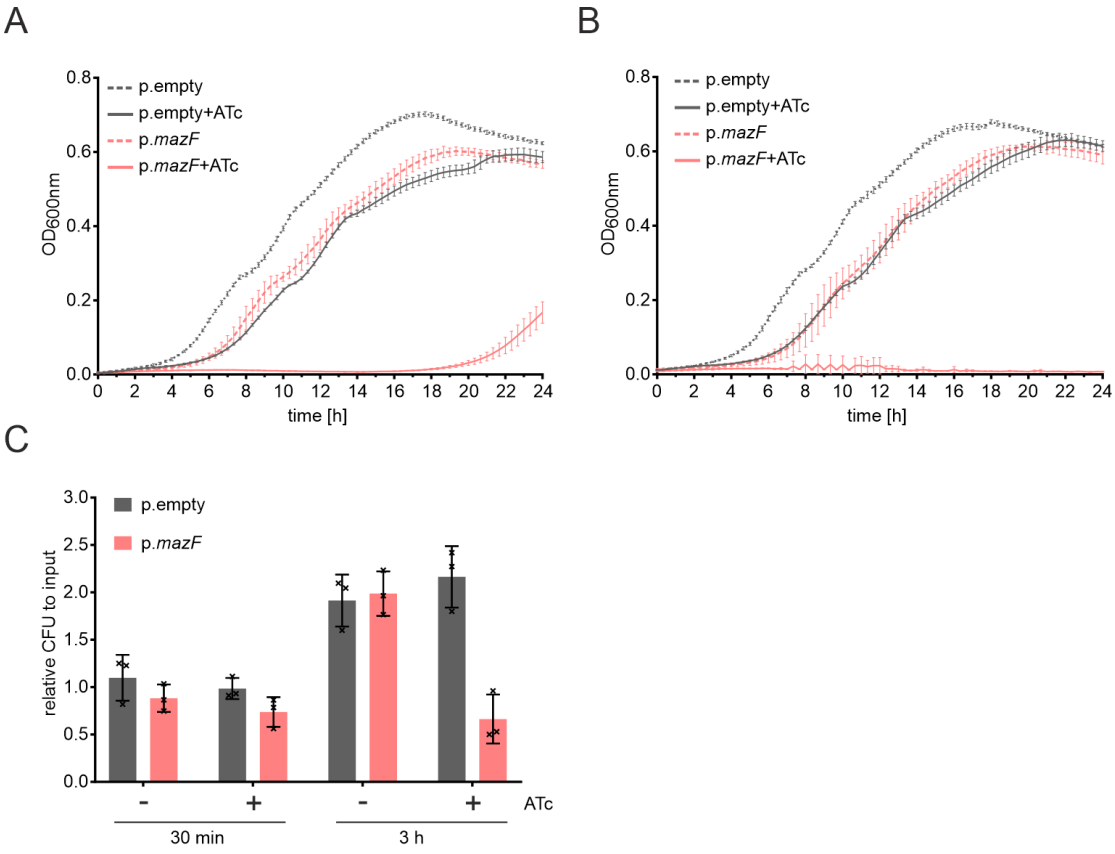

Figure 3.5 | Testing MazF expression as a counter-selection marker.
Measurement of growth over 24 h using the optical density (OD_{600nm}) as readout. MazF-expression was induced by the addition of ATc at 0 h. The experiment was carried out in the absence of thiamphenicol. The experiment was carried out in the absence (**A**) or presence (**B**) of thiamphenicol to ensure plasmid maintenance. An empty vector control (p.empty) served as control in all cases and was treated the same as p.*mazF*. **C.** Quantification of CFUs 30 min and 3 h post-induction of MazF (p.*mazF*) with ATc compared to the input count. An ATc concentration of 100 ng/ml was used in all panels. The data show the average of three biological replicates with the SD. This figure has been adapted from Ponath et al. (2022b)

Moreover, the quantification of viable cells after *mazF* induction revealed a reduction of ∼30 to 40% of viable cells after 30 min and 3 h, respectively, while

the control increased its cell number by ∼two-fold after 3 h (Fig. 3.5C).

Hence, the combined results indicate that the inducible expression of the MazF toxin is a promising form of counter-selection to successfully select double-crossover events during homologous recombination in *F. nucleatum*.

3.5.2 Proof-of-concept use of the inducible *mazF* gene deletion system

In order to validate the *mazF*-system as a means for successful counter-selection a fusobacterial suicide vector was created by removing the ORI_{FN} of p.*mazF* resulting in pVoPo-04 (Fig. 3.6). The gene *fadA* was selected as a test candidate since previous works have shown it to be non-essential (Casasanta et al., 2020; Han et al., 2005) and to be monocistronic (see Chapter 2.2; Fig. 2.7).

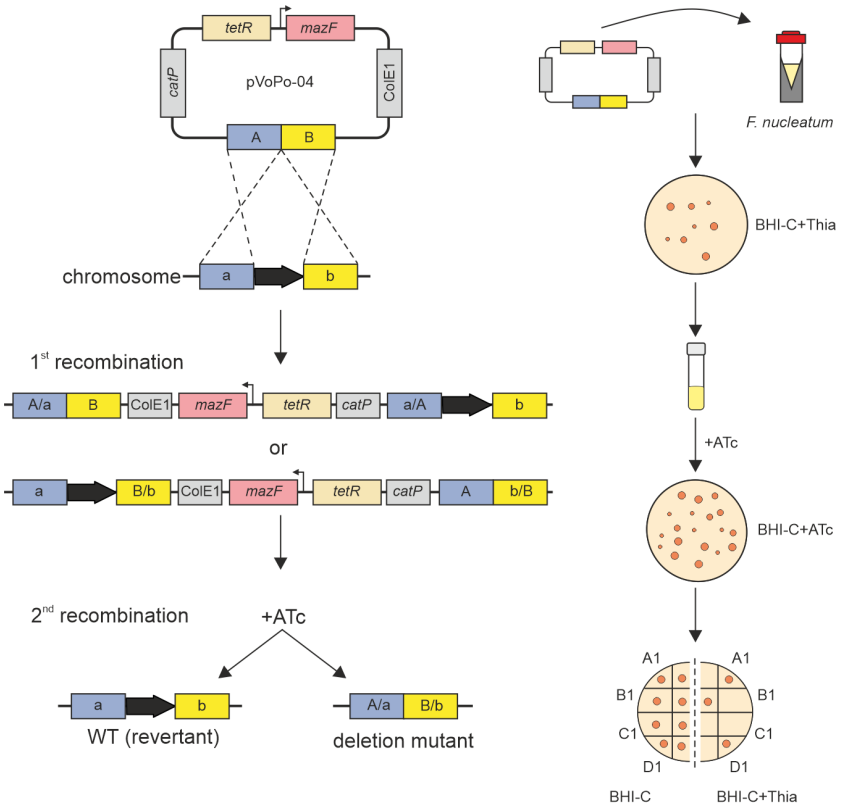

Figure 3.6 | Overview of scarless-deletion using pVoPo-04.
Schematic representation of allelic exchange (left side) and experimental workflow (right side) using the pVoPo-04 system to generate unmarked deletion strains. Gene specific 1 kB up- and downstream regions are inserted into pVoPo-04. This construct is transformed into *F. nucleatum* via electroporation and plated on BHI-C plates containing thiamphenicol (Thia). Colonies that grow successfully have integrated the suicide vector into the genome (marked as 1^{st} recombination). The 2^{nd} recombination step is induced by inoculating a single colony from the previous step. After overnight growth, the culture is diluted in fresh media containing ATc, grown for 4 h and serial dilutions are plated on BHI-C plates containing ATc. Subsequently, colonies are re-streaked on BHI-C plates with or without thiamphenicol. Colonies that only grow on the plates lacking antibiotic are further selected to verify if the 2^{nd} recombination led to a revertant or the desired deletion via PCR. This figure has been adapted from Ponath et al. (2022b).

To achieve deletion of *fadA*, ~1 kB long fragments corresponding to up- and downstream regions of the adhesin-encoding gene were amplified and placed into pVoPo-04 (Fig. 3.6). Five micrograms of the suicide vector were transformed into electro-competent *F. nucleatum* and plated on antibiotic-containing agar plates. Only bacteria that chromosomally integrated the plasmid (1st recombination) were able to grow. Next, a single colony was inoculated in media without antibiotic selection pressure to allow the second recombination step to take place. After 24 h, the culture was diluted 1:50 and the MazF expression induced through the addition of ATc. Serial dilutions were plated 4 h post-induction on agar plates containing ATc to further repress growth of bacteria still harboring the plasmid. To verify

that colonies grown on the ATc-containing plasmids indeed had excised the plasmid and thus lost the antibiotic resistance cassette, the colonies were re-streaked on agar plates with and without antibiotic. Only colonies that were unable to grow on the antibiotic-containing plates were selected for PCR verification of the loss of the *fadA* gene. The deletion of *fadA* could further be verified via northern blot analysis probing for the mRNA (Fig. 3.7). Thus, the successful deletion of the adhesin *fadA* using the inducible *mazF* system validates the approach of using this toxin as a method for counter-selection in *F. nucleatum*. Importantly, this system enables the markerless gene deletion in a wild-type fusobacterial strain.

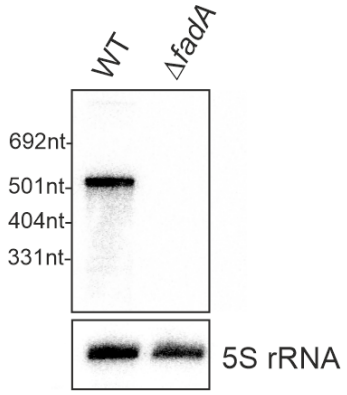

Figure 3.7 | Verification of Δ*fadA*.
Northern blot detecting the *fadA* mRNA in wild-type (WT) and Δ*fadA* strain. The 5S rRNA served as loading control. This figure has been adapted from Ponath et al. (2022b).

3.6 Reporter systems

3.6.1 Transcriptional reporter systems

Transcriptional reporter systems have been developed to easily measure the transcriptional regulation of a selected target during different conditions or in a genetic backgrounds and can also be applied to high-throughput approaches (Belliveau et al., 2018; Kinney et al., 2010). Generally, a reporter such as a fluorescent protein is placed under the control of the promoter of interest and the level of transcriptional activity can be assessed by measuring the emitted fluorescence.

To adapt this system for the application in *F. nucleatum*, a codon-optimized mScarlet-I was placed in pVoPo-00 to function as reporter. By inserting a promoter region of interest, the resulting expression of mScarlet-I under different conditions can be detected. Due to the requirement of molecular oxygen for correct

folding of the common fluorescent proteins, direct measurement of the fluorescence signal during growth cannot be readily achieved. Thus, the cells were first fixed using paraformaldehyde (PFA), stopping all cellular activity, and kept overnight at 4°C to allow maturation of mScarlet-I prior to measuring the fluorescence intensity.

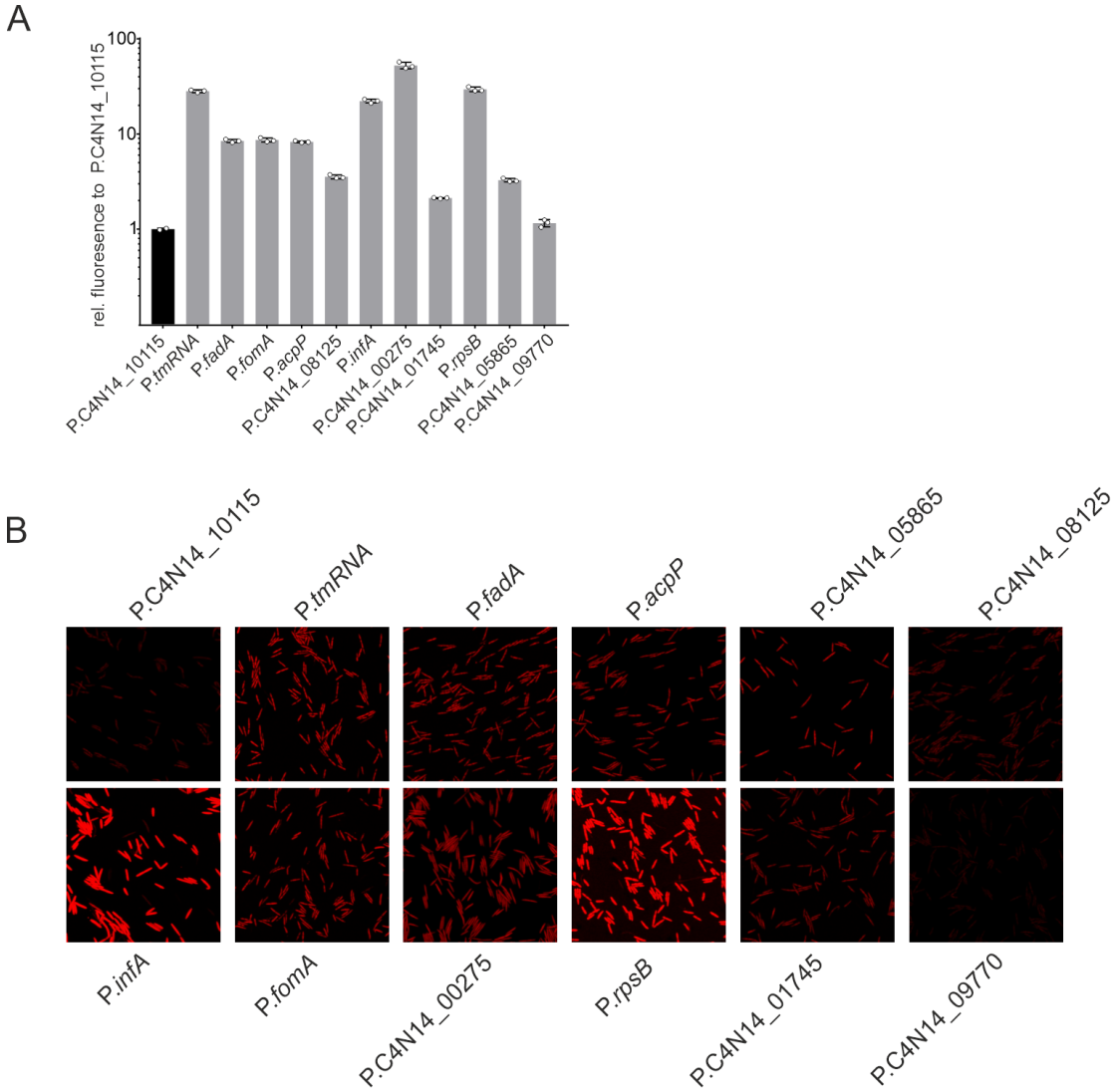

Figure 3.8 | Analysis of different promoters driving mScarlet-I expression. **A.** Quantification via flow cytometry for different promoter regions driving expression of mScarlet-I. The data show the relative fluorescence intensity and SD to the P.*accD* for three biological replicates. 50.000 DAPI$^+$ cells were measured. **B.** Representative microscopy images of each construct captured with the same setting laser intensity based on the signal from the p.*accD* construct.

As a test, 14 promoters showing high to medium expression in the dRNA-seq data were selected to construct transcriptional fusions. Next, the individual transcriptional reporters were measured via flow cytometry for samples derived

from the mid-exponential phase. For all reporters, mScarlet-I expression lead to a strong increase of fluorescence signal compared to the control vector using the promoter region of *accD*. (Fig. 3.8A). The promoter region for ribosomal proteins *rspB* (P.*rspB*) showed the strongest signal together with promoter driving transcription of the translational initiation factor IF1 (P.*infA*). The constructs carrying the promoters for the tmRNA and the abundant porin FomA also showed a $\geq$10 fold signal compared to P.*accD* inline with high expression levels detected in the RNA-seq data (Supplementary data 8.2). Importantly, the transcriptional reporter for a putative alanine racemase (C4N14_09770) displayed similar fluorescence levels to P.*accD* while the reporter for e.g. *fadA* displayed in between levels. The could further be corroborated by confocal microscopy (Fig. 3.8B). Overall, the different transcriptional fusions reflect the general abundance of their transcripts and thus the system can be used as transcriptional reporter in *F. nucleatum*.

To generate an broadly usable transcriptional reporter system, mScarlet-I was replaced by mCherry which also allows detection by western blot. The final vector pVoPo-01 further carries a cloning site for inserting a promoter of interest upstream of *mCherry*.

3.6.2 Translational reporter

Translational reporter systems are a fundamental tool to investigate the post-transcriptional regulation of a gene of interest. These systems take advantage of creating a translational fusion of a reporter gene and a gene of interest, using the 5'UTR and an additional 30 to 150 bp of the target gene. Such a system using GFP as the reporter has been originally used to study the post-transcriptional regulation of sRNAs and its targets originally in *Salmonella typhimurium* (Urban and Vogel, 2007) and has been adapted for use in additional organisms since (Ferrara et al., 2021; Peschek et al., 2019; Klähn et al., 2015).

To establish this system in *F. nucleatum*, the dRNA-seq data were searched for genes that are constitutively expressed during growth at a medium level compared to the expression levels of all other CDS. The metabolic gene *accD* (acetyl-coenzyme A carboxylase) fit both criteria and its promoter region was placed upstream of mCherry in the pVoPo-01 backbone to drive stable expression of the reporter gene. Translational fusion for a gene of interest can then be created by cloning the 5'-region in-frame with *mCherry* using added restrictions site at the

+1 site of the promoter and after the start codon of *mCherry*. The successful establishment of the translational fusions will be shown in chapter 4.

3.6.3 Visualizing *F. nucleatum* using different fluorescent proteins.

Previously, the detection of *F. nucleatum* relied on classical approaches such as (qRT-)PCR, culturing or 16S rRNA sequencing, all of which are labor-intensive and do not allow temporal or spatial analysis of biological processes. Alternatively, one can temporarily monitor *F. nucleatum* by labeling them with membrane dyes for short term experiments, as the dye signal diminishes with each bacterial division (Casasanta et al., 2020). However, fluorescently tagged bacteria represent a powerful tool for monitoring cell numbers, infection processes or their localization in infected tissues.

Since both mCherry and mScarlet-I could be stably expressed in *F. nucleatum* additional codon-optimized genes encoding for superfolder GFP (GFP) or mNeonGreen (mNG) were generated. The genes encoding for all four fluorescent proteins were introduced into the pVoPo-backbone. Further, the promoter region of the *acpP* gene and the short synthetic 5'UTR were combined to drive the expression of the different fluorescent proteins based on the strong and stable expression observed for mScarlet-I (see Chapter 3.6.1).

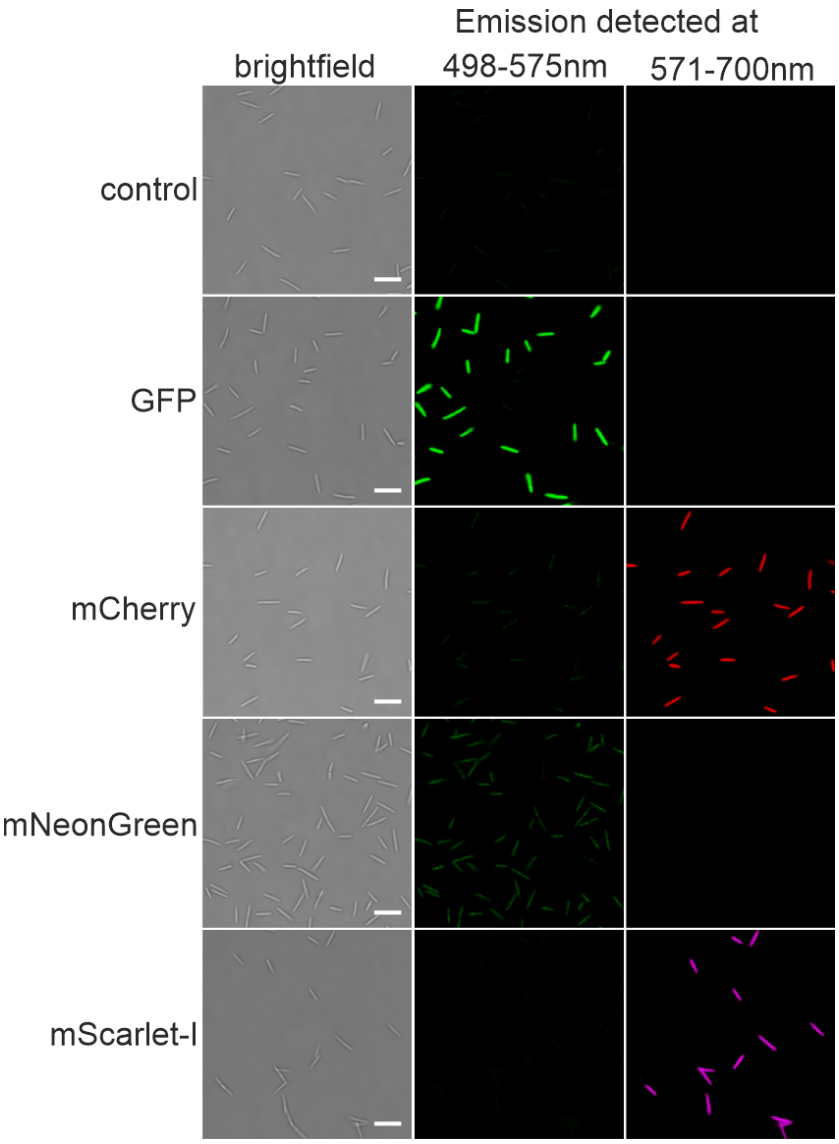

Figure 3.9 | Expression of different fluorescent proteins in *F. nucleatum*. Representative confocal microscopy images for the expression of superfolder GFP (GFP), mCherry, mScarlet-I and mNeonGreen in *F. nucleatum*. The scale corresponds to 5 µm.

Of note, no such construct could be generated for mNeonGreen as the expression levels were likely toxic for the intermediate *E. coli* host. Instead, the promoter of the metabolic *accD* gene was placed upstream of mNeonGreen which allowed the succesful construction of the vector in intermediary host *E. coli*. These vectors were named pVoPo-FP with FP abbreviating for the different fluorescent proteins.

All constructs displayed clear levels of uniformly distributed fluorescence detected at their respective emission after PFA-fixation and maturation (Fig. 3.9) as mentioned above. Importantly, in all cases the fluorescence signal is distinguishable from the empty vector control. This includes bacteria expressing the

alternative mNeonGreen construct, which yields a comparatively lower fluorescence signal than e.g. GFP.

Thus, the pVoPo-FP vectors represent another important research tool especially in the investigation of *F. nucleatum*'s interaction with the host.

4 Investigation of the σ^E regulon in *Fusobacterium nucleatum*

Parts of this chapter have been published before:

- Ponath, F., Tawk, C., Zhu, Y., Barquist, L., Faber, F., & Vogel, J. (2021). RNA landscape of the emerging cancer-associated microbe *Fusobacterium nucleatum*. *Nature Microbiology*, 6(8), 1007–1020. https://doi.org/10.1038/s41564-021-00927-7

- Ponath, F., Zhu, Y., Cosi, V., & Vogel, J. (2022). Expanding the genetic toolkit helps dissect a global stress response in the early-branching species *Fusobacterium nucleatum*. *Proceedings of the National Academy of Sciences of the United States of America*, 119(40), e2201460119. https://doi.org/10.1073/pnas.2201460119

Further, parts of the work in this chapter have been performed in collaboration with the following people:

- Yan Zhu (Institute for Molecular Infection Biology, University Würzburg) generated the inducible *rpoE* construct and performed the growth curve of the FoxI and FoxI-3C overexpression.

- Yan Zhu (Institute for Molecular Infection Biology, University Würzburg) and Valentina Cosi (Helmholtz Institute for RNA-based Infection Research Würzburg, Helmholtz Centre for Infection Research Braunschweig) supported the qRT-PCR measurements.

4.1 A putative σ^E homolog in *F. nucleatum*

As outlined in Chapter 1, bacteria need to adapt to continuously changing environments; especially when encountering adverse conditions. This also holds true for *F. nucleatum* when translocating to extra-oral sites, such as the CRC tissue, which is vastly different from its primary oral niche when considering the

state of inflammation and different immune cell populations (Hiam-Galvez et al., 2021).

However, our knowledge on how *F. nucleatum* responds to these changes is severely limited. Only recently the role of the two TCS was identified in lysine metabolism and co-aggregation (Wu et al., 2021) or response to oxidative stress (Scheible et al., 2022). Yet, we are lacking knowledge on potential global stress responses in this bacterium including the one that supports the bacterial envelope as the first-line of defense in Gram-negative bacteria (Rowley et al., 2006).

Interestingly, *F. nucleatum* encodes a conserved ECF (Fig. 4.1A; Fig. 2.5A) that could enable this anaerobe bacterium to monitor its environment (see Chapter 1.3.1). Further, this ECF could be involved in maintaining membrane homeostasis as this is a common role for these alternative sigma factors (Helmann, 2002).

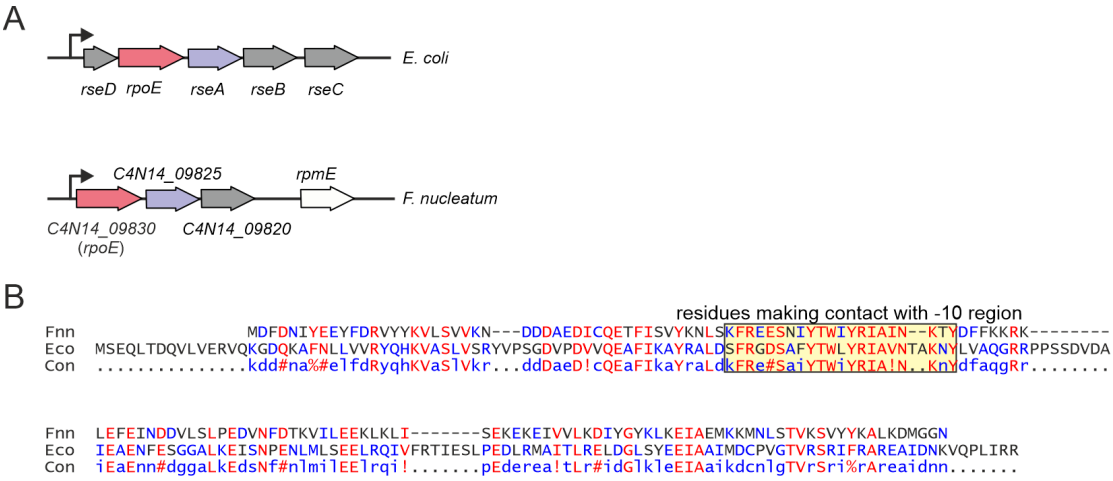

Figure 4.1 | Comparing *rpoE* between *E. coli* and *F. nucleatum*.
A. Schematic representation of the *rpoE* operon structure in *E. coli* and *F. nucleatum*. *rpoE* genes and the anti-sigma factor genes are indicated in red and purple, respectively, while the remaining genes of the operon are marked in grey. **B.** Sequence alignment between σ^E of *F. nucleatum* (Fnn) and *E. coli* (Eco). The region making contact with the -10 box is indicated. Panel **A** figures has been adapted from Ponath et al. (2022b).

The current genome annotations of *F. nucleatum* mark this gene as *rpoE* suggesting similarities to the well-studied envelope stress response regulator σ^E in *E. coli* (Erickson and Gross, 1989; Hews et al., 2019). Yet, a recent analysis of ECFs placed the fusobacterial σ^E in an ECF group of unknown function, which only includes proteins found in members of Clostridia or Fusobacteria (Casas-Pastor et al., 2021). A sequence comparison of the σ^E proteins from *F. nucleatum* and *E. coli* did not reveal any similarities beyond conserved amino acids important for contacting the -10 promoter region (Fig. 4.1B) (Todor et al., 2020; Campagne

et al., 2014). The downstream gene of the fusobacterial σ^E is further annotated as an anti-sigma factor resembling the operon structure in *E. coli* (Fig. 4.1A). Similar to findings of the ECF protein, this protein does not exhibit sequence conservation with its counterpart RseA in *E. coli*. However, both proteins share a transmembrane domain indicating that σ^E in *F. nucleatum* might be regulated by its downstream gene product similarly as in *E. coli*.

4.2 The σ^E regulon

4.2.1 Transcriptional changes upon σ^E-pulse expression

The above stated similarities suggest that the fusobacterial gene C4N14_09830, further referred to as *rpoE*, encodes a transcriptional regulator functionally similar to σ^E in *E. coli*. To investigate this possibility, the *rpoE* of *F. nucleatum* was placed into the inducible expression system pVoPo-03. *rpoE* expression was induced for 30 min after which samples for RNA extraction were collected. Next, the extracted RNA was used as input for cDNA library preparation and subsequent RNA-seq analysis. Compared to the equally treated empty vector control, a global gene expression analysis identified a total of 170 differentially expressed genes (-1 $\leq \log_2$ fold change ($\log_2$FC) ≥ 1; false-discovery-rate (FDR) ≤ 0.05) upon the σ^E expression (Fig. 4.2A). The expression analysis clearly marks the ECF as a transcriptional activator as the majority of differentially expressed genes was upregulated (147/170) while only 23 were downregulated (Fig. 4.2A; Supplementary data 8.9).

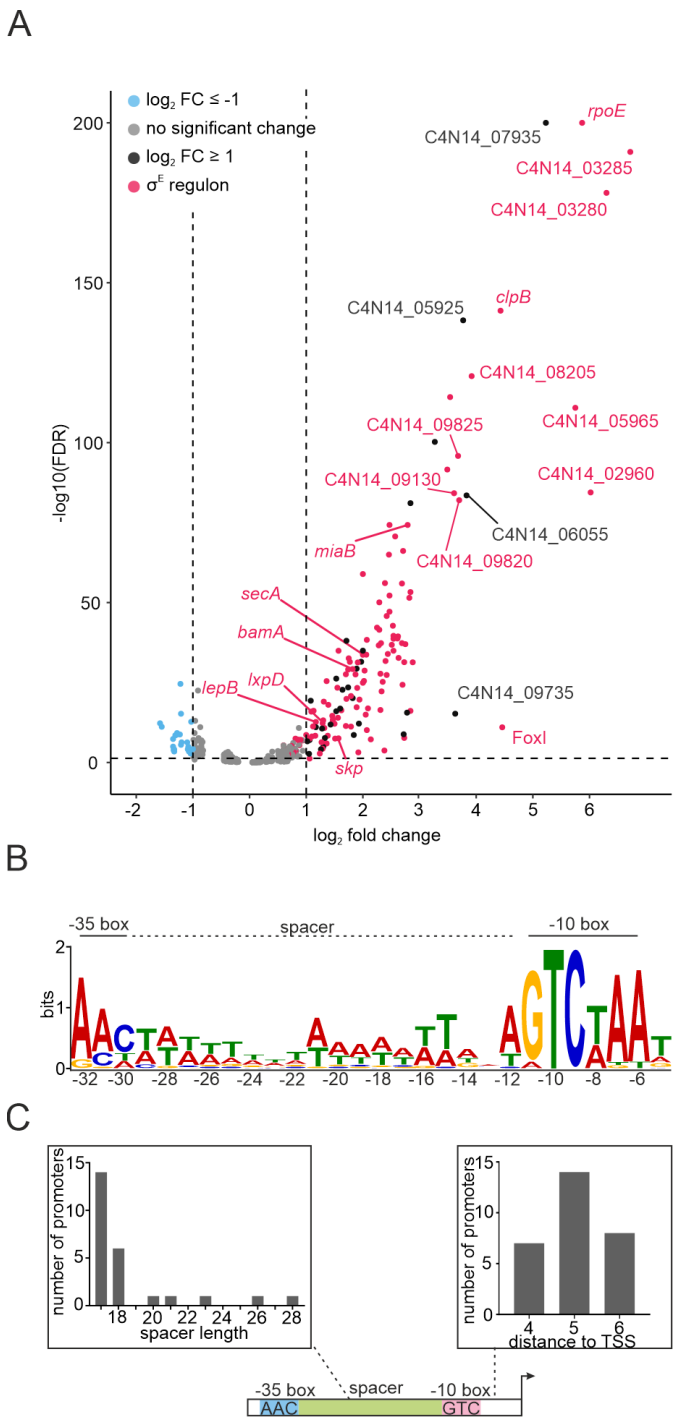

Figure 4.2 | Transcriptional changes upon σ^E pulse-expression.
A. Volcano plot displaying transcriptional changes after 30 min of σ^E expression upon ATc induction (100 ng/ml) compared to the similarly treated empty vector control. Genes were considered significantly regulated when they displayed a false-discovery rate (FDR) $\leq$0.05 (vertical dashed line) and a $\log_2$ fold change of $\leq$-1 (blue) or $\geq$1 (black) (horizontal dashed line). Genes that are part of the σ^E regulon are marked in red. **B.** Result of the MEME motif analysis for all upregulated transcriptional units. The common σ^E motif in the -10 and -35 box is indicated. **B.** Distance of the -10 box from the TSS and -35 box found for the σ^E motif among all cistrons. This figure was modified from Ponath et al. (2022b).

4.2.2 Promoter analysis of differentially expressed genes

The 170 differentially expressed genes consist of a total of 68 transcriptional units. To identify the a potential shared promoter motif, the 50 nt upstream region for all transcriptional units was used as input for a MEME analysis (Bailey et al., 2015). This analysis revealed a common motif with a 'GTCWAA' in the -10 box separated by an A+T-rich spacer region from a less distinct 'AAC' as the -35 box for 28 operons (Fig. 4.2B). This motif showed a striking similarity to the consensus sequence bound by recognized by σ^E in *E. coli* (Rhodius and Mutalik, 2010) or the homolog AlgU in *Pseudomonas aeruginosa* (Schulz et al., 2015). Additionally, the spacing of the -10 region from the TSS and -35 box is consistent with that found in *E. coli* (Fig. 4.2C) (Rhodius et al., 2006).

The 28 operons consist out of 127 genes and account for 116/147 of the up-regulated genes. Additionally, transcription for 14 of the motif-associated operons were initiated as a sub-operon (Supplementary data 8.9).

Importantly, the motif was also found upstream of the *rpoE* operon, suggesting that *rpoE* is likely subject to auto-regulatory as common among ECF (Todor et al., 2020; Rouvière et al., 1995). The proposed role of the transcriptional activator is further supported by the fact that this motif is found exclusively upstream of transcriptional units that are upregulated upon σ^E expression (Fig. 4.2A).

4.2.3 The σ^E regulon

4.2.3.1 Validation of σ^E target genes

The above introduced transcriptional reporter system (see Chapter 3.6.1) was used as an orthogonal method to validate a subset of the putative σ^E targets. For this, four transcriptional reporters were constructed including *rpoE* itself to confirm the proposed auto-regulation. The conserved cytosine in the -10 box of the promoter is required for recognition by σ^E in *E. coli* (Campagne et al., 2014). Thus, an alternative version for each reporter construct was generated harboring a C-to-G mutation (Fig. 4.3A). Western blot analysis showed that all native transcriptional reporters led to expression of mCherry (Fig. 4.3B). However, insertion of the point mutation in the promoter sequence completely abrogated the expression of the reporter. Overall, these results indicate that the tested genes and others associated with the identified promoter motif depend on σ^E for their transcription.

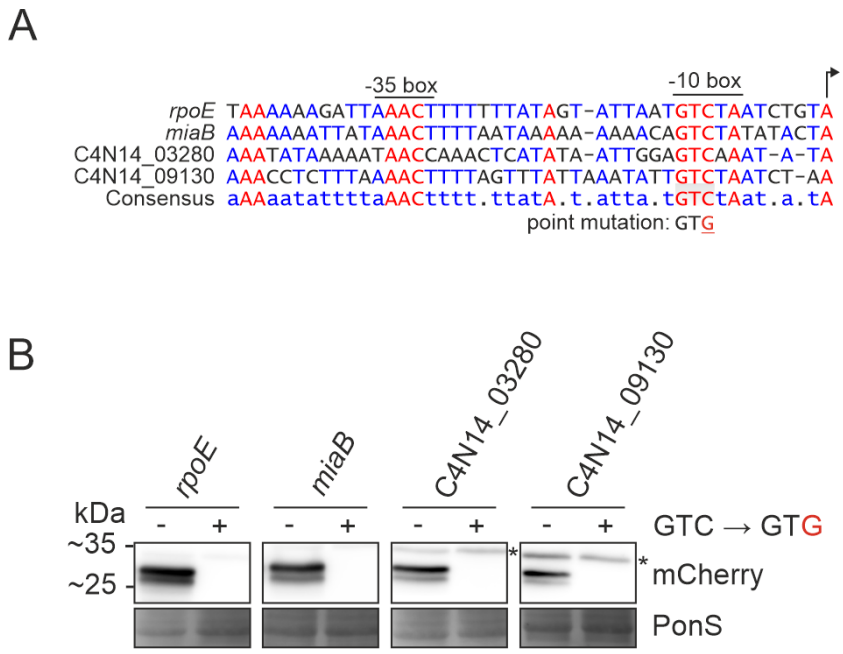

Figure 4.3 | Verification of σ^E-dependence using transcriptional reporters. **A.** Sequence alignment of the promoter region for different σ^E-dependent genes. The conserved σ^E recognition motif is indicated. **B.** Western blot analysis detecting mCherry. Transcription of the protein is driven by the WT or mutant promoter region of the indicated genes. Ponceau S served as a loading control. An asterisk marks an unspecific band. This figure has been modified from Ponath et al. (2022b).

4.2.3.2 A conserved role in envelope homeostasis for the σ^E regulon

The regulon of σ^E is commonly involved in maintaining envelope homeostasis (Rowley et al., 2006; Hews et al., 2019). Interestingly, several genes of the fusobacterial σ^E regulon seem to fulfill a similar function. Such is the case for the chaperone Skp and OMP BamA which cooperatively ensure the proper insertion of unfolded OMPs into the outer membrane (Schäfer et al., 1999; Knowles et al., 2009). Similarly, both LpxD and LptB are important for the biosynthesis and transport of LPS, respectively, which represent another vital component of the envelope of Gram-negative bacteria (Kelly et al., 1993; Sperandeo et al., 2007). Beyond functional parallels, all four genes are also controlled by σ^E in *E. coli* (Anderson et al., 1985; Dartigalongue et al., 2001; Knowles et al., 2009; Rhodius et al., 2006; Sklar et al., 2007) suggesting a conserved core regulon across phyla.

Beyond the commonalities with the σ^E regulon of *E. coli*, three genes are shared between the regulon of σ^E in *F. nucleatum* and that of the AlgU in *P. aeruginosa*: the stress chaperone *clpB*, the signal recognition particle receptor *ftsY*, and the rod-shape determining factor *mreB* (Schulz et al., 2015). Upregu-

lation of *mreB* might help *F. nucleatum* maintain its shape under envelope stress (Errington, 2015). Increased expression of ClpB might suggest a broader role as it is commonly part of a general stress response but can also be involved in the regulation of metabolism and growth through its chaperone activity (Alam et al., 2021).

The σ^E-dependent regulation of *ftsY* is especially interesting beyond its shared regulation in *P. aeruginosa*. The IM protein FtsY is part of the signal recognition particle (SRP) complex and a member of the SEC-dependent protein translocation pathway (Tsirigotaki et al., 2017). Additionally, σ^E also upregulated *secA* and *lepB* expression that are further members of the translocation pathway. With regards to this process, FtsY and SecA aid in the co- or post-translational translocation of proteins while the signal peptidase LepB releases the translocated proteins into the periplasm (Tsirigotaki et al., 2017). The increased translocation capacity could thus act in synergy with enhanced OMP insertions as the SEC-system is the major transport pathway for OMPs across the IM (Leyton et al., 2012; Tsirigotaki et al., 2017).

An additional 24 of the σ^E-induced genes are of unknown function. These genes include also the dicistronic operon C4N14_03280-C4N14_03285 which was among the most highly upregulated genes (Fig. 4.2A). This strong regulation of the operon suggests that both genes could play an important role in maintaining envelope homeostasis or protein translocation. However, both genes and the remaining 22 genes of unknown function might yet fulfill entirely new functions in the σ^E-dependent stress response in *F. nucleatum*.

4.2.3.3 Conditions that might activate σ^E

Past works in different bacteria showed that σ^E responds to various stimuli such as osmotic stress, heat shock, unfolded OMPs, oxidative stress or singlet oxygen (Erickson and Gross, 1989; Mecsas et al., 1993; Testerman et al., 2002; Rowley et al., 2006; Anthony et al., 2005). The identification of activation conditions for σ^E in *F. nucleatum* might give more insights into the biological role of the ECF. To uncover this stimuli, expression levels of the *rpoE* mRNA were monitored via qRT-PCR after subjecting the bacterium to different stress conditions consisting of envelope stress (polymyxin B; lysozyme, bile), hyperosmotic shock (NaCl), oxidative stress (H_2O_2; diamide; S-nitroglutathione (GNSO)), DNA damage (mit-

omycin C) heat shock (42°C) or exposure to atmospheric oxygen concentrations.

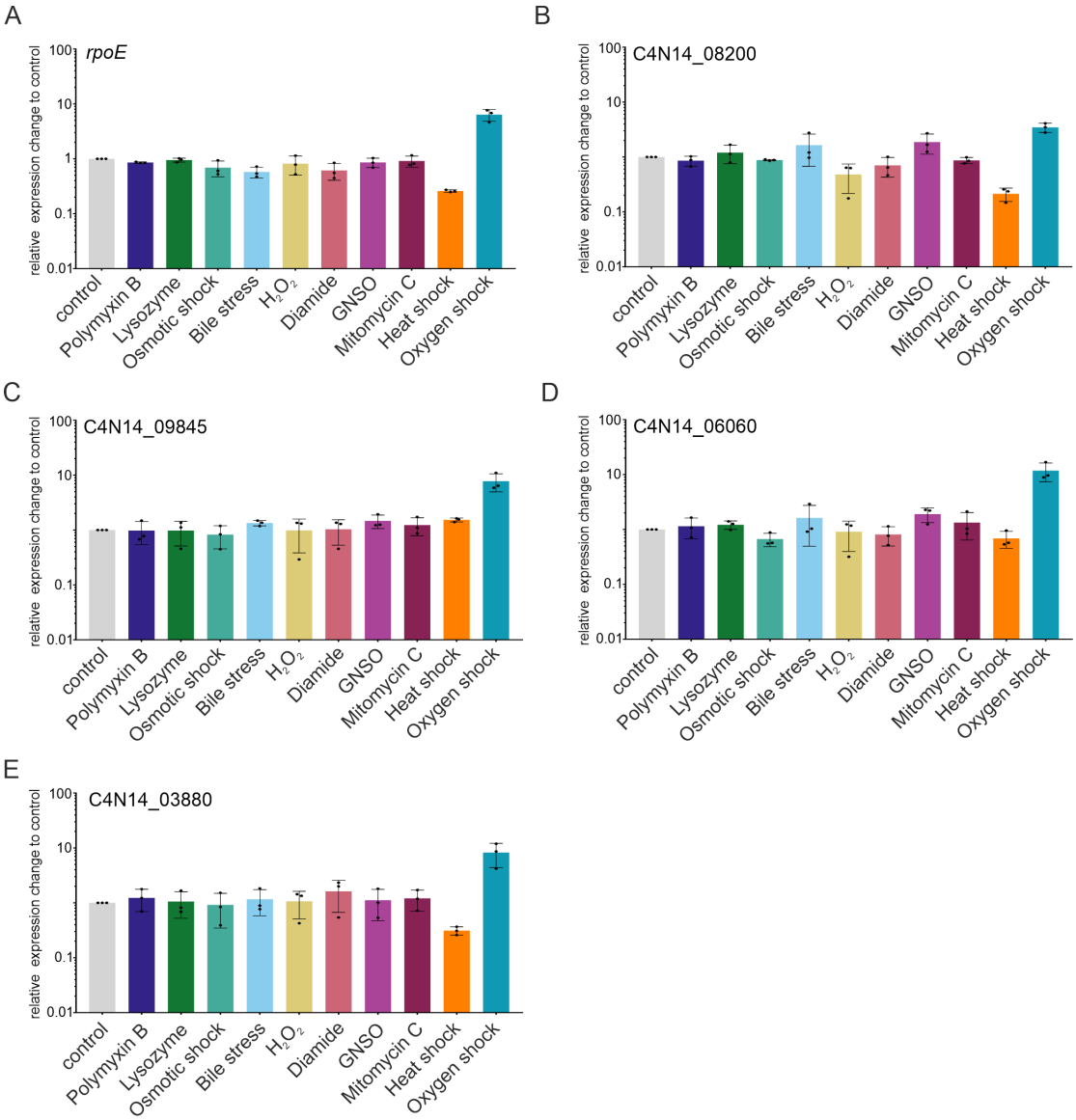

Figure 4.4 | Oxygen exposure leads to upregulation of the σ^E regulon.
qRT-PCR analysis of *rpoE* (**A**) and indicated members of the σ^E regulon (**B-E**) after 60 min of exposure to different stress conditions. Shown is the relative change of gene expression compared to the untreated control. The 5S rRNA served as a reference gene. The data represent the average of three biological replicates and the SD. This figure has been modified from Ponath et al. (2022b).

Surprisingly, the envelope disrupting antibiotic polymyxin B did not trigger σ^E activation (Fig. 4.4). This is despite the activity of the antimicrobial against Gram-negative bacteria such as *F. nucleatum* (Nang et al., 2021). Instead, the anaerobe bacterium displayed induction of the *rpoE* mRNA only when exposed to oxygen (Fig. 4.4A). This oxygen-dependent effect was further validated by

measuring the RNA-levels of four additional σ^E-dependent genes which showed a similar increase after oxygen-treatment as *rpoE* compared to the control (Fig. 4.4B-E).

4.2.3.4 The global response to oxygen

Oxygen is not a known activator of σ^E and a link to the possible role of this ECF in envelope homeostasis is unclear. To better understand the global response of *F. nucleatum* to oxygen, the bacterium was exposed to atmospheric oxygen for 20 min followed by RNA-seq analysis. Similarly, *F. nucleatum* was treated with polymyxin B for 20 min to evaluate a classical envelope response in comparison to that of oxygen. This antibiotic is active against *F. nucleatum* (Nang et al., 2021) and as evident by the delayed growth of the bacterium at the used concentrations (Fig. 4.5).

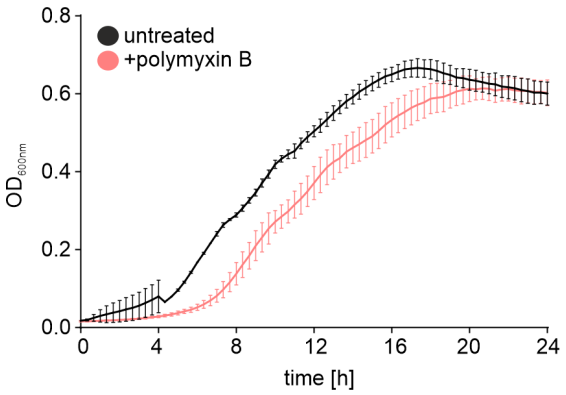

Figure 4.5 | Growth impact of polymyxin B on *F. nucleatum*.
Measurement of growth over 24 h using the optical density (OD_{600nm}) as readout. Growth media contained either 400 ng/ml polymyxin B (red) or was unmodified (black). The data represent the average and SD of three biological replicates.

Differential gene expression analysis compared to the untreated control revealed a total of 289 significantly regulated genes (FDR $\leq$0.5) (Fig. 4.6A). The 174 upregulated genes ($\log_2$ fold change $\geq$1) include the *rpoE* operon and the sRNA FoxI confirming their induction upon oxygen exposure. In addition, 19 members of the σ^E regulon were also among the upregulated genes. As observed with the σ^E induction (Fig. 4.2), both genes of the dicistronic operon C4N14_03280-C4N14_03285 displayed the strongest upregulation of the σ^E regulon, underpinning an important function in the σ^E response.

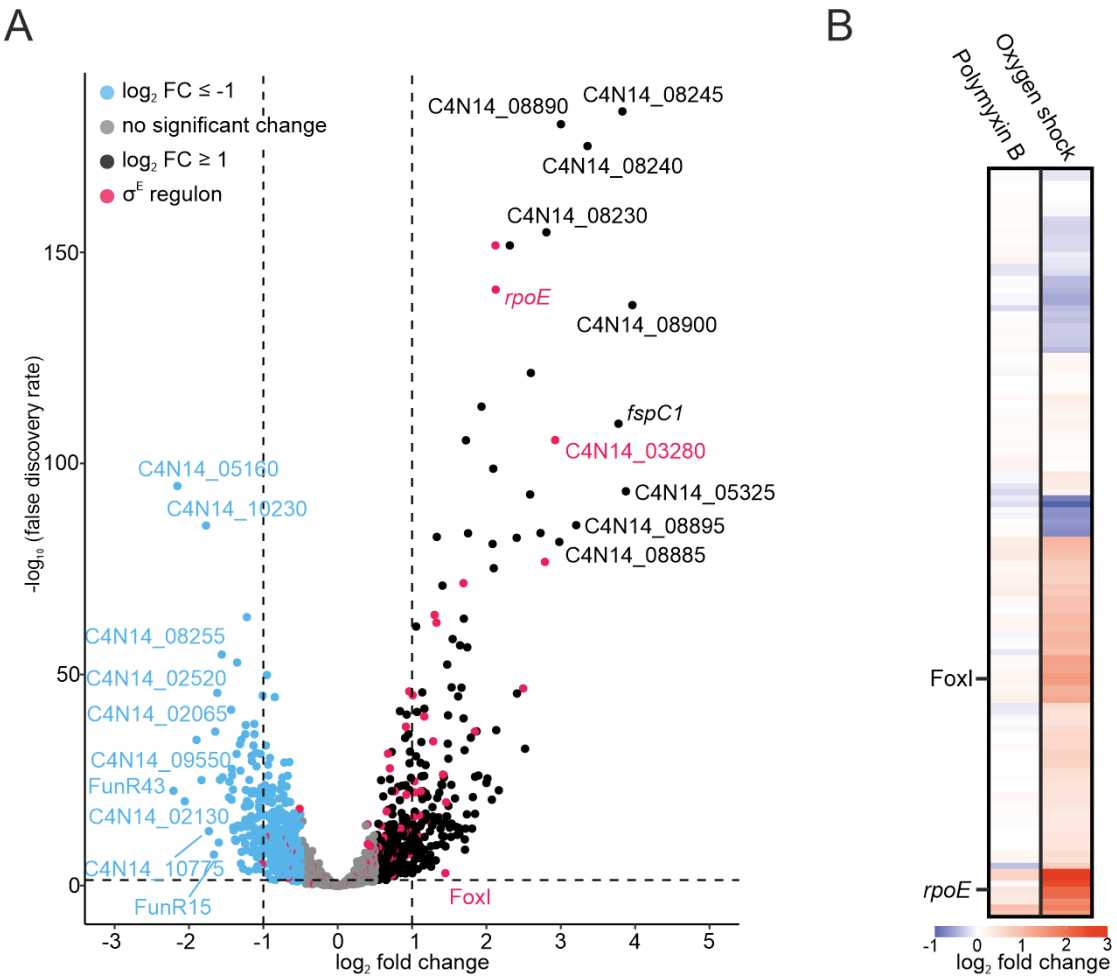

Figure 4.6 | Global transcriptional response to oxygen exposure.
A. Volcano plot displaying the transcriptional changes after 20 min of oxygen exposure compared to the untreated control. Genes were considered significantly regulated when they displayed a false-discovery rate (FDR) ≤0.05 (vertical dashed line) and a $\log_2$ fold change of ≤-1 (blue) or ≥1 (black) (horizontal dashed line). Genes that are part of the σ^E regulon are marked in red. **B.** Heatmap displaying the $\log_2$ fold changes for genes of the σ^E regulon after 20 min of oxygen exposure or treatment with polymyxin B (400 ng/ml). This figure has been adapted from Ponath et al. (2022b).

Additionally, 115/289 significantly regulated genes were downregulated ($\log_2$ fold change ≤-1). Strikingly, the 6S RNA, often found upregulated under stress in other bacteria (Wassarman, 2018), was repressed upon oxygen exposure. In addition, a number of metabolic genes such as *pfk* or ABC transporters including *mglB* as part of a putative galactose uptake operon were downregulated.

Additionally, a total of 15 transcriptional regulators were differentially expressed overall underpinning a strong global response of *F. nucleatum* when exposed to oxygen. This also includes the increased expression of genes involved in a putative response to oxidative damage including a thiol-disulfide oxidoreductase system, peroxiredoxins, flavodoxins and a metal-uptake operon (Supplementary data 8.10).

A comparison of expression changes for the σ^E regulon between the oxygen exposure and polymyxin B treatment clearly showed that this set of genes responds to oxygen rather than the antibiotic (Fig. 4.6B). Interestingly, polymyxin B treatment caused the differential expression of only three genes, despite delaying growth at the used concentration (Fig. 4.7). All three genes belong to a putative antimicrobial efflux operon, C4N14_05770-C4N14_05755, and are upregulated, further showing that *F. nucleatum* does respond to polymyxin B. Taken together, the RNA-seq results show a distinguished response to polymyxin B and oxygen while only the latter leads to σ^E activation.

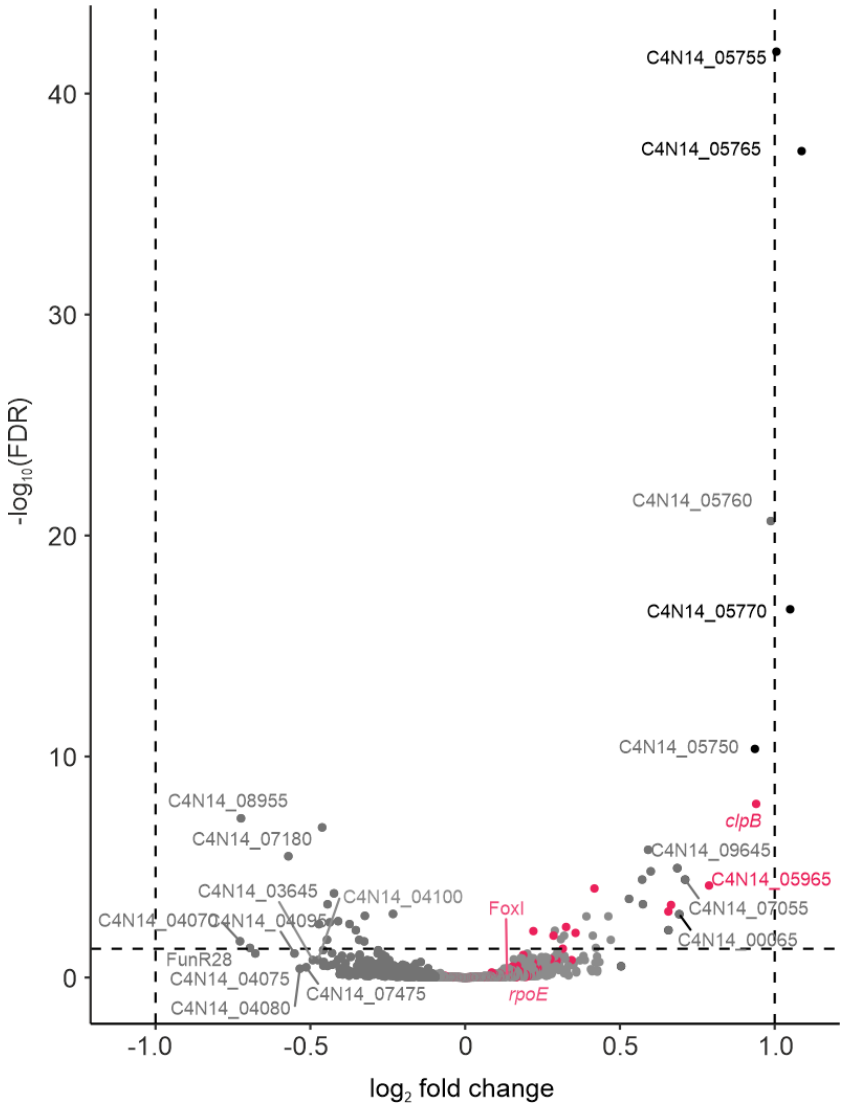

Figure 4.7 | Global transcriptional response to polymyxin B.
Volcano plot displaying transcriptional changes after 20 min of treatment with 400 ng/ml polymyxin B compared to the untreated control. Genes were considered significantly regulated when they displayed a false-discovery rate (FDR) ≤0.05 (vertical dashed line) and a $\log_2$ fold change of ≤-1 (blue) or ≥1 (black) (horizontal dashed line).

4.3 The σ^E-dependent sRNA FoxI

As described earlier (see Chapter 1.3.1.1), σ^E also induces sRNAs serving as the ECF's negative regulatory arm in *E. coli* and other bacteria (see Chapter 1.1; (Fröhlich and Gottesman, 2018)). Interestingly, the sRNA FoxI was upregulated in both the σ^E expression and oxygen exposure data set (Fig. 4.2; Fig. 4.6). This raised the question if FoxI would act akin to its σ^E-dependent counterparts MicA and RybB in *E. coli* by regulating multiple OMPs (Udekwu et al., 2005; Johansen et al., 2006; Papenfort et al., 2006; Pfeiffer et al., 2007; Papenfort et al.,

2010; Thompson et al., 2007; Gogol et al., 2011) or rather specifically acting on one target, as shown for MicL (Guo et al., 2014).

4.3.1 Conservation of the oxygen activated sRNA FoxI

Based on the conservational analysis (see Chapter 2.3, Fig. 2.15C), the ∼87 nt long sRNA FoxI can be found in all members of *F. nucleatum* as well as the related species *F. hwasooki* and *F. periodonticum*. A comparison of the surrounding genes of FoxI revealed a strong conservation in terms of its genomic synteny as the sRNA is located in a genomic island between the genes for the rRNA methyltransferase *rsmB* and the tryptophan synthase *trpB* (Fig. 4.8A). The conserved position of FoxI is in stark contrast to the changes in the flanking genomic regions found between the different subspecies. Together with its broad distribution, this argues that the FoxI sRNA plays an important regulatory role for *F. nucleatum* and closely related fusobacterial species.

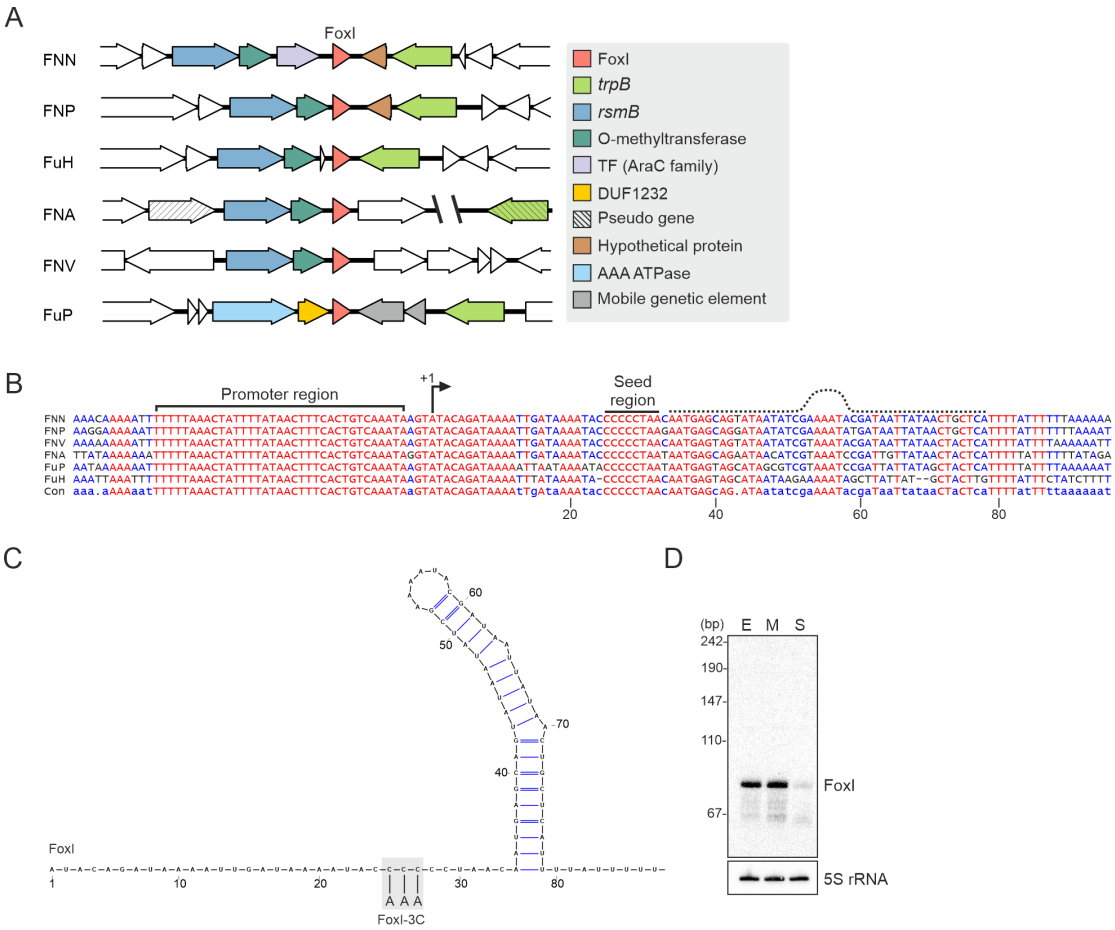

Figure 4.8 | The conserved sRNA FoxI.
A. Schematic representation of the genomic synteny for the sRNA FoxI (red) in the vicinity of *rsmB* (dark blue) and *trpB* (green). **B.** Sequence alignment of FoxI for representative members of the different subspecies and related species. The promoter region and the seed region of the sRNA are indicated. **C.** Secondary structure prediction of FoxI. Mutations introduced in FoxI-3C are indicated. **D.** Northern blot detection of FoxI across the different growth stages. The 5S rRNA served as a loading control. This figure has been adapted from Ponath et al. (2021).

A multiple-sequence alignment (MSA) of FoxI shows that only the 5'-region and a cytosine-stretch are conserved among Fusobacteria while a shared ρ-terminator is only found at the level of the secondary structure (Fig. 4.8B). An *in silico* RNA folding, using mfold (Zuker, 2003), predicts FoxI to consist out of a 33 nt long single-stranded region ended by the ρ-independent terminator stem-loop and a short U-rich stretch (Fig. 4.8C). In fact, the secondary structure of FoxI is similar to that of the σ^E-dependent sRNA MicA from *E. coli* (Udekwu et al., 2005; Rasmussen et al., 2005). The sRNA MicA base-pairs with its target via the seed-region contained in the linear stretch of its secondary structure supported by the RBP Hfq (Udekwu et al., 2005; Rasmussen et al., 2005). While Fusobacteria lack Hfq or known alternative RBPs, the single-stranded stretch of FoxI might function

as the sRNAs interaction site with mRNA targets. Particularly the cytosine-rich stretch could base-pair with guanine-rich RBS of mRNAs (see 2.2.2; Fig. 2.4) and possibly inhibit translation.

In comparison to the only partially conserved sequence of the sRNA itself, the MSA identified a highly conserved promoter region for FoxI (Fig. 4.8B). This suggest that the transcription of FoxI is tightly regulated and either activated or repressed under specific conditions. The dRNA-seq data suggest that FoxI is expressed primarily during the exponential growth phases, but decreases in the stationary phase which could be validated via northern blot analysis (Fig. 4.8D).

4.3.2 Evidence of FoxI dependency on σ^E

A comparison of the promoter region for the auto-regulatory *rpoE* and FoxI clearly showed the identified σ^E recognition motif as a conserved part in the promoter of the sRNA (Fig. 4.9A). Together with the observed upregulation of FoxI upon *rpoE*-induction, these findings suggest that the sRNA depends on σ^E for its transcription. To validate this, the same approach as above was applied (see Chapter 4.2.2) by placing the FoxI promoter region in the transcriptional reporter system pVoPo-01 with the native sequence or harboring the C-to-G mutation in the -10 box of the promoter region (Fig. 4.9A). The construct with the wild-type promoter showed a clear expression of mCherry as detected via western blot analysis while the mutated version lacked any detectable signal (Fig. 4.9B).

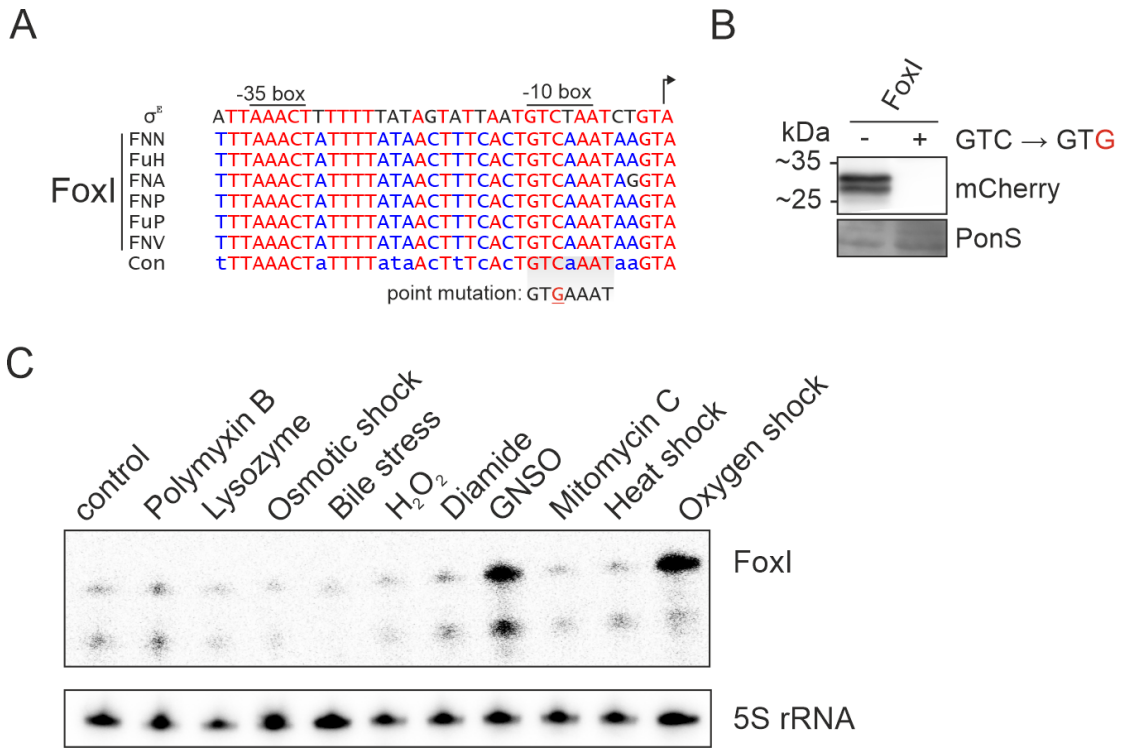

Figure 4.9 | FoxI - a σ^E-dependent sRNA.
A. Sequence alignment for the promoter region of *rpoE* in *Fnn* and FoxI for representative strains. The conserved σ^E motif is indicated alongside the introduced point mutation (see **B**). **B.** Western blot detection of mCherry of the transcriptional reporter for FoxI carrying either a point C-to-G point mutation in the promoter or the native sequence. Ponceau S (PonS) served as loading control. **C.** Northern blot detection of FoxI for RNA samples treated with the indicated stress conditions also used in Fig. 4.4. The 5S rRNA served as loading control. Panel **B** and **C** of this figure have been adapted from Ponath et al. (2022b).

The apparent dependence of FoxI on σ^E for its transcription raised the question if the sRNA would also be specifically upregulated upon oxygen-exposure. To this end, FoxI levels were analyzed via northern blot for the same stress panel as tested for *rpoE* (see Chapter 4.2.3.3), which showed a clear increase of the sRNA upon oxygen exposure (Fig. 4.9C). Hence, the sRNA was named as **F**usobacterial **ox**ygen-**i**nduced sRNA (FoxI). While GNSO treatment also led to elevated levels of FoxI, the abundance of the sRNA did not increase e.g. by polymyxin B, paralleling the results of the *rpoE* expression (Fig. 4.9C).

Overall, the result reveal FoxI to be a σ^E-dependent sRNA further suggesting that sRNA might act as the ECF's non-coding arm to repress e.g. OMPs. Still, the alternative activation by nitrosative stress (GNSO) suggests that FoxI might yet be controlled by an additional transcriptional regulator.

4.4 FoxI as the non-coding arm of σ^E

A large body of work has shown that sRNAs regulate their mRNA targets by different mechanisms, such as translational repression, translational activation, modulation of RNA stability, transcriptional attenuation or RNA sponging (Hör et al., 2020b). Yet, we do not know if sRNAs in Fusobacteria work in similar fashion as they lack the major RNA-chaperones Hfq, ProQ or CsrA commonly binding sRNAs (Holmqvist and Vogel, 2018).

4.4.1 FoxI as the regulator of the major outer membrane protein FomA

In the past, overexpression of sRNAs has been successfully used to identify mRNA targets of the regulatory RNAs. Thus, the approach was applied to FoxI by generating an overexpression construct using the pEcoFus-backbone (see Chapter 3.2). Since, the cytosine-stretch in the single-stranded portion of FoxI was a suspected seed region, a mutant version FoxI-3C (Fig. 4.8B), carrying three consecutive C-to-T mutations, was overexpressed in parallel. Both sRNAs could be stably detected via northern blot analysis (Fig. 4.10A). Interestingly, FoxI overexpression prevented *F. nucleatum* to reach similar optical density as the empty vector control (control) or the FoxI-3C overexpression bacteria (Fig. 4.10B). This observed growth phenotype indicated that FoxI is biologically active but requires an uninterrupted cytosine-stretch for its activity. Of note, no morphological changes were observed upon overexpressing either sRNA variant.

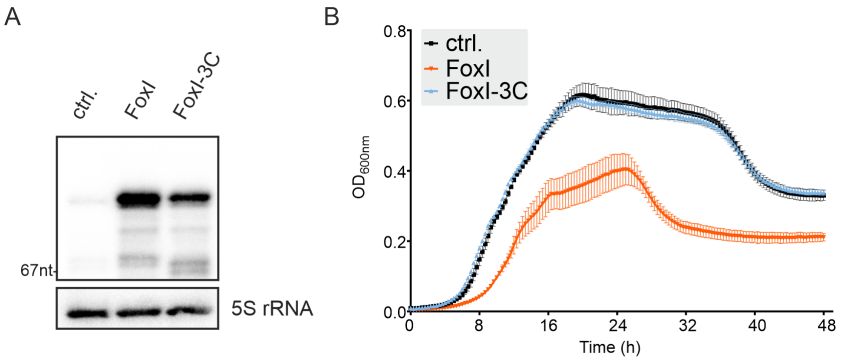

Figure 4.10 | Stable overexpression of FoxI and FoxI-3C.
A. Northern blot detection of FoxI for RNA samples of an empty vector control (ctrl.), pEcoFus-backbone overexpressing either FoxI or the seed-region mutant FoxI-3C. The 5S rRNA served as loading control. **B.** Measurement of growth by monitoring the optical density (OD_{600nm}) over 48 h for *F. nucleatum* carrying either the empty vector control (ctrl.), the FoxI or FoxI-3C overexpression vector. This figure has been adapted from Ponath et al. (2021).

To investigate if the overexpression of FoxI also causes apparent changes on protein synthesis, total protein profiles from the empty vector control and the sRNA overexpression were compared by denaturing SDS-PAGE. This revealed the depletion of an abundant protein in the size range of ∼35 to ∼55 kDa (Fig. 4.11A). A subsequent subcellular fractionation identified the downregulated protein to be primarily located in the outer membrane (OM) (Fig. 4.11A). In order to identify the affected protein, all bands excised from the gel from both the total protein samples (control and FoxI/FoxI-3C overexpression) and the OM fraction of the control were subjected to analysis via mass spectrometry.

The mass spectrometry analysis of the total protein and OM fraction for the control identified the major outer membrane porin FomA to be the most abundant protein in all samples (Fig. 4.11B). Interestingly, the abundance of FomA was strongly reduced in the samples for the FoxI overexpression (Fig. 4.11B). To further verify these results, an anti-FomA antibody was generated and used for western blot analysis of the samples. Complimentary to the mass spectrometry results, the FoxI overexpression led to a strongly decreased levels of the OMP using the FomA-specific antibody (Fig. 4.11A). In contrast to FoxI, the overexpression of FoxI-3C did not cause a decrease in FomA levels, suggesting that this activity is also related to the cytosine-stretch.

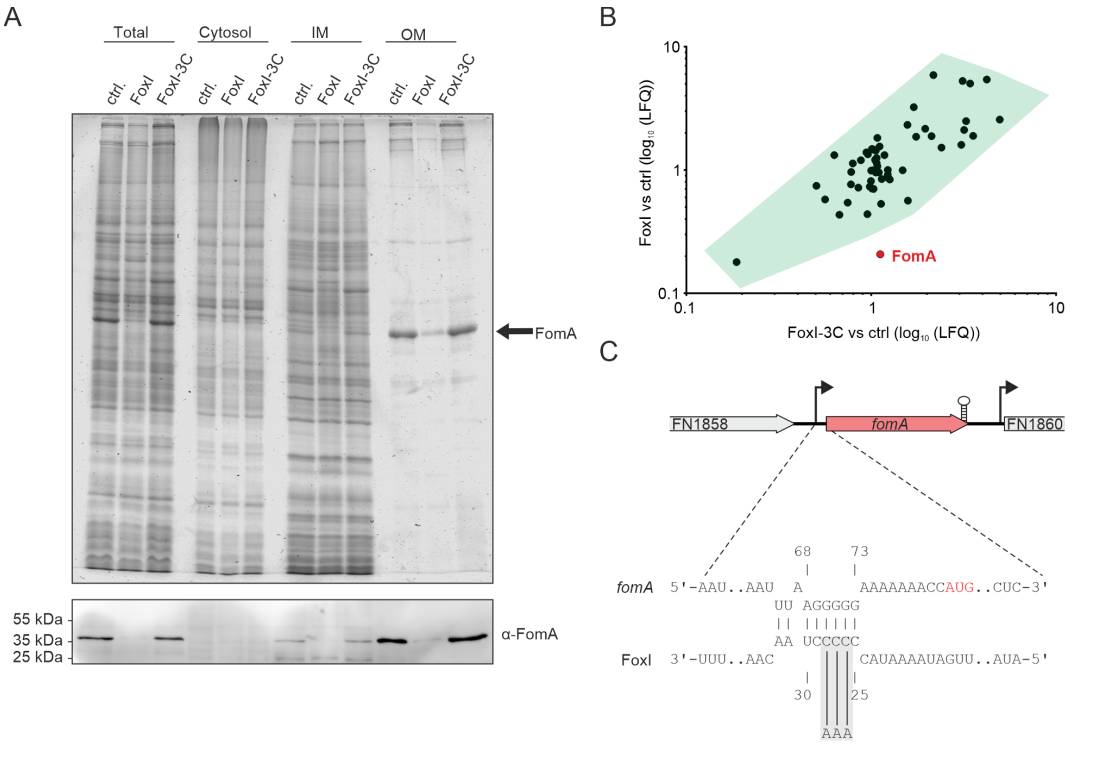

Figure 4.11 | FoxI inhibits translation of the *fomA* mRNA.
A. (top) Denaturing SDS-page analysis of protein samples for *F. nucleatum* harboring either an empty vector control (ctrl.), the FoxI or FoxI-3C overexpression vector. Equal OD_{600nm} of a subcellular fractionation were loaded for total protein (Total) and the cytosolic fraction (Cytosol). The amount loaded was increased 10x for samples of the inner membrane fraction (IM) and outer membrane fraction (OM) to allow proper visualization of protein content via Coomassie staining. (bottom) Western blot detection of FomA in the samples using an anti-FomA antibody. For this, only half of the OD_{600nm} compared to the Coomassie gel was loaded. **B.** Comparison of the ratios for the label-free quantification (LFQ) from the LC-MS/MS analysis of the FoxI and FoxI-3C overexpression samples relative to the empty vector control. The displayed proteins were filtered for a protein size between 35 kDa and 55 kDa. **C.** Prediction of FoxI-*fomA* base pairing using IntaRNA (Mann et al., 2017). The AUG start codon (red) and FoxI-3C mutation (grey) are indicated. **D.** Sequence alignment of *fomA* for different representative strains. This figure has been modified from Ponath et al. (2021).

To examine if the reduced FomA levels resulted from a direct RNA-RNA interaction between the OMP mRNA and FoxI an *in silico* target prediction was carried out using the IntaRNA (Mann et al., 2017). This analysis predicted an interaction between the RBS of the *fomA* mRNA consisting of an 8 bp long stretch including a bulged adenine (Fig. 4.11C). On the side of FoxI, the interaction site includes the conserved cytosine-stretch of the sRNA. Thus, it is likely that FoxI represses

the translation of FomA by sequestering the RBS of the mRNA. Importantly, an MSA of the *fomA* mRNA for fusobacterial strains harboring FoxI showed that this region is highly conserved further supporting this hypothesis (Fig. 4.11D).

4.4.1.1 *In vitro* analysis of the FoxI-*fomA* interaction

As a first step to understand the interaction between FoxI and the *fomA* mRNA primer extension for the OMP mRNA was carried out. Primer extension can be used to map the 5'-end of an RNA but can also be used to investigate processing events and quantify transcript levels by the specific reverse transcription of an target RNA (Carbon et al., 1975). The same amount of RNA from samples carrying the empty vector control, FoxI or FoxI-3C overexpression was used as input for the reverse transcription with a primer specific to a region ∼150 nt downstream of the TSS. Correlating the detected full-length transcript (*fomA*-L) to the gene-specific ladder confirmed the TSS annotated in the dRNA-seq analysis. Further, it also showed that the RNA levels between all samples were comparable as evident by *fomA*-L signal. Curiously, FoxI overexpression caused the absence of a shorter isoform of *fomA* mRNA (*fomA*-S) (Fig. 4.12). This effect was neither observed for the control or FoxI-3C overexpression. Based on the dRNA-seq result, this *fomA* isoform could represent a processed version of the mRNA. While it seems very likely that FoxI inhibits formation of the isoform or selectively causes its degradation, it remains unclear how or if this is related to the effect on translation of *fomA* as the majority of mRNA seems to be unaffected.

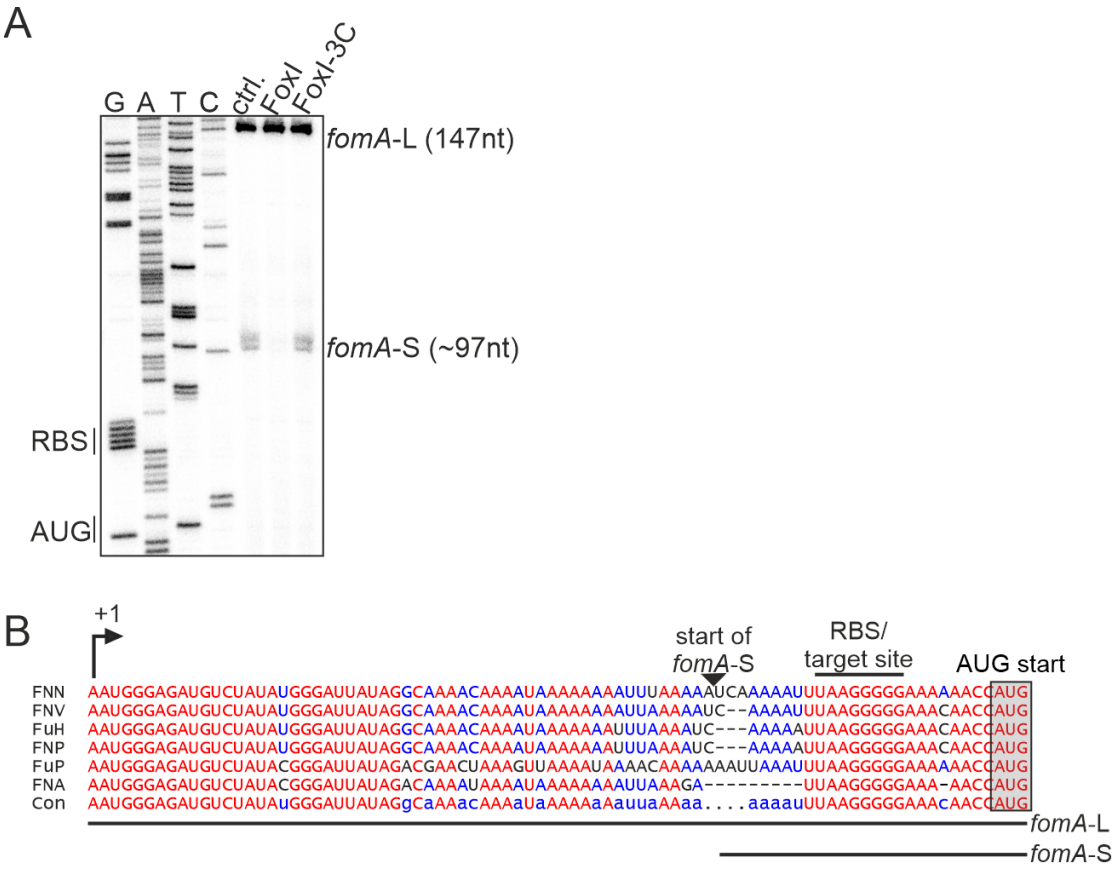

Figure 4.12 | Detection of *fomA* isoforms regulated by FoxI.
A. Results of a primer extension experiment for total RNA-samples isolated from *F. nucleatum* harboring either an empty vector control (ctrl.), the FoxI or the FoxI-3C overexpression vector. Ten microgram of DNA-free RNA for each sample was used as input together with a radioactively labeled primer specific for *fomA*. The DNA ladder for the fragment is indicated as well as the RBS of *fomA*. Two signals could be detected matching the full-length 5'-end of the *fomA* mRNA (*fomA*-L) and a shorter isoform (*fomA*-S). **B.** Sequence alignment of the 5'-region of the *fomA* mRNA is shown in **A**. *fomA*-L and *fomA*-S are indicated next to the RBS and AUG start codon.

As the second approach to probe the FoxI-*fomA* interaction, *in vitro* transcribed radioactively labeled *fomA* mRNA was used together with unlabeled FoxI RNA for an Electrophoretic Mobility Shift Assay (EMSA). A stable interaction of both RNAs would result in a decreased mobility of the RNA-RNA complex through the gel, evident by an upward shift of the signal. However, only a very minor shift of the labeled *fomA* mRNA could be observed in the presence of increasing concentrations of FoxI but not FoxI-3C (Fig. 4.13). These results suggest that the assay conditions possibly did not support proper RNA-RNA interaction or that an unidentified RBP could be needed to facilitate a stable mRNA-sRNA interaction.

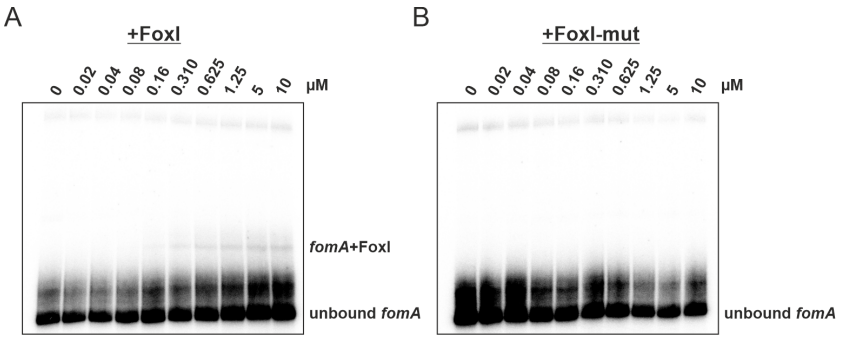

Figure 4.13 | FoxI-*fomA* interaction cannot be recapitulated by EMSA.
EMSA gels of labeled *fomA* mRNA fragment carrying the FoxI binding site. The mRNA was incubated with increasing concentrations of unlabeled (**A**) FoxI or (**B**) FoxI-3C RNA. Unbound *fomA* and a potential complex with FoxI are indicated.

Lastly, based upon the *in silico* prediction and supported by the results of the overexpression experiment, FoxI binds to the RBS of the *fomA* mRNA. In order to investigate if this interaction can impact translation initiation, the *in vitro*-synthesized 5'-region of the *fomA* mRNA was used together with 30S ribosomal subunits in a toeprinting experiment. If the initiation complex forms in the presence of the initiator *N*-formylmethionine tRNA ($fMet^{tRNA}$) and the 30S ribosomal subunits, the reverse transcription of the mRNA will be abrogated with a distance of 14-16 nt downstream of the +1 site (Hartz et al., 1989). This stop signal was detected when incubating the synthesized *fomA* transcript together with the $fMet^{tRNA}$ and the 30S ribosomal subunit (Fig. 4.14). The addition of increasing amounts of the *in vitro* transcribed FoxI led to the decreased formation of this complex indicating that the sRNA prevents the formation of the translation initiation complex. Interestingly, a second stop signal further upstream could be detected which overlaps with the predicted binding site for FoxI on *fomA*. Furthermore, FoxI-3C showed only weaker impact on the 30S ribosomal subunit binding and did not cause the secondary stop. Combined, these data show that the FoxI-*fomA* interaction depends on the cytosine-stretch further referred to as the sRNA seed region.

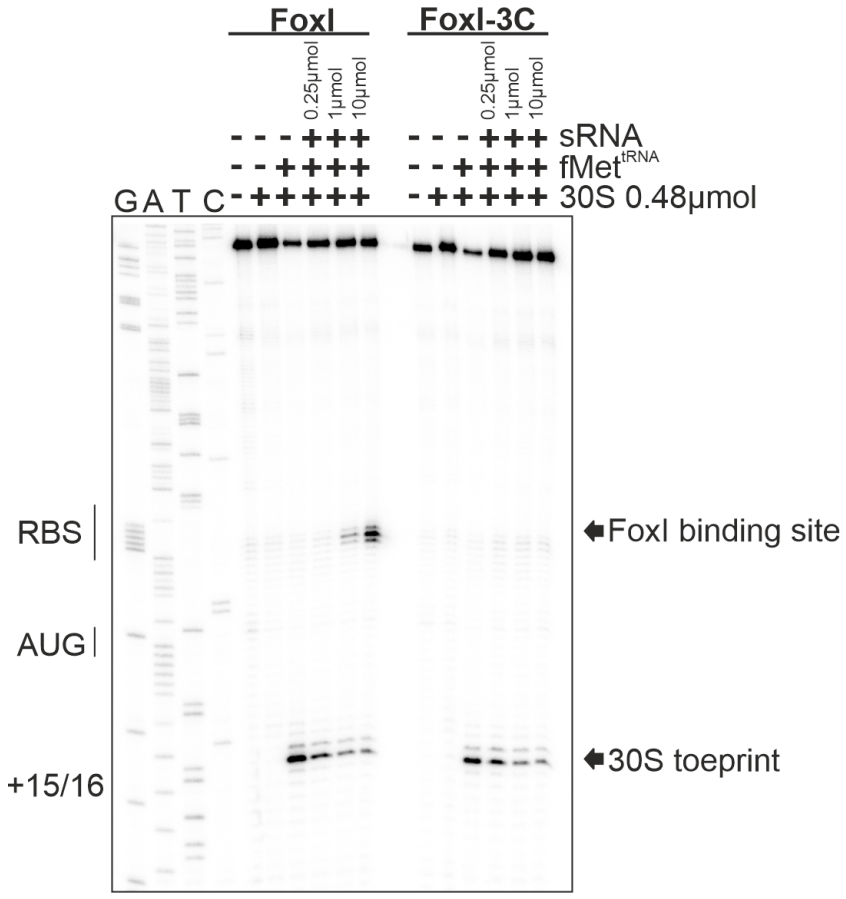

Figure 4.14 | FoxI-binding can impact initiation of *fomA* translation .
Toeprinting assay for the radioactively labeled 5'-end of the *fomA* mRNA. The labeled RNA was incubated alone, with purified 30S ribosomal subunit (*E. coli*), initiator tRNA ($fMet^{tRNA}$) and different concentrations of sRNA (FoxI/ FoxI-3C). The stop signal caused by the binding of the 30S ribosomal subunit as well as the RBS, AUG start and toeprint region are indicated.

To summarize, the *in vitro* analysis of the FoxI-*fomA* interaction supports a model of FoxI acting via translational inhibition in base-pair specific manner to decrease FomA levles. However, the effect on *fomA* processing as well as the failure to recapitulate the RNA-RNA interaction via EMSA require further investigation and point to a possible role of an RBP in this process.

4.4.2 Pulse expression of FoxI

Next to overexpression, the concept of pulse-expressing an sRNA represents another powerful approach to discover the targetome of sRNAs (Massé et al., 2005; Papenfort et al., 2006; Pfeiffer et al., 2007; Hör et al., 2018). Further, this approach also excludes possible confounding factors such as the induction of potential additional sRNAs, as might the case during the pulse-expression of σ^E. Prior to this, a $\Delta foxI$ strain was generated using the pVoPo-04 system (see

Chapter 3.5) which was validated for the absence of the sRNA via northern blot analysis (Fig. 4.15A). Next, inducible expression constructs for FoxI, FoxI-3C and FoxI-C4A, harboring only a single mutation in the seed region, were generated and introduced into the Δ*foxI* background. All three sRNAs were pulse-expressed for 20 min through the addition of ATc and extracted RNA was analyzed by RNA-seq. The global expression analysis identified 30 significantly downregulated genes ($-0.5 \leq \log_2\text{FC} \geq 0.5$; FDR ≤ 0.05) (Fig. 4.15B; Supplementary data 8.11).

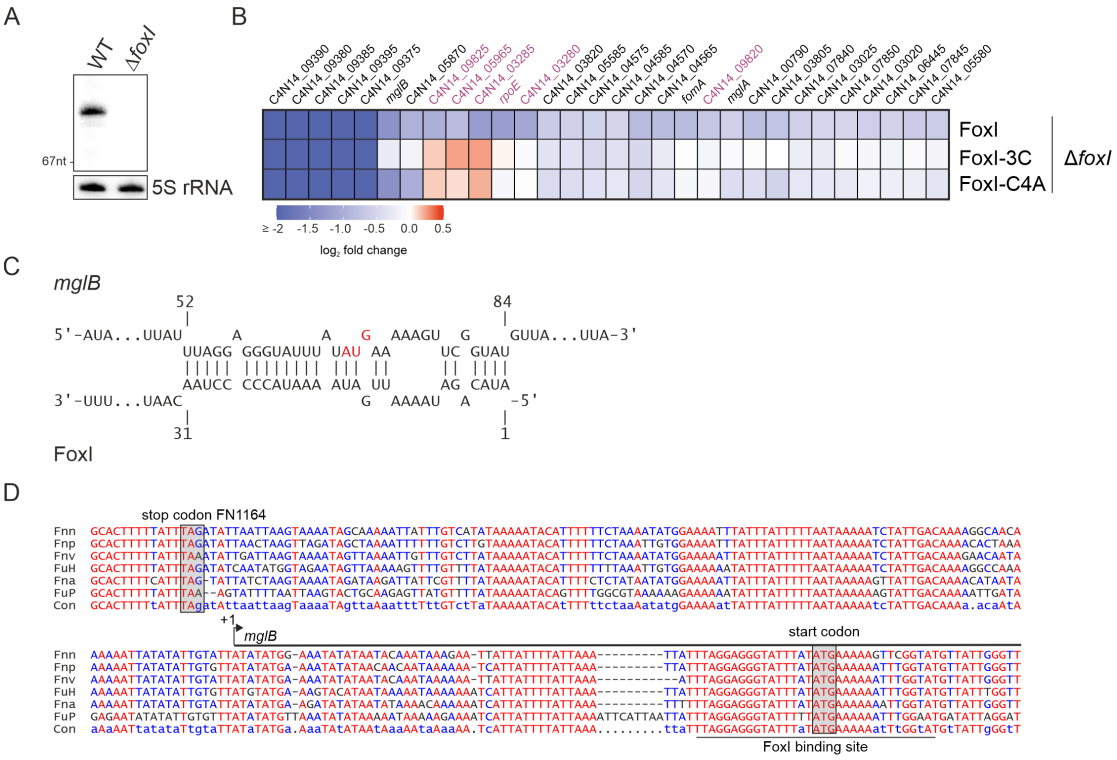

Figure 4.15 | Pulse expression of FoxI identifies *mglB* as a new target.
A. Northern blot verification of Δ*foxI* strain generated using the pVoPo-04 system. The 5S rRNA served as loading a control. **B.** Heatmap displaying the $\log_2$ fold changes for all genes significantly downregulated ($\log_2\text{FC} \leq -0.5$; FDR ≤ 0.05) after 20 min of pulse expressing FoxI, FoxI-3C or FoxI-C4A compared to an empty vector control in the Δ*foxI* background. Members of the σ^E regulon are marked in purple. **C.** Base-pairing prediction between *mlgB* and FoxI. The start codon of *mglB* is marked in red. **D.** Sequence alignment of *mglB* for different representative fusobacterial strains. The FoxI binding site as well as relevant start and stop codons are indicated. This figure has been modified from Ponath et al. (2022b).

The top downregulated genes were all part of an operon of unknown function (C4N14_09375-C4N14_09395) (Fig. 4.15B). However, as this effect was also found with the induction of both FoxI seed region mutants, the operon is unlikely to be directly targeted by FoxI.

Interestingly, the expression of FoxI caused the downregulation of several members of the identified σ^E regulon, including the *rpoE* operon itself and the top-

upregulated dicistronic operon C4N14_03280-C4N14_03285 (see 4.2.2). Importantly, this could not be observed with the pulse-expression of any of the two seed region mutants. An *in silico* target prediction indicated only poor interactions between FoxI and any of the downregulated σ^E-dependent genes (Supplementary data 8.12). This suggests that their downregulation could be an indirect consequence achieved through a FoxI-mediated relief of basal σ^E activation in the Δ*foxI* background.

Further, FoxI pulse-expression led to clear downregulation of the *mglB* and *mglA*, the leading genes of the *mglBAC* operon likely functioning as an IM galactose uptake system (Fig. 4.15B) (Harayama et al., 1983). The *mglBAC* operon remained unchanged with expression of FoxI-3C, while induction of FoxI-C4A still lowered *mglB* RNA levels similar to the wild-type sRNA. In comparison, downregulation of the *fomA* mRNA could only observed with expression of FoxI but neither of the seed region mutants. These results suggest that the interaction between FoxI and *mglB* might be more robust than that with the *fomA* mRNA, which is comprised almost exclusively of the cytosine-rich stretch (Fig. 4.11).

Indeed, the RNA-RNA prediction for FoxI-*mglB* suggests an over 20 nt long interaction site covering both the RBS and the start codon of the mRNA (Fig. 4.15C). Thus, a single interruption of the seed region of FoxI might not be sufficient to abrogate regulation of *mglB* whereas the binding of *fomA* is shorter and possibly easier to disrupt. As with *fomA*, also the FoxI binding site is conserved for *mglB* across the different Fusobacteria (Fig. 4.15D) further supporting a relevant interaction.

4.4.3 Repressed genes upon σ^E pulse expression

The σ^E expression did not only activate gene expression, as focused on in chapter 4.2.2, but also led to the downregulation of 23 mRNAs (Fig. 4.2). Interestingly, most of the downregulated transcripts encode for membrane or membrane-associated proteins. The majority of the repressed mRNAs belong to three similar multicistronic operons containing paralogs of a FadA-domain containing protein, an OmpA family protein and divergent type 5a autotransporters. Noteworthy, the FoxI target *fomA* did not pass the selected cutoff however the mRNA levels of the porin showed a clear decrease upon induction of σ^E ($\log_2$FC of -0.74). However, σ^E expression clearly downregulated the putative FoxI target *mglB*, thus raising

the question if FoxI target spectrum includes additional mRNAs downregulated with σ^E expression, thereby supporting the role of FoxI as the ECF's non-coding arm.

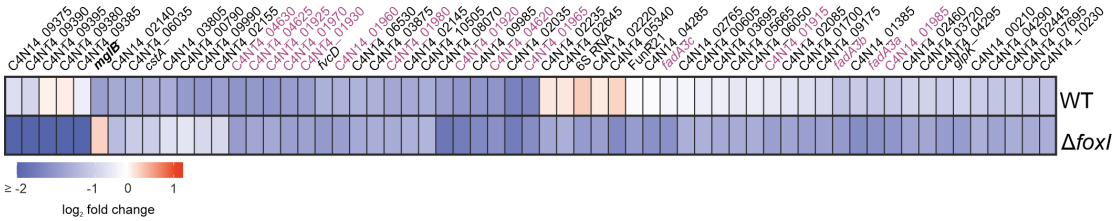

Figure 4.16 | Downregulated genes in wild-type and Δ*foxI* background upon σ^E expression.
Heatmap displaying the $\log_2$ fold changes of significantly downregulated genes ($\log_2$ ≤-1; FDR ≤0.05) after σ^E pulse expression in wild-type (WT) or Δ*foxI* background compared to the empty vector control. Genes that are part of the FadA-paralog associated operons are marked in purple. *mglB* is marked in bold. This figure has been modified from Ponath et al. (2022b).

To test this hypothesis, the σ^E induction was repeated in the FoxI-deletion background and analyzed via RNA-seq. In the absence of FoxI σ^E expression did not downregulate *mglB* or *fomA* mRNA levels in agreement with the pulse-expression data (Fig. 4.16). At the same time, no other mRNA showed differential expression in the Δ*foxI* background. This could indicate two scenarios which are not mutually exclusive: (i) FoxI primary mode of action is the blocking of translation with a minor effect on RNA stability and/or (ii) σ^E in *F. nucleatum* controls a second unidentified sRNA with a redundant function to FoxI as found in *E. coli* and *V. cholerae* (Peschek et al., 2019; Gogol et al., 2011).

4.4.4 Identification and validation of FoxI targets using the translational reporter system

The RNA-seq-based approach of target identification for FoxI revealed the *mglB* mRNA as an additional target of the sRNA. To validate this interaction, the translational reporter system (see Chapter 3.6.2) was used to create a translational fusion for *mglB* (*mglB*::mCherry). The plasmid contained either the fusion product alone (control) or in addition the expression cassette for FoxI or FoxI-3C (Fig. 4.17A). Similar constructs were generated for *fomA* as a validated FoxI target and C4N14_09375, which does not contain a predicted FoxI binding site. Subsequent western blot analysis showed a striking reduction of MglB::mCherry levels only with the expression of FoxI compared to the control (Fig. 4.17B). Importantly, the *fomA*::mCherry fusion also displayed a strong reduction in the

presence of FoxI while C4N14_09375::mCherry was unaffected. The quantification of the fluorescence signal via flow cytometry for the different mCherry fusion products further confirmed the results (Fig. 4.17C).

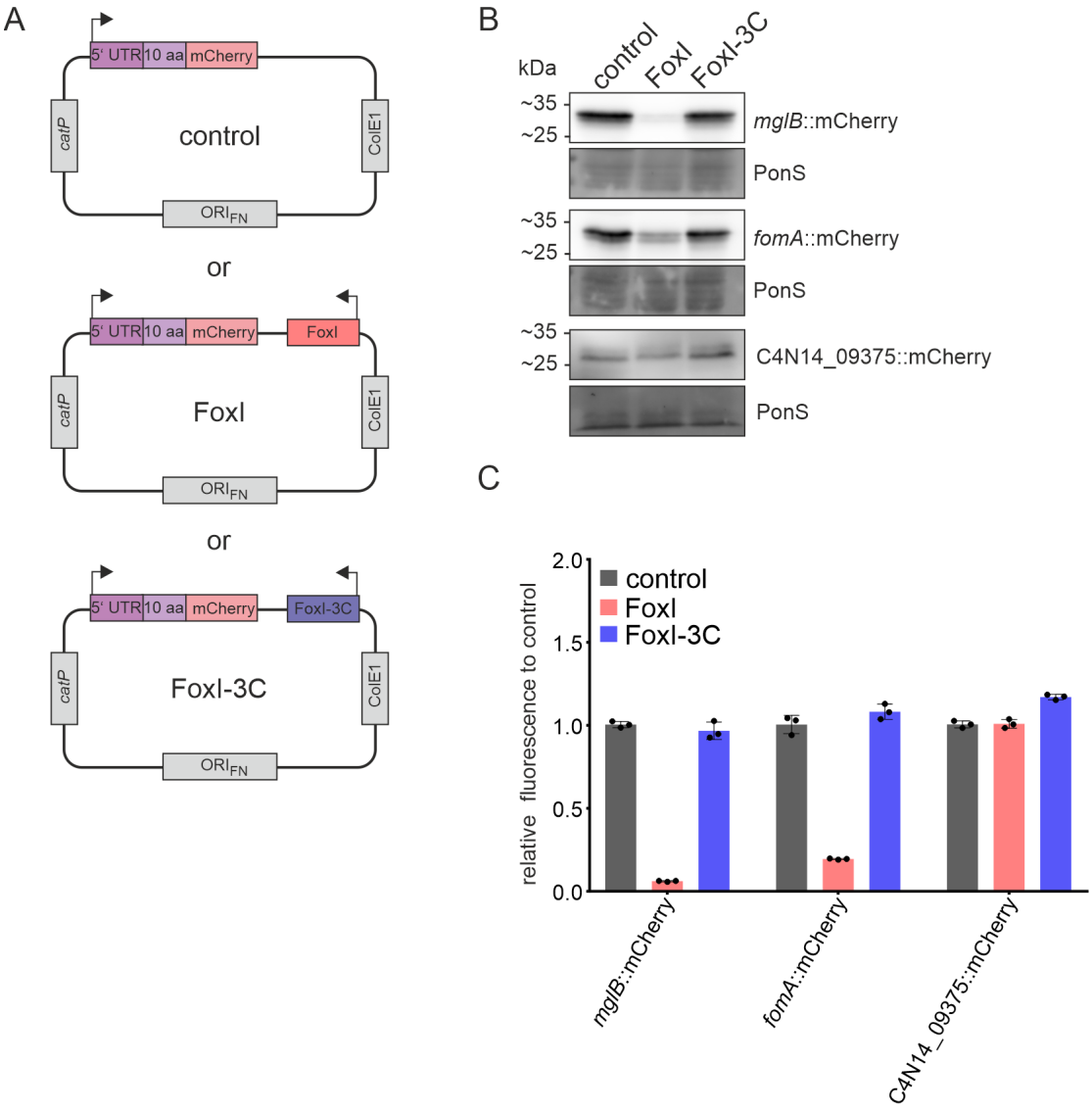

Figure 4.17 | FoxI represses translation of *mlgB*.
A. Schematic overview of the different translational reporter plasmids used. mCherry was fused to the 5'end including 5'UTR and 30 nt of a target gene. This alone served as the control vector while additional constructs were included harboring a FoxI or FoxI-3C overexpression cassette. **B.** Western blot detection of mCherry for indicated translational reporter constructs. The individual translational fusions were either expressed alone (control), together with FoxI (FoxI) or FoxI-3C (FoxI-3C). C4N14_09375::mCherry served as control fusion as it does not harbor a FoxI-binding site. Ponceau S (PonS) served as loading control. **C.** Quantification of the relative fluorescence intensity compared to the control for the constructs shown in **B**. Shown are the average and SD for three biological replicates. This figure has been adapted from Ponath et al. (2022b).

The application of the translational reporter system further supports the hy-

pothesis that the regulation of FoxI targets is a consequence of direct base-pairing as the seed region mutant FoxI-3C failed to repress translation of the fusion product. Additionally, C4N14_09375::mCherry does not contain a FoxI binding site and was not effected by co-expression of the wild-type sRNA, thereby making an unspecific effect of the FoxI expression unlikely. Thus, the results confirm *fomA* and *mglB* as direct targets of FoxI. Importantly, this also demonstrate the application of the translational reporter system for the investigation of sRNA-mediated post-transcriptional regulation in *F. nucleatum*.

The combined results for the FoxI targets *fomA* and *mglB* demonstrate that the sRNA can drastically repress translation, while weaker effects on the mRNA levels are observed. This leads to the hypothesis that FoxI might generally act by blocking translation of its targets without causing rapid RNA decay. Interestingly, several of the membrane-associated genes downregulated upon σ^E-expression contain promising FoxI binding sites based on *in silico* target prediction while the effect on the RNA level was only minor (Fig. 4.16). This suggests that FoxI could repress their translation as shown for *fomA* and *mglB*.

To testing this, five additional translational fusions were generated for genes displaying a predicted FoxI binding site: two of the three paralogous FadA domain-containing proteins (*fadA3a*; *fadA3b*), the type 5c autotransporter *fvcD*, and two genes encoding for IM proteins of unknown function (C4N14_02035; C4N14_00275) (Fig. 4.18A). Both western analysis and measurement of fluorescence intensity showed a reduction for the mCherry fusions levels of all five tested genes with FoxI co-expression, with *fadA3a* and *fadA3b* displaying the strongest repression (Fig. 4.18B, C). Just as for *fomA* and *mglB*, FoxI likely blocks the translational of all five targets through direct base-pairing which was disrupted in the seed region mutant FoxI-3C.

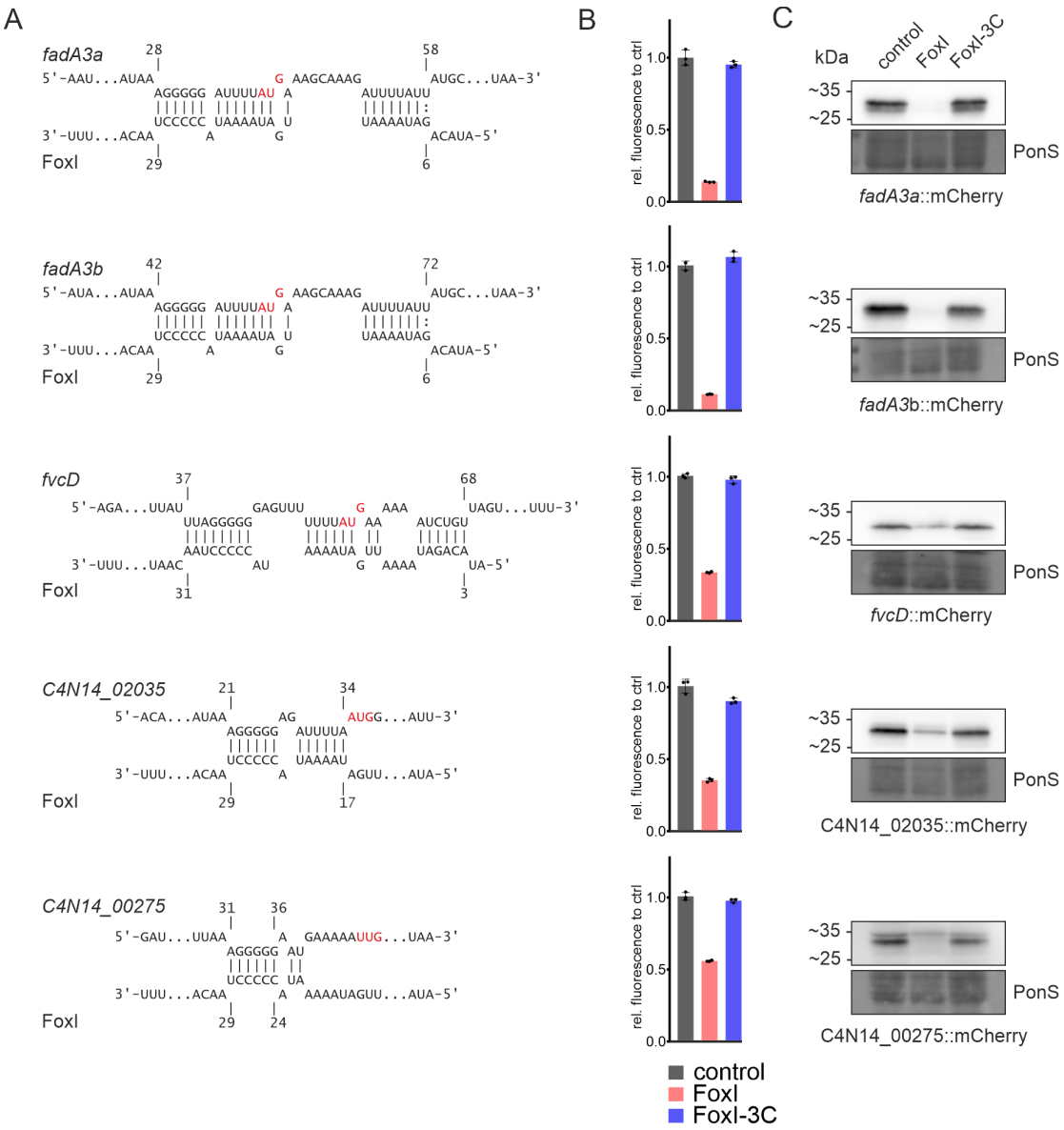

Figure 4.18 | FoxI targets multiple envelope proteins.
A. Base-pairing prediction between FoxI and the indicated mRNAs. The start codon of each target mRNA is marked in red. **B.** Quantification of the relative fluorescence intensity compared to the control of the translational reporter for mRNAs shown in **A**. Displayed are the average and SD for three biological replicates. **C.** Western blot detection of mCherry for the different translational reporter constructs of targets shown in **A**. Ponceau S (PonS) served as loading control. This figure has been adapted from Ponath et al. (2022b).

Overall, the validation of FoxI targets via the translational reporter system supports the hypothesis that the sRNA is a multi-targeting repressor of translation through direct interactions with the target mRNA. The above uncovered targetome of FoxI further highlights similarities to other σ^E-dependent sRNAs, such as MicA or RybB, by primarily regulating the expression of envelope proteins. Furthermore, targets that are also downregulated with σ^E suggest that FoxI

represents the non-coding arm of the σ^E-response in *F. nucleatum*.

4.5 The σ^E response in *F. nucleatum*

Our understanding of regulatory networks for *F. nucleatum* is still very limited. Yet, the here uncovered σ^E response in the anaerobe bacterium displays surprising parallels to envelope stress response in well-studied in γ-Proteobacteria such as *E. coli* (Hews et al., 2019; Rowley et al., 2006) supporting the following model (Fig. 4.19):

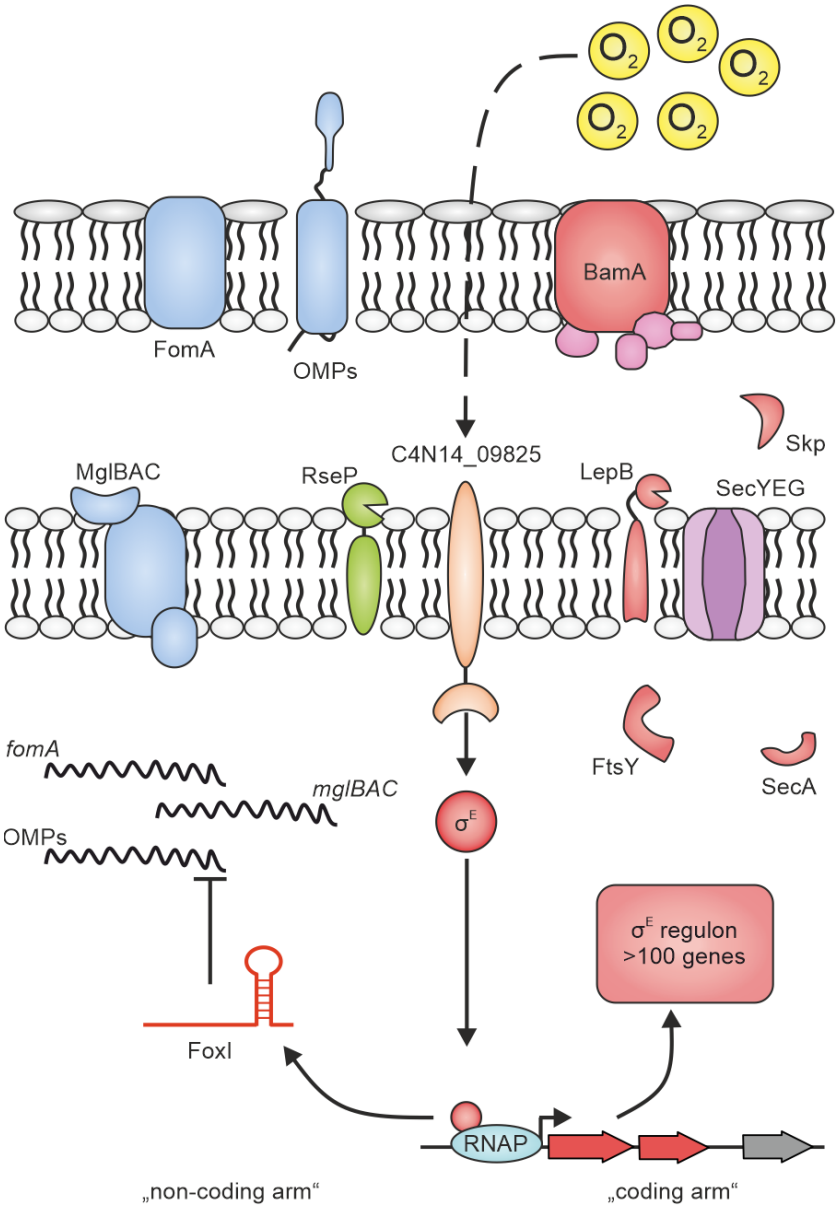

Figure 4.19 | Model of σ^E activity in *F. nucleatum*
This figure has been adapted from Ponath et al. (2022b).

When exposed to oxygen, *F. nucleatum* senses this environmental condition through an yet unknown mechanism leading to the release of σ^E from its cognate anti-sigma factor. This enables the ECF to associate with the RNAP leading to the transcription initiation of target genes harboring the conserved σ^E recognition motif. At least a subset of genes, as part of the σ^E coding arm, augments the translocation of OMPs and their insertion into the OM. Additionally, σ^E activates the transcription of its non-coding arm, consisting of at least of FoxI, which subsequently represses the synthesis of several envelope proteins to possibly decrease the burden on the translocation system. Altogether, both regulatory arms of σ^E might act together to maintain envelope homeostasis when *F. nucleatum* faces increased levels of oxygen threatening the anaerobe's survival.

5 Discussion

5.1 The RNA landscape of *F. nucleatum*

F. nucleatum is part of a poorly explored phylum despite its relevance to human health (Brennan and Garrett, 2019). One reason for this paucity of information on *F. nucleatum* is its distinctive phylogenetic distance from other bacteria, such as the common model organisms *E. coli* (Coleman et al., 2021), which limits simple knowledge transfer. Any advances in understanding *F. nucleatum* are further impeded by the lack of genetic tools and general difficulty working with anaerobic bacteria. Combined, this led to the lack of knowledge on the basic molecular biology and RNA biology of this medically important bacterium.

5.1.1 An RNA-guided update of the annotation for *F. nucleatum*

For example, this work describes the first promoter motif in *F. nucleatum*, which drives transcription of ∼2/3 of all genes, with a strong resemblance to the σ^{70} promoter motif found e.g. in *E. coli* (Hawley and Mcclure, 1983) or *H. pylori* (Sharma et al., 2010). To that end, the generation of the single-nucleotide maps for representative members of the *Fusobacterium* species will be an important resource for future studies complementing the recently improved annotation (Sanders et al., 2018). The identification of TSS and 5'UTRs will be invaluable to correctly design and assess global approaches (further discussed in Chapter 5.3.3), for instance to determine polar effects in high-throughput transposon screens (Hutchison et al., 2019).

An initial low-throughput transposon screen had been used to identify the virulence factor Fap2 (Coppenhagen-Glazer et al., 2015). The adhesin and other OMPs such as FadA (Rubinstein et al., 2013) are of particular interest as these proteins thus far represent the main factors in the interaction with cancer cells (Brennan and Garrett, 2019). Consequently, improving our knowledge on their

gene structure will aid in understanding their function. In addition to FadA and Fap2, a pTSS could be assigned to ∼200 putative virulence factors (Kumar et al., 2016). However, the transcriptomic data for these genes showed that most of the virulence factors are constitutively expressed throughout growth in stark contrast to virulence genes in other human pathogens. In the case of *Salmonella* or *Yersinia*, these factors are expressed in a highly regulated manner upon contact with the host (Erhardt and Dersch, 2015). This raises the possibility that *F. nucleatum* might regulate its virulence genes through other mechanisms, including post-transcriptional regulation which is typical in *Enterobacteriaceae* (Heroven et al., 2017).

The hot-spot of bacterial post-transcriptional regulation is the 5'-end of an mRNA that can harbor e.g. riboregulatory elements (further discussed below) or binding sites for sRNAs. The latter are usually strongly conserved and thus knowledge of the exact 5'UTRs for a gene of interest can identify important stretches of the untranslated region. This is highlighted by example of the sequence conservation of the RybB sRNA binding site on the *ompC* mRNA across different γ-proteobacteria (Papenfort et al., 2010). Leveraging such evolutionary comparison between the different *Fusobacterium* species could thus aid in identifying regulatory features or support sRNA target predictions.

One example is the ∼100 nt long 5'UTR of the *hutH* gene that only contains two conserved regions: the RBS and a further imperfect palindrome (see Chapter 2.4.4). Interestingly, expression of the *hutH* gene in *B. subtilis* is regulated by the RNA-binding anti-terminator HutP (Oda et al., 2004), raising the possibility that this post-transcriptional regulatory mechanism is shared between both species. The conserved repetitive sequence element, in particular, could serve as the binding site for a functional analog of HutP in *F. nucleatum* to regulate the histidine metabolism.

Transcriptional regulation in *F. nucleatum* is unexplored including promoter motifs or transcription factor binding sites. While previous work on the tryptophanase operon identified the first TSS in *F. nucleatum* (Sasaki-Imamura et al., 2010), this does not allow for a global extrapolation on common promoter features in the bacterium. This is has been remedied through the here generated global TSS annotation for the five fusobacterial strains, which globally identi-

fied a common promoter motif with an extended -10 box and less prevalent -35 box in all strains. Both boxes are highly similar to the regions recognized by the principal σ^{70} and bear an overall similarity to the promoter structure found for members of diverse families such as *Helicobacteraceae* (Sharma et al., 2010), *Campylobacteraceae* (Dugar et al., 2013), *Enterococcaceae* (Michaux et al., 2020) and *Corynebacteriaceae* (Pátek et al., 2003; Pfeifer-Sancar et al., 2013) including an A+T-rich spacer region. This strongly suggest that σ^{70} drives basic transcription in *Fusobacterium* as also found for other bacterial species.

The lack of additional motifs is likely due to two factors: firstly, the analysis was only conducted during growth in rich media; secondly, *F. nucleatum* only encodes 4-5 sigma factors (Fig. 2.5). The only member of the σ^{54} family, RpoN, is absent in *F. nucleatum* subsp. *nucleatum* but present in all other analyzed strains. RpoN is involved in the regulation of nitrogen metabolism, but can regulate other metabolic pathways or virulence genes (Damron et al., 2012; Leang et al., 2009; Hirschman et al., 1985). Conversely, no promoter motif could be detected for RpoN in any of the additional fusobacterial strains and the RNA-seq analysis showed only in *F. periodonticum* any transcription of this gene. These findings suggest that RpoN is only expressed and active under specific conditions in these fusobacterial strains. Yet, it remains unclear why *F. nucleatum* subsp. *nucleatum* is lacking the *rpoN* gene, while it still encodes for associated activating proteins of the σ factor.

Overall, this still leaves the transcriptional regulation in *F. nucleatum* as an almost blank canvas. At the same time, the TSS and promoter regions might provide guidance in finding the direct targets of the two-component systems CarRS and ModRS (Wu et al., 2021; Scheible et al., 2022) as their immediate regulons are unknown as of now.

5.1.2 Discovery of small ORFs

Small proteins in bacteria have garnered more attention as they can often regulate larger protein complexes (Hemm et al., 2020), as in the case of the sORF ArcZ modulating antibiotic resistance through the AcrAB-TolC complex (Hobbs et al., 2012). However, the limited amount of information encoded in the ORF of small proteins represents a challenge for computational predictions with the presence of a strongly conserved sORF and an RBS often being the most promising

identifiers of a functional sORF (Hemm et al., 2010).

Combining the dRNA-seq data with such a comparative genomics approach identified three new high confidence candidates for small proteins which all carry an RBS and an AUG start codon otherwise lacking in the 22 putative sORFs contained in the automated annotation of the reference strain (see Chapter 2.2.5). All three putative small proteins harbor predicted α-helices indicating a potential membrane association also observed for small proteins in other organisms (Gray et al., 2022), such as MgrB promoting PhoP inactivation of the *E. coli* PhoPQ system during magnesium limitation (Salazar et al., 2016).

While exact functions for either of these proteins are currently unknown, their constitutive expression and strong conservation suggest a role at least during growth for *F. nucleatum*. In any case, their translation needs to be validated first by established methods such as via epitope-tagging (Hemm et al., 2020; Storz et al., 2014).

5.1.3 Housekeeping RNAs

The presence of the 4.5S RNA, tmRNA and M1 RNA reflects the essential role of these housekeeping RNAs for virtually all bacteria; including Fusobacteria. The additional identification of the highly abundant 6S RNA in this phylum displays that these RNAs likely emerged early on in evolution of the bacteria given the early divergence of the Fusobacteriota from other bacterial phyla (Coleman et al., 2021).

The 6S RNA serves as a template to the RNAP, when interacting with the complex. This generates the pRNA which is important for the function of the 6S RNA to disassociate from the RNAP (Wassarman and Saecker, 2006). As the pRNA signifies the formation of a 6S RNA-RNAP complex, the detection of the short transcript originating in *cis* to the 6S RNA candidate argues for the ncRNA to be a true 6S RNA homolog (Wassarman, 2018). The transcription pattern of the fusobacterial 6S RNA shows further similarities to its counterpart in *E. coli* by accumulating towards the stationary phase. This suggests that the biological function of the 6S RNA is similar between both species.

In *E. coli*, the 6S RNA particularly impacts transcription of genes from promoters with a weak -35 box especially in the presence of an extended -10 box (Cavanagh et al., 2008). Interestingly, the promoter analysis for *F. nucleatum*

shows that a strong -35 box is generally less prevalent while the extended -10 box associated with the majority of pTSS, suggesting a possible similar global impact.

Surprisingly, the 6S RNA is downregulated in response to oxygen exposure in *F. nucleatum*, which contradicts a common observation that the RNA is typically upregulated in different bacteria when stressed (Wassarman, 2018). This could be specific to oxygen stress in *F. nucleatum* but indicates a possible different biological role in the bacterium.

5.1.4 Riboregulatory elements

RNA-based regulators modulate gene expression by undergoing structural changes upon sensing a particular cue, such as temperature for RNA-thermometers of a small molecule for riboswitches. In particular, riboswitches can respond to vast number of different signals ranging from ions to larger molecules like the cobalamin (Breaker, 2022). The latter class is also the most abundant class of riboswitches in *F. nucleatum* indicating the requirement to tie cobalamin metabolism to the levels of the molecule itself in the bacterium.

Interestingly, RNA thermometers are absent in *F. nucleatum* as the bacterium likely has no need to have a temperature-regulated expression of virulence genes such as pathogens like *Listeria monocytogenes* (Johansson et al., 2002; Kortmann and Narberhaus, 2012).

5.1.5 A suite of sRNAs

In addition to the above mentioned RNAs, the combined computational and experimental approach identified for the first time a suite of sRNAs in *F. nucleatum*. The validation further showed that the bacterium encodes sRNAs of the four common categories (intergenic; 5'/3'UTR-derived; antisense) as found in a broad range of bacteria (Sharma et al., 2010; Ryan et al., 2020; Fuchs et al., 2021; Michaux et al., 2020; Hör et al., 2020b; Bischler et al., 2015).

Strikingly, most of these sRNAs are highly conserved within *F. nucleatum* despite the divergence between different subspecies (Kook et al., 2017). In fact, the sRNAs present in only a few strains of *F. nucleatum* are primarily transposase-associated and likely part of mobile genetic elements. Thus, the thorough analysis for sRNAs in *Fnn* likely identified the core set of sRNAs in *F. nucleatum*.

Beyond *F. nucleatum*, the identified sRNAs are only conserved in the closer related species *F. periodonticum* and *F. hwasookii*. This indicates that selection pressure likely caused the evolution of novel sRNAs in the other *Fusobacterium* species such as *F. necrophorum*. One could speculate that this might be related to the capability of active invasion which places *F. nucleatum*, *hwasookii* and *periodonticum* in a single lineage and distinguishes them from most remaining *Fusobacterium* species (Manson McGuire et al., 2014).

This clear disparity between the different species might reflect their different host association as *F. hwasooki*, *nucleatum* and *periodonticum* are primarily found in humans while *F. necrophorum* is also commonly found in animals.

Speculating on the function of the short RNAs, 3'UTR-derived sRNAs are often functioning in the pathway of their parental mRNA (Ponath et al., 2022a). This is especially interesting for FunR16, found to be co-transcribed with a stress operon containing the chaperone DnaK. The sRNA is primarily conserved at the secondary structure level in *F. nucleatum*, *F. hwasookii* and *F. periodonticum*. Remarkably, the sRNA Tpke11 is encoded in the same genomic location downstream of *dnak* in *E. coli*. While little is known about the positional ortholog Tpke11 beyond three putative targets (Mihailovic et al., 2018), it is possible that both Tpke11 and FunR16 are working in direct synergy or antagonism to DnaK.

A TSS could be detected for all sRNAs but FunR16 which enabled a targeted analysis of their promoter region. Interestingly, 12 ncRNAs possess the same σ^{70} motif identified for the majority of the CDS. This group of sRNAs includes FunR35 with constitutively high expression levels or FunR12 which decreases towards the stationary phase. This already suggest that despite the presence of the σ^{70} motif the expression of both sRNAs is likely controlled by additional factors.

sRNAs such as the σ^{E}-dependent MicA or the Fur-dependent RybD often harbor strong promoter sites for their respective regulator and depend on them for their transcription. At the very least, this principle is also valid for the sRNA FoxI in *F. nucleatum*. Thus, the conserved promoter regions of the fusobacterial sRNAs can be used as a starting point to computationally identify new regulatory networks for further experimental validation.

A recent RNA-seq approach combining 5'- and 3'-end sequencing uncovered a new class of sRNAs derived from within the ORF of mRNAs (Adams et al., 2021).

The dRNA-seq method alone is unable to discern such sRNAs and consequently additional regulatory RNAs might be hidden in the CDSs of *F. nucleatum*.

5.2 σ^E - a deeply rooted regulon

The majority of knowledge on the σ^E and its regulon has primarily been garnered from representative members of the γ-proteobacteria and overall defined the ECF's central role in defense against envelope stress (Hews et al., 2019). In this protective role σ^E upregulates genes that mediate the maintenance of the bacterial envelope by supporting proper protein folding and insertion of OMPs or lipoproteins into the OM. Taking the large phylogenetic distance between Proteobacteria and Fusobacteria into consideration (Coleman et al., 2021), it was thus surprising to discover several similarities between both σ^E architectures (Rhodius et al., 2006; Gogol et al., 2011), including genes involved in envelope homeostasis and the upregulation of a non-coding arm represented by the sRNA FoxI. At the same time these parallels appear discordant when considering that the 'classical' envelope stressor polymyxin B does not activate σ^E in *F. nucleatum* but rather oxygen highly induces the ECF in this anaerobic bacterium.

5.2.1 The coding arm

5.2.1.1 Regulating the outer layer

The global RNA-seq analysis of σ^E expression identified a large regulon (127 genes) with a similar size to its homolog in *E. coli* ($\sim$130 genes (Rhodius et al., 2006; Guo et al., 2014; Gogol et al., 2011; Rezuchova et al., 2003; Dartigalongue et al., 2001)).

Despite being small in numbers, the overlap of homologs between the two σ^E regulons is notable. In both species the expression of the periplasmic chaperone *skp* and the outer membrane assembly factor *bamA* is induced by σ^E and thus enables both proteins to cooperatively secure insertion of OMPs into the outer membrane (Schäfer et al., 1999; Knowles et al., 2009). LPS is another part of the envelope regulated by σ^E and also here the ECF regulon shares homologs as it upregulates both LpxD and LptB which are involved in generating of the outermost protective layer (Kelly et al., 1993; Sperandeo et al., 2007). Addition-

ally, despite the absence of sequence homology between *E. coli* and *F. nucleatum* another operon involved in LPS biogenesis is activated by σ^E *(rfaD-waaF-waaC-waaL* in *E. coli* (Dartigalongue et al., 2001); C4N14_05930-C4N14_09555 in *F. nucleatum* (Supplementary data 8.9). This highlights that regulation of different components of the outer part of the bacterial envelope is a shared function of the ECF in both species.

Interestingly, the fusobacterial σ^E shares three targets with its counterpart AlgU in *P. aeruginosa*: the rod-shape determining factor *mreB*, the general stress chaperone *clpB* and the signal recognition receptor *ftsY* (Schulz et al., 2015). While the latter suggest a role in protein translocation (see below), increased expression of ClpB and MreB might link σ^E into to broader stress response with their regulatory role in metabolism and chaperone activity (Alam et al., 2021) or in maintaining the cell shape (Errington, 2015), respectively.

In addition, our understanding of gene function and σ^E regulons in other more divergent species, such as the actinobacterium *Streptomyces coelicolor* is far less developed. However, an MreB homolog is also regulated by the ECF in this Gram-positive bacterium (Tran et al., 2019), albeit with a role in hyphae and spore formation (Mazza et al., 2006).

Nonetheless, these smaller overlaps among diverse phyla highlight a broad functional conservation of the σ^E regulon which may only be expanded by further research into the molecular biology of non-model bacteria such as *F. nucleatum*. Additionally, the result also suggest that the σ^E-mediated regulation, or parts of it, might have been already present in the LBCA when considering the early evolutionary divergence of the Fusobacteriota (Coleman et al., 2021).

5.2.1.2 Improving protein translocation

Next to *ftsY*, the additional members of the SEC-dependent protein translocation pathway, *secA* and *lepB*, were upregulated. Together, these proteins can synergistically increase translocation capacity of OMPs across the inner membrane and thus support the adaption and maintenance of the envelope (Leyton et al., 2012; Tsirigotaki et al., 2017). Yet, *lepB* or *secA* have not been found to be controlled by σ^E in other bacteria.

Further, the SEC-pathway is essential for the crossing of type 5 autotransporter into the cytoplasm where they can insert into the outer membrane (Tsirigotaki

et al., 2017). As Fap2 and other autotransporters are a vital part in the interaction with host cells (Coppenhagen-Glazer et al., 2015; Gur et al., 2015; Umaña et al., 2019; Casasanta et al., 2020) σ^E might further play a role in the fusobacterial pathogenesis as it does in different Enterobacteria (Hews et al., 2019).

5.2.1.3 Defense against oxygen

While σ^E is primarily involved in maintaining envelope homeostasis in other bacteria, yet different forms of membrane stress and alternative conditions did not activate the ECF in *F. nucleatum*. Rather atmospheric levels of oxygen led to rapid increase of σ^E activity which is further reflected in the shared upregulation of 23 genes between ECF induction and oxygen exposure. This overlap includes the dicistronic *ccdA-msrAB* which are both involved in the repair of oxidatively damaged proteins as part of a thiol-disulfide oxidoreductase system (Saleh et al., 2013). In this pathway, both proteins are required in the defense against reactive oxygen species (ROS) in different bacteria (Jalal and Lee, 2020; Saleh et al., 2013; Nasreen et al., 2021). Thus, it is likely that *F. nucleatum* uses σ^E to control expression of *ccdA-msrAB* as a response against the sensed oxygen that can damage or inactivate sensitive proteins (Lu and Imlay, 2021). Intriguingly, a paralogous operon of *ccdA-msrAB* regulated by the TCS ModRS plays an important role in protecting *F. nucleatum* from H_2O_2 but is not activated upon exposure to oxygen (Scheible et al., 2022). This suggests that *F. nucleatum* senses oxygen and H_2O_2 by different mechanisms.

Furthermore, *F. nucleatum* mounts an extensive response to oxygen including the differential expression of 15 transcription factors next to σ^E indicating that the ECF is not the only regulator of the oxygen-response. The transcriptional response to oxygen further includes the upregulation of peroxiredoxins as well as flavodoxins that synergistically support the resistance towards ROS (Kawasaki et al., 2005; Poole, 2005; Si et al., 2014). These genes could be especially relevant for *F. nucleatum*, as the bacterium does not harbor any catalases or superoxide dismutases (Kapatral et al., 2002). Additionally, oxygen highly induces a putative metal-uptake operon that could support the repair of metal-dependent proteins damaged under the stress (Lu and Imlay, 2021).

While *F. nucleatum* appears to employ known anti-ROS mechanisms to resist oxygen, the oxygen-dependent activation of σ^E may indicate an additional defense system against the molecule.

5.2.1.4 Activation of σ^E

The oxygen-specific activation σ^E raises the question of how *F. nucleatum* perceives this signal and ultimately activates the ECF. In *E. coli*, the protease DegS senses unfolded OMPs as a sign of envelope damage. After this, DegS initiates the proteolytic cascade and liberates σ^E from the anti-sigma factor RseA (Walsh et al., 2003; Ades et al., 1999; Alba et al., 2002). However, *F. nucleatum* lacks a DegS homolog and only encodes RseP which relies on prior cleavage by DegS for its own activity in *E. coli* (Alba et al., 2002; Kanehara et al., 2002; Li et al., 2009). While not excluding a role for the RseP homolog in the activation of σ^E in *F. nucleatum*, this fact suggests that the bacterium uses an alternative mechanism for sensing oxygen and relaying this environmental cue.

Such alternative activation pathways for σ^E are present in other bacteria: As an example, *V. parahaemolyticus* uses the phosphorylation of the σ factor EcfP to enable its association with the RNAP complex (Iyer et al., 2020). In contrast to EcfP, the proposed negatively charged 'DAED' motif important for the interaction with RNAP is present in the fusobacterial σ^E and thus likely does not require a substituting phosphorylation for its activity.

The two-component system CseBC in *Streptomyces coelicolor* senses envelope stress and activates σ^E (Paget et al., 1999; Tran et al., 2019). However, in *F. nucleatum* the two characterized TCS, CarRS and ModRS, are not activated by oxygen (Supplementary data 8.10). Similarly, the four remaining two-component systems in the bacterium were also unaffected by oxygen exposure but their role in σ^E activation cannot be entirely ruled out.

The anti-sigma factor ChrR of *Rhodobacter sphaeroides* has a dual-function to sense the ROS species singlet-oxygen and subsequently to activate σ^E (Anthony et al., 2005; Campbell et al., 2007; Greenwell et al., 2011). Intriguingly, the putative fusobacterial anti-sigma factor (C4N14_09825) contains several conserved amino acid residues in the N-terminus, which are likely exposed to the cytoplasm and could be involved in the sensing of oxygen. As ChrR in *R. sphaeroides* contains an entire domain dedicated to this process, it seems unlikely that a similar mechanism is involved in the activation of σ^E in *F. nucleatum*.

More common in bacteria is the use of co-factors, such as $[4Fe\text{-}4S]^{2+}$ or heme, to directly sense oxygen (Green et al., 2009). Thus, an additional protein partner containing such a co-factor could interact with the conserved N-terminal region of C4N14_09825 or act in an alternative indirect manner to control σ^E activation.

5.2.2 FoxI: the non-coding arm of σ^E

σ^E activates the transcription of sRNAs in Proteobacteria (see 1.1). In return, these sRNAs represent the non-coding arm of the σ^E response (Gogol et al., 2011) by repressing the translation of membrane proteins. The findings for the sRNA FoxI playing a similar role in *F. nucleatum* was thus surprising for two reasons: the early divergence of the *F. nucleatum* from other bacteria (Coleman et al., 2021) and the lack of any known sRNAs chaperones.

Despite the frequent role of sRNAs in regulatory pathways, the Fur-regulon and associated sRNAs currently present the only example of such a broad conserved regulatory circuit (Salvail and Massé, 2012). The widespread occurrence of Fur is likely linked to its role in controlling iron metabolism which is essential for all bacteria (Andrews et al., 2003). Additionally, the Fur-dependent sRNAs like RyhB in *E. coli* (Massé and Gottesman, 2002) and FsrA in *B. subtilis* (Gaballa et al., 2008) further extend the regulatory potential of Fur during iron-limitation (Carpenter et al., 2009; Oglesby-Sherrouse and Murphy, 2013).

In comparison, σ^E-dependent sRNAs have been found only in the phylum of Proteobacteria thus far (Table 1.1). The discovery of FoxI as a σ^E-dependent sRNA in *F. nucleatum* now adds an additional phylum to this regulatory principle. Further, FoxI represses the translation of abundant envelope proteins such as FomA, the putative MglBAC galactose uptake operon and, the leading genes of paralogous operons containing type 5a autotransporters. This is paralleling the function of the σ^E-dependent sRNAs such as MicA or RybB, which block translational of several OMPs (Gogol et al., 2011). While the biological function of FoxI activity is yet to be fully understood, the validated targets of the sRNA suggest that repression of the different envelope proteins lessens the burden on the SEC- and BAM-dependent pathway for insertion of proteins into the outer membrane. Supporting this is a reduced σ^E activity represented by decreased expression of the *rpoE* operon and the σ^E-dependent operon C4N14_03280-C4N14_03285 when expressing FoxI in the $\Delta foxI$ background (Fig. 4.15).

Overall these findings indicate that FoxI acts synergistically with σ^E in maintaining envelope homeostasis in *F. nucleatum*. Importantly, the shown role for FoxI parallels the established non-coding arm of σ^E in *E. coli* suggesting that this regulatory principle could represent a second widespread stress response in addition to the Fur regulon.

However, the observation that gene repression by σ^E did not solely rely on FoxI indicates that *F. nucleatum* harbors additional σ^E-dependent sRNAs with an overlapping targetome. A likely candidate is the FunR52, due to the presence of a conserved σ^E promoter motif (Fig. 5.1A). In fact, FunR52 harbors a cytosine-rich stretch similar to FoxI and target predictions suggest promising binding sites with validated FoxI target mRNAs like *fomA* or C4N14_02035 (Fig. 5.1B).Thus, further work in FunR52 might uncover a similar functional redundancy with FoxI as shown for the σ^E-dependent sRNAs in *E. coli* (MicA & RybB (Gogol et al., 2011)) and *V. cholerae* (VrrA & MicV (Peschek et al., 2019)).

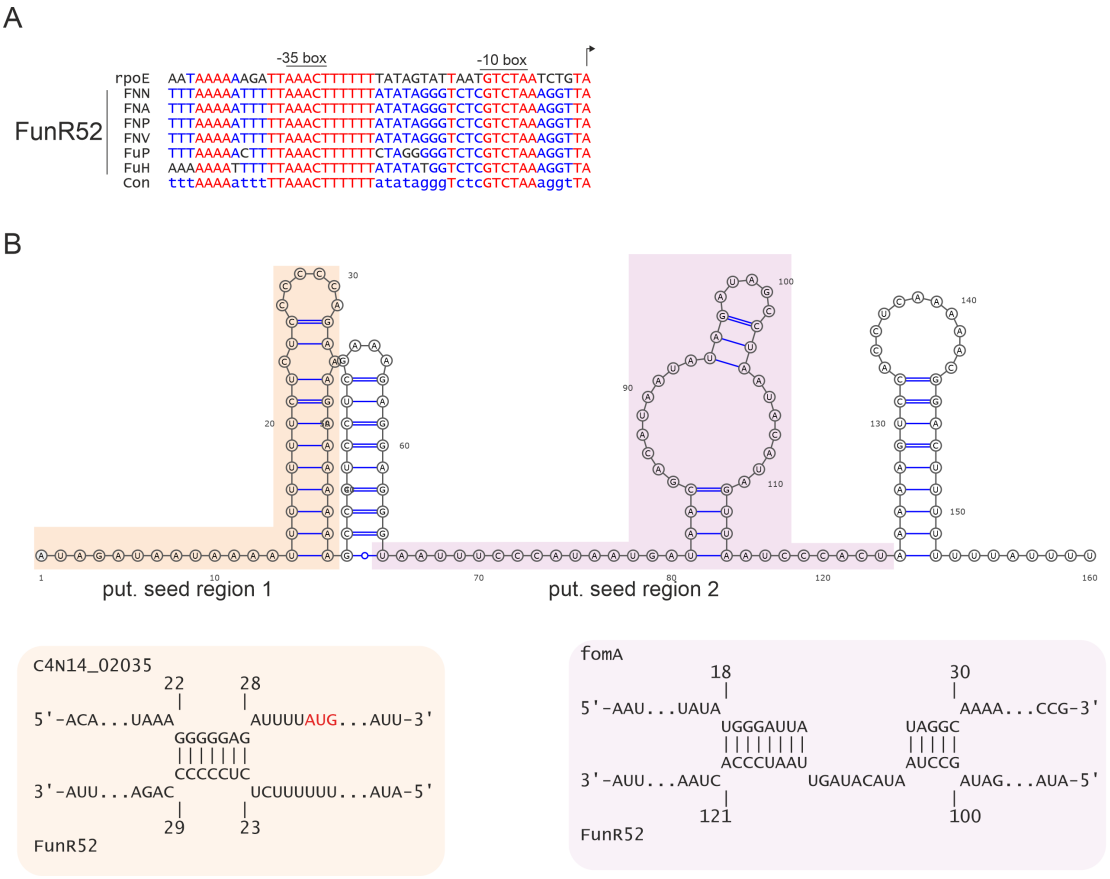

Figure 5.1 | FunR52 - a 2^{nd} σ^E-dependent sRNA?
A. Sequence alignment of the promoter region for representative sequences of FunR52 and *rpoE* of *Fnn*. **B.** Prediction of the secondary structure of FunR52. Below the predicted RNA-RNA interaction of FunR52 with C4N14_02035 and *fomA* mRNA is shown with the interacting region on FunR52 indicated in orange and purple, respectively.

5.2.2.1 The mode-of-action of FoxI

Based on the translational reporter system, FoxI regulates its targets such as FomA or MlgB primarily on the translational level with only a weaker impact on the mRNA level. This is in contrast to sRNAs in other organisms where sRNA-mediated translational repression of a target typically also causes degradation of the mRNA as a secondary effect (Caron et al., 2010). For this these sRNAs often act together with the RNA chaperone Hfq which not only supports the sRNA-mRNA interaction but also boosts RNA degradation by recruiting RNase E and the degradasome (De Lay et al., 2013). Consequently, the absence of Hfq or other sRNA-associated proteins in *F. nucleatum* may explain why FoxI-mediated post-transcriptional regulation is unable to promote strong RNA decay.

The cytosine-rich stretch of FoxI is crucial for the sRNA to bind the identified targets as disruption of the seed region abrogates the sRNA's activity. Such cytosine-rich stretches can also be found for additional fusobacterial sRNAs including FunR12, FunR47 and FunR52 that are possibly similarly involved in the base-pairing with their respective targets as found for FoxI. In fact, cytosine-rich motifs are also important in other bacteria's sRNAs for recognizing their mRNA targets (Geissmann et al., 2009; Papenfort et al., 2008; De Lay and Gottesman, 2009), implying that these stretches are elementary seed regions.

5.3 The importance of genetic tools

The ability to genetically manipulate bacteria plays a central role in deciphering their molecular biology and further our understanding of their general biology. Initially, tools that facilitate this have primarily been established in the model-organism *E. coli* (Gagarinova and Emili, 2012) and this member of the γ-proteobacteria is still the an important cornerstone of developing new methods to study bacteria such as CRISPRi (Qi et al., 2013; Bikard et al., 2013), base-editing (Banno et al., 2018; Zheng et al., 2018; Tong et al., 2021), RNA-guided transposon insertion (Strecker et al., 2019) or cellular recording (Schmidt et al., 2018; Tang and Liu, 2018). While these tools can often be successfully applied in other Proteobacteria, genetic manipulation of non-model species has always presented a hurdle in the respective fields. For instance the anaerobe *B. thetaiotaomicron* only recently became model organism of the human gut commensals due to advances in the genetics of the bacterium (Horn et al., 2016; Whitaker et al., 2017; Lim et al., 2017).

5.3.1 Understanding post-transcriptional regulation

Despite the increasing interest in *F. nucleatum*, we were still lacking basic components in its genetic toolbox (Brennan and Garrett, 2019). Only after more than 15 years since the first reported gene inactivation in the bacterium (Haake et al., 2000) has a new deletion system been developed (Casasanta et al., 2020; Wu et al., 2018). However, removal of a gene represents only one part in dissecting gene function with constitutive or controlled gene expression being additional important methods towards reaching this goal. As applied for the sRNA FoxI in this work, exogenous expression of sRNAs has been widely used to identify their respective mRNA targets (Massé et al., 2005; Papenfort et al., 2006; Rasmussen et al., 2009;

Ryan et al., 2020; Chen et al., 2004).

Defining the sRNA-mRNA interaction was made easier by development of a translational reporter system (Urban and Vogel, 2007). Applied to *F. nucleatum*, this principle has also enabled the understanding of the mode-of-action for the FoxI sRNA and expanded its targetome. As FoxI-mediated regulation does not cause rapid RNA decay for the identified targets using the here established translational reporter will be important to further investigate sRNAs and their post-transcriptional regulation in *F. nucleatum*.

However, post-transcriptional regulation in bacteria is not exclusively carried out by sRNAs. A variety of RBPs such as Hfq, ProQ or CsrA but also ribonucoleases and RNA helicases are important post-transcriptional regulators (Arraiano et al., 2010; Vakulskas et al., 2014; Holmqvist and Vogel, 2018). And while all common sRNA-associated proteins like Hfq are absent in *Fusobacterium*, a putative RNase adaptor protein with similarities to RapZ from *E. coli* can be found in the genus and is likely involved in post-transcriptional regulation together with an RNase (Göpel et al., 2013; Durica-Mitic et al., 2020). *F. nucleatum* also encodes the ubiquitous transcription termination associated proteins Rho and NusA which can act as broad transcriptional attenuators (Holmqvist and Vogel, 2018). Thus, the development of the translation reporter opens new possibilites for understanding post-transcriptional regulation in *F. nucleatum*.

5.3.2 Fluorescent proteins

Fluorescent proteins, such as GFP, have revolutionized our scientific approach and helped us to understand different aspects of bacterial biology including persister formation (Helaine et al., 2014; Stapels et al., 2018) or the interaction with host cells (Westermann et al., 2016; Westermann and Vogel, 2021). The only limitations of GFP and related fluorescent proteins is their requirement for molecular oxygen to form a functional fluorophore (Tsien, 1998). This impedes the application of these proteins in anaerobe bacteria as the fluorescence signal can only be measured after fluorophore maturation in an aerobic environment requiring either aerotolerance of the bacterium or prior fixation. Only a few oxygen-independent alternatives are currently available, such the FMN-binding LOV-domain proteins (Drepper et al., 2007), FAST proteins (Plamont et al., 2016) or RNA aptamers (Ouellet, 2016) which do not reach the brightness of the oxygen-dependent fluorescent proteins and in case of the last two examples require a ligand.

Nonetheless, the establishment of fluorescent proteins in *F. nucleatum* repre-

sents an important advance that goes beyond investigating post-transcriptional regulation. As translational fusions with mCherry are feasible in *F. nucleatum*, the fluorescence protein could be used to determine cellular localization or interaction of proteins within the bacterium (Shapiro et al., 2009; Alexeeva et al., 2010).

Monitoring individual gene activity or entire signaling pathways is a prominent application of fluorescent proteins since its early use to understand virulence regulation in *Salmonella* (Valdivia and Falkow, 1996; Lee et al., 2000; Kim et al., 2003). As such, further use of the fluorescent proteins might similarly allow a deeper understanding of regulation of *F. nucleatum* during its interaction with cancer cells and the tumor-environment which only three studies have touched upon so far (Cochrane et al., 2020; Parhi et al., 2020; Despins et al., 2021). However, all studies performed transcriptomics on the entire cell population thereby not differentiating between free, attached or intracellular *F. nucleatum*. Lacking this discrimination poses a limitation on what information we can extract from these types of experiments, such as if host-associated *F. nucleatum* expresses a possible virulence signature to modulate the host, which is lost when sequencing the entire bacterial population. Thus, using fluorescence-activated cell sorting (FACS) to discriminate infected cells from non-infected cells, termed dualRNA-seq and pioneered in *Salmonella* (Westermann et al., 2016), can broaden our understanding of such a specific fusobacterial response to the host and similarly inform us on transcriptional events in the host (Westermann and Vogel, 2021).

In this regard, fluorescently tagged *F. nucleatum* will also improve our detection of the bacterium during infection studies which currently rely on exogenous labeling (Casasanta et al., 2020) thereby limiting especially long-term experiments. These fluorescently-tagged bacteria could also be used for mice experiments allowing the clear localization of *F. nucleatum* in the tissue and enable powerful approaches such as tissue-dualRNA-seq to gain insights into the host-pathogen interplay in a complex model (Damron et al., 2016; Nuss et al., 2017). In the future, dualRNA-seq could further be improved to allow single-cell resolution (Westermann and Vogel, 2021), reaching an unprecedented understanding of the established heterogeneity for the host (Longo et al., 2021) and bacterial cells (Imdahl et al., 2020; Kuchina et al., 2021; Blattman et al., 2020) alike during the infection process.

While the use of fluorescent proteins in *F. nucleatum* is promising, there is currently one major limitation for live-cell approaches: the requirement for oxygen (Heim et al., 1994). However, fluorescent alternatives exist, including the oxygen-independent reporters such as flavin-binding fluorescent proteins (Mukherjee et al., 2012), Fluorescence-Activating and absorption-Shifting Tag proteins (FAST) (Plamont et al., 2016) or RNA-aptamers (Zhang et al., 2015). In addition, *F. nucleatum* can also grow under micro-aerobic conditions (Diaz et al., 2002) and therefore experimental conditions, such as those for infection studies, could be adjusted to accommodate the maturation of oxygen-dependent fluorescent proteins.

5.3.3 Towards high-throughput approaches

Large forward genetic studies are a powerful approach to probe for gene function in the larger biological context in an unbiased manner. This includes transposon mutagenesis which is a broadly applicable method (Reznikoff, 2008) and has been applied in *F. nucleatum* to uncover the role of FtsX and EnvC in cell division (Wu et al., 2018), the two TCS CarRS and ModRS (Wu et al., 2021; Scheible et al., 2022) or the adhesin Fap2 (Coppenhagen-Glazer et al., 2015).

More recently CRISPR-Cas-based technologies have revolutionized forward genetic approaches with a particular role for CRISPR interference (CRISPRi) (Todor et al., 2021). However, constitutive expression of the required Cas nuclease is usually toxic for the bacteria (Zhang and Voigt, 2018). Overcoming this hurdle may be provided by the here developed inducible gene expression system that could allow the application of CRISPR-based methodologies in *F. nucleatum* due to its tight regulation (see Chapter 3.4). The use of e.g. CRISPRi could thus enable the global screening for functions of essential and non-essential genes under various conditions thereby greatly accelerating knowledge-gain in this non-model bacterium (Todor et al., 2021; Call and Andrews, 2022).

Yet, achieving higher throughput with transposon-inserted sequencing has not been conducted in this bacterium (Cain et al., 2020). This is in part due to *F. nucleatum* being recalcitrant to genetic transformation (Brennan and Garrett, 2019). As highlighted in a recent pre-print (Umaña et al., 2022), transformation efficiency can be improved by overcoming the restriction modification system of *F. nucleatum* and thus might allow more efficient generation of large libraries.

6 Conclusion & Outlook

This work generated fundamental knowledge on the organization of transcriptome for *F. nucleatum* by globally identifying TSS, 5'UTRs and sRNAs. This RNA-based map further enabled the elucidation of the σ^E response in the bacterium and with it provides evidence of a broadly conserved regulatory principle for the ECF. The latter could only be accomplished through the herein developed genetic tools that allow the probing of single genes or larger regulatory networks.

RNA-seq has become a standard method in studying bacteria and their gene function by allowing to place observations into a global cellular context. While RNA-seq presents an invaluable technique, the proper interpretation of the data requires a complete understanding of all transcribed elements of an organism which is currently impossible by computational predictions alone. As such, single-nucleotide resolution present an important resource for the study of pathogens and commensal (Sharma et al., 2010; Sharma and Vogel, 2014; Ryan et al., 2020; Fuchs et al., 2021; Michaux et al., 2020; Kröger et al., 2018; Gill et al., 2018; Wicke et al., 2021; Cervantes-Rivera et al., 2020). As RNA-seq is increasingly applied to investigate *F. nucleatum* (Mutha et al., 2018; Cochrane et al., 2020; Wu et al., 2021; Despins et al., 2021; Scheible et al., 2022), the here generated RNA-guided annotation will benefit future transcriptomics studies. Knowledge on the transcript boundaries and operon structures will also aid in correctly evaluating forward genetic approaches such as transposon screens. While 3'UTR can be predicted through the presence of terminators, a global profiling of 3'-ends by Term-seq (Dar et al., 2016) or alternative approaches (Shishkin et al., 2015; Yan et al., 2018; Ju et al., 2019; Fuchs et al., 2021) would further improve the current annotation.

The up- and downstream region of TSS in bacteria contain valuable information regarding their transcriptional and post-transcriptional regulation. Thus, the generated TSS annotation now enables a full investigation of these processes *F. nucleatum* as demonstrated by uncovering the σ^E regulon acting on both the

transcriptional and post-transcriptional level. This will especially aid in probing regulatory networks of *F. nucleatum* in the interaction with cancer cells, which remain largely unknown.

In light of missing known RNA chaperones in *F. nucleatum*, it is surprising that the sRNA-mediated regulation operates similar as in e.g. *E. coli* in which the RBP Hfq plays a crucial role for the function of sRNAs (Holmqvist and Vogel, 2018). This observation suggests that *F. nucleatum* harbors an alternative RBP working cooperatively with its sRNAs. The validated sRNA could be utilized to identify such RBPs by using them as bait in pulldown experiments typically applied for sRNA-target identification (Lalaouna et al., 2015; Tomasini et al., 2017; Correia Santos et al., 2021). Alternatively, global approaches such as Grad-seq or GradR (Smirnov et al., 2016; Gerovac et al., 2020) present an unbiased approach to uncover such RBP in association to provide insights into RNA-Protein and Protein-Protein complexes (Gerovac et al., 2021). The discovery of an RBP would be quite valuable as it enables the use of powerful approaches to globally dissect post-transcriptional regulation in *F. nucleatum* (Hör et al., 2020b) and with it accelerate our knowledge gain for these networks. Interestingly, *F. nucleatum* encodes for a homolog of KhpB, which represents an emerging class of sRNA-associated RBPs (Zheng et al., 2017; Hör et al., 2020a; Lamm-Schmidt et al., 2021; Olejniczak et al., 2021), and thus might be the missing sRNA chaperone in this species.

RBP-independent approaches such GRIL- or HiGRIL-seq could also be applied in *F. nucleatum* to globally explore the targetome of sRNAs (Han et al., 2016; Zhang et al., 2017).

The new genetic tools present another important advances for probing *F. nucleatum* by providing the means to investigate individual regulatory networks as shown in this work but also to gain a more system-wide understanding. Regarding the latter, applying either reporter system to the Sort-seq approach (Belliveau et al., 2018) could identify specific sequence elements for transcriptional and post-transcriptional regulation in a high-throughput manner. Furthermore, the inducible expression system could facilitate the study of gene regulation and function in an unprecedented manner for *F. nucleatum* by allowing the application of the ever-expanding list of CRISPR-Cas-based tools (Barrangou and van Pijkeren, 2016; Vigouroux and Bikard, 2020; Pickar-Oliver and Gersbach, 2019).

Simultaneously resolving the transcriptomic changes of the bacteria and host by dualRNA-seq can provide a detailed picture of the physiological state of bacteria and host cells during the infection (Westermann et al., 2016, 2017). This can now be similarly achieved for *F. nucleatum* as the fluorescently tagged bacteria allow the differentiation of infected cells from by-standers and thus could help to identify the specific networks activated in either organism during the infection. Additionally, the approach could also be expanded to triple RNA-seq (Seelbinder et al., 2020),or potentially multi-species RNA-seq, to study the transcriptional responses of *F. nucleatum* and members of the microbiome readily binding to the cancer-associated bacterium (Bradshaw et al., 1998; Wu et al., 2015; Mutha et al., 2018; Engevik et al., 2020).

And with ongoing works into overcoming the hurdle of the genetic intractability of *F. nucleatum* (Umaña et al., 2022), the tools and technologies will become valuable in investigating also highly invasive strains of *F. nucleatum* subsp. *animalis* (Strauss et al., 2011) and enable detailed comparative approaches for the diverse species of *F. nucleatum* (Kook et al., 2017; Manson McGuire et al., 2014).

Importantly, such a deeper understanding of *F. nucleatum*'s biology might open new avenues into specifically removing the bacterium from diseased body sites by applying targeted antisense oligonucleotide technologies (Vogel, 2020), programmable phage therapy (Hatfull et al., 2022) or small molecules (Johnson and Abramovitch, 2017).

In the context of above mentioned possibilities, this work does not only shed light on aspects of *F. nucleatum*'s molecular biology but also presents a valuable foundation for future studies into the bacterium, while also presenting a potential 'blueprint' for approaching similarly unknown members of the human microbiome.

7 Material & methods

7.1 Materials

7.1.1 Instruments

Table 7.1 | List of used instruments.

Instruments	Manufacturer
Anaerobic chamber	Coy
Cam12 hydrogen/oxygen detector	Coy
Centrifuge Eppendorf 5415R	Eppendorf
Centrifuge Eppendorf 5424	Eppendorf
Centrifuge Hereaus Multifuge X3R	Thermo Scientific
Desiccant/Catalyst Stak-Pak for atmospheric control	
Electroporator MicroPulser	Bio-Rad
Epoch 2 Microplate Spectrophotometer	BioTek
FastPrep-24	MP Biomedicals
Forced Air Incubator for bacterial plates	Coy
GelStick Imager	Intas
Heat block Thermomixer comfort	Eppendorf
Horizontal gel electrophoresys Perfect Blue Mini S, M, L	PeqLab
Hybridization oven HP-1000	UVP
Hydrogen Sulfide Removal Column	Coy
ImageQuant LAS 4000 imaging system	GE Healthcare
Incubator for bacterial plates	Memmert
Incubator for eukaryotic plates HERAcell 150i	Thermo Scientific
NovoCyte Quanteon flow cytometer	Agilent

Table 7.1 *continued from previous page*

Instrument and software	Manufacturer
Phosphoimager Typhoon FLA 7000	GE Healthcare
Power supply peqPOWER E250, E300	PeqLab
RealTime CFX96 System	Bio-Rad
Research plus pipettes	Eppendorf
Rotator SB2	Stuart
Scale 572	Kern
Semi-dry electroblotter Perfect Blue SEDEC M	PeqLab
Shaking incubator Innova 44	New Brunswick Scientific
Sonicator Sonoplus HD 70	Bandelin
Spectrophotometer NanoDrop 2000	PerkinElmer
Spectrophotometer Ultraspec 10 Cell Density Meter	Amersham Biosciences
Table-top ultracentrifuge optima MAX-XP	Beckman Coulter
TCS SP5 confocal microscope	Leica
Thermal cycler MJ Mini	Bio-Rad
Ultracentrifuge Optima XP-80	Beckman Coulter
Ultracentrifuge rotor SW 40 Ti	Beckman Coulter
Ultracentrifuge rotor SW 60 Ti	Beckman Coulter
UV crosslinker (254 nm)	Vilber
Vertical gel electrophoresis Perfect Blue Twin	PeqLab
Vortex Genie 2	Scientific Industries
Waterbath 1092	GFL
Western blot imaging system Amersham ImageQuant 800	Cytiva

7.1.2 Consumables

Table 7.2 | List of used consumables.

Consumables	Manufacturer
0.45µm polyethersulfone membrane	Millipore
12-well plates	Corning

6-well plates	Corning
96-well plates	Nunc
Bolt 4–12% Bis-Tris gels	Thermo Fischer Scientific
Centrifuge tubes	Sarstedt
Cuvettes	Sarstedt
Disposale glass pipettes	Kimble
Electroporation cuvettes	Cell projects
FastPrep tubes	MP Biomedicals
Fisherbrand cell scrapers	Fischer Scientific
G-25 MicroSpin columns	GE Healthcare
G-50 MicroSpin columns	GE Healthcare
Glass beads 0.1 mm	Roth
Glass bottles	Schott
Hard-Shell 96-Well PCR Plates	Bio-Rad
Hybond-XL membranes	GE Healthcare
PCR tubes	Thermo Fischer Scientific
Petri dishes	Corning
Phase lock gel tubes 2ml	5 Prime
Phosphor screen	Fujifilm
Pipetboy acu-jet pro	BRAND
Pipette tips	Sarstedt
PVDF membrane	GE Healthcare
Safe-lock tubes 1.5 mL, 2 mL	Eppendorf
Sierological pipettes (plastic)	Greiner bio-one
Ultracentrifugation tubes	Seton

7.1.3 Chemicals and reagents

Table 7.3 | List of chemicals and reagents used.

Chemicals and reagents	Manufacturer
γ -32P-ATP	Hartmann Analytic
1,4-Dithiothreitol (DTT)	Roth
Acetone	Roth
Anhydrotetracylcin	Roth
bile acids	Sigma-Aldrich
Brain-Heart-Infusion (BHI)	Gibco
Chloramphenicol	Roth
Columbia Broth	Gibco
diamide	Sigma-Aldrich
Dimethyl dulfoxide (DMSO)	Roth
DNA loading buffer (6x)	Thermo Fisher Scientific
dNTPs	Thermo Fisher Scientific
ECL western blot detection reagent	GE Healthcare
EDTA	Roth
Ethanol	Roth
Ethanol absolute	Merck
Fetal calf serum (FCS)	Biochrom
Formaldehyde	Roth
Gel loading buffer II	Ambion
GeneRuler 1 kb DNA Ladder	Fermentas
Glycerol (99%)	Sigma
GlycoBlue	Thermo Fisher Scientific
hydrogen peroxide	AppliedChem GmbH
Isopropanol	Roth
LDS sample buffer	Thermo Fischer Scientific
Methanol	Roth
Milk powder	Roth
mitomycin C	Sigma-Aldrich
N-lauroylsarcosine sodium salt	Sigma-Aldrich
PageRuler Plus prestained protein ladder	Thermo Fischer Scientific
PBS	Gibco
Polymyxin B	Sigma-Aldrich

Table 7.3 *continued from previous page*

Chemicals and reagents	Manufacturer
Propidium iodide	Sigma
pUC mix marker, 8	Fermentas
RedSafe	ChemBio
RNA ladder high and low range	Fermentas
Roti-Aqua P/C/I	Roth
Roti-Blue	Roth
Roti-Hybri-Quick	Roth
Rotiphorese gel 40 (19:1)	Roth
Rotiphorese gel 40 (37.5:1)	Roth
SimplyBlue Coomassie	Thermo Fischer Scientific
S-nitrosoglutathione	Sigma-Aldrich
Sodium desoxycholate	Sigma-Aldrich
SYBR Gold Nucleic Acid Gel Stain	Invitrogen
Thiamphenicol	Roth
Trichloroacetic acid	Sigma
Triton X-100	Sigma

7.1.4 Enzymes and kits

Table 7.4 | List of used enzymes.

Enzyme	Manufacturer
Calf intestinal phosphatase (CIP)	NEB
DNaseI	Thermo Fischer Scientific
Lysozyme	Roth
Phusion DNA polymerase	NEB
Polynucleotide kinase (PNK)	Thermo Fischer Scientific
Restriction enzymes	Thermo Fischer Scientific
RNase inhibitor	Thermo Fischer Scientific
SUPERaseIN RNase inhibitor	Ambion
SuperScript II /IIIreverse transcriptase	Thermo Fischer Scientific
T4 DNA ligase	NEB
Taq DNA polymerase	NEB

Trypsin-EDTA	Gibco

Table 7.5 | List of used commercial kits.

Commercial kits	Manufacturer
DNA Cycle Sequencing kit	Jena Bioscience
MEGAscript T7 Kit	Ambion
M-MLV Reverse Transcriptase	Thermo Fisher
NEBuilder HIFI DNA Assembly Mix	NEB
NucleoSpin Gel and PCR clean-up	Macherey-Nagel
NucleoSpin Plasmid EasyPure	Macherey-Nagel
Takyon No ROX SYBR 2X MasterMix blue dTTP	Eurogentec

7.1.5 Antibodies

Table 7.6 | List of used antibodies.

Antibody (source)	Working dilution	Provider
	1:200 in 3% BSA	Thermo Scientific
	1:3,000 in 3% BSA	Thermo Scientific
	1:10,000 in 3% BSA	Thermo Scientific
	1:10,000 in 3% BSA	Eurogentec

Table 7.7 | List of used solutions and buffers.

Solution/Buffer	Components
Buffer for the isolation of the inner membrane	lysis buffer with 0.5% N-lauroylsarcosyn
Buffer for the isolation of the outer membrane	ysis buffer with 2% triton X-100
Columbia broth (Col. B.)	37g Columbia broth; H_2O fill to 1l
DNA loading dye (5x stock)	10 mM Tris-HCl pH 7.6; 60% (v/v) glycerol; 60 mM EDTA pH 8.0; 0.025% (w/v) bromophenol blue
Ethanol/sodium acetate 30:1	29 ml ethanol; 1 ml 3M sodium acetate pH 6.5

Table 7.7 *continued from previous page*

Solution/Buffer	**Components**
Lennox broth (LB)	10 g tryptone; 5 g yeast extract; 5 g NaCl; H_2O fill to 1 l
Lysis buffer for cell fractionation	20 mM Tris HCl pH 7.5; 150 mM KCl; 1 mM $MgCl_2$; 1 mM PMSF
modified Brain-Heart-Infusion agar (BHI-C)	38g BHI; 5g yeast extract; 20g agar; 1% (w/V) glucose; 10ml hemin (5µg/ml);10ml FCS; H_2O fill to 1l
PAA running gel electrophoresis solution; for proteins	3.75 ml lower buffer; 3 or 3.25 ml Rotiphorese gel 40 (37.5:1) and 3.25 or 3 ml H_2O for 12 or 15% gels; 75 μl 10% (w/v) APS; 7.5 μl TEMED
PAA stacking gel electrophoresis solution; for proteins	1.25 ml upper buffer; 1 ml Rotiphorese gel 40 (37.5:1); 7.5 ml H_2O; 90 μl 10% (w/v) APS; 9 μl TEMED
PBS (10x)	2 g KCl; 2.4 g KH_2PO_4; 80 g NaCl; 14.4 g Na_2HPO_4; adjust the pH to 7.4; H_2O fill to 1 l
Polyacrilamide (PAA) gel electrophoresis solution (6%); for RNA	100 ml 10x TBE; 420 g urea; 150 or 100 ml Rotiphorese gel 40 (19:1) for 6% or 4% gels; H_2O fill to 1 l
Protein loading dye (5x)	15 g SDS; 46.95 ml 1M Tris-HCl; pH 6.8; 75 ml glycerol; 11.56 g DTT; 0.075 g bromophenol blue; H_2O fill to 150 ml
RNA loading dye (2x)	0.025% (w/v) bromophenol blue; 0.025% (w/v) xylene cyanol; 18 μM EDTA pH 8; 0.13% (w/v) SDS; 95% formamide
SDS running buffer (10x)	30.275 g Tris base; 144 g glycine; 10 g SDS; H_2O fill to 1 l
SSC buffer (20x stock)	3M NaCl; 0.3M sodium citrate pH 7
STOP mix	95% ethanol; 5% acidic phenol
TAE buffer (50x stock)	242 g Tris base; 51.7 ml acetic acid; 10mM EDTA pH 8; H_2O fill to 1 l
TBE buffer (10x stock)	108 g Tris base; 55 g boric acid; 20mM EDTA pH 8; H_2O fill to 1 l
TBS buffer (10x stock)	24.11 g Tris base; 72.6 g NaCl; adjust to pH 7.4 with HCl; H_2O fill to 1 l

Table 7.7 *continued from previous page*

Solution/Buffer	Components
TBS-T buffer (10x stock)	1 TBS; 0.1% (v/v) Tween-20
Transfer buffer (10x stock)	3 g Tris base; 14.4 g glycine; 200 ml methanol; H_2O fill to 1 l
Tris running gel solution	1.5 M Tris-HCl; pH 8.8; 0.4% (w/v) SDS
Tris stacking gel solution	0.5M Tris-HCl pH 6.8; 0.4% (w/v) SDS

Table 7.8 | List of used plasmids.

Name	Description
pEcoFus	empty vector for expression of gene of interest; under control of 4.5S RNA promoter
pEcoFus-FoxI	contstitutive overexpression of FoxI
pEcoFus-FoxI-3C	contstitutive overexpression of FoxI-3C
pEcoFus-FunR12	contstitutive overexpression of FunR12
pEcoFus-FunR16	contstitutive overexpression of FunR16
pEcoFus-FunR47	contstitutive overexpression of FunR47
pEcoFus-FunR7	contstitutive overexpression of FunR7
pEcoFus-FunR7-ext. 3'end	contstitutive overexpression of FunR7+ext. 3'end
pFP14	fusobacterial origin of replication replacing CDIFF origin of replication in pRPF185
pFP275	transcriptional fusion *miaB*::mCherry
pFP276	transcriptional fusion *miaB*::mCherry point mutation in -10 (C to G)
pFP277	transcriptional fusion C4N14_09130::mCherry
pFP278	transcriptional fusion C4N14_09130::mCherry point mutation in -10 (C to G)
pFP279	transcriptional fusion C4N14_03280::mCherry
pFP280	transcriptional fusion C4N14_03280::mCherry point mutation in -10 (C to G)

Table 7.8 *continued from previous page*

Name	Description
pFP281	transcriptional fusion *rpoE*::mCherry
pFP282	transcriptional fusion *rpoE*::mCherry point mutation in -10 (C to G)
pFP283	transcriptional fusion FoxI::mCherry
pFP284	transcriptional fusion FoxI::mCherry point mutation in -10 (C to G)
pFP287	C4N14_09375::mCherry 5UTR+10aa
pFP288	C4N14_09375::mCherry 5UTR+10aa + FoxI OE
pFP289	C4N14_09375::mCherry 5UTR+10aa + FoxI-3C OE
pFP290	mglB::mCherry 5UTR+10aa
pFP291	mglB::mCherry 5UTR+10aa + FoxI OE
pFP292	mglB::mCherry 5UTR+10aa + FoxI-3C OE
pFP296	fadA3a::mCherry 5UTR+10aa
pFP297	fadA3a::mCherry 5UTR+10aa + FoxI OE
pFP298	fadA3a::mCherry 5UTR+10aa + FoxI-3C OE
pFP299	fadA3b::mCherry 5UTR+10aa
pFP300	fadA3b::mCherry 5UTR+10aa + FoxI OE
pFP301	fadA3b::mCherry 5UTR+10aa + FoxI-3C OE
pFP305	C4N14_02035::mCherry 5UTR+10aa
pFP306	C4N14_02035::mCherry 5UTR+10aa + FoxI OE
pFP307	C4N14_02035::mCherry 5UTR+10aa + FoxI-3C OE
pFP308	fvcD::mCherry 5UTR+10aa
pFP309	fvcD::mCherry 5UTR+10aa + FoxI OE
pFP310	fvcD::mCherry 5UTR+10aa + FoxI-3C OE
pFP311	C4N14_00275::mCherry 5UTR+10aa
pFP312	C4N14_00275::mCherry 5UTR+10aa + FoxI OE
pFP313	C4N14_00275::mCherry 5UTR+10aa + FoxI-3C OE
pFP314	fomA::mCherry 5UTR+10aa
pFP315	fomA::mCherry 5UTR+10aa + FoxI OE
pFP316	fomA::mCherry 5UTR+10aa + FoxI-3C OE
pFP423	P2 generated with ssOligo JVO-20857 (constitutive mScarlet expression)
pFP424	P3 generated with ssOligo JVO-20858 (constitutive mScarlet expression)

Table 7.8 *continued from previous page*

Name	Description
pFP425	P4 generated with ssOligo JVO-20859 (constitutive mScarlet expression)
pFP426	P5 generated with ssOligo JVO-20860 (constitutive mScarlet expression)
pFP427	P6 generated with ssOligo JVO-20861 (constitutive mScarlet expression)
pFP428	P7 generated with ssOligo JVO-20862 (constitutive mScarlet expression)
pFP429	P8 generated with ssOligo JVO-20863 (constitutive mScarlet expression)
pFP430	P9 generated with ssOligo JVO-20864 (constitutive mScarlet expression)
pFP431	P10 generated with ssOligo JVO-20865 (constitutive mScarlet expression)
pFP432	P11 generated with ssOligo JVO-20866 (constitutive mScarlet expression)
pFP433	P12 generated with ssOligo JVO-20867 (constitutive mScarlet expression)
pFP434	P13 generated with ssOligo JVO-20868 (constitutive mScarlet expression)
pFP435	P15 generated with ssOligo JVO-20870 (constitutive mScarlet expression)
pORI92	source for fusobacterial origin of replication
pRPF185	backbone for generating pEcoFus
pVoPo-00	improved backbone
pVoPo-01	base vector for transcriptional reporter
pVoPo-02	base vector for translational reporter
pVoPo-03	base vector for inducible expression system
pVoPo-03-FoxI	inducible expression of FoxI
pVoPo-03-FoxI-3C	inducible expression of FoxI-3C
pVoPo-03-FoxI-C4A	inducible expression of FoxI-C4A
pVoPo-03-mazF	inducible expression of *mazF*

Table 7.8 *continued from previous page*

Name	**Description**
pVoPo-03-mCherry	inducible expression of mCherry
pVoPo-03-rpoE	inducible expression of *rpoE*
pVoPo-04	base vector for gene deletion system
pVoPo-GFP	vector for constitutive expression of sfGFP in *F. nucleatum*
pVoPo-mCh	vector for constitutive expression of mCherry in *F. nucleatum*
pVoPo-mNG	vector for constitutive expression of mNeonGreen in *F. nucleatum*
pVoPo-mSc	vector for constitutive expression of mScarlet-I in *F. nucleatum*

Table 7.9 | List of used oligonucleotides.

Name	Sequence (5'→3'	Description
JVO-15976	TTTTGTGCTAGTGCTTGTCT	northern blot probe for detection of fadA
JVO-16094	gccctaaaagatgatgaaat	Northern blot probe for FunR3/FunR5
JVO-16466	taacttagaggtggtttgtggctcattgcc	Northern blot probe for FunR2
JVO-16468	agataatctttgatagctaagccccaccta	Northern blot probe for FunR7
JVO-16469	attttattagggaggggtttatctatgttc	Northern blot probe for FunR12
JVO-16470	aaatagcacctactggcataggtgctattt	Northern blot probe for FunR14
JVO-16471	gcagggaatacctgccagagtaaccccaca	Northern blot probe for FunR15
JVO-16472	CAAAGCGAGTTTCTCTGTTTTGCAGCGAAT	Northern blot probe for FunR16
JVO-16473	AACCAAAAAAGGCTATCACAAATCAATGAT	Northern blot probe for FunR17
JVO-16474	atatgaatagttcttatccacaggcttgaa	Northern blot probe for FunR19
JVO-16475	tccttccagatctgacacggttcgacaata	Northern blot probe for 4.5S RNA
JVO-16476	gatggtactcccttcgttggcattatccaa	Northern blot probe for FunR21
JVO-16477	tatttgtttcattgcgacttcctccttttg	Northern blot probe for FunR23
JVO-16478	tagataagctctccagaggttcgtccaata	Northern blot probe for FunR24
JVO-16480	atctcaacgggcctgtccctcaatctctct	Northern blot probe for FunR27
JVO-16481	tgattatactgctcactacactatgagcat	Northern blot probe for FunR28
JVO-16482	ttcacacccttgttagcctctctggactct	Northern blot probe for FunR29
JVO-16483	caaccccatgacgacattagacccaaaacc	Northern blot probe for 6S RNA
JVO-16484	CAGAGGAGATAGCAGCCCTAAAAGATGATA	Northern blot probe for FunR33

Table 7.9 *continued from previous page*

Name	Sequence (5'→3')	Description
JVO-16485	cctcacaaccccgtcgaaacctttgcatcc	Northern blot probe for tmRNA
JVO-16486	aagggagtgttaaaggcaaaactacactcc	Northern blot probe for FunR35
JVO-16487	ttatggttaaccgtagggacacccttccat	Northern blot probe for FunR39
JVO-16488	aaggtactcaagtatgtgaactgcacccaa	Northern blot probe for FunR40
JVO-16490	aatagactatgtctaaatttaaattctaat	Northern blot probe for FunR43
JVO-16491	agaaggttatctataccttctttctgggaa	Northern blot probe for FunR47
JVO-16701	GAATTAACTCAAAACTCATTTAAATTC	Northern blot probe for CRISPR spacer 1
JVO-16703	CATCTGTAATACTTATTTAAATTC	Northern blot probe for CRISPR spacer 3
JVO-16705	CCTGGAATAGTTACATTTAAATTC	Northern blot probe for CRISPR spacer 5
JVO-16711	GAATTACCAGTTGTAGTATATTTAAATTC	Northern blot probe for CRISPR spacer 11
JVO-16714	CTCCTATCAGAAAAATTACCTTTTTTAATATTTAAATTC	Northern blot probe for CRISPR spacer 14
JVO-16717	TTGGGGTCAGTACATTTTAAATACATAAATTTATTC	Northern blot probe for CRISPR terminal spacer
JVO-16933	TCGTATTTTCGATATTATACTGCTCATTGTTAGG	northern blot probe for detection of FoxI
JVO-16933	TCGTATTTTCGATATTATACTGCTCATTGTTAGG	Northern blot probe for FoxI

Table 7.9 *continued from previous page*

Name	Sequence (5'→3')	Description
JVO-16934	TTGTAGAAATTAAACTATAATTAAGGAAACGATAC	Northern blot probe for FunR45
JVO-17207	GACAcgatcgTTCTAGATATTCTTTTTTAAAC	for. amplifcation of origin of replication for F. nucleatum +PvuI-site
JVO-17248	GATAgcggccgcGGTACTGGTAAGCCTGAGAG	rev. amplifcation of origin of replication for F. nucleatum +NotI-site
JVO-17251	GATAgcggccgcCTCCCAACATAAACATATATTGC	rev. Open pRPF185 and remove CD-IFF origin +NotI site
JVO-17252	GACAcgatcgGTATTAGTAAGCATTGGAAATGCT	for. Open pRPF185 and remove CD-IFF origin +PvuI site
JVO-17257	gataggtaccttaaaaacaaagataagagaaagaataaaaag	GB for. 4.5S RNA promoter for pEcoFus +KpnI-site
JVO-17259	aatatatattatacagataaaattgataaaatacccc	GB for. FoxI for pEcoFus
JVO-17302	aagttttattaaaacttataggatccaaaaaataaaatgagcagttataattatc	GB rev. FoxI for pEcoFus
JVO-17306	aatatatattgctatgttaattatagaacatattttaataaattg	GB for. FunR47 for pEcoFus
JVO-17307	aagttttattaaaacttataggatccatgaaaagaaaaaagaaggtatc	GB rev. FunR47 for pEcoFus
JVO-17537	GATCctcgagAGGAGCTCAGATCTGTTAACGCTAC	PCR primer for. amplification of removal of GUS and insertion of XhoI resitrction site in pVoPo-03

Table 7.9 *continued from previous page*

Name	Sequence (5'→3')	Description
JVO-17538	TCctcgagTTTGGATCCTATAAGTTTTAATAAAAC	PCR primer rev. amplification of removal of GUS and insertion of XhoI resitrction site in pVoPo-03
JVO-17543	aatatatattatgattattagcaacctattatatagag	GB for. FunR7 for pEcoFus
JVO-17544	aagttttattaaaacttataggatccagctaccaaagataatctttg	GB rev. FunR7 for pEcoFus
JVO-17545	aagttttattaaaacttataggatccaataaaacctcctaacattcatc	GB rev. FunR7+extended 3'end for pEcoFus
JVO-17549	aatatatattatataagggaacatagataaacc	GB for. FunR12 for pEcoFus
JVO-17550	aagttttattaaaacttataggatccaagaatacaaaaaaccacaag	GB rev. FunR12 for pEcoFus
JVO-17558	aatatatattaatataaagggactgttgatatg	GB for. FunR16 for pEcoFus
JVO-17559	aagttttattaaaacttataggatccaactaaaactataatggggac	GB rev. FunR16 for pEcoFus
JVO-17603	GATActcgagGAAGGGAAGATTGCCCAAGGTGAGTACTAATTGTG	PCR primer for. amplification of mazF from gDNA Clostridium difficile with XhoI/BamHI
JVO-17604	GAATggatccTTAAAAAGTAAACAATCCTAAACTTATTTGAAG	PCR primer rev. amplification of mazF from gDNA Clostridium difficile with XhoI/BamHI
JVO-17611	gataGGATCCaatatatattatcatttttattc	GB rev. 4.5S RNA promoter for pEcoFus +BamHI-site

Table 7.9 *continued from previous page*

Name	Sequence (5'→3')	Description
JVO-18069	cgatcgagtgctagccaatcattttgcggccgcTCCAGTAATGACCTCAGAACTCC	PCR primer for. amplification of ColE
JVO-18070	ccctaaaatccataagaattcGCCAGCAAAAGGCCAGGAAC	PCR primer rev. amplification of ColE
JVO-18071	gaattcttatgGATTTTAGGGTAACAAAAAACAC	PCR primer for. amplification of catP
JVO-18072	gtaattttaaATGGTATTTGAAAAAATTGATAAAAATAG	PCR primer rev. amplification of catP
JVO-18073	caaataccatTTAAAATTACCTCCTTAAAATATACTG	PCR primer for. amplification of 5'UTR of C4N14_09865
JVO-18074	gcggccgcaaaatgattggctagcactcgatcgATAAAATATAAGTAAAAAATTTACATAACTTTAAAAATAC	PCR primer rev. amplification of 5'UTR of C4N14_09865
JVO-18075	tattggtgagaatccaagcactagtcataaaaataagaagcctgc	PCR primer for. amplification of transfer tetR-mazF into pFP76 to generate pVoPo-04
JVO-18076	ttaagggattttggtcatgactagttttttatttagggatttctcacataaaatag	PCR primer rev. amplification of transfer tetR-mazF into pFP76 to generate pVoPo-04
JVO-18101	ttttacttatattttatcgatAAATTTGTACAAAAAATATAGGGAAATTATTTC	PCR for pVoPo-04 FoxI 1000nt upstream_fwd
JVO-18102	ttttaaatttTTTTATTTTATTTCACTAAATCCTCTAATATTAAATAAAAAAAC	PCR for pVoPo-04 FoxI 1000nt upstream_rev
JVO-18103	gaaataaaataaaaAAATTTAAAACTTTTATAAATGTTTAAAAGTCTAAC	PCR for pVoPo-04 FoxI downstream 1000nt_fwd

Table 7.9 *continued from previous page*

Name	Sequence (5'→3')	Description
JVO-18104	atgattggctagcactcgatATGCGCAATGGTCAATTG	PCR for pVoPo-04 FoxI downstream 1000nt_rev
JVO-18105	ttttacttatattttatcgatGCAAAAGAAGCTCAATATACAAATTATG	PCR for pVoPo-04 fadA 1000nt upstream_fwd
JVO-18106	tcaaaatttaTTTTGTTACCTCCCAAATTAAATTATAATAAATTATTTC	PCR for pVoPo-04 fadA 1000nt upstream_rev
JVO-18107	ggtaacaaaaTAAATTTTGAAAAAATGCTAGCATG	PCR for pVoPo-04 fadA downstream 1000nt_fwd
JVO-18108	atgattggctagcactcgatCAGCATAATCAAGTCCTG	PCR for pVoPo-04 fadA downstream 1000nt_rev
JVO-18273	ccttttgctggcgaattc	PCR primer for. opening of pVoPo-00
JVO-18274	ccaggaaccgtaaaaagg	PCR primer for. opening of pVoPo-00
JVO-18337	CTTTTCACCCTTACCTTAATGGCGGTCTATTTTCT	Northern blot probe for M1 RNA
JVO-18338	agcaacgcggccttttttacggttcctggctagtttaatatttaaggaggagaataatggtatctaaaggagaagaagataatatg	ss Oligo gibson assembly stitching: 5'UTR of acpP
JVO-18339	atggtatctaaaggagaagaag	PCR primer for. amplification of mCherry
JVO-18340	aagaattcgccagcaaaagggagctaaagaggtccctag	PCR primer rev. amplification of mCherry
JVO-18341	agtttaataatttaaggaggagaataatgg	PCR primer for. opening of pVoPo-01 construction

Table 7.9 *continued from previous page*

Name	Sequence (5'→3')	Description
JVO-18342	agccaggaaccgtaaaaaggccgcgttg	PCR primer for. opening of pVoPo-01 construction
JVO-18344	agcaacgcggccttttttacggttcctggctaaaaaaaaataaaaata ttgaaaaaacttagaaaatataatagaattaagtagtttaataatt taaggaggagaataatgg	ss Oligo gibson assembly stitching: accD promoter region
JVO-18346	GGATCCTATAAGTTTTAATAAAACTTTAAATAG	PCR primer rev. opening pVoPo-03 for CDS insertion
JVO-18355	TATTTTAATTATACTCTATCAATGATAGAG	PCR primer rev. opening pVoPo-03 for 5'UTR
JVO-18371	gttcagaacgctcggttgccttctgcatcaagctagcataaaaata ag	PCR primer for. amplification of tetR-GUS from pRPF185
JVO-18372	accaataaaaaacgcccggcacatcaccgacgagcaag	PCR primer rev. amplification of tetR-GUS from pRPF185
JVO-18543	TCCTGATTTTGAACTTTATAGTGAAGGCG	PCR primer for. verify FoxI deletion
JVO-18544	GTAAAAAATAATGGACTTGTATTAAGTTCGCC	PCR primer rev. verify FoxI deletion
JVO-18548	GAAGATTAAAAAATAATAAAAAAAATTGGAGGGG	PCR primer for. verify fadA deletion
JVO-18549	GTAAAAGATAACAAATTTCTTATTTTAGACTTCC	PCR primer rev. verify fadA deletion
JVO-18871	TTTAGCTGATGCAACTGATCCAACTAC	in vitro transc. fomA +150nt rev. ;primer extension
JVO-19090	GTATCTAAAGGAGAAGAAGATAATATG	PCR primer for. opening of pVoPo-01
JVO-19091	ACTTAATTCTATTATATTTTCTAAGTTTTTTC	PCR primer for. opening of pVoPo-01

Table 7.9 *continued from previous page*

Name	Sequence (5'→3')	Description
JVO-19275	AGGATCCTATAAGTTTTAATAAAAC	PCR primer for. opening pVoPo-02
JVO-19276	TTATTCTCCTCCTTAAATTATTAAAC	PCR primer rev. opening pVoPo-02
JVO-19279	tttaataatttaaggaggagaataaATGGTTTCTAAAGGAGAAG	PCR primer for. amplify mNeonGreen for insertion into pVoPo-02
JVO-19280	gttttattaaaacttataggatcctTTATTTGTACAATTCATCCATTC	PCR primer rev. amplify mNeonGreen for insertion into pVoPo-02
JVO-19601	atggattttgataacatttatgaagaatattttg	PCR primer for. opening pVoPo-03 for 5'UTR
JVO-19605	cactctatcattgatagagtataattaaaaGTATATTAGAAAGGAGGAATATATAatggattttgataacatttatgaag	ssOligo for pVoPo-03 containing short 5'UTR
JVO-19859	GGATCCTATAAGTTTTAATAAAACTTTAAATAGAAAAAG	PCR primer rev. removing rpoE from p.rpoE
JVO-19860	TATATATTCCTCCTTTCTAATATACTTTTAATTATAC	PCR primer for. removing rpoE from p.rpoE
JVO-19865	ttagaaaggaggaatatataATGGTATCTAAAGGAGAAGAAGATAATATG	PCR primer for. amplification of mCherry for pVoPo-03
JVO-19866	tattaaaacttataggatccTTATTTATATAATTCATCCATACCTCCTG	PCR primer rev. amplification of mCherry for pVoPo-03
JVO-19869	ttagaaaggaggaatatataATGGTATCAAAAGGTGAAGCAG	PCR Primer for. amplify mScarlet-I for p.rpoE backbone

Table 7.9 *continued from previous page*

Name	Sequence (5'→3')	Description
JVO-19870	tattaaaacttataggatccTTATTTGTATAATTCATCCATACCACC	PCR Primer rev. amplify mScarlet-I for p.rpoE backbone
JVO-20099	ACAGATGATTTTATAATAAATTATTTTAGAAATGTGAGG	PCR primer rev. C4N14_09375::mCherry 5UTR+10aa
JVO-20101	ATATATGGAAATATATAATACAAATAAAGAATTATTATTTTATTAAATTATTTAGG	PCR primer rev. mglB::mCherry 5UTR+10aa
JVO-20105	ATAATTTATTTATAAAAAAAGTATAAAGGGG	PCR primer rev. fadA3a::mCherry 5UTR+10aa
JVO-20107	ATATAAAAAAGTTTTATCAATTTTTACTCAATAAAGTATAAAGG	PCR primer rev. fadA3b::mCherry 5UTR+10aa
JVO-20119	ACAGGACAATATAAAAATAAAGGG	PCR primer rev. C4N14_02035::mCherry 5UTR+10aa
JVO-20121	AGATAAGTTATGTATTCGTTAAAAATAAAATTTTATTTAGGG	PCR primer rev. fvcD::mCherry 5UTR+10aa
JVO-20125	GATAACAAGTTTTTTAAACGAAAAAATTAAAGGG	PCR primer rev. C4N14_00275::mCherry 5UTR+10aa
JVO-20127	AATGGGAGATGTCTATATGGGATTATAGGC	PCR primer rev. fomA::mCherry 5UTR+10aa
JVO-20214	cttagaaaatataatagaattaagtACTagtttaataatttaaggaggagaataatgCTCGAGtattatcttcttctcctttagatac	ss Oligo gibson assembly stitching: insert ScaI resitrction site in in-frame XhoI site in pVoPo-02

Table 7.9 *continued from previous page*

Name	Sequence (5'→3')	Description
JVO-20215	aatactcgagTAAAGTCAAAATTAAAAAAATCTTTTTC	PCR primer for. C4N14_09375::mCherry 5UTR+10aa
JVO-20216	aatactcgagTGAACCCAATAACATACCGAAC	PCR primer for. mglB::mCherry 5UTR+10aa
JVO-20218	aatactcgagCATTGAGCATAATAAAATC	PCR primer for. fadA3a::mCherry 5UTR+10aa
JVO-20219	aatactcgagCATTGAGCATAATAAAATCTTTGC	PCR primer for. fadA3b::mCherry 5UTR+10aa
JVO-20225	aatactcgagAGCAAAATATAGACTTC	PCR primer for. C4N14_02035::mCherry 5UTR+10aa
JVO-20226	aatactcgagGATCAATTTTAAACTAACAGATTTTTTC	PCR primer for. fvcD::mCherry 5UTR+10aa
JVO-20228	aatactcgagTGTAAGTATCAACCAAAATACATATCCC	PCR primer for. C4N14_00275::mCherry 5UTR+10aa
JVO-20229	aatactcgagTGAACCTAATACTAATGCTAATTTTTTCATGG	PCR primer for. fomA::mCherry 5UTR+10aa
JVO-20401	agcaacgcggccttttacggttcctggctAAAAAATTATAAACTTTTAATAAAAAAAAACAGTCTATATACTATATTGAagtttaataatttaaggaggagaataatgg	ssOligo gibson assembly stitching: transcriptional fusion miaB::mCherry

Table 7.9 *continued from previous page*

Name	Sequence (5'→3')	Description
JVO-20402	agcaacgcggccttttttacggttcctggctAAAAAATTATAAACTTTTAATAAAAAAAAACAGTgTATATACTATATTGAagtttaataatttaaggaggagaataatgg	ssOligo gibson assembly stitching: transcriptional fusion miaB::mCherry point mutation in -10
JVO-20403	agcaacgcggccttttttacggttcctggctCATATAAACCTCTTTAAAACTTTTAGTTTATTAAATATTGTCTAATCTAAagtttaataatttaaggaggagaataatgg	ssOligo gibson assembly stitching: transcriptional fusion C4N14_09130::mCherry
JVO-20404	agcaacgcggccttttttacggttcctggctCATATAAACCTCTTTAAAACTTTTAGTTTATTAAATATTGTgTAATCTAAagtttaataatttaaggaggagaataatgg	ssOligo gibson assembly stitching: transcriptional fusion C4N14_09130::mCherry point mutation in -10
JVO-20405	agcaacgcggccttttttacggttcctggctTGTTTGATATAAAAATAACCAAACTCATATAATTGGAGTCAAATATATTAagtttaataatttaaggaggagaataatgg	ssOligo gibson assembly stitching: transcriptional fusion C4N14_03280::mCherry
JVO-20406	agcaacgcggccttttttacggttcctggctTGTTTGATATAAAAATAACCAAACTCATATAATTGGAGTgAAATATATTAagtttaataatttaaggaggagaataatgg	ssOligo gibson assembly stitching: transcriptional fusion C4N14_03280::mCherry point mutation in -10
JVO-20407	agcaacgcggccttttttacggttcctggctGAAAATAAAAAAGATTAAACTTTTTTTATAGTATTAATGTCTAATCTGTAagtttaataatttaaggaggagaataatgg	ssOligo gibson assembly stitching: transcriptional fusion rpoE::mCherry

Table 7.9 *continued from previous page*

Name	Sequence (5'→3')	Description
JVO-20408	agcaacgcggccttttttacggttcctggctGAAAATAAAAAAGATTAAACTTTTTTTATAGTATTAATGTgTAATCTGTAagtttaataatttaaggaggagaataatgg	ssOligo gibson assembly stitching: transcriptional fusion rpoE::mCherry point mutation in -10
JVO-20409	agcaacgcggccttttttacggttcctggctAAAACAAAAATTTTTTTAAACTATTTTATAACTTTCACTGTCAAATAAGTagtttaataatttaaggaggagaataatgg	ssOligo gibson assembly stitching: transcriptional fusion FoxI::mCherry
JVO-20410	agcaacgcggccttttttacggttcctggctAAAACAAAAATTTTTTTAAACTATTTTATAACTTTCACTGTgAAATAAGTagtttaataatttaaggaggagaataatgg	ssOligo gibson assembly stitching: transcriptional fusion FoxI::mCherry point mutation in -10
JVO-20852	GTATATTAGAAAGGAGGAATATATAATGG	PCR primer for. opening p.mSc for insertion of promoter
JVO-20853	actagtgcttggattctcacc	PCR primer rev. opening p.mSc for insertion of promoter
JVO-20856	gttttttattggtgagaatccaagcactagtAAAATAAATTTTAAAACCTTGAATTTTATTTTAAATATGTTATAATATTTgtatattagaaaggaggaatatataatgg	ssOligo stitching: 50bp promtoer region for mScarlet: 6S RNA
JVO-20857	gttttttattggtgagaatccaagcactagtTCTTCTAAAAAATCCTTGCTAAAATTTAATAATAGTGCTATAATTAAGTCgtatattagaaaggaggaatatataatgg	ssOligo stitching: 50bp promtoer region for mScarlet: tmRNA
JVO-20858	gttttttattggtgagaatccaagcactagtGATAATAAAAAAAACATTGTAATTTTTTAAAATTTTATGTATAATTGTCAgtatattagaaaggaggaatatataatgg	ssOligo stitching: 50bp promtoer region for mScarlet: fadA

Table 7.9 *continued from previous page*

Name	Sequence (5'→3')	Description
JVO-20859	gttttttattggtgagaatccaagcactagtAAGCATAAAAAAAAG TTGGATAATATACTAAAATAGTGTATAATAAATATgtatattagaa aggaggaatatataatgg	ssOlgio for generating pVoPo-mSc
JVO-20859	gttttttattggtgagaatccaagcactagtAAGCATAAAAAAAAG TTGGATAATATACTAAAATAGTGTATAATAAATATgtatattagaa aggaggaatatataatgg	ssOligo stitching: 50bp promtoer region for mScarlet: acpP
JVO-20860	gttttttattggtgagaatccaagcactagtTTTTATAATAAAATA CTTGATTTTTTTTGAAAATAAAGTATAATAAATACgtatattagaa aggaggaatatataatgg	ssOligo stitching: 50bp promtoer region for mScarlet: C4N14_05865
JVO-20861	gttttttattggtgagaatccaagcactagtTTTACCTAAAACACT TGAAAAAACTTGTAATATATGATATAAAATATTGTgtatattagaa aggaggaatatataatgg	ssOligo stitching: 50bp promtoer region for mScarlet: C4N14_08125
JVO-20862	gttttttattggtgagaatccaagcactagtAAACAAAAAATATAT TTACAAATATAAAAAAATGTAGTATACTAAGTTCAgtatattagaa aggaggaatatataatgg	ssOligo stitching: 50bp promtoer region for mScarlet: C4N14_03770/fomA
JVO-20863	gttttttattggtgagaatccaagcactagtTTTAAAAAAAATTTG TAGATTTTTTTTATTTGTTATGATATAATAATATGgtatattagaa aggaggaatatataatgg	ssOligo stitching: 50bp promtoer region for mScarlet: C4N14_05690
JVO-20864	gttttttattggtgagaatccaagcactagtATATAATATTTTTTC TTGACATTAAAGAATTTTTATAGTAGAATGCGAAAgtatattagaa aggaggaatatataatgg	ssOligo stitching: 50bp promtoer region for mScarlet: C4N14_07045
JVO-20865	gttttttattggtgagaatccaagcactagtTTTTAGGGAAAAATT GATACAAAATAATTTTTTTATGTTATAATGAATCTgtatattagaa aggaggaatatataatgg	ssOligo stitching: 50bp promtoer region for mScarlet: C4N14_00275
JVO-20866	gttttttattggtgagaatccaagcactagtTTACAAGAGAGAATA ATTGACAAAATCTGAAAAATATGGTAGAATTCTTTgtatattagaa aggaggaatatataatgg	ssOligo stitching: 50bp promtoer region for mScarlet: C4N14_00660

Table 7.9 *continued from previous page*

Name	Sequence (5'→3')	Description
JVO-20867	gttttttattggtgagaatccaagcactagtACAATAATATTTTTCTTGACAAGTTCTAAAAAATACTATATACTTATATGgtatattagaaaggaggaatatataatgg	ssOligo stitching: 50bp promtoer region for mScarlet: C4N14_05025
JVO-20868	gttttttattggtgagaatccaagcactagtATATTTATTTTTTGTTGAATTAATATATATAAAATGGTATATTTATATGTgtatattagaaaggaggaatatataatgg	ssOligo stitching: 50bp promtoer region for mScarlet: C4N14_01745
JVO-20869	gttttttattggtgagaatccaagcactagtTATAAATAAATACTTGACAAAATATCAAATAACTTGTAAAATTCTAAACTgtatattagaaaggaggaatatataatgg	ssOligo stitching: 50bp promtoer region for mScarlet: RNase P
JVO-20870	gttttttattggtgagaatccaagcactagtATTTAATAATATTTATCATTAAAATCAATTTTTTATGTTATAATTATAAAgtatattagaaaggaggaatatataatgg	ssOligo stitching: 50bp promtoer region for mScarlet: C4N14_09770
JVO-21079	tatatattcctccttttctaatatacATATTTATTATACAC	PCR primer for. opening pVoPo-FP backbone
JVO-21080	ttagaaaggaggaatatataatgagaaaaggagaagaacttttttac	PCR primer for. amplify sfGFP for insertion into pVoPo-FP
JVO-21081	tattaaaacttataggatccttatttatacaattcatccattccatgag	PCR primer rev. amplify sfGFP for insertion into pVoPo-FP
JVO-21084	ttagaaaggaggaatatataatggtatctaaaggagaagaagataatatg	PCR primer for. amplify mCherry for insertion into pVoPo-FP
JVO-21085	tattaaaacttataggatccttatttatataattcatccatacctcctg	PCR primer rev. amplify mCherry for insertion into pVoPo-FP

7.1.6 Strains

Table 7.10 | List of used strains.

Name	Species	Description
FPS-189	*F. n. nucleatum* ATCC 23726	*fadA* deletion strain
FPS-227	*F. n. nucleatum* ATCC 23726	*foxI* deletion strain
JVS-11527	*F. n. nucleatum* ATCC 25586	*F. n. nucleatum* ATCC 25586 (reference strain)
JVS-11541	*F. n. nucleatum* ATCC 23726	*F. n. nucleatum* ATCC 23726 wild-type (WT)
JVS-11855	*F. n. polymorphum* ATCC 10953	*F. n. polymorphum* ATCC 10953
JVS-12771	*F. periodonticum* 2_1_31	*F. periodonticum* 2_1_31
JVS-12772	*F.n. vincentii* 3_1_36A2	*F.n. vincentii* 3_1_36A2
JVS-12774	*F.n. animalis* 7_1	*F.n. animalis* 7_1

7.1.7 Used computational software and tools

Table 7.11 | Used software and tools.

Name	Application	Reference
ANNOgesic	Analysis of dRNA-seq data.	Yu et al. (2018b)
Blast	Search for homologous sequences.	Altschul et al. (1990); States and Gish (1994); Camacho et al. (2009)
CorelDRAW 2018	Vector graphics editor.	Corel Corporation
Excel 365 ProPlus	Spreadsheet editor	Microsoft Corporation
ggplot2	R package for generating volcano plots.	Hadley (2016)
ggtree	R package for generation of the phylogenetic tree.	Yu et al. (2018a)
IGV 2.8.2	Genome browser	Thorvaldsdottir et al. (2013)
ImageJ 1.52u	Image processing and quantification	Schneider et al. (2012)
IntaRNA	sRNA-mRNA target prediction.	Mann et al. (2017)
MiKTex 5.5	LATEX distribution	Donald Arseneau
Multalign	Alignment of DNA and amino acid sequences using the DNA or Blosum62 parameter, respectively.	Corpet (1988)
NovoExpress Software	Analysis of flow cytometry data	Agilent
Prism 8.4.0	Graphing	GraphPad Software, Inc.
R	Running different visualization packages.	Team (2020)
READemption	Analysis of RNA-seq. This included mapping and differential gene expression analysis.	Förstner et al. (2014)
superheat	R packages for generating heat maps.	Barter and Yu (2018)
TeXstudio 4.3.1	LATEX editor	Benito van der Zander
VARNA 3-93	Visualization of RNA secondary structures	Darty et al. (2009)

7.2 Methods

7.2.1 Bacterial culture

The fusobacterial strains *F. nucleatum* subsp. *nucleatum* ATCC 25586 and *F. nucleatum* subsp. *polymorphum* ATCC 10953 were obtained from the the German Collection of Microorganisms and Cell Culture (DSMZ, Germany). *F. nucleatum* subsp. *nucleatum* ATCC 23726 was acquired from the American Type Culture Collection (ATCC, USA). *F. nucleatum* subsp.*animalis* 7_1, *F. nucleatum* subsp. vincentii 3_1_36A2 and F. periodonticum 2_1_31 were provided as a gift by Emma Allen-Vercoe (University of Guelph, Canada). Routine culturing for all strains was performed in Columbia broth (Col. B.) and supplemented Brain-Heart-Infusion (BHI) 2% agar plates supplemented with 1% (w/V) yeast-extract, 1% (w/V) glucose, 5µg/ml hemin and 1% (V/V) FBS referred to as BHI-C. Single colonies were used to inoculate 24 h pre-cultures which were diluted 1:50 to yield the final working cultures. The working cultures were used as start point for all experiments. Plasmid retention was maintained by adding thiamphenicol to both BHI-C plates (5 µg/ml) and Col. B. (2.5 µg/ml). All culturing steps were performed in an anaerobe environment (N_2:H_2:CO_2; 80:10:10) at 37°C if not stated otherwise. This also extents all all larger volume of liquids and agar plates which were placed in the chamber the night before use. If the experiments indicate a certain growth this corresponds to the following OD_{600nm}: early-exponential: OD_{600nm}0.2; mid-exponential: OD_{600nm}0.4-0.6;early-stationary: OD_{600nm}0.8-1.0. Other modifications to the medium or treatment are given in following sections.

7.2.2 Preparation of electro-competent *F. nucleatum*

F. nucleatum was grown to mid-exponential phase (OD_{600nm}∼0.4) and spun down for 5 min at 4.500 xg and 4°C. All following steps were performed on ice. The media was decanted and the cell pellet was resuspended in pre-chilled 10% (V/V) glycerol solution. The bacteria were spun down again for 5 min at 4.500 xg and 4°C prior to a second round of resuspension in the glycerol solution. These steps were repeated for a total of five times. Afterwards cells were concentrated to OD_{600nm}∼5/80µl and stored at -80°C until use.

7.2.3 Electroporation of *F. nucleatum*

Electro-competent *F. nucleatum* were thawed on ice before mixing 80µl with 100ng-200ng (replicative plasmids) or 10µg (suicide vectors) of plasmid DNA. In case of the latter, salts were removed by placing 20µl of DNA on a hydrophobic membrane (MF-Millipore 25nm) floating in double distilled water (ddH_2O) for 1h. Next, the cells were electroporated using a BioRad MicroPulser using the following: electroporation cuvettes (1mm gap size) and voltage set to 2.0 kV. Afterwards the bacteria were recovered for 2 h or 4 h in 900µl Col. B. for replicate and suicide vectors, respectively. Last, 100µl of bacterial suspension were plated on BHI-C containing thiamphenicol.

7.2.4 DNA & cloning methods

7.2.4.1 Polymerase chain reaction (PCR)

DNA was amplified using the Phusion DNA polymerase except for colony PCR for which *Taq* polymerase was used. DNA products were purified using the NucleoSpin Gel and PCR clean-up kit according to the manufacturer's protocol. All oligonucleotides used are listed in (Table 7.9).

7.2.4.2 Agarose gel electrophoresis

DNA products in DNA loading buffer were analyzed on 0.8 to 2% (w/V) agarose gels prepared with TAE buffer. The DNA was visualized by staining with ethidium bromide for 10 min.

7.2.4.3 Cloning via restriction digest

2µg of vector and 500ng to 2µg of PCR product were digested with the appropriate restriction enzymes for 1 h at 37°C. All products were purified after agarose gel electrophoresis. 50ng of linearized vector were ligated with the PCR in 3x molar excess using 1 U if T4 DNA ligase overnight at 4°C. The reaction was then transformed into electro-competent *E. coli* cells.

7.2.4.4 Cloning via Gibson assembly

Vectors were either opened via inverse PCR or restriction digest. The PCR products were amplified containing 20-25nt overlapping regions to the desired neighboring region in the final product. All products were purified via agarose gel

electrophoresis. 50ng of backbone was used with PCR products in a molar ratio of 5:1 for products ≤200nt or 3:1 for all others. In case a single stranded oligonucleotide (ssOligo) was used only 10pmol of the ssOligo were used together with 50ng of the vector. All fragments were combined in 10µl reaction with the NEBuilder® HiFi DNA Assembly master according to the provided protocol. Assembly took place for 1h at 50°C after which 5µl were transformed into electro-competent *E. coli*.

7.2.5 Generation of genetic tools

7.2.5.1 Cloning of shuttle vector pEcoFus and sRNA overexpression

pRPF185 (Fagan and Fairweather, 2011) was used as the backbone as it was developed as an *E. coli* - *C. difficil*e shuttle vector with the latter showing similar G+C content and codon usage. First, the *repA* and *orfB* genes were removed by inverse PCR (JVO-17251/17252) which also introduced restriction sites for PvuI and NotI. The fusobacterial origin of replication ORI_{FN} was amplified from pORI92 (a gift from Gilad Bachrac, Israel; JVO-17207/17248) which added matching restriction sites to the backbone. The final product after restriction digest and ligation resulted in pFP14. The 100nt promoter region of the fusobacterial 4.5S RNA was amplified and inserted into pFP14 digested with KpnI and BamHI via Gibson assembly. The finished vector presented pEcoFus as the 1^{st} generation shuttle vector for *F. nucleatum*.

Overexpression constructs for the different sRNAs were generated by amplifying the full-length sRNAs with flanking region towards the BamHI restriction site in pEcoFus. The PCR products were used for DNA assembly with BamHI-digested pEcoFus.

7.2.5.2 Construction of shuttle vector pVoPo-00

The following steps were performed to yield a smaller and improved basic vector for *F. nucleatum*: ColE1 (JVO-18069/18070) and the *catP* chloramphenicol (JVO-18071/18072) resistance cassette were amplified from pEcoFus and included two multiple cloning sites. The promoter region and 5'UTR of the constitutive expressed flavodoxin C4N14_09865 from *F. nucleatum* were also amplified (JVO-18073/18074) and all three fragments were assembled via Gibson assembly into pFP76. To allow plasmid maintenance in *F. nucleatum*, ORI_{FN} was taken from pEcoFus and placed in between the PvuI and NotI site in pFP76 via restriction

digest and ligation. The final product was the improved *E. coli-F. nucleatum* shuttle vector pVoPo-00.

7.2.5.3 Construction of pVoPo-01: a transcriptional reporter system

As a first step codon-optimized mCherry was amplified from pDSW1728 (Ransom et al., 2015) (JVO-18339/18340). pVoPo-00 was opened via inverse PCR (JVO-18273/18274) and both fragments together with an ssOligo for the 5'UTR of fusobacterial *acpP* (JVO-18338) were used for an DNA assembly reaction. The resulting product was opened via inverse PCR (JVO-18341/18342) and assembled together with the ssOligo containing the 50nt-promoter region for the constitutive expressed *accD* gene (C4N14_10115; JVO-18344) generating pFP119. After verifying successful expression of mCherry for pFP119 in *F. nucleatum*, the promoter region was replaced by a XhoI/ScaI restriction site to insert a promoter of interest via restriction digest. For this the above vector was opened via inverse PCR (JVO-18342/21139) and stitched to together with an ssOligo containing the restriction sites (JVO-21140). This resulted in pVoPo-01.

7.2.5.4 Construction of pVoPo-02: a translational reporter system

pFP119 was opened by inverse PCR (JVO-19090/19091) and stitched with an ssOligo (JVO-20214). The ssOligo includes a ScaI restriction site at transcriptional start of the *accD* promoter while the XhoI site is in-frame with the coding sequence of mCherry. This yielded pVoPo-02 to easily create translational fusions for a gene of interest.

7.2.5.5 Construction of pVoPo-03: an inducible gene expression system

A codon optimized cassette encoding for the tetracycline-dependent repressor TetR was amplified from pRPF185 (Fagan and Fairweather, 2011) (JVO-18371/18372). The PCR product was assembled into the SpeI site of pVoPo-00. Inverse PCR of this vector was used to leave a BamHI restriction site at the +1 site of the vector (JVO-17537/17538). The final product, pVoPo-03 could thus be used to clone a gene of interest at the transcriptional start of the TetR-regulated promoter.

7.2.5.6 Construction of pVoPo-04: a scarless gene deletion system

The *mazF* gene including its 5'UTR was amplified from genomic DNA of *C.difficile* 630 including XhoI and BamHI restriction sites (JVO-17603/17604) to be inserted into the similarly digested pVoPo-03, named pFP96. The *tetR-mazF* expression cassette was subsequently transferred into the SpeI site of pFP76 to yield pVoPo-04, a fusobacterial suicide vector (JVO-18075/18076). Homologous arms flanking the region of interest to remove can be inserted into the PvuI and NotI site of the vector.

7.2.5.7 Generating inducible σ^E expression vector

As a first step pFP96 was opened via inverse PCR (JVO-18355/18346) and the fusobacterial *rpoE* amplified (JVO-18356/18357). Both products were assembled via Gibson assembly. Next, this vector was opened (JVO-19601/18355) to replace the 5'UTR with a short synthetic UTR optimized (Xiao et al., 2020) for translation encoded on an ssOligo (JVO-19605). This yielded the final inducible σ^E expression vector (p.*rpoE*).

7.2.5.8 Generating different transcriptional reporter driving mScarlet-I expression

p.*rpoE* was used as basis to generate different mScarlet-I expression vectors. First, the *rpoE* gene was replaced by mScarlet-I by opening the vector (JVO-19859/19860) and inserting the latter (JVO-19869/19870) via DNA assembly. Next, the *tetR* gene was removed via inverse PCR (JVO-20854/20855) yielding a promoterless mScarlet-I cassette. Finally, gene-specific promoter were inserted by opening the vector (JVO-20852/20853) and performing Gibson assembly with respective ssOligos (Table 7.9).

7.2.6 RNA extraction

7.2.6.1 Hot phenol extraction protocol

RNA was extracted from 6 OD_{600nm} units of *F. nucleatum* by fixating them through the addition of 0.2 volumes of ice-cold STOP mix (5% phenol and 95% ethanol). The samples were then immediately snap-frozen in liquid nitrogen and kept at -80°C until further processing. On the day of extraction,the samples were thawed on ice, spun down for 10 min at 4.500xg and 4°C before removing

the supernatant. The pellet was resuspended in 600µl of TE-buffer containing 50µg/ml lysozyme. Afterwards 60µl of 10% (w/V) SDS was added. The samples were incubated at 64°C for 2 min. Next, 66µl of 3M sodium acetate (pH 5.2) was added, the tubes inverted and incubated before mixing it with 750µl of acidic phenol. Organic and aqueous phases were then separated through centrifugation at 13.000 xg and at 4°C for 15 min. The aqueous phase was transferred into phase-lock-gel (PLG) tubes. The addition of 750µl chloroform followed by vigorous mixing of the tubes and repeat of the centrifugation step allowed the transfer of a re-extracted aqueous phase into fresh 2ml tubes. The addition of 1.4ml 30:1 mix (29 parts ethanol; 1 part 3M sodium acetate (pH 6.5)) was used to precipitate nucleic acids at -20°C overnight. Subsequent centrifugation for 30 min at 13.000 xg and 4°C led to the pellet formation of precipitated nucleic acids. The pellet was washed with 75% (V/V) of ethanol prior to drying at RT for 15 min. The pellet was resuspended in water followed by incubation at 65°C shaking at 1.000 rpm for 5 min and either stored at -20°C or directly used for DNase I digest.

7.2.6.2 DNase I digest

40µg of extracted nucleic acids were used as input together with RNase-free DNase I 2000 U, DNase I buffer and 0.5 U of RNase inhibitor. The mix was incubated for 1 h at 37°C before adding 2 volumes of water and transfer to PLG tubes. The same volume was added of P/C/I, the tubes vigorously mixed and centrifuged 15 min at 13.000 xg at 4°C. The aqueous phase was precipitated as above using 2 volumes of 30:1 mix and incubation overnight at -20°C. This step was followed by centrifugation, an ethanol wash, drying and resuspension as described above resulting in DNA-free RNA.

7.2.7 Northern blot analysis

For northern blot analysis, 3-10µg of DNase I-treated total RNA was separated on a 6% polyacrylamide (PAA) gel containing 7 M urea from three biological replicates for each time point or strain. After transferring the RNA to Hybond-XL membranes, hybridization with ^{32}P-ATP end-labeled deoxyribonucleotide probes took place overnight at 42°C. For signal visualization, a Typhoon FLA 7000 phosphorimager (GE Healthcare) was used.

7.2.8 Radioactive labeling of nucleic acids

For labeling of oligonucleotides, 10pmol of the individual oligonucleotide were used as input together with 0.5µl of T4 polynucleotide kinase and 10µCi ^{32}P-γ-ATP for the labling reaction. After 45 min incubation at 37°C, the probe was purified G-25 columns according to the manufacturer's instructions to remove unlabeled oligonucleotides.

To label *in vitro* transcribed RNA, 50pmol of RNA were treated with 10 U of calf intestinal phosphatase (CIP) at 37°C for 1 h. CIP-treated RNA was purified following the P/C/I extraction as described for the DNase I digest. 20pmol of the dephosphorylated RNA were used input for the labeling reaction as described above but using G-50 columns. After purification the RNA was separated on a denaturing 6% PAA gel. The correct band was visualized by detecting the radioactive signal and afterwards excised and purified as described for the *in vitro* transcription.

7.2.9 Sample collection and analysis for translation reporter experiments

After reaching mid-exponential phase, cells from *F. nucleatum* expressing the individual translational fusion alone or in conjunction with FoxI/FoxI-3C were spun down for 3 min at 4.000 xg. All subsequent processes were carried out outside the anaerobic chamber. Next, the bacteria were fixed using 4% PFA (w/V) at 4°C for 20 min. After removing the PFA solution, the cells were incubated in PBS containing 100ng/ml DAPI for 5 min at RT before performing a single washing step in PBS followed by overnight incubation in PBS at 4°C which allows for full fluorescent protein maturation. Measurement of fluorescence intensity was performed the next day using Novocyte flow cytometer (Agilent). For each sample 50.000 DAPI-positive cells were analyzed for their signal intensity at 615 to 620nm.

For western blot analysis the samples were spun down and snap-frozen prior to resuspending the pellet in protein loading buffer.

7.2.10 Sample collection and analysis for transcriptional reporter experiments

All used strains were grown to mid-exponential phase before being spun down and snap-frozen. The resulting pellets were resuspended in protein loading buffer.

7.2.11 Sample collection for RNA-seq analysis after treatment with oxygen and polymyxin B

F. nucleatum was grown in biological triplicate to mid-exponential phase and exposed to either 400ng/ml polymyxin B or molecular oxygen at 37°C for 20 min. For the latter the culture was poured into a Petri dish and kept in an incubator outside the anaerobic chamber. The control was left untreated in the anaerobic chamber. At the end of the treatment the samples were fixed by the addition of STOP mix and RNA extracted.

7.2.12 Procedure for pulse-expression experiments.

In all experiments, three biological experiments were grown to mid-exponential phase. Next, cultures were split into a treated and untreated sample. Treated samples received 100ng/ml anhydrotetracycline (ATc) for the pulse expression of σ^E or sRNAs and indicated concentrations for the expression of mCherry. Pulse-expression of proteins was performed for 30 min while sRNAs were induced for 20 min before fixing samples for further processing.

7.2.13 Stress conditions for σ^E activation

Three biological triplicates of *F. nucleatum* were grown to mid-exponential phase. 1ml of a 4x solution containing the stressor in Col. B. was added to 3ml of culture and treated for 60 min at 37°C. The different used stressors and the final concentration were as follows: polymyxin B (400ng/ml), lysozyme (125µg/ml), NaCl (600mM), bile (0.05% (w/V), H_2O_2 (400µM), diamide (125µM), S-nitrosoglutathione (GNSO; 250µM), mitomycin C (625ng/ml). For the heat shock samples 1 ml of Col. B. was added prior to incubation at 42°C for 60 min. The oxygen shock was facilitated by pouring the culture into a Petri dish outside and incubating the bacteria at 37°C outside of the chamber. 1 ml of Col. B was added to the control. After 60 min, 20% (V/V) of STOP mix was added, the samples snap-frozen and proceeded with RNA extraction.

7.2.14 qRT-PCR

cDNA was generated using the M-MLV reverse transcriptase according to the manufacturer's instruction. 1µg of DNase I digested RNA was used as input and reverse transcribed using random hexamer primers. The equivalent of 10ng of RNA

in cDNA was used as input the qPCR analysis. For this, the Takyon Master Mix was used according to the manufacturer's instructions with gene specific primers. The $2^{-\Delta\Delta Ct}$ method was used to calculate the fold changes (Livak and Schmittgen, 2001) compared to the control. The 5S rRNA served as a reference gene.

7.2.15 Generation of cDNA libraries for dRNA-seq

All cDNA libraries were generated by Vertis Biotechnology AG (Munich, Germany) and used 5µg of DNase I-digested RNA as input. No RNA-size fractionation step was performed. Technical duplicates for each sample were provided. One representing the entire cellular RNA pool was left untreated (TEX -) while the other one was enriched for primary transcript through TEX-treatment (TEX +). Afterwards the samples were polyadenylated before removing all 5'-triphosphate groups using tobacco acid pyrophosphatase (TAP) and subsequent 5'-adapter ligation. cDNA synthesis was initiated by using M-MLV reverse transcriptase with an oligo(dT)-adapter specific primer. The cDNA was further amplified to a concentration of 20-30 ng/µl using a high fidelity polymerase including 5'-sequencing adapter also allowing de-multiplexing of samples later on. The different cDNA libraries were pooled equimolar and sequenced on an Illumina NextSeq 500 (75bp single-end reads). The sequencing was performed with Vertis Biotechnology AG for *Fnn* and *Fnp* or the Core Unit SysMed (University Würzburg) for *Fna*, *Fnv* and *FuP*.

7.2.16 Generation of cDNA libraries for RNA-seq

All cDNA libraries were generated by Vertis Biotechnology AG (Munich, Germany) and used 5µg of DNase I-digested RNA as input. rRNA was depleted followed by fragmentation (30 s at 4°C). Adapters were then ligated to the 3'-end of the fragmented RNA. Next, the M-MLV reverse transcriptase was used for first-strand synthesis in combination with a primer complementary to the introduced 3'end adapter. Next, 5'-Illumina TrueSeq adapters were added to the cDNA which before amplifying it to 10-20 ng /µl with a high fidelity DNA polymerase and adapter specific primers. The amplified cDNA was then purified via the Agencourt AMPure XP kit (Beckman Coulter Genomics) before the evaluation via capillary electrophoresis to ensure a size distribution of 200 nt to 600 nt. Sequencing of the finished libraries was performed by the Core Unit SysMed (University Würzburg) on an Illumina NextSeq 500 (75bp single-end reads).

7.2.17 Mapping of RNA-seq data and differential gene expression

Adapater sequences and low quality reads were trimmed off of the raw FASTQ files using the FASTX toolkit (version 0.10.1; http://hannonlab.cshl.edu/fastx_toolkit/) and its fastq_quality_trimmer function. These were used as input for the READemption tool (version 0.4.3) (Förstner et al., 2014) which was used for all subsequent analysis. To align reads to the individual genome sequences for the used strains. Poly(A)-tails as well as all reads ≤12nt or falling below a 95% mapping accuracy were removed. Generated coverage plots were used for the gene quantification of annotated features. The yielded read counts were used for the differential expression analysis with the DEseq2 package (version 1.18.01) (Love et al., 2014) using the data for three biological triplicates used in each experiment. In all experiments genes were considered significantly expressed with a false-discovery rate (FDR) ≤0.05 and a minimal average read count of 50. Fold changes limits are indicated in the respective text sections in Chapter 2 and 4.

7.2.18 Re-annotation of CDS

The sequence of all coding genes including 20 nt upstream of the ORF start site were checked for an AUG start codon and a Shine Dalgarno sequence. All ORFs missing either feature were subjected for manual inspection surveying the region for in-frame up- or downstream start codons with a Shine-Dalgarno sequence. Re-annotation was performed on CDS for which the corrected start was supported by the existence of a start codon (AUG, GUG, or UUG), a Shine-Dalgarno sequence, and sequence conservation in other different *F. nucleatum* susbp. *nucleatum* strains.

7.2.19 Sequence analysis via MEME suite

In case of the promoter analysis of the dRNA-seq data, the +50 nt region for all TSS was extracted using BEDtools (Quinlan, 2014) and used as input for MEME (version. 4.12.0; (Bailey et al., 2015)) either all together or divided into the different TSS classes. This was similarly done for the identification of σ^E promoter motif. Here, the regions for all differentially expressed genes were also manually inspected to identify TSS which were not detected in the dRNA-seq data.

Sequence features in the 5'UTR were identified by using all their sequences as input for MEME.

In all cases, MEME ran searching all window sizes from 3 nt to 50 nt.

7.2.20 RNA-based annotation and prediction via ANNOgesic

The ANNOgesic tool (Yu et al., 2018b) was used with to predict the following features: TSS, 5'UTRs, operons, sRNAs, CRISPR locus, regulatory RNA-elements, sORFs and terminators. A TSS had to be detected in all replicates of a single condition to considered. Predicted secondary TSS were excluded if they were $\leq$7nt apart from a primary TSS. The automated annotation was further curated by manul inspection. The annotation of 5'UTRs was performed by allowing a total maximum length of 300nt in addition to a 25nt extension to connect it with a detected transcript. sRNA candidates were predicted taking TSS, overall coverage, presence of terminator and detected promoter sequences into account. Rho-independent terminators prediction was carried out using ANNOgesic which combines two heuristic prediction algorithms: TransTermHP111 analysis and a detection of sharp coverage decreases around the predicted terminator sequence. All other parameters were unchanged and the remaining features were also annotated with the default settings.

7.2.21 sRNA conservation

All sRNA candidate sequences were used as input for a BlastN analysis against a custom database including all *Fusobacterium* strains with a completed genome (Table 7.12). These hits were filtered for a nucleotide identity $\geq$75% over $\geq$75% of the length of the query sRNA. In case of sRNAs in *Fnn* the up- and downstream genes were also used for a Blast analysis with the same criteria to evaluate a conserved genomic synteny. This was further expanded for FoxI were the region between *rsmB* and *trpB* was further considered.

Table 7.12 | Used strains for the conservational analysis.

Name	Accession number
F.n. nucleatum ATCC 25586	NC_003454.1
F.n. animalis 21_1A	NZ_CM002368.1
F.n. animalis 4_8	NC_021281.1
F.n. animalis 7_1	NZ_CP007062.1
F.n. animalis KCOM 1279	NZ_CP012713.1
F.n. animalis KCOM 1325	NZ_CP012715.1
F.n. nucleatum ATCC 23726	NZ_CP028109.1
F.n. nucleatum ChDCF316	NZ_CP012716.1
F.n. nucleatum ChDCF317	NZ_CP022122.1
F.n. polymorphum ATCC 10953	NZ_CM000440.1
F.n. polymorphum ChDCF218	NZ_CP021934.1
F.n. polymorphum ChDCF306	NZ_CP013121.1
F.n. polymorphum KCOM 1275	NZ_CP022123.1
F.n. polymorphum NCTC10562	NZ_LN831027.1
F.n. vincentii 3_1_27	NZ_CP007065.1/NZ_CP007064.1
F.n. vincentii 3_1_36A2	NC_022196.1
F.n. vincentii KCOM 2931	NZ_CP024749.1
F. gonidiaformans ATCC 25563	NZ_CP028106.1
F. hwasookii ChDCF174	NZ_CP013331.1
F. hwasookii ChDCF206	NZ_CP013336.1
F. hwasookii ChDCF300	NZ_CP013334.1
F. mortiferum ATCC 9817	NZ_CP028102.1
F. necrophorum 1_1_36S	NZ_CP028107.1
F. periodonticum 2_1_31	NZ_CP028108.1
F. periodonticum KCOM 1261	NZ_CP024699.1
F. periodonticum KCOM 1262	NZ_CP024731.1
F. periodonticum KCOM 1263	NZ_CP024700.1
F. periodonticum KCOM 1277	NZ_CP024701.1
F. periodonticum KCOM 1282	NZ_CP024702.1
F. periodonticum KCOM 1283	NZ_CP024698.1
F. periodonticum KCOM 2305	NZ_CP024703.1
F. periodonticum KCOM 2555	NZ_CP024704.1
F. periodonticum KCOM 2653	NZ_CP024705.1
F. ulcerans ATCC 49185	NZ_CP028105.1
F. varium ATCC 27725	NZ_CP028103.1
F. varium Fv113-g1	NZ_AP017968.1

7.2.22 *In silico* target prediction

All predictions of sRNA-mRNA interactions were performed using the IntaRNA algorithm (version 2.0.4) (Mann et al., 2017). The input consisted out of the sRNA and full-length mRNA sequence including the 5'UTR. In cases were no 5'UTR was annotated the CDS was extended by 50nt. IntaRNA calculations ran in the heuristic mode and allowing a seed region of ≥3nt with all other parameters set to default.

7.2.23 *In vitro* transcription

A sequence of interest was amplified with a T7 promoter by PCR. *In vitro* transcription was performed using the MEGAscript T7 transcription kit using 400ng PCR product as template following the manufacturer's protocol. The transcription react incubated at 37°C for 16 h. The RNA product was verified on a denaturing PAA gel followed by ethidium bromide staining. Correct products were excised from the gel and placed into RNA elution buffer (0.1M sodium acetate, 0.1% SDS, 10mM EDTA at 4°C) shaking at 1.000 rpm overnight. The aqueous solution followed the same P/C/I extraction as described for the DNase I digest.

7.2.24 30S subunit toeprinting analysis

The assay was conducted similar as described in Hartz et al. (1988) and Smirnov et al. (2017). In SB buffer without magnesium, 0.2pmol *in vitro*-transcribed *fomA* (first 150nt) and 5'-labeled oligo JVO-18871 were denatured at 95°C for 1min. After incubation on ice for 5 min, the annealing mixture consisting out of 5mM dNTP mixture, SB with magnesium (60mM) and were indicated different molarities of *in vitro*-transcribed FoxI and FoxI-3C were added. This was followed by an incubation step at 37°C for 5 min. Next, the mix was incubated with 0.48µmol of 30S subunits in SB buffer with magnesium (10mM) for an additional 5 min at 37°C for corresponding samples. Pre-activated uncharged $fMet^{tRNA}$ (10pmol) was added to the corresponding samples and incubated for at 37°C for 15 min. This was followed by the reverse transcription reaction through the addition of 100 U SuperScript II reverse transcriptase and incubation for 20 min at 37°C after which the reaction was stopped by adding 100 µl of toeprint stop buffer. Nucleic acids were extracted through P/C/I extraction. Remaining RNA was degraded by treatment with 5 µl of 3M KOH at 90°C for 5 min. The remaining cDNA was precipitated by adding 10 µl of 3 M acetic acid, 1µl of GlycoBlue and

300 µl of 30:1 mix followed by overnight incubation at -20°C. After pelleting the cDNA, the solution was removed, the pellet washed with 70% (V/V) ethanol and finally dissolved in 10 µl of 1x RNA loading buffer. Prior to loading the samples were denatured for 3 min at 90°C. The samples were separated on a denaturing 8% sequencing gel running for 3 h at 40 W. A *fomA*-specific sequencing ladder was generated using the DNA Cycle Sequencing kit following the manufacturer's instructions and included when running the samples. Afterwards the gels were dried and exposed on a phosphor screen. For signal visualization, a Typhoon FLA 7000 phosphorimager (GE Healthcare) was used.

7.2.25 Electrophoretic mobility shift assays (EMSA)

0.04pmol of labeled in vitro-transcribed RNA was denatured for 1 min at 95°C. After a 5 min incubation on ice, 1x Structure Buffer (SB; 10 mM Tris-HCl pH 7.0, 0.1 M KCl, 10 mM $MgCl_2$) was added followed by re-naturation of the RNA for 37°C for 10 min. Next 1µg of Yeast RNA (Ambion) was added. This mixture was added to tubes containing increasing concentrations of the corresponding sRNA. Formation of the mRNA-sRNA complex was allowed to take place for 20 min at 37°C and stopped by adding 5x RNA native loading buffer. The samples were loaded on a native 6% polyacrylamide gel in 0.5% TBE at 4°C and constant current (40mA) for 3 h. Afterwards the gels were dried and exposed on a phosphor screen. For signal visualization, a Typhoon FLA 7000 phosphorimager (GE Healthcare) was used.

7.2.26 Primer extension

10µg of DNase I digested RNA were used as input for each sample. Reverse transcription buffer was added to this containing First Strand buffer (FS), 5mM DTT, 0.5mM dNTP mix and 0.5 U of RNase inhibitor. 1µl of labeled JVO-18871 was added to the elongation mix and denatured for 1 min at 95°C before incubation for 5 min on ice. After elongation mix was incubated for 5 min at 42°C, 1 U of SuperScript II was added and reverse transcription took place for 1 h at 50°C. The enzyme was deactivated by raising the temperature to 70°C for 15 min. Afterwards, RNA was removed by incubating the elongation mix with 1 U of RNase H for 15 min at 37°C. Lastly, the RNA loading buffer was added and the samples were run on a denaturing 6% PAA gel. Afterwards the gels were dried and exposed on a phosphor screen. For signal visualization, a Typhoon FLA 7000

phosphorimager (GE Healthcare) was used.

7.2.27 Coomassie staining and western blot detection

Protein detection was carried out by separating 0.2 OD_{600nm} units (resuspended in protein loading buffer; denatured for 5 min at 95°C) on denaturing 12% SDS-polyacrylamide gels. For direct protein detection the Roti-Blue (Roth) Coomassie was used according to the manufacturer's instruction. In preparation for mass-spectrometry analysis, protein samples were loaded onto Bolt 4 to 12% Bis-Tris gels and stained with SimplyBlue Comassie staining.

For western blot analysis, proteins were transferred to a polyvinylidene fluoride (PVDF) membranes in a semi-dry blot chamber prior to staining with Ponceau S solution. This staining was further used as loading control. After removing the stain by incubating the membranes with 1M NaOH for 1 min, the membranes were blocked in 5% (w/V) skim milk (in TBS-T) for 1 h at room temperature (RT). The membranes were then rinsed and incubated with the primary antibody (diluted in 1% BSA/TBS-T (w/V); 1 h at RT or overnight at 4°C) and appropiate secondary antibody (diluted in TBS-T; 1 h at RT). After each incubation step the membranes were washed thrice in TBS-T for 15 min each. The signal was detected using the ECL chemilumiescent solution using on the Fuji LAS-4000 imager.

7.2.28 Subcellular fractionation

The protocol for the subcellular fractionation was adapted as described by Knoke et al. (2020) for two biological replicates. Mid-exponential phase cultures were collected with 50 OD_{600nm} units by pelleting the cells at 4.000 xg for 10 minutes at 4°C. Total protein samples were obtained after the cell pellet was re-suspended in lysis buffer (10 mM Tris-HCl (pH 7.5), 0.2 mM PMSF, and 1 mM $MgCl_2$). Next, cells were lysed by sonication (20 s burst for 10 s; 15 repeats) after which an lysate samples was obtained after clearing cell debris with another centrifugation step. This was followed by 1 h of centrifugation at 15.000 xg at 4°C which separated the cytosolic (supernatant) from envelope fraction (pellet). The pellet was resuspended in lysis buffer containing 0.5% N-lauroylsarcosine under stirring overnight at 4°C. The next day, the inner membrane fraction (supernatant) was separated from the outer membrane fraction (pellet) through another centrifugation step at 100.000 xg for 2 h at 4°C. The pellet was dissolved in lysis buffer including 0.5% Triton X-100. For the protein analysis 0.1 or 1.0 OD_{600nm}

units were used for total/ cytosol and inner/ outer membrane fraction.

7.2.29 nanoLC-MS/MS analysis of protein samples

For the identification of FomA from Figure 4.11A the region of the decreased signal with FoxI overexpression was excised from total protein samples (all in biological duplicate) and the OM fraction (only control). Prior to trypsin digestion, the samples were destained in 100 mM ammonium bicarbonate containing 30% acetonitrile and shrunk in 100% acetonitrile (0.1g in 100mM ammonium bicarbonate, overnight at 37°C). After that, the samples were diluted in 5% formic acid for nanoLC-MS/MS analysis. NanoLC-MS/MS analysis was performed on an Orbitrap Fusion (Thermo Fisher Scientific) together with a PicoView Ion source (New Objective) and an Easy-nLC 1000 (Thermo Fisher Scientific). The dissolved samples were separated on PicoFrit capillary columns (30cm x 150 M ID, New Objective) packed with ReproSil-Pur 120 C18-AQ (1.9 M, Dr. Maisch) running a linear gradient (3-30% acetonitrile, 0.1% formic acid) at a flow rate of 500nl/min for 140 minutes. The Orbitrap was used for both MS (60,000 scans; AGC target value: 2x105) and MS/MS (7,500 scans; AGC target value: 5x104) analysis, with the HCD fragmentation set to 35% normalized collision energy. For the Top Speed data-dependent MS/MS approach, a fixed cycle time of 3 s was utilized, with an additional exclusion of 1 repetition count every minute. Precursors were chosen at a minimum threshold of 50.000, with single charged ones excluded. An internal calibration with Easy-IC was applied to optimize mass-to-charge ratio assignment. The data was then analyzed using MaxQuant (version 1.5.7.4) with integrated Andromeda and compared to the Uniprot database for *F. nucleatum* subsp. *nucleatum* ATCC 25586 (FUSNN).

7.2.30 Analysis of fluorescent protein expression by confocal microscopy

The pVoPo-FP-carrying *F. nucleatum* was grown to mid-exponential phase. The next steps were carried out outside of the anaerobic chamber. 1 ml of each culture was pelleted and washed in 1ml of PBS. Next, the bacteria were fixed using 4% PFA (w/V) at 4°C for 20 min. This was followed by single washing step in PBS prior to overnight incubation in PBS at 4°C which allows for complete fluorescent protein maturation. The samples were imaged the next day using ibdi chambered coverslips and a Leica SP5 laser scanning confocal microscope (Leica

Microsystems), which captured the fluorescence signal at the designated wave lengths.

Bibliography

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., and Bachrach, G. (2016). Fap2 Mediates Fusobacterium nucleatum Colorectal Adenocarcinoma Enrichment by Binding to Tumor-Expressed Gal-GalNAc. *Cell Host Microbe*, 20(2):215–225.

Acuña, L. G., Barros, M. J., Peñaloza, D., Rodas, P. I., Paredes-Sabja, D., Fuentes, J. A., Gil, F., and Calderón, I. L. (2016). A feed-forward loop between SroC and MgrR small RNAs modulates the expression of eptB and the susceptibility to polymyxin B in Salmonella Typhimurium. *Microbiology (United Kingdom)*, 162(11):1996–2004.

Adams, P. P., Baniulyte, G., Esnault, C., Chegireddy, K., Singh, N., Monge, M., Dale, R. K., Storz, G., and Wade, J. T. (2021). Regulatory roles of Escherichia coli 5' UTR and ORF-internal RNAs detected by 3' end mapping. *eLife*, 10:1–33.

Ades, S. E. (2008). Regulation by destruction: design of the sigmaE envelope stress response. *Curr Opin Microbiol*, 11(6):535–540.

Ades, S. E., Connolly, L. E., Alba, B. M., and Gross, C. A. (1999). The Escherichia coli sigma(E)-dependent extracytoplasmic stress response is controlled by the regulated proteolysis of an anti-sigma factor. *Genes Dev*, 13(18):2449–2461.

Agarwala, R., Barrett, T., Beck, J., Benson, D. A., Bollin, C., Bolton, E., Bourexis, D., Brister, J. R., Bryant, S. H., Canese, K., Charowhas, C., Clark, K., Dicuccio, M., Dondoshansky, I., Federhen, S., Feolo, M., Funk, K., Geer, L. Y., Gorelenkov, V., Hoeppner, M., Holmes, B., Johnson, M., Khotomlianski, V., Kimchi, A., Kimelman, M., Kitts, P., Klimke, W., Krasnov, S., Kuznetsov, A., Landrum, M. J., Landsman, D., Lee, J. M., Lipman, D. J., Lu, Z., Madden, T. L., Madej, T., Marchler-Bauer, A., Karsch-Mizrachi, I., Murphy, T., Orris, R., Ostell, J., O'sullivan, C., Panchenko, A., Phan, L., Preuss, D., Pruitt, K. D., Rodarmer, K., Rubinstein, W., Sayers, E., Schneider, V., Schuler, G. D., Sherry, S. T., Sirotkin, K., Siyan, K., Slotta, D., Soboleva, A., Soussov, V., Starchenko, G., Tatusova, T. A., Todorov, K., Trawick, B. W., Vakatov, D., Wang, Y., Ward, M., Wilbur, W. J., Yaschenko, E., and Zbicz, K. (2016). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, 44(D1):D7–D19.

Akopian, D., Shen, K., Zhang, X., and Shan, S. O. (2013). Signal recognition particle: an essential protein-targeting machine. *Annu Rev Biochem*, 82:693–

721.

Al-Hinai, M. A., Fast, A. G., and Papoutsakis, E. T. (2012). Novel System for Efficient Isolation of Clostridium Double-Crossover Allelic Exchange Mutants Enabling Markerless Chromosomal Gene Deletions and DNA Integration. *Applied and Environmental Microbiology*, 78(22):8112–8121.

Alam, A., Bröms, J. E., Kumar, R., and Sjöstedt, A. (2021). The Role of ClpB in Bacterial Stress Responses and Virulence. *Frontiers in Molecular Biosciences*, 8:283.

Alba, B. M., Leeds, J. A., Onufryk, C., Lu, C. Z., and Gross, C. A. (2002). DegS and YaeL participate sequentially in the cleavage of RseA to activate the sigma(E)-dependent extracytoplasmic stress response. *Genes Dev*, 16(16):2156–2168.

Alexeeva, S., Gadella, T. W., Verheul, J., Verhoeven, G. S., and Den Blaauwen, T. (2010). Direct interactions of early and late assembling division proteins in Escherichia coli cells resolved by FRET. *Molecular Microbiology*, 77(2):384–398.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403–410.

Amann, E., Ochs, B., and Abel, K. J. (1988). Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in Escherichia coli. *Gene*, 69(2):301–315.

Amitay, E. L., Werner, S., Vital, M., Pieper, D. H., Hofler, D., Gierse, I. J., Butt, J., Balavarca, Y., Cuk, K., Brenner, H., Höfler, D., Gierse, I. J., Butt, J., Balavarca, Y., Cuk, K., Brenner, H., Hofler, D., Gierse, I. J., Butt, J., Balavarca, Y., Cuk, K., and Brenner, H. (2017). Fusobacterium and colorectal cancer: causal factor or passenger? Results from a large colorectal cancer screening study. *Carcinogenesis*, 38(8):781–788.

Anderson, M. S., Bulawa, C. E., and Raetz, C. R. (1985). The biosynthesis of gram-negative endotoxin. Formation of lipid A precursors from UDP-GlcNAc in extracts of Escherichia coli. *J Biol Chem*, 260(29):15536–15541.

Andrews, S. C., Robinson, A. K., and Rodríguez-Quiñones, F. (2003). Bacterial iron homeostasis. *FEMS Microbiology Reviews*, 27(2-3):215–237.

Ang, M. Y., Dutta, A., Wee, W. Y., Dymock, D., Paterson, I. C., and Choo, S. W. (2016). Comparative Genome Analysis of Fusobacterium nucleatum. *Genome Biology and Evolution*, 8(9):2928–2938.

Anthony, J. R., Warczak, K. L., and Donohue, T. J. (2005). A transcriptional response to singlet oxygen, a toxic byproduct of photosynthesis. *Proc Natl Acad Sci U S A*, 102(18):6502–6507.

Aoyama, J. J., Raina, M., Zhong, A., and Storz, G. (2022). Dual-function Spot 42 RNA encodes a 15-amino acid protein that regulates the CRP transcription

factor. *Proceedings of the National Academy of Sciences of the United States of America*, 119(10):e2119866119.

Arraiano, C. M., Andrade, J. M., Domingues, S., Guinote, I. B., Malecki, M., Matos, R. G., Moreira, R. N., Pobre, V., Reis, F. P., Saramago, M., Silva, I. J., and Viegas, S. C. (2010). The critical role of RNA processing and degradation in the control of gene expression.

Arsène, F., Tomoyasu, T., and Bukau, B. (2000). The heat shock response of Escherichia coli. In *International Journal of Food Microbiology*, volume 55, pages 3–9. Int J Food Microbiol.

Babu, M. M. and Teichmann, S. A. (2003a). Evolution of transcription factors and the gene regulatory network in Escherichia coli.

Babu, M. M. and Teichmann, S. A. (2003b). Functional determinants of transcription factors in Escherichia coli: Protein families and binding sites.

Bachrach, G., Haake, S. K., Glick, A., Hazan, R., Naor, R., Andersen, R. N., and Kolenbrander, P. E. (2004a). Characterization of the novel Fusobacterium nucleatum plasmid pKH9 and evidence of an addiction system. *Applied and Environmental Microbiology*, 70(12):6957–6962.

Bachrach, G., Rosen, G., Bellalou, M., Naor, R., and Sela, M. N. (2004b). Identification of a Fusobacterium nucleatum 65 kDa serine protease. *Oral Microbiology and Immunology*, 19(3):155–159.

Bader, M. W., Sanowar, S., Daley, M. E., Schneider, A. R., Cho, U., Xu, W., Klevit, R. E., Le Moual, H., and Miller, S. I. (2005). Recognition of antimicrobial peptides by a bacterial sensor kinase. *Cell*, 122(3):461–472.

Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The MEME Suite. *Nucleic Acids Res*, 43(W1):W39–49.

Bak, G., Lee, J. Y. J., Suk, S., Kim, D., Lee, J. Y. J., Kim, K. S., Choi, B. S., Lee, Y., Young Lee, J., Kim, K. S., Choi, B. S., and Lee, Y. (2015). Identification of novel sRNAs involved in biofilm formation, motility, and fimbriae formation in Escherichia coli. *Sci Rep*, 5:15287.

Balleza, E., López-Bojorquez, L. N., Martínez-Antonio, A., Resendis-Antonio, O., Lozada-Chávez, I., Balderas-Martínez, Y. I., Encarnación, S., and Collado-Vides, J. (2009). Regulation by transcription factors in bacteria: beyond description. *FEMS Microbiology Reviews*, 33(1):133–151.

Banno, S., Nishida, K., Arazoe, T., Mitsunobu, H., and Kondo, A. (2018). Deaminase-mediated multiplex genome editing in Escherichia coli. *Nature Microbiology*, 3(4):423–429.

Barnhill, E. C., Crucello, A., Houserova, D., King, V. M., Amin, S. V., Roberts, J. T., Zambrano, M. E., DeMeis, J. D., Dahmer, D. J., Ijaz, Z., Barchie, A. A., Watters, B. C., Prusak, J. E., Dean, M. A., Holton, N. W., Ferreira-Filho,

J. A., Sant'Ana, A. S., Spector, M. P., and Borchert, G. M. (2019). Characterization of novel small RNAs (sRNAs) contributing to the desiccation response of Salmonella enterica serovar Typhimurium. *RNA Biology*, 16(11):1643–1657.

Barrangou, R. and Horvath, P. (2017). A decade of discovery: CRISPR functions and applications.

Barrangou, R. and van Pijkeren, J. P. (2016). Exploiting CRISPR-Cas immune systems for genome editing in bacteria.

Barter, R. L. and Yu, B. (2018). Superheat: An R Package for Creating Beautiful and Extendable Heatmaps for Visualizing Complex Data. *Journal of Computational and Graphical Statistics*, 27(4):910–922.

Beck, H. J. and Moll, I. (2018). Leaderless mRNAs in the Spotlight: Ancient but Not Outdated! *Microbiology Spectrum*.

Belliveau, N. M., Barnes, S. L., Ireland, W. T., Jones, D. L., Sweredoski, M. J., Moradian, A., Hess, S., Kinney, J. B., and Phillips, R. (2018). Systematic approach for dissecting the molecular mechanisms of transcriptional regulation in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 115(21):E4796–E4805.

Bender, R. A. (2012). Regulation of the Histidine Utilization (Hut) System in Bacteria. *Microbiology and Molecular Biology Reviews*, 76(3):565–584.

Berghoff, B. A., Glaeser, J., Sharma, C. M., Vogel, J., and Klug, G. (2009). Photooxidative stress-induced and abundant small RNAs in Rhodobacter sphaeroides. *Mol Microbiol*, 74(6):1497–1512.

Bertram, R. and Hillen, W. (2008). The application of Tet repressor in prokaryotic gene regulation and expression. *Microb Biotechnol*, 1(1):2–16.

Bi, D., Zhu, Y., Gao, Y., Li, H., Zhu, X., Wei, R., Xie, R., Cai, C., Wei, Q., and Qin, H. (2022). Profiling Fusobacterium infection at high taxonomic resolution reveals lineage-specific correlations in colorectal cancer. *Nature Communications*, 13(1):1–12.

Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., and Marraffini, L. A. (2013). Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Research*, 41(15):7429–7437.

Bischler, T., Tan, H. S., Nieselt, K., and Sharma, C. M. (2015). Differential RNA-seq (dRNA-seq) for annotation of transcriptional start sites and small RNAs in Helicobacter pylori. *Methods*, 86:89–101.

Blattman, S. B., Jiang, W., Oikonomou, P., and Tavazoie, S. (2020). Prokaryotic single-cell RNA sequencing by in situ combinatorial indexing. *Nat Microbiol*, 5(10):1192–1201.

Bobonis, J., Mitosch, K., Mateus, A., Karcher, N., Kritikos, G., Selkrig, J., Zietek, M., Monzon, V., Pfalz, B., Garcia-Santamarina, S., Galardini, M., Sueki, A., Kobayashi, C., Stein, F., Bateman, A., Zeller, G., Savitski, M. M., Elfenbein, J. R., Andrews-Polymenis, H. L., and Typas, A. (2022). Bacterial retrons encode phage-defending tripartite toxin–antitoxin systems. *Nature*, 609(7925).

Bonner, E. R., D'Elia, J. N., Billips, B. K., and Switzer, R. L. (2001). Molecular recognition of pyr mRNA by the Bacillus subtilis attenuation regulatory protein PyrR. *Nucleic Acids Res*, 29(23):4851–4865.

Borukhov, S. and Nudler, E. (2008). RNA polymerase: the vehicle of transcription.

Bossi, L., Maloriol, D., and Figueroa-Bossi, N. (2008). Porin biogenesis activates the σE response in Salmonella hfq mutants. *Biochimie*, 90(10):1539–1544.

Bradshaw, D. J., Marsh, P. D., Keith Watson, G., and Allison, C. (1998). Role of Fusobacterium nucleatum and coaggregation in anaerobe survival in planktonic and biofilm oral microbial communities during aeration. *Infection and Immunity*, 66(10):4729–4732.

Braun, V. (1995). Energy-coupled transport and signal transduction through the Gram-negative outer membrane via TonB-ExbB-ExbD-dependent receptor proteins. *FEMS Microbiology Reviews*, 16(4):295–307.

Braun, V. (2005). Bacterial iron transport related to virulence.

Breaker, R. R. (2022). The Biochemical Landscape of Riboswitch Ligands. *Biochemistry*, 61(3):137–149.

Brennan, C. A. and Garrett, W. S. (2019). Fusobacterium nucleatum - symbiont, opportunist and oncobacterium. *Nat Rev Microbiol*, 17(3):156–166.

Brewer, M. L., Dymock, D., Brady, R. L., Singer, B. B., Virji, M., and Hill, D. J. (2019). Fusobacterium spp. target human CEACAM1 via the trimeric autotransporter adhesin CbpF. *Journal of Oral Microbiology*, 11(1).

Bronzato, J. D., Bomfim, R. A., Edwards, D. H., Crouch, D., Hector, M. P., and Gomes, B. P. (2020). Detection of Fusobacterium in oral and head and neck cancer samples: A systematic review and meta-analysis.

Brook, I. (1994). Fusobacterial infections in children. *Journal of Infection*, 28(2):155–165.

Browning, D. F. and Busby, S. J. W. (2016). Local and global regulation of transcription initiation in bacteria. *Nature Reviews Microbiology*, 14(10):638–650.

Browning, D. F., Butala, M., and Busby, S. J. (2019). Bacterial Transcription Factors: Regulation by Pick "N" Mix.

Bullman, S., Pedamallu, C. S., Sicinska, E., Clancy, T. E., Zhang, X., Cai, D., Neuberg, D., Huang, K., Guevara, F., Nelson, T., Chipashvili, O., Hagan, T.,

Walker, M., Ramachandran, A., Diosdado, B., Serna, G., Mulet, N., Landolfi, S., Ramon, Y. C. S., Fasani, R., Aguirre, A. J., Ng, K., Élez, E., Ogino, S., Tabernero, J., Fuchs, C. S., Hahn, W. C., Nuciforo, P., Meyerson, M., Elez, E., Ogino, S., Tabernero, J., Fuchs, C. S., Hahn, W. C., Nuciforo, P., and Meyerson, M. (2017). Analysis of Fusobacterium persistence and antibiotic response in colorectal cancer. *Science*, 358(6369):1443–1448.

Butler, E. B., Xiong, Y., Wang, J., and Strobel, S. A. (2011). Structural basis of cooperative ligand binding by the glycine riboswitch. *Chemistry and Biology*, 18(3):293–298.

Cain, A. K., Barquist, L., Goodman, A. L., Paulsen, I. T., Parkhill, J., and van Opijnen, T. (2020). A decade of advances in transposon-insertion sequencing. *Nat Rev Genet*, 21(9):526–540.

Call, S. N. and Andrews, L. B. (2022). CRISPR-Based Approaches for Gene Regulation in Non-Model Bacteria. *Frontiers in Genome Editing*, 4.

Calogero, R. A., Pon, C. L., Canonaco, M. A., and Gualerzi, C. O. (1988). Selection of the mRNA translation initiation region by Escherichia coli ribosomes. *Proceedings of the National Academy of Sciences of the United States of America*, 85(17):6427–6431.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, 10.

Campagne, S., Marsh, M. E., Capitani, G., Vorholt, J. A., and Allain, F. H.-T. (2014). Structural basis for -10 promoter element melting by environmentally induced sigma factors. *Nature Structural & Molecular Biology*, 21(3):269–276.

Campbell, E. A., Greenwell, R., Anthony, J. R., Wang, S., Lim, L., Das, K., Sofia, H. J., Donohue, T. J., and Darst, S. A. (2007). A conserved structural module regulates transcriptional responses to diverse stress signals in bacteria. *Mol Cell*, 27(5):793–805.

Cao, H., Ma, Q., Chen, X., and Xu, Y. (2019). DOOR: a prokaryotic operon database for genome analyses and functional inference. *Brief Bioinform*, 20(4):1568–1577.

Carbon, J., Shenk, T. E., and Berg, P. (1975). Biochemical procedure for production of small deletions in simian virus 40 DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 72(4):1392–1396.

Caron, M. P., Lafontaine, D. A., and Massé, E. (2010). Small RNA-mediated regulation at the level of transcript stability.

Carpenter, B. M., Whitmire, J. M., and Merrell, D. S. (2009). This is not your mother's repressor: the complex role of fur in pathogenesis. *Infect Immun*, 77(7):2590–2601.

Carte, J., Wang, R., Li, H., Terns, R. M., and Terns, M. P. (2008). Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes and Development*, 22(24):3489–3496.

Casas-Pastor, D., Müller, R. R., Jaenicke, S., Brinkrolf, K., Becker, A., Buttner, M. J., Gross, C. A., Mascher, T., Goesmann, A., and Fritz, G. (2021). Expansion and re-classification of the extracytoplasmic function (ECF) σ factor family. *Nucleic Acids Res*, 49(2):986–1005.

Casasanta, M. A., Yoo, C. C., Smith, H. B., Duncan, A. J., Cochrane, K., Varano, A. C., Allen-Vercoe, E., and Slade, D. J. (2017). A chemical and biological toolbox for Type Vd secretion: Characterization of the phospholipase A1 autotransporter FplA from Fusobacterium nucleatum. *Journal of Biological Chemistry*, 292(49):20240–20254.

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., and 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(641).

Castellarin, M., Warren, R. L., Freeman, J. D., Dreolini, L., Krzywinski, M., Strauss, J., Barnes, R., Watson, P., Allen-Vercoe, E., Moore, R. A., and Holt, R. A. (2012). Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. *Genome Res*, 22(2):299–306.

Cavanagh, A. T., Chandrangsu, P., and Wassarman, K. M. (2010). 6S RNA regulation of relA alters ppGpp levels in early stationary phase. *Microbiology*, 156(12):3791–3800.

Cavanagh, A. T., Klocko, A. D., Liu, X., and Wassarman, K. M. (2008). Promoter specificity for 6S RNA regulation of transcription is determined by core promoter sequences and competition for region 4.2 of σ ⁷⁰. *Molecular Microbiology*, 67(6):1242–1256.

Cervantes-Rivera, R., Tronnet, S., and Puhar, A. (2020). Complete genome sequence and annotation of the laboratory reference strain Shigella flexneri serotype 5a M90T and genome-wide transcriptional start site determination. *BMC Genomics*, 21(1).

Chao, Y. and Vogel, J. (2016). A 3' UTR-Derived Small RNA Provides the Regulatory Noncoding Arm of the Inner Membrane Stress Response. *Mol Cell*, 61(3):352–363.

Chen, H. and Jiang, W. (2014). Application of high-throughput sequencing in understanding human oral microbiome related with health and disease. *Frontiers in Microbiology*, 5(SEP).

Chen, S., Zhang, A., Blyn, L. B., and Storz, G. (2004). MicC, a second small-RNA regulator of Omp protein expression in Escherichia coli. *J Bacteriol*,

186(20):6689–6697.

Chen, Y.-W., Camacho, M. I., Chen, Y., Bhat, A. H., Chang, C., Peluso, E. A., Wu, C., Das, A., and Ton-That, H. (2022). Genetic Determinants of Hydrogen Sulfide Biosynthesis in Fusobacterium nucleatum Are Required for Bacterial Fitness, Antibiotic Sensitivity, and Virulence. *mBio*.

Citron, D. M. (2002). Update on the taxonomy and clinical aspects of the genus fusobacterium. *Clin Infect Dis*, 35(Suppl 1):S22–7.

Cochrane, K., McGuire, A. M., Priest, M. E., Abouelleil, A., Cerqueira, G. C., Lo, R., Earl, A. M., and Allen-Vercoe, E. (2016). Complete genome sequences and analysis of the Fusobacterium nucleatum subspecies animalis 7-1 bacteriophage phi-Funu1 and phi-Funu2. *Anaerobe*, 38:125–129.

Cochrane, K., Robinson, A. V., Holt, R. A., and Allen-Vercoe, E. (2020). A survey of Fusobacterium nucleatum genes modulated by host cell infection. *Microb Genom*, 6(2).

Coleman, G. A., Davín, A. A., Mahendrarajah, T. A., Szánthó, L. L., Spang, A., Hugenholtz, P., Szöllősi, G. J., and Williams, T. A. (2021). A rooted phylogeny resolves early bacterial evolution. *Science*, 372(6542).

Cook, J., Baverstock, T. C., McAndrew, M. B., Stansfeld, P. J., Roper, D. I., and Crow, A. (2020). Insights into bacterial cell division from a structure of EnvC bound to the FtsX periplasmic domain. *Proceedings of the National Academy of Sciences of the United States of America*, 117(45):28355–28365.

Coppenhagen-Glazer, S., Sol, A., Abed, J., Naor, R., Zhang, X., Han, Y. W., and Bachrach, G. (2015). Fap2 of Fusobacterium nucleatum is a galactose-inhibitable adhesin involved in coaggregation, cell adhesion, and preterm birth. *Infection and Immunity*, 83(3):1104–1113.

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Research*, 16(22):10881–10890.

Correia Santos, S., Bischler, T., Westermann, A. J., and Vogel, J. (2021). MAPS integrates regulation of actin-targeting effector SteC into the virulence control network of Salmonella small RNA PinT. *Cell Reports*, 34(5).

Creecy, J. P. and Conway, T. (2015). Quantitative bacterial transcriptomics with RNA-seq.

Cromie, M. J., Shi, Y., Latifi, T., and Groisman, E. A. (2006). An RNA Sensor for Intracellular Mg2+. *Cell*, 125(1):71–84.

Cui, W., Han, L., Cheng, J., Liu, Z., Zhou, L., Guo, J., and Zhou, Z. (2016). Engineering an inducible gene expression system for Bacillus subtilis from a strong constitutive promoter and a theophylline-activated synthetic riboswitch. *Microbial Cell Factories*, 15(1):199.

Dalebroux, Z. D. and Miller, S. I. (2014). Salmonellae PhoPQ regulation of the outer membrane to resist innate immunity.

Damron, F. H., Oglesby-Sherrouse, A. G., Wilks, A., and Barbier, M. (2016). Dual-seq transcriptomics reveals the battle for iron during Pseudomonas aeruginosa acute murine pneumonia. *Scientific Reports*, 6(1):1–12.

Damron, F. H., Owings, J. P., Okkotsu, Y., Varga, J. J., Schurr, J. R., Goldberg, J. B., Schurr, M. J., and Yu, H. D. (2012). Analysis of the Pseudomonas aeruginosa regulon controlled by the sensor kinase KinB and sigma factor RpoN. *Journal of Bacteriology*, 194(6):1317–1330.

Dar, D., Shamir, M., Mellin, J. R., Koutero, M., Stern-Ginossar, N., Cossart, P., and Sorek, R. (2016). Term-seq reveals abundant ribo-regulation of antibiotics resistance in bacteria. *Science*, 352(6282).

Dartigalongue, C., Missiakas, D., and Raina, S. (2001). Characterization of the Escherichia coli sigma E regulon. *J Biol Chem*, 276(24):20866–20875.

Darty, K., Denise, A., and Ponty, Y. (2009). VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, 25(15):1974–1975.

De Las Peñas, A., Connolly, L., and Gross, C. A. (1997). The sigmaE-mediated response to extracytoplasmic stress in Escherichia coli is transduced by RseA and RseB, two negative regulators of sigmaE. *Mol Microbiol*, 24(2):373–385.

De Lay, N. and Gottesman, S. (2009). The crp-activated small noncoding regulatory RNA CyaR (RyeE) links nutritional status to group behavior. *Journal of Bacteriology*, 191(2):461–476.

De Lay, N., Schu, D. J., and Gottesman, S. (2013). Bacterial small RNA-based negative regulation: Hfq and its accomplices.

Despins, C. A., Brown, S. D., Robinson, A. V., Mungall, A. J., Allen-Vercoe, E., and Holt, R. A. (2021). Modulation of the Host Cell Transcriptome and Epigenome by Fusobacterium nucleatum. *mBio*, 12(5).

Desvaux, M., Khan, A., Beatson, S. A., Scott-Tucker, A., and Henderson, I. R. (2005). Protein secretion systems in Fusobacterium nucleatum: genomic identification of Type 4 piliation and complete Type V pathways brings new insight into mechanisms of pathogenesis. *Biochim Biophys Acta*, 1713(2):92–112.

Diaz, P. I., Zilm, P. S., and Rogers, A. H. (2002). Fusobacterium nucleatum supports the growth of Porphyromonas gingivalis in oxygenated and carbon-dioxide-depleted environments. *Microbiology*, 148(2):467–472.

Didilescu, A. C., Rusu, D., Anghel, A., Nica, L., Iliescu, A., Greabu, M., Bancescu, G., and Stratul, S. I. (2012). Investigation of six selected bacterial species in endo-periodontal lesions. *International Endodontic Journal*, 45(3):282–293.

Ding, Y., Davis, B. M., and Waldor, M. K. (2004). Hfq is essential for Vib-

rio cholerae virulence and downregulates sigma expression. *Mol Microbiol*, 53(1):345–354.

Doron, L., Coppenhagen-Glazer, S., Ibrahim, Y., Eini, A., Naor, R., Rosen, G., and Bachrach, G. (2014). Identification and characterization of fusolisin, the Fusobacterium nucleatumautotransporter serine protease. *PLoS ONE*, 9(10).

Doron, S., Melamed, S., Ofir, G., Leavitt, A., Lopatina, A., Keren, M., Amitai, G., and Sorek, R. (2018). Systematic discovery of antiphage defense systems in the microbial pangenome. *Science*, 359(6379).

Drepper, T., Eggert, T., Circolone, F., Heck, A., Krauß, U., Guterl, J. K., Wendorff, M., Losi, A., Gärtner, W., and Jaeger, K. E. (2007). Reporter proteins for in vivo fluorescence without oxygen. *Nature Biotechnology*, 25(4):443–445.

Dugar, G., Herbig, A., Forstner, K. U., Heidrich, N., Reinhardt, R., Nieselt, K., and Sharma, C. M. (2013). High-resolution transcriptome maps reveal strain-specific regulatory features of multiple Campylobacter jejuni isolates. *PLoS Genet*, 9(5):e1003495.

Durica-Mitic, S., Göpel, Y., Amman, F., and Görke, B. (2020). Adaptor protein RapZ activates endoribonuclease RNase e by protein-protein interaction to cleave a small regulatory RNA. *RNA*, 26(9):1198–1215.

Ellis, M. J., Trussler, R. S., and Haniford, D. B. (2015). A cis-encoded sRNA, Hfq and mRNA secondary structure act independently to suppress IS200 transposition. *Nucleic Acids Research*, 43(13):6511–6527.

Engevik, M., Danhof, H. A., Auchtung, J., Endres, B. T., Ruan, W., Bassères, E., Engevik, A. C., Wu, Q., Nicholson, M., Luna, R. A., Garey, K. W., Crawford, S. E., Estes, M. K., Lux, R., Yacyshyn, M. B., Yacyshyn, B., Savidge, T., Britton, R. A., and Versalovic, J. (2020). Fusobacterium nucleatum adheres to Clostridioides difficile via the RadD adhesin to enhance biofilm formation in intestinal mucus. *Gastroenterology*.

Epshtein, V., Mironov, A. S., and Nudler, E. (2003). The riboswitch-mediated control of sulfur metabolism in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 100(9):5052–5056.

Erhardt, M. and Dersch, P. (2015). Regulatory principles governing Salmonella and Yersinia virulence.

Erickson, J. W. and Gross, C. A. (1989). Identification of the sigma E subunit of Escherichia coli RNA polymerase: a second alternate sigma factor involved in high-temperature gene expression. *Genes Dev*, 3(9):1462–1471.

Errington, J. (2015). Bacterial morphogenesis and the enigmatic MreB helix.

Escolar, L., Pérez-Martín, J., and De Lorenzo, V. (1999). Opening the iron box: Transcriptional metalloregulation by the fur protein.

Evguenieva-Hackenberg, E. (2022). Riboregulation in bacteria: From general principles to novel mechanisms of the trp attenuator and its sRNA and peptide products.

Fagan, R. P. and Fairweather, N. F. (2011). Clostridium difficile has two parallel and essential Sec secretion systems. *J Biol Chem*, 286(31):27483–27493.

Fan, Y. and Pedersen, O. (2021). Gut microbiota in human metabolic health and disease. *Nat Rev Microbiol*, 19(1):55–71.

Fardini, Y., Chung, P., Dumm, R., Joshi, N., and Han, Y. W. (2010). Transmission of diverse oral bacteria to murine placenta: Evidence for the oral microbiome as a potential source of intrauterine infection. *Infection and Immunity*, 78(4):1789–1796.

Feklístov, A., Sharon, B. D., Darst, S. A., Gross, C. A., Feklistov, A., Sharon, B. D., Darst, S. A., and Gross, C. A. (2014). Bacterial sigma factors: a historical, structural, and genomic perspective. *Annual Review of Microbiology*, 68(1):357–376.

Feng, Q. (2015). Gut microbiome development along the colorectal adenoma-carcinoma sequence. *Nat. Commun.*, 6.

Ferrara, S., Carloni, S., Fulco, R., Falcone, M., Macchi, R., and Bertoni, G. (2015). Post-transcriptional regulation of the virulence-associated enzyme AlgC by the $\sigma(22)$ -dependent small RNA ErsA of Pseudomonas aeruginosa. *Environ Microbiol*, 17(1):199–214.

Ferrara, S., Carrubba, R., Santoro, S., and Bertoni, G. (2021). The Small RNA ErsA Impacts the Anaerobic Metabolism of Pseudomonas aeruginosa Through Post-Transcriptional Modulation of the Master Regulator Anr. *Frontiers in Microbiology*, 12.

Figueroa-Bossi, N., Lemire, S., Maloriol, D., Balbontín, R., Casadesús, J., and Bossi, L. (2006). Loss of Hfq activates the sigmaE-dependent envelope stress response in Salmonella enterica. *Mol Microbiol*, 62(3):838–852.

Fimlaid, K. A. and Shen, A. (2015). Diverse mechanisms regulate sporulation sigma factor activity in the Firmicutes.

Flores-Kim, J. and Darwin, A. J. (2016). The Phage Shock Protein Response. *Annual Review of Microbiology*, 70(1):83–101.

Flynn, J. M., Levchenko, I., Sauer, R. T., and Baker, T. A. (2004). Modulating substrate choice: The SspB adaptor delivers a regulator of the extracytoplasmic-stress response to the AAA+ protease ClpXP for degradation. *Genes and Development*, 18(18):2292–2301.

Flynn, J. M., Levchenko, I., Seidel, M., Wickner, S. H., Sauer, R. T., and Baker, T. A. (2001). Overlapping recognition determinants within the ssrA degradation tag allow modulation of proteolysis. *Proceedings of the National Academy of*

Sciences of the United States of America, 98(19):10584–10589.

Fontaine, F., Fuchs, R. T., and Storz, G. (2011). Membrane localization of small proteins in Escherichia coli. *Journal of Biological Chemistry*, 286(37):32464–32474.

Förstner, K. U., Vogel, J., and Sharma, C. M. (2014). READemption-a tool for the computational analysis of deep-sequencing-based transcriptome data. *Bioinformatics*, 30(23):3421–3423.

Fox, K. A., Ramesh, A., Stearns, J. E., Bourgogne, A., Reyes-Jara, A., Winkler, W. C., and Garsin, D. A. (2009). Multiple posttranscriptional regulatory mechanisms partner to control ethanolamine utilization in Enterococcus faecalis. *Proceedings of the National Academy of Sciences*, 106(11):4435–4440.

Fröhlich, K. S. and Gottesman, S. (2018). Small Regulatory RNAs in the Enterobacterial Response to Envelope Damage and Oxidative Stress. *Microbiol Spectr*, 6(4).

Fröhlich, K. S., Papenfort, K., Fekete, A., and Vogel, J. (2013). A small RNA activates CFA synthase by isoform-specific mRNA stabilization. *EMBO Journal*, 32(22):2963–2979.

Fuchs, M., Lamm-Schmidt, V., Sulzer, J., Ponath, F., Jenniches, L., Kirk, J. A., Fagan, R. P., Barquist, L., Vogel, J., and Faber, F. (2021). An RNA-centric global view of Clostridioides difficile reveals broad activity of Hfq in a clinically important gram-positive bacterium. *Proceedings of the National Academy of Sciences of the United States of America*, 118(25).

Gaballa, A., Antelmann, H., Aguilar, C., Khakh, S. K., Song, K. B., Smaldone, G. T., and Helmann, J. D. (2008). The Bacillus subtilis iron-sparing response is mediated by a Fur-regulated small RNA and three small, basic proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 105(33):11927–11932.

Gagarinova, A. and Emili, A. (2012). Genome-scale genetic manipulation methods for exploring bacterial molecular biology.

Gaiser, R. A., Halimi, A., Alkharaan, H., Lu, L., Davanian, H., Healy, K., Hugerth, L. W., Ateeb, Z., Valente, R., Fernández Moro, C., Del Chiaro, M., and Sällberg Chen, M. (2019). Enrichment of oral microbiota in early cystic precursors to invasive pancreatic cancer. *Gut*, 68(12):2186–2194.

Galaski, J., Shhadeh, A., Umaña, A., Yoo, C. C., Arpinati, L., Isaacson, B., Berhani, O., Singer, B. B., Slade, D. J., Bachrach, G., and Mandelboim, O. (2021). Fusobacterium nucleatum CbpF Mediates Inhibition of T Cell Function Through CEACAM1 Activation. *Frontiers in Cellular and Infection Microbiology*, 11.

García Véscovi, E., Soncini, F. C., and Groisman, E. A. (1996). Mg2+ as an

extracellular signal: environmental regulation of Salmonella virulence. *Cell*, 84(1):165–174.

Garsin, D. A. (2010). Ethanolamine utilization in bacterial pathogens: roles and regulation. *Nature Reviews Microbiology*, 8(4):290–295.

Gay, P., Le Coq, D., Steinmetz, M., Berkelman, T., and Kado, C. I. (1985). Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. *Journal of Bacteriology*, 164(2):918–921.

Geissmann, T., Marzi, S., and Romby, P. (2009). The role of mRNA structure in translational control in bacteria.

Georg, J. and Hess, W. R. (2018). Widespread Antisense Transcription in Prokaryotes. *Microbiol Spectr*, 6(4).

Gerovac, M., El Mouali, Y., Kuper, J., Kisker, C., Barquist, L., and Vogel, J. (2020). Global discovery of bacterial RNA-binding proteins by RNase-sensitive gradient profiles reports a new FinO domain protein. *Rna*, 26(10):1448–1463.

Gerovac, M., Vogel, J., and Smirnov, A. (2021). The World of Stable Ribonucleoproteins and Its Mapping With Grad-Seq and Related Approaches. *Frontiers in Molecular Biosciences*, 8.

Gilbert, J. A., Blaser, M. J., Caporaso, J. G., Jansson, J. K., Lynch, S. V., and Knight, R. (2018). Current understanding of the human microbiome. *Nature Medicine*, 24(4):392–400.

Gill, E. E., Chan, L. S., Winsor, G. L., Dobson, N., Lo, R., Ho Sui, S. J., Dhillon, B. K., Taylor, P. K., Shrestha, R., Spencer, C., Hancock, R. E., Unrau, P. J., and Brinkman, F. S. (2018). High-throughput detection of RNA processing in bacteria. *BMC Genomics*, 19(1).

Gimpel, M. and Brantl, S. (2017). Dual-function small regulatory RNAs in bacteria. *Molecular Microbiology*, 103(3):387–397.

Giuliodori, A. M., Di Pietro, F., Marzi, S., Masquida, B., Wagner, R., Romby, P., Gualerzi, C. O., and Pon, C. L. (2010). The cspA mRNA Is a Thermosensor that Modulates Translation of the Cold-Shock Protein CspA. *Molecular Cell*, 37(1):21–33.

Glasscock, C. J., Biggs, B. W., Lazar, J. T., Arnold, J. H., Burdette, L. A., Valdes, A., Kang, M. K., Tullman-Ercek, D., Tyo, K. E., and Lucks, J. B. (2021). Dynamic Control of Gene Expression with Riboregulated Switchable Feedback Promoters. *ACS Synthetic Biology*, 10(5):1199–1213.

Gogol, E. B., Rhodius, V. A., Papenfort, K., Vogel, J., and Gross, C. A. (2011). Small RNAs endow a transcriptional activator with essential repressor functions for single-tier control of a global stress regulon. *Proc Natl Acad Sci U S A*, 108(31):12875–12880.

Göpel, Y., Papenfort, K., Reichenbach, B., Vogel, J., and Görke, B. (2013). Targeted decay of a regulatory small RNA by an adaptor protein for RNase E and counteraction by an anti-adaptor RNA. *Genes and Development*, 27(5):552–564.

Gottesman, S. (2019). Trouble is coming: Signaling pathways that regulate general stress responses in bacteria.

Gottesman, S. and Storz, G. (2011). Bacterial small RNA regulators: Versatile roles and rapidly evolving variations. *Cold Spring Harbor Perspectives in Biology*, 3(12).

Gray, T., Storz, G., and Papenfort, K. (2022). Small Proteins; Big Questions. In Henkin, T. M., editor, *Journal of Bacteriology*, volume 204. American Society for Microbiology.

Green, J., Crack, J. C., Thomson, A. J., and LeBrun, N. E. (2009). Bacterial sensors of oxygen. *Current Opinion in Microbiology*, 12(2):145–151.

Greenwell, R., Nam, T. W., and Donohue, T. J. (2011). Features of Rhodobacter sphaeroides ChrR required for stimuli to promote the dissociation of σ(E)/ChrR complexes. *J Mol Biol*, 407(4):477–491.

Griffen, A. L., Beall, C. J., Campbell, J. H., Firestone, N. D., Kumar, P. S., Yang, Z. K., Podar, M., and Leys, E. J. (2012). Distinct and complex bacterial profiles in human periodontitis and health revealed by 16S pyrosequencing. *Isme j*, 6(6):1176–1185.

Groisman, E. A. (2001). The pleiotropic two-component regulatory system PhoP-PhoQ. *Journal of Bacteriology*, 183(6):1835–1842.

Gross, C. A., Chan, C., Dombroski, A., Gruber, T., Sharp, M., Tupy, J., and Young, B. (1998). The functional and regulatory roles of sigma factors in transcription. In *Cold Spring Harbor Symposia on Quantitative Biology*, volume 63, pages 141–155. Cold Spring Harbor Laboratory Press.

Guisbert, E., Rhodius, V. A., Ahuja, N., Witkin, E., and Gross, C. A. (2007). Hfq modulates the sigmaE-mediated envelope stress response and the sigma32-mediated cytoplasmic stress response in Escherichia coli. *J Bacteriol*, 189(5):1963–1973.

Guo, L., Shokeen, B., He, X., Shi, W., and Lux, R. (2017). Streptococcus mutans SpaP binds to RadD of Fusobacterium nucleatum ssp. polymorphum. *Molecular Oral Microbiology*, 32(5):355–364.

Guo, M. S., Updegrove, T. B., Gogol, E. B., Shabalina, S. A., Gross, C. A., and Storz, G. (2014). MicL, a new sigmaE-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. *Genes Dev*, 28(14):1620–1634.

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., Miklic, K., Jonjic, S., Brennan, C. A., Garrett, W. S., Bachrach, G., Mandelboim, O., Miklić, K., Jonjic, S., Brennan, C. A., Garrett, W. S., Bachrach, G., and Mandelboim, O. (2015). Binding of the Fap2 protein of Fusobacterium nucleatum to human inhibitory receptor TIGIT protects tumors from immune cell attack. *Immunity*, 42(2):344–355.

Gur, C., Maalouf, N., Shhadeh, A., Berhani, O., Singer, B. B., Bachrach, G., and Mandelboim, O. (2019). Fusobacterium nucleatum supresses anti-tumor immunity by activating CEACAM1. *OncoImmunology*, 8(6).

Haake, S. K., Yoder, S. C., Attarian, G., and Podkaminer, K. (2000). Native plasmids of Fusobacterium nucleatum: Characterization and use in development of genetic systems. *Journal of Bacteriology*, 182(4):1176–1180.

Hadley, W. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Han, K., Tjaden, B., and Lory, S. (2016). GRIL-seq provides a method for identifying direct targets of bacterial small regulatory RNA by in vivo proximity ligation. *Nature Microbiology*, 2.

Han, Y. W. (2015). Fusobacterium nucleatum: a commensal-turned pathogen. *Curr Opin Microbiol*, 23:141–147.

Han, Y. W., Ikegami, A., Rajanna, C., Kawsar, H. I., Zhou, Y., Li, M., Sojar, H. T., Genco, R. J., Kuramitsu, H. K., and Deng, C. X. (2005). Identification and characterization of a novel adhesin unique to oral fusobacteria. *J Bacteriol*, 187(15):5330–5340.

Hantke, K. (1984). Cloning of the repressor protein gene of iron-regulated systems in Escherichia coli K12. *MGG Molecular and General Genetics*, 197(2):337–341.

Harayama, S., Bollinger, J., Iino, T., and Hazelbauer, G. L. (1983). Characterization of the mgl operon of Escherichia coli by transposon mutagenesis and molecular cloning. *Journal of Bacteriology*, 153(1):408–415.

Hartmann, E. and Hartmann, R. K. (2003). The enigma of ribonuclease P evolution. *Trends Genet*, 19(10):561–569.

Hartz, D., McPheeters, D. S., and Gold, L. (1989). Selection of the initiator tRNA by Escherichia coli initiation factors. *Genes and Development*, 3(12 A):1899–1912.

Hartz, D., McPheeters, D. S., Traut, R., and Gold, L. (1988). Extension Inhibition Analysis of Translation Initiation Complexes. *Methods in Enzymology*, 164(C):419–425.

Hatfull, G. F., Dedrick, R. M., and Schooley, R. T. (2022). Phage Therapy for Antibiotic-Resistant Bacterial Infections. *Annual Review of Medicine*, 73(1):197–211.

Hawley, D. K. and Mcclure, W. R. (1983). Compilation and analysis of Escherichia coli promoter DNA sequences. *Nucleic Acids Research*, 11(8):2237–2255.

Heim, R., Prasher, D. C., and Tsien, R. Y. (1994). Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proceedings of the National Academy of Sciences*, 91(26):12501–12504.

Helaine, S., Cheverton, A. M., Watson, K. G., Faure, L. M., Matthews, S. A., and Holden, D. W. (2014). Internalization of salmonella by macrophages induces formation of nonreplicating persisters. *Science*, 343(6167):204–208.

Helmann, J. D. (2002). The extracytoplasmic function (ECF) sigma factors. *Adv Microb Physiol*, 46:47–110.

Hemm, M. R., Paul, B. J., Miranda-Ríos, J., Zhang, A., Soltanzad, N., and Storz, G. (2010). Small stress response proteins in Escherichia coli: proteins missed by classical proteomic studies. *Journal of bacteriology*, 192(1):46–58.

Hemm, M. R., Weaver, J., and Storz, G. (2020). Escherichia coli Small Proteome . *EcoSal Plus*, 9(1).

Heroven, A. K., Nuss, A. M., and Dersch, P. (2017). RNA-based mechanisms of virulence control in Enterobacteriaceae. *RNA Biology*, 14(5):471–487.

Hershberg, R., Yeger-Lotem, E., and Margalit, H. (2005). Chromosomal organization is shaped by the transcription regulatory network.

Hews, C. L., Cho, T., Rowley, G., and Raivio, T. L. (2019). Maintaining Integrity Under Stress: Envelope Stress Response Regulation of Pathogenesis in Gram-Negative Bacteria. *Front Cell Infect Microbiol*, 9:313.

Hiam-Galvez, K. J., Allen, B. M., and Spitzer, M. H. (2021). Systemic immunity in cancer.

Hindley, J. (1967). Fractionation of 32P-labelled ribonucleic acids on polyacrylamide gels and their characterization by fingerprinting. *Journal of Molecular Biology*, 30(1).

Hirschman, J., Wong, P. K., Sei, K., Keener, J., and Kustu, S. (1985). Products of nitrogen regulatory genes ntrA and ntrC of enteric bacteria activate glnA transcription in vitro: Evidence that the ntrA product is a σ factor. *Proceedings of the National Academy of Sciences of the United States of America*, 82(22):7525–7529.

Ho, T. D. and Ellermeier, C. D. (2019). Activation of the extracytoplasmic function σ factor σ(V) by lysozyme. *Mol Microbiol*, 112(2):410–419.

Hobbs, E. C., Yin, X., Paul, B. J., Astarita, J. L., and Storz, G. (2012). Conserved small protein associates with the multidrug efflux pump AcrB and differentially affects antibiotic resistance. *Proceedings of the National Academy of Sciences of the United States of America*, 109(41):16696–16701.

Holmqvist, E. and Vogel, J. (2018). RNA-binding proteins in bacteria. *Nat Rev Microbiol*, 16(10):601–615.

Hör, J., Garriss, G., Di Giorgio, S., Hack, L. M., Vanselow, J. T., Förstner, K. U., Schlosser, A., Henriques-Normark, B., Vogel, J., Henriques-Normark, B., Vogel, J., Hor, J., Garriss, G., Di Giorgio, S., Hack, L. M., Vanselow, J. T., Forstner, K. U., Schlosser, A., Henriques-Normark, B., and Vogel, J. (2020a). Grad-seq in a Gram-positive bacterium reveals exonucleolytic sRNA activation in competence control. *EMBO J*, 39(9):e103852.

Hör, J., Gorski, S. A., Vogel, J., Hör, J., Gorski, S. A., and Vogel, J. (2018). Bacterial RNA Biology on a Genome Scale. *Mol Cell*, 70(5):785–799.

Hör, J., Matera, G., Vogel, J., Gottesman, S., Storz, G., Hör, J., Matera, G., Vogel, J., Gottesman, S., and Storz, G. (2020b). Trans-Acting Small RNAs and Their Effects on Gene Expression in Escherichia coli and Salmonella enterica. *EcoSal Plus*, 9(1).

Horn, N., Carvalho, A. L., Overweg, K., Wegmann, U., Carding, S. R., and Stentz, R. (2016). A novel tightly regulated gene expression system for the human intestinal symbiont Bacteroides thetaiotaomicron. *Frontiers in Microbiology*, 7(JUL).

Hunke, S., Keller, R., and Müller, V. S. (2012). Signal integration by the Cpx-envelope stress system. *FEMS Microbiology Letters*, 326(1):12–22.

Hutchison, C. A., Merryman, C., Sun, L., Assad-Garcia, N., Richter, R. A., Smith, H. O., and Glass, J. I. (2019). Polar effects of transposon insertion into a minimal bacterial genome. *Journal of Bacteriology*, 201(19).

Imdahl, F., Vafadarnejad, E., Homberger, C., Saliba, A. E., and Vogel, J. (2020). Single-cell RNA-sequencing reports growth-condition-specific global transcriptomes of individual bacteria. *Nat Microbiol*.

Ito, M., Kanno, S., Nosho, K., Sukawa, Y., Mitsuhashi, K., Kurihara, H., Igarashi, H., Takahashi, T., Tachibana, M., Takahashi, H., Yoshii, S., Takenouchi, T., Hasegawa, T., Okita, K., Hirata, K., Maruyama, R., Suzuki, H., Imai, K., Yamamoto, H., and Shinomura, Y. (2015). Association of Fusobacterium nucleatum with clinical and molecular features in colorectal serrated pathway. *International Journal of Cancer*, 137(6):1258–1268.

Iyer, S. C., Casas-Pastor, D., Kraus, D., Mann, P., Schirner, K., Glatter, T., Fritz, G., and Ringgaard, S. (2020). Transcriptional regulation by σ factor phosphorylation in bacteria. *Nature Microbiology*, 5(3):395–406.

Jackson, S. A., McKenzie, R. E., Fagerlund, R. D., Kieper, S. N., Fineran, P. C., and Brouns, S. J. (2017). CRISPR-Cas: Adapting to change. *Science*, 356(6333).

Jalal, N. and Lee, S. F. (2020). The MsrAB reducing pathway of Streptococcus gordonii is needed for oxidative stress tolerance, biofilm formation, and oral

colonization in mice. *PLoS ONE*, 15(2).

Johansen, J., Rasmussen, A. A., Overgaard, M., and Valentin-Hansen, P. (2006). Conserved small non-coding RNAs that belong to the sigmaE regulon: role in down-regulation of outer membrane proteins. *J Mol Biol*, 364(1):1–8.

Johansson, J., Mandin, P., Renzoni, A., Chiaruttini, C., Springer, M., and Cossart, P. (2002). An RNA thermosensor controls expression of virulence genes in Listeria monocytogenes. *Cell*, 110(5):551–561.

Johnsen, M., Christensen, T., Dennis, P. P., and Fiil, N. P. (1982). Autogenous control: ribosomal protein L10-L12 complex binds to the leader sequence of its mRNA. *The EMBO Journal*, 1(8):999–1004.

Johnson, B. K. and Abramovitch, R. B. (2017). Small Molecules That Sabotage Bacterial Virulence.

Ju, X., Li, D., and Liu, S. (2019). Full-length RNA profiling reveals pervasive bidirectional transcription terminators in bacteria. *Nature Microbiology*, 4(11):1907–1918.

Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E. P., Rivas, E., Eddy, S. R., Bateman, A., Finn, R. D., and Petrov, A. I. (2018). Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res*, 46(D1):D335–D342.

Kanehara, K., Ito, K., and Akiyama, Y. (2002). YaeL (EcfE) activates the σE pathway of stress response through a site-2 cleavage of anti-σE, RseA. *Genes and Development*, 16(16):2147–2155.

Kapatral, V., Anderson, I., Ivanova, N., Reznik, G., Los, T., Lykidis, A., Bhattacharyya, A., Bartman, A., Gardner, W., Grechkin, G., Zhu, L., Vasieva, O., Chu, L., Kogan, Y., Chaga, O., Goltsman, E., Bernal, A., Larsen, N., D'Souza, M., Walunas, T., Pusch, G., Haselkorn, R., Fonstein, M., Kyrpides, N., and Overbeek, R. (2002). Genome Sequence and Analysis of the Oral Bacterium Fusobacterium nucleatum Strain ATCC 25586. *Journal of Bacteriology*, 184(7):2005–2018.

Kaplan, C. W., Lux, R., Haake, S. K., and 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(1):35–47.

Kaplan, C. W., Lux, R., Huynh, T., Jewett, A., Shi, W., and Haake, S. K. (2005). Fusobacterium nucleatum apoptosis-inducing outer membrane protein. *J Dent Res*, 84(8):700–704.

Kaplan, C. W., Ma, X., Paranjpe, A., Jewett, A., Lux, R., Kinder-Haake, S., and Shi, W. (2010). Fusobacterium nucleatum outer membrane proteins Fap2 and RadD induce cell death in human lymphocytes. *Infect Immun*, 78(11):4773–

4778.

Kawasaki, S., Watamura, Y., Ono, M., Watanabe, T., Takeda, K., and Niimura, Y. (2005). Adaptive responses to oxygen stress in obligatory anaerobes Clostridium acetobutylicum and Clostridium aminovalericum. *Applied and Environmental Microbiology*, 71(12):8442–8450.

Kazmierczak, M. J., Wiedmann, M., and Boor, K. J. (2005). Alternative Sigma Factors and Their Roles in Bacterial Virulence. *Microbiology and Molecular Biology Reviews*, 69(4):527–543.

Kelly, T. M., Stachula, S. A., Raetz, C. R., and Anderson, M. S. (1993). The firA gene of Escherichia coli encodes UDP-3-O-(R-3-hydroxymyristoyl)-glucosamine N-acyltransferase. The third step of endotoxin biosynthesis. *J Biol Chem*, 268(26):19866–19874.

Kesavalu, L., Sathishkumar, S., Bakthavatchalu, V., Matthews, C., Dawson, D., Steffen, M., and Ebersole, J. L. (2007). Rat model of polymicrobial infection, immunity, and alveolar bone resorption in periodontal disease. *Infection and Immunity*, 75(4):1704–1712.

Kim, C. C., Monack, D., and Falkow, S. (2003). Modulation of virulence by two acidified nitrite-responsive loci of Salmonella enterica serovar Typhimurium. *Infection and Immunity*, 71(6):3196–3205.

Kinder Haake, S., Yoder, S., and Gerardo, S. H. (2006). Efficient gene transfer and targeted mutagenesis in Fusobacterium nucleatum. *Plasmid*, 55(1):27–38.

Kinney, J. B., Murugan, A., Callan, C. G., and Cox, E. C. (2010). Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. *Proceedings of the National Academy of Sciences of the United States of America*, 107(20):9158–9163.

Klähn, S., Schaal, C., Georg, J., Baumgartner, D., Knippen, G., Hagemann, M., Muro-Pastor, A. M., and Hess, W. R. (2015). The sRNA NsiR4 is involved in nitrogen assimilation control in cyanobacteria by targeting glutamine synthetase inactivating factor IF7. *Proceedings of the National Academy of Sciences of the United States of America*, 112(45):E6243–E6252.

Knoke, L. R., Abad Herrera, S., Gotz, K., Justesen, B. H., Gunther Pomorski, T., Fritz, C., Schakermann, S., Bandow, J. E., and Aktas, M. (2020). Agrobacterium tumefaciens Small Lipoprotein Atu8019 Is Involved in Selective Outer Membrane Vesicle (OMV) Docking to Bacterial Cells. *Front Microbiol*, 11:1228.

Knowles, T. J., Scott-Tucker, A., Overduin, M., and Henderson, I. R. (2009). Membrane protein architects: the role of the BAM complex in outer membrane protein assembly. *Nat Rev Microbiol*, 7(3):206–214.

Kolenbrander, P. E., Andersen, R. N., Blehert, D. S., Egland, P. G., Foster, J. S., and Palmer, R. J. (2002). Communication among Oral Bacteria. *Microbiology*

and Molecular Biology Reviews, 66(3):486–505.

Kolenbrander, P. E., Palmer, R. J., Periasamy, S., and Jakubovics, N. S. (2010). Oral multispecies biofilm development and the key role of cell-cell distance.

Kook, J. K., Park, S. N., Lim, Y. K., Cho, E., Jo, E., Roh, H., Shin, Y., Paek, J., Kim, H. S. H. S., Kim, H. S. H. S., Shin, J. H., and Chang, Y. H. (2017). Genome-Based Reclassification of Fusobacterium nucleatum Subspecies at the Species Level. *Curr Microbiol*, 74(10):1137–1147.

Kortmann, J. and Narberhaus, F. (2012). Bacterial RNA thermometers: molecular zippers and switches. *Nat Rev Microbiol*, 10(4):255–265.

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., and Garrett, W. S. (2013). Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe*, 14(2):207–215.

Kostic, A. D., Gevers, D., Pedamallu, C. S., Michaud, M., Duke, F., Earl, A. M., Ojesina, A. I., Jung, J., Bass, A. J., Tabernero, J., Baselga, J., Liu, C., Shivdasani, R. A., Ogino, S., Birren, B. W., Huttenhower, C., Garrett, W. S., and Meyerson, M. (2012). Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. *Genome Res*, 22(2):292–298.

Kreft, J., Bernhard, K., and Goebel, W. (1978). Recombinant plasmids capable of replication in B. subtilis and E. coli. *MGG Molecular & General Genetics*, 162(1):59–67.

Kröger, C., MacKenzie, K. D., Alshabib, E. Y., Kirzinger, M. W., Suchan, D. M., Chao, T. C., Akulova, V., Miranda-CasoLuengo, A. A., Monzon, V. A., Conway, T., Sivasankaran, S. K., Hinton, J. C., Hokamp, K., and Cameron, A. D. (2018). The primary transcriptome, small RNAs and regulation of antimicrobial resistance in Acinetobacter baumannii ATCC 17978. *Nucleic Acids Research*, 46(18):9684–9698.

Kuchina, A., Brettner, L. M., Paleologu, L., Roco, C. M., Rosenberg, A. B., Carignano, A., Kibler, R., Hirano, M., DePaolo, R. W., and Seelig, G. (2021). Microbial single-cell RNA sequencing by split-pool barcoding. *Science*, 371(6531).

Kumar, A., Thotakura, P. L., Tiwary, B. K., and Krishna, R. (2016). Target identification in Fusobacterium nucleatum by subtractive genomics approach and enrichment analysis of host-pathogen protein-protein interactions. *BMC Microbiol*, 16:84.

Kwon, M. and Strobel, S. A. (2008). Chemical basis of glycine riboswitch cooperativity. *Rna*, 14(1):25–34.

Lalaouna, D., Carrier, M. C., Semsey, S., Brouard, J. S., Wang, J., Wade, J. T., and Massé, E. (2015). A 3' external transcribed spacer in a tRNA transcript

acts as a sponge for small RNAs to prevent transcriptional noise. *Molecular Cell*, 58(3):393–405.

Lamm-Schmidt, V., Fuchs, M., Sulzer, J., Gerovac, M., Hör, J., Dersch, P., Vogel, J., and Faber, F. (2021). Grad-seq identifies KhpB as a global RNA-binding protein in Clostridioides difficile that regulates toxin production. *microLife*, 2.

Leang, C., Krushkal, J., Ueki, T., Puljic, M., Sun, J., Juárez, K., Núñez, C., Reguera, G., DiDonato, R., Postier, B., Adkins, R. M., and Lovley, D. R. (2009). Genome-wide analysis of the RpoN regulon in Geobacter sulfurreducens. *BMC Genomics*, 10.

Lee, A. K., Detweiler, C. S., and Falkow, S. (2000). OmpR regulates the two-component system SsrA-SsrB in Salmonella pathogenicity island 2. *Journal of Bacteriology*, 182(3):771–781.

Lee, W. S., Jean, S. S., Chen, F. L., Hsieh, S. M., and Hsueh, P. R. (2020). Lemierre's syndrome: A forgotten and re-emerging infection.

Leyton, D. L., Rossiter, A. E., and Henderson, I. R. (2012). From self sufficiency to dependence: mechanisms and factors important for autotransporter biogenesis. *Nat Rev Microbiol*, 10(3):213–225.

Li, J., Gálvez, E. J. C., Amend, L., Almási, É., Iljazovic, A., Lesker, T. R., Bielecka, A. A., Schorr, E. M., and Strowig, T. (2021). A versatile genetic toolbox for Prevotella copri enables studying polysaccharide utilization systems. *Embo j*, 40(23):e108287.

Li, L., Huang, D., Cheung, M. K., Nong, W., Huang, Q., and Kwan, H. S. (2013). BSRD: A repository for bacterial small regulatory RNA. *Nucleic Acids Research*, 41(D1):D233.

Li, X., Wang, B., Feng, L., Kang, H., Qi, Y., Wang, J., and Shi, Y. (2009). Cleavage of RseA by RseP requires a carboxyl-terminal hydrophobic amino acid following DegS cleavage. *Proceedings of the National Academy of Sciences of the United States of America*, 106(35):14837–14842.

Li, Y.-Y. (2016). Association of Fusobacterium nucleatum infection with colorectal cancer in Chinese patients. *World J. Gastroenterol.*, 22:3227–3233.

Lim, B., Zimmermann, M., Barry, N. A., and Goodman, A. L. (2017). Engineered Regulatory Systems Modulate Gene Expression of Human Commensals in the Gut. *Cell*, 169(3):547–558.e15.

Livak, K. J. and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, 25(4):402–408.

Loh, E., Dussurget, O., Gripenland, J., Vaitkevicius, K., Tiensuu, T., Mandin, P., Repoila, F., Buchrieser, C., Cossart, P., and Johansson, J. (2009). A trans-acting riboswitch controls expression of the virulence regulator PrfA in Listeria

monocytogenes. *Cell*, 139(4):770–779.

Loh, E., Righetti, F., Eichner, H., Twittenhoff, C., and Narberhaus, F. (2018). RNA Thermometers in Bacterial Pathogens. *Microbiol Spectr*, 6(2).

Lonetto, M., Gribskov, M., and Gross, C. A. (1992). The sigma 70 family: sequence conservation and evolutionary relationships. *Journal of Bacteriology*, 174(12):3843–3849.

Longo, S. K., Guo, M. G., Ji, A. L., and Khavari, P. A. (2021). Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics.

Lorenz, R., Bernhart, S. H., Höner Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., and Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Algorithms Mol Biol*, 6:26.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, 15(12):550.

Lu, Z. and Imlay, J. A. (2021). When anaerobes encounter oxygen: mechanisms of oxygen toxicity, tolerance and defence.

Lutz, R. and Bujard, H. (1997). Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res*, 25(6):1203–1210.

Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J. J., Charpentier, E., Cheng, D., Haft, D. H., Horvath, P., Moineau, S., Mojica, F. J. M., Scott, D., Shah, S. A., Siksnys, V., Terns, M. P., Venclovas, Č., White, M. F., Yakunin, A. F., Yan, W., Zhang, F., Garrett, R. A., Backofen, R., van der Oost, J., Barrangou, R., and Koonin, E. V. (2020). Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology*, 18(2):67–83.

Makino, K., Shinagawa, H., Amemura, M., and Nakata, A. (1986). Nucleotide sequence of the phoB gene, the positive regulatory gene for the phosphate regulon of Escherichia coli K-12. *Journal of molecular biology*, 190(1):37–44.

Mann, M., Wright, P. R., and Backofen, R. (2017). IntaRNA 2.0: enhanced and customizable prediction of RNA-RNA interactions. *Nucleic Acids Res*, 45(W1):W435–W439.

Manson McGuire, A., Cochrane, K., Griggs, A. D., Haas, B. J., Abeel, T., Zeng, Q., Nice, J. B., MacDonald, H., Birren, B. W., Berger, B. W., Allen-Vercoe, E., and Earl, A. M. (2014). Evolution of Invasion in a Diverse Set of Fusobacterium Species. *mBio*, 5(6):e01864.

Martin-Verstraete, I., Peltier, J., and Dupuy, B. (2016). The regulatory networks that control Clostridium difficile toxin synthesis.

Mascher, T., Helmann, J. D., and Unden, G. (2006). Stimulus Perception in Bacterial Signal-Transducing Histidine Kinases. *Microbiology and Molecular Biology Reviews*, 70(4):910–938.

Massé, E. and Gottesman, S. (2002). A small RNA regulates the expression of genes involved in iron metabolism in Escherichia coli. *Proc Natl Acad Sci U S A*, 99(7):4620–4625.

Massé, E., Vanderpool, C. K., and Gottesman, S. (2005). Effect of RyhB small RNA on global iron use in Escherichia coli. *J Bacteriol*, 187(20):6962–6971.

Mazza, P., Noens, E. E., Schirner, K., Grantcharova, N., Mommaas, A. M., Koerten, H. K., Muth, G., Flärdh, K., Van Wezel, G. P., and Wohlleben, W. (2006). MreB of Streptomyces coelicolor is not essential for vegetative growth but is required for the integrity of aerial hyphae and spores. *Molecular Microbiology*, 60(4):838–852.

McCown, P. J., Corbino, K. A., Stav, S., Sherlock, M. E., and Breaker, R. R. (2017). Riboswitch diversity and distribution. *Rna*, 23(7):995–1011.

McCoy, A. (2013). Fusobacterium is associated with colorectal adenomas. *PLOS ONE*, 8.

McGinn, J. and Marraffini, L. A. (2019). Molecular mechanisms of CRISPR–Cas spacer acquisition.

Mecsas, J., Rouviere, P. E., Erickson, J. W., Donohue, T. J., and Gross, C. A. (1993). The activity of sigma E, an Escherichia coli heat-inducible sigma-factor, is modulated by expression of outer membrane proteins. *Genes Dev*, 7(12b):2618–2628.

Melamed, S., Adams, P. P., Zhang, A., Zhang, H., and Storz, G. (2020). RNA-RNA Interactomes of ProQ and Hfq Reveal Overlapping and Competing Roles. *Mol Cell*, 77(2):411–425 e7.

Meng, Q., Gao, Q., Mehrazarin, S., Tangwanichgapong, K., Wang, Y., Huang, Y., Pan, Y., Robinson, S., Liu, Z., Zangiabadi, A., Lux, R., Papapanou, P. N., Guo, X. E., Wang, H., Berchowitz, L. E., and Han, Y. W. (2021). Fusobacterium nucleatum secretes amyloid-like FadA to enhance pathogenicity . *EMBO reports*, 22(7).

Merritt, J., Niu, G., Okinaga, T., and Qi, F. (2009). Autoaggregation response of Fusobacterium nucleatum. *Appl Environ Microbiol*, 75(24):7725–7733.

Michaux, C., Hansen, E. E., Jenniches, L., Gerovac, M., Barquist, L., and Vogel, J. (2020). Single-Nucleotide RNA Maps for the Two Major Nosocomial Pathogens Enterococcus faecalis and Enterococcus faecium. *Frontiers in Cellular and Infection Microbiology*, 10:600325.

Mihailovic, M. K., Vazquez-Anderson, J., Li, Y., Fry, V., Vimalathas, P., Herrera, D., Lease, R. A., Powell, W. B., and Contreras, L. M. (2018). High-throughput

in vivo mapping of RNA accessible interfaces to identify functional sRNA binding sites. *Nature Communications*, 9(1):1–16.

Milewski, S. (2002). Glucosamine-6-phosphate synthase - The multi-facets enzyme.

Millman, A., Bernheim, A., Stokar-Avihail, A., Fedorenko, T., Voichek, M., Leavitt, A., Oppenheimer-Shaanan, Y., and Sorek, R. (2020). Bacterial Retrons Function In Anti-Phage Defense. *Cell*.

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., and Ogino, S. (2016). Fusobacterium nucleatum in colorectal carcinoma tissue and patient prognosis. *Gut*, 65(12):1973–1980.

Missiakas, D., Mayer, M. P., Lemaire, M., Georgopoulos, C., and Raina, S. (1997). Modulation of the Escherichia coli sigmaE (RpoE) heat-shock transcription-factor activity by the RseA, RseB and RseC proteins. *Mol Microbiol*, 24(2):355–371.

Mitchell, A. M. and Silhavy, T. J. (2019). Envelope stress responses: balancing damage repair and toxicity. *Nat Rev Microbiol*, 17(7):417–428.

Mitchell, J. E., Zheng, D., Busby, S. J., and Minchin, S. D. (2003). Identification and analysis of 'extended -10' promoters in Escherichia coli. *Nucleic Acids Research*, 31(16):4689–4695.

Mitsuhashi, K., Nosho, K., Sukawa, Y., Matsunaga, Y., Ito, M., Kurihara, H., Kanno, S., Igarashi, H., Naito, T., Adachi, Y., Tachibana, M., Tanuma, T., Maguchi, H., Shinohara, T., Hasegawa, T., Imamura, M., Kimura, Y., Hirata, K., Maruyama, R., Suzuki, H., Imai, K., Yamamoto, H., and Shinomura, Y. (2015). Association of Fusobacterium species in pancreatic cancer tissues with molecular features and prognosis. *Oncotarget*, 6(9):7209–7220.

Miyakoshi, M., Chao, Y., and Vogel, J. (2015). Cross talk between ABC transporter m RNA s via a target m RNA -derived sponge of the G cv B small RNA . *The EMBO Journal*, 34(11):1478–1492.

Mohanraju, P., Saha, C., van Baarlen, P., Louwen, R., Staals, R. H., and van der Oost, J. (2022). Alternative functions of CRISPR–Cas systems in the evolutionary arms race.

Møller, T., Franch, T., Udesen, C., Gerdes, K., and Valentin-Hansen, P. (2002). Spot 42 RNA mediates discoordinate expression of the E. coli galactose operon. *Genes Dev*, 16(13):1696–1706.

Moore, S. D. and Sauer, R. T. (2007). The tmRNA system for translational surveillance and ribosome rescue. *Annu Rev Biochem*, 76:101–124.

Moore, W. E. C. and Moore, L. V. H. (1994). The bacteria of periodontal diseases.

Periodontology 2000, 5(1):66–77.

Morfeldt, E., Taylor, D., Von Gabain, A., and Arvidson, S. (1995). Activation of alpha-toxin translation in Staphylococcus aureus by the trans-encoded antisense RNA, RNAIII. *EMBO Journal*, 14(18):4569–4577.

Morrison, J. M., Miller, E. W., Benson, M. A., Alonzo, F., Yoong, P., Torres, V. J., Hinrichs, S. H., and Dunmana, P. M. (2012). Characterization of SSR42, a Novel virulence factor regulatory RNA that contributes to the pathogenesis of a Staphylococcus aureus USA300 representative. *Journal of Bacteriology*, 194(11):2924–2938.

Mukherjee, A., Weyant, K. B., Walker, J., and Schroeder, C. M. (2012). Directed evolution of bright mutants of an oxygen-independent flavin-binding fluorescent protein from Pseudomonas putida. *Journal of Biological Engineering*, 6(1):20.

Müller, K. M., Berghoff, B. A., Eisenhardt, B. D., Remes, B., and Klug, G. (2016). Characteristics of Pos19 - A Small Coding RNA in the Oxidative Stress Response of Rhodobacter sphaeroides. *PLoS One*, 11(9):e0163425.

Mutalik, V. K., Nonaka, G., Ades, S. E., Rhodius, V. A., and Gross, C. A. (2009). Promoter strength properties of the complete sigma E regulon of Escherichia coli and Salmonella enterica. *J Bacteriol*, 191(23):7279–7287.

Mutha, N. V. R., Mohammed, W. K., Krasnogor, N., Tan, G. Y. A., Choo, S. W., and Jakubovics, N. S. (2018). Transcriptional responses of Streptococcus gordonii and Fusobacterium nucleatum to coaggregation. *Mol Oral Microbiol*, 33(6):450–464.

Nakamura, Y., Gojobori, T., and Ikemura, T. (2000). Codon usage tabulated from international DNA sequence databases: Status for the year 2000.

Nang, S. C., Azad, M. A. K., Velkov, T., Zhou, Q. T., and Li, J. (2021). Rescuing the Last-Line Polymyxins: Achievements and Challenges. *Pharmacol Rev*, 73(2):679–728.

Nasreen, M., Fletcher, A., Hosmer, J., Zhong, Q., Essilfie, A.-T., McEwan, A. G., and Kappler, U. (2021). The Alternative Sigma Factor RpoE2 Is Involved in the Stress Response to Hypochlorite and in vivo Survival of Haemophilus influenzae. *Frontiers in Microbiology*, 12:62.

Nejman, D., Livyatan, I., Fuks, G., Gavert, N., Zwang, Y., Geller, L. T., Rotter-Maskowitz, A., Weiser, R., Mallel, G., Gigi, E., Meltser, A., Douglas, G. M., Kamer, I., Gopalakrishnan, V., Dadosh, T., Levin-Zaidman, S., Avnet, S., Atlan, T., Cooper, Z. A., Arora, R., Cogdill, A. P., Khan, M. A. W., Ologun, G., Bussi, Y., Weinberger, A., Lotan-Pompan, M., Golani, O., Perry, G., Rokah, M., Bahar-Shany, K., Rozeman, E. A., Blank, C. U., Ronai, A., Shaoul, R., Amit, A., Dorfman, T., Kremer, R., Cohen, Z. R., Harnof, S., Siegal, T., Yehuda-Shnaidman, E., Gal-Yam, E. N., Shapira, H., Baldini, N., Langille, M. G. I., Ben-Nun, A., Kaufman, B., Nissan, A., Golan, T., Dadiani, M., Levanon, K.,

Bar, J., Yust, S. K., Barshack, I., Peeper, D. S., Raz, D. J., Segal, E., Wargo, J. A., Sandbank, J., Shental, N., Straussman, R., Yust-Katz, S., Barshack, I., Peeper, D. S., Raz, D. J., Segal, E., Wargo, J. A., Sandbank, J., Shental, N., and Straussman, R. (2020). The human tumor microbiome is composed of tumor type-specific intracellular bacteria. *Science*, 368(6494):973–980.

Nicoloff, H., Gopalkrishnan, S., and Ades, S. E. (2017). Appropriate regulation of the σEd-ependent envelope stress response is necessary to maintain cell envelope integrity and stationary-phase survival in Escherichia coli. *Journal of Bacteriology*, 199(12).

Nitzan, M., Rehani, R., and Margalit, H. (2017). Integration of Bacterial Small RNAs in Regulatory Networks.

Nuss, A. M., Beckstette, M., Pimenova, M., Schmühl, C., Opitz, W., Pisano, F., Heroven, A. K., and 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 of the United States of America*, 114(5):E791–E800.

Oda, M., Kobayashi, N., Fujita, M., Miyazaki, Y., Sadaie, Y., Kurusu, Y., and Nishikawa, S. (2004). Analysis of HutP-dependent transcription antitermination in the Bacillus subtilis hut operon: Identification of HutP binding sites on hut antiterminator RNA and the involvement of the N-terminus of HutP in binding of HutP to the antiterminator RNA. *Molecular Microbiology*, 51(4):1155–1168.

Oglesby-Sherrouse, A. G. and Murphy, E. R. (2013). Iron-responsive bacterial small RNAs: variations on a theme. *Metallomics*, 5(4):276.

Olejniczak, M., Jiang, X., Basczok, M. M., and Storz, G. (2021). KH domain proteins: Another family of bacterial RNA matchmakers? *Mol Microbiol*.

Ouellet, J. (2016). RNA fluorescence with light-Up aptamers.

Paget, M. S., Leibovitz, E., and Buttner, M. J. (1999). A putative two-component signal transduction system regulates sigmaE, a sigma factor required for normal cell wall integrity in Streptomyces coelicolor A3(2). *Mol Microbiol*, 33(1):97–107.

Papenfort, K., Bouvier, M., Mika, F., Sharma, C. M., and Vogel, J. (2010). Evidence for an autonomous 5' target recognition domain in an Hfq-associated small RNA. *Proc Natl Acad Sci U S A*, 107(47):20435–20440.

Papenfort, K., Pfeiffer, V., Lucchini, S., Sonawane, A., Hinton, J. C. D., and Vogel, J. (2008). Systematic deletion of Salmonella small RNA genes identifies CyaR, a conserved CRP-dependent riboregulator of OmpX synthesis. *Molecular Microbiology*, 68(4):890–906.

Papenfort, K., Pfeiffer, V., Mika, F., Lucchini, S., Hinton, J. C. D., and Vogel, J. (2006). σE-dependent small RNAs of Salmonella respond to membrane stress by accelerating global omp mRNA decay. *Mol Microbiol*, 62(6):1674–1688.

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., and Bachrach, G. (2020). Breast cancer colonization by Fusobacterium nucleatum accelerates tumor growth and metastatic progression. *Nat Commun*, 11(1):3259.

Pátek, M., Nešvera, J., Guyonvarch, A., Reyes, O., and Leblon, G. (2003). Promoters of Corynebacterium glutamicum. *Journal of Biotechnology*, 104(1-3):311–323.

Payne, L. J., Todeschini, T. C., Wu, Y., Perry, B. J., Ronson, C. W., Fineran, P. C., Nobrega, F. L., and Jackson, S. A. (2021). Identification and classification of antiviral defence systems in bacteria and archaea with PADLOC reveals new system types. *Nucleic Acids Research*, 49(19):10868–10878.

Pelechano, V. and Steinmetz, L. M. (2013). Gene regulation by antisense transcription.

Pérez-Rueda, E. and Collado-Vides, J. (2000). The repertoire of DNA-binding transcriptional regulators in Escherichia coli K-12. *Nucleic Acids Research*, 28(8):1838–1847.

Peschek, N., Hoyos, M., Herzog, R., Förstner, K. U., and Papenfort, K. (2019). A conserved RNA seed-pairing domain directs small RNA-mediated stress resistance in enterobacteria. *Embo j*, 38(16):e101650.

Pfeifer-Sancar, K., Mentz, A., Rückert, C., and Kalinowski, J. (2013). Comprehensive analysis of the Corynebacterium glutamicum transcriptome using an improved RNAseq technique. *BMC Genomics*, 14(1):888.

Pfeiffer, V., Sittka, A., Tomer, R., Tedin, K., Brinkmann, V., and Vogel, J. (2007). A small non-coding RNA of the invasion gene island (SPI-1) represses outer membrane protein synthesis from the Salmonella core genome. *Mol Microbiol*, 66(5):1174–1191.

Pickar-Oliver, A. and Gersbach, C. A. (2019). The next generation of CRISPR–Cas technologies and applications.

Plamont, M. A., Billon-Denis, E., Maurin, S., Gauron, C., Pimenta, F. M., Specht, C. G., Shi, J., Quérard, J., Pan, B., Rossignol, J., Morellet, N., Volovitch, M., Lescop, E., Chen, Y., Triller, A., Vriz, S., Le Saux, T., Jullien, L., and Gautier, A. (2016). Small fluorescence-activating and absorption-shifting tag for tunable protein imaging in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, 113(3):497–502.

Polak, D., Wilensky, A., Shapira, L., Halabi, A., Goldstein, D., Weiss, E. I., and Houri-Haddad, Y. (2009). Mouse model of experimental periodontitis induced by Porphyromonas gingivalis / Fusobacterium nucleatum infection: bone loss and host response. *Journal of Clinical Periodontology*,

36(5):406–410.

Ponath, F., Hör, J., and Vogel, J. (2022a). An overview of gene regulation in bacteria by small RNAs derived from mRNA 3' ends. *FEMS Microbiology Reviews*.

Ponath, F., Tawk, C., Zhu, Y., Barquist, L., Faber, F., and Vogel, J. (2021). RNA landscape of the emerging cancer-associated microbe Fusobacterium nucleatum. *Nature Microbiology*.

Ponath, F., Zhu, Y., Cosi, V., and Vogel, J. (2022b). Expanding the genetic toolkit helps dissect a global stress response in the early-branching species Fusobacterium nucleatum. *Proceedings of the National Academy of Sciences*, 119(40):e2201460119.

Poole, L. B. (2005). Bacterial defenses against oxidants: Mechanistic features of cysteine-based peroxidases and their flavoprotein reductases.

Potter, K. D., Merlino, N. M., Jacobs, T., and Gollnick, P. (2011). TRAP binding to the Bacillus subtilis trp leader region RNA causes efficient transcription termination at a weak intrinsic terminator. *Nucleic Acids Research*, 39(6):2092–2102.

Prelich, G. (2012). Gene overexpression: Uses, mechanisms, and interpretation.

Prévost, K., Salvail, H., Desnoyers, G., Jacques, J. F., Phaneuf, É., and Massé, E. (2007). The small RNA RyhB activates the translation of shiA mRNA encoding a permease of shikimate, a compound involved in siderophore synthesis. *Molecular Microbiology*, 64(5):1260–1273.

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., and Lim, W. A. (2013). Repurposing CRISPR as an RNA-γuided platform for sequence-specific control of gene expression. *Cell*, 152(5):1173–1183.

Quinlan, A. R. (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics*, 47:11.12.1–34.

Raffa, R. G. and Raivio, T. L. (2002). A third envelope stress signal transduction pathway in Escherichia coli. *Molecular Microbiology*, 45(6):1599–1611.

Raivio, T. L. (2014). Everything old is new again: An update on current research on the Cpx envelope stress response. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1843(8):1529–1541.

Ranquet, C., Ollagnier-de Choudens, S., Loiseau, L., Barras, F., and Fontecave, M. (2007). Cobalt stress in Escherichia coli: The effect on the iron-sulfur proteins. *Journal of Biological Chemistry*, 282(42):30442–30451.

Ransom, E. M., Ellermeier, C. D., and Weiss, D. S. (2015). Use of mCherry Red fluorescent protein for studies of protein localization and gene expression in Clostridium difficile. *Appl Environ Microbiol*, 81(5):1652–1660.

Rasmussen, A. A., Eriksen, M., Gilany, K., Udesen, C., Franch, T., Petersen, C.,

and Valentin-Hansen, P. (2005). Regulation of ompA mRNA stability: The role of a small regulatory RNA in growth phase-dependent control. *Molecular Microbiology*, 58(5):1421–1429.

Rasmussen, A. A., Johansen, J., Nielsen, J. S., Overgaard, M., Kallipolitis, B., and Valentin-Hansen, P. (2009). A conserved small RNA promotes silencing of the outer membrane protein YbfM. *Mol Microbiol*, 72(3):566–577.

Rest, J. S. and Mindell, D. P. (2003). Retroids in Archaea: Phylogeny and lateral origins. *Molecular Biology and Evolution*, 20(7):1134–1142.

Reznikoff, W. S. (2008). Transposon Tn5.

Rezuchova, B., Miticka, H., Homerova, D., Roberts, M., and Kormanec, J. (2003). New members of the Escherichia coli sigmaE regulon identified by a two-plasmid system. *FEMS Microbiol Lett*, 225(1):1–7.

Rhodius, V. A. and Mutalik, V. K. (2010). Predicting strength and function for promoters of the Escherichia coli alternative sigma factor, sigmaE. *Proc Natl Acad Sci U S A*, 107(7):2854–2859.

Rhodius, V. A., Suh, W. C., Nonaka, G., West, J., and Gross, C. A. (2006). Conserved and variable functions of the sigmaE stress response in related genomes. *PLoS Biol*, 4(1):e2.

Richter, H., Zoephel, J., Schermuly, J., Maticzka, D., Backofen, R., and Randau, L. (2012). Characterization of CRISPR RNA processing in Clostridium thermocellum and Methanococcus maripaludis. *Nucleic Acids Res*, 40(19):9887–9896.

Rivas, E., Klein, R. J., Jones, T. A., and Eddy, S. R. (2001). Computational identification of noncoding RNAs in E. coli by comparative genomics. *Current Biology*, 11(17):1369–1373.

Rosano, G. L. and Ceccarelli, E. A. (2014). Recombinant protein expression in Escherichia coli: Advances and challenges.

Rouvière, P. E., De Las Peñas, A., Mecsas, J., Lu, C. Z., Rudd, K. E., and Gross, C. A. (1995). rpoE, the gene encoding the second heat-shock sigma factor, sigma E, in Escherichia coli. *Embo j*, 14(5):1032–1042.

Rowley, G., Spector, M., Kormanec, J., and Roberts, M. (2006). Pushing the envelope: extracytoplasmic stress responses in bacterial pathogens. *Nat Rev Microbiol*, 4(5):383–394.

Rubinstein, M. R., Wang, X., Liu, W., Hao, Y., Cai, G., and Han, Y. W. (2013). Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating E-cadherin/beta-catenin signaling via its FadA adhesin. *Cell Host Microbe*, 14(2):195–206.

Ryan, D., Jenniches, L., Reichardt, S., Barquist, L., and Westermann, A. J. (2020). A high-resolution transcriptome map identifies small RNA regulation

of metabolism in the gut microbe Bacteroides thetaiotaomicron. *Nat Commun*, 11(1):3557.

Sakanaka, A., Kuboniwa, M., Shimma, S., Alghamdi, S. A., Mayumi, S., Lamont, R. J., Fukusaki, E., and Amano, A. (2022). Fusobacterium nucleatum Metabolically Integrates Commensals and Pathogens in Oral Biofilms. *mSystems*.

Salazar, M. E., Podgornaia, A. I., and Laub, M. T. (2016). The small membrane protein MgrB regulates PhoQ bifunctionality to control PhoP target gene expression dynamics. *Molecular Microbiology*, 102(3):430–445.

Saleh, M., Bartual, S. G., Abdullah, M. R., Jensch, I., Asmat, T. M., Petruschka, L., Pribyl, T., Gellert, M., Lillig, C. H., Antelmann, H., Hermoso, J. A., and Hammerschmidt, S. (2013). Molecular architecture of streptococcus pneumoniae surface thioredoxin-fold lipoproteins crucial for extracellular oxidative stress resistance and maintenance of virulence. *EMBO Molecular Medicine*, 5(12):1852–1870.

Salvail, H. and Massé, E. (2012). Regulating iron storage and metabolism with RNA: an overview of posttranscriptional controls of intracellular iron homeostasis. *Wiley Interdisciplinary Reviews: RNA*, 3(1):26–36.

Sanders, B. E., Umana, A., Lemkul, J. A., and Slade, D. J. (2018). FusoPortal: an Interactive Repository of Hybrid MinION-Sequenced Fusobacterium Genomes Improves Gene Identification and Characterization. *mSphere*, 3(4).

Santos-Zavaleta, A., Salgado, H., Gama-Castro, S., Sánchez-Pérez, M., Gómez-Romero, L., Ledezma-Tejeida, D., García-Sotelo, J. S., Alquicira-Hernández, K., Muñiz-Rascado, L. J., Peña-Loredo, P., Ishida-Gutiérrez, C., Velázquez-Ramírez, D. A., Del Moral-Chávez, V., Bonavides-Martínez, C., Méndez-Cruz, C. F., Galagan, J., and Collado-Vides, J. (2019). RegulonDB v 10.5: Tackling challenges to unify classic and high throughput knowledge of gene regulation in E. coli K-12. *Nucleic Acids Research*, 47(D1):D212–D220.

Saraogi, I. and ou Shan, S. (2014). Co-translational protein targeting to the bacterial membrane.

Sasaki-Imamura, T., Yano, A., and Yoshida, Y. (2010). Production of indole from L-tryptophan and effects of these compounds on biofilm formation by Fusobacterium nucleatum ATCC 25586. *Appl Environ Microbiol*, 76(13):4260–4268.

Schäfer, U., Beck, K., and Müller, M. (1999). Skp, a molecular chaperone of gram-negative bacteria, is required for the formation of soluble periplasmic intermediates of outer membrane proteins. *J Biol Chem*, 274(35):24567–24574.

Scheible, M., Nguyen, C. T., Luong, T. T., Lee, J. H., Chen, Y.-W., Chang, C., Wittchen, M., Camacho, M. I., Tiner, B. L., Wu, C., Tauch, A., Das, A., and Ton-That, H. (2022). The Fused Methionine Sulfoxide Reductase MsrAB Promotes Oxidative Stress Defense and Bacterial Virulence in Fusobacterium

nucleatum. *mBio*, 13(3):e0302221.

Schmidt, F., Cherepkova, M. Y., and Platt, R. J. (2018). Transcriptional recording by CRISPR spacer acquisition from RNA. *Nature*, 562(7727):380–385.

Schmidt, K. L., Peterson, N. D., Kustusch, R. J., Wissel, M. C., Graham, B., Phillips, G. J., and Weiss, D. S. (2004). A Predicted ABC Transporter, FtsEX, Is Needed for Cell Division in Escherichia coli. *Journal of Bacteriology*, 186(3):785–793.

Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat Methods*, 9(7):671–675.

Schulz, S., Eckweiler, D., Bielecka, A., Nicolai, T., Franke, R., Dötsch, A., Hornischer, K., Bruchmann, S., Düvel, J., and Häussler, S. (2015). Elucidation of sigma factor-associated networks in Pseudomonas aeruginosa reveals a modular architecture with limited and function-specific crosstalk. *PLoS Pathog*, 11(3):e1004744.

Seelbinder, B., Wallstabe, J., Marischen, L., Weiss, E., Wurster, S., Page, L., Löffler, C., Bussemer, L., Schmitt, A. L., Wolf, T., Linde, J., Cicin-Sain, L., Becker, J., Kalinke, U., Vogel, J., Panagiotou, G., Einsele, H., Westermann, A. J., Schäuble, S., and Loeffler, J. (2020). Triple RNA-Seq Reveals Synergy in a Human Virus-Fungus Co-infection Model. *Cell Reports*, 33(7).

Settem, R. P., El-Hassan, A. T., Honma, K., Stafford, G. P., and Sharma, A. (2012). Fusobacterium nucleatum and tannerella forsythia induce synergistic alveolar bone loss in a mouse periodontitis model. *Infection and Immunity*, 80(7):2436–2443.

Shapiro, L., McAdams, H. H., and Losick, R. (2009). Why and how bacteria localize proteins.

Sharma, C. M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiß, S., Sittka, A., Chabas, S., Reiche, K., Hackermüller, J., Reinhardt, R., Stadler, P. F., and Vogel, J. (2010). The primary transcriptome of the major human pathogen Helicobacter pylori. *Nature*, 464(7286):250–255.

Sharma, C. M. and Vogel, J. (2014). Differential RNA-seq: the approach behind and the biological insight gained. *Curr Opin Microbiol*, 19:97–105.

Shine, J. and Dalgarno, L. (1974). The 3'-Terminal Sequence of Escherichia coli 16S Ribosomal RNA: Complementarity to Nonsense Triplets and Ribosome Binding Sites. *Proceedings of the National Academy of Sciences*, 71(4):1342–1346.

Shishkin, A. A., Giannoukos, G., Kucukural, A., Ciulla, D., Busby, M., Surka, C., Chen, J., Bhattacharyya, R. P., Rudy, R. F., Patel, M. M., Novod, N., Hung, D. T., Gnirke, A., Garber, M., Guttman, M., and Livny, J. (2015). Simultaneous generation of many RNA-seq libraries in a single reaction. *Nature*

Methods, 12(4):323–325.

Shokeen, B., Park, J., Duong, E., Rambhia, S., Paul, M., Weinberg, A., Shi, W., and Lux, R. (2020). Role of FAD-I in fusobacterial interspecies interaction and biofilm formation. *Microorganisms*, 8(1).

Si, M. R., Zhang, L., Yang, Z. F., Xu, Y. X., Liu, Y. B., Jiang, C. Y., Wang, Y., Shen, X. H., and Liu, S. J. (2014). NrdH redoxin enhances resistance to multiple oxidative stresses by acting as a peroxidase cofactor in Corynebacterium glutamicum. *Applied and Environmental Microbiology*, 80(5):1750–1762.

Simon, A. J., Ellington, A. D., and Finkelstein, I. J. (2019). Retrons and their applications in genome engineering.

Sklar, J. G., Wu, T., Kahne, D., and Silhavy, T. J. (2007). Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in Escherichia coli. *Genes Dev*, 21(19):2473–2484.

Skovierova, H., Rowley, G., Rezuchova, B., Homerova, D., Lewis, C., Roberts, M., and Kormanec, J. (2006). Identification of the sigmaE regulon of Salmonella enterica serovar Typhimurium. *Microbiology (Reading)*, 152(Pt 5):1347–1359.

Smirnov, A., Forstner, K. U., Holmqvist, E., Otto, A., Gunster, R., Becher, D., Reinhardt, R., and Vogel, J. (2016). Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. *Proc Natl Acad Sci U S A*, 113(41):11591–11596.

Smirnov, A., Wang, C., Drewry, L. L., and Vogel, J. (2017). Molecular mechanism of mRNA repression in trans by a ProQ-dependent small RNA . *The EMBO Journal*, 36(8):1029–1045.

Song, T., Mika, F., Lindmark, B., Liu, Z., Schild, S., Bishop, A., Zhu, J., Camilli, A., Johansson, J., Vogel, J., and Wai, S. N. (2008). A new Vibrio cholerae sRNA modulates colonization and affects release of outer membrane vesicles. *Mol Microbiol*, 70(1):100–111.

Soutourina, O. A. and Bertin, P. N. (2003). Regulation cascade of flagellar expression in Gram-negative bacteria.

Sperandeo, P., Cescutti, R., Villa, R., Di Benedetto, C., Candia, D., Dehò, G., and Polissi, A. (2007). Characterization of lptA and lptB, two essential genes implicated in lipopolysaccharide transport to the outer membrane of Escherichia coli. *Journal of Bacteriology*, 189(1):244–253.

Stapels, D. A., Hill, P. W., Westermann, A. J., Fisher, R. A., Thurston, T. L., Saliba, A. E., Blommestein, I., Vogel, J., and Helaine, S. (2018). Salmonella persisters undermine host immune defenses during antibiotic treatment. *Science*, 362(6419):1156–1160.

States, D. J. and Gish, W. (1994). Combined Use of Sequence Similarity and Codon Bias for Coding Region Identification. *Journal of Computational Biology*,

1(1):39–50.

Storz, G., Vogel, J., and Wassarman, K. M. (2011). Regulation by Small RNAs in Bacteria: Expanding Frontiers. *Molecular Cell*, 43(6):880–891.

Storz, G., Wolf, Y. I., and Ramamurthi, K. S. (2014). Small proteins can no longer be ignored. *Annu Rev Biochem*, 83:753–777.

Strauss, J., Kaplan, G. G., Beck, P. L., Rioux, K., Panaccione, R., DeVinney, R., Lynch, T., and Allen-Vercoe, E. (2011). Invasive potential of gut mucosa-derived fusobacterium nucleatum positively correlates with IBD status of the host. *Inflammatory Bowel Diseases*, 17(9):1971–1978.

Strecker, J., Ladha, A., Gardner, Z., Schmid-Burgk, J. L., Makarova, K. S., Koonin, E. V., and Zhang, F. (2019). RNA-guided DNA insertion with CRISPR-associated transposases. *Science*, 364(6448):48–53.

Swidsinski, A., Dörffel, Y., Loening-Baucke, V., Theissig, F., Rückert, J. C., Ismail, M., Rau, W. A., Gaschler, D., Weizenegger, M., Kühn, S., Schilling, J., and Dörffel, W. V. (2011). Acute appendicitis is characterised by local invasion with Fusobacterium nucleatum/necrophorum. *Gut*, 60(1):34–40.

Tahara, T., Shibata, T., Kawamura, T., Okubo, M., Ichikawa, Y., Sumi, K., Miyata, M., Ishizuka, T., Nakamura, M., Nagasaka, M., Nakagawa, Y., Ohmiya, N., Arisawa, T., and Hirata, I. (2015). Fusobacterium Detected in Colonic Biopsy and Clinicopathological Features of Ulcerative Colitis in Japan. *Digestive Diseases and Sciences*, 60(1):205–210.

Tang, W. and Liu, D. R. (2018). Rewritable multi-event analog recording in bacterial and mammalian cells. *Science*, 360(6385).

Taylor, R. G., Lambert, M. A., Sexsmith, E., Sadler, S. J., Ray, P. N., Mahuran, D. J., and McInnes, R. R. (1990). Cloning and expression of rat histidase. Homology to two bacterial histidases and four phenylalanine ammonia-lyases. *Journal of Biological Chemistry*, 265(30):18192–18199.

Team, R. C. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

Ternes, D., Tsenkova, M., Pozdeev, V. I., Meyers, M., Koncina, E., Atatri, S., Schmitz, M., Karta, J., Schmoetten, M., Heinken, A., Rodriguez, F., Delbrouck, C., Gaigneaux, A., Ginolhac, A., Nguyen, T. T. D., Grandmougin, L., Frachet-Bour, A., Martin-Gallausiaux, C., Pacheco, M., Neuberger-Castillo, L., Miranda, P., Zuegel, N., Ferrand, J.-Y., Gantenbein, M., Sauter, T., Slade, D. J., Thiele, I., Meiser, J., Haan, S., Wilmes, P., and Letellier, E. (2022). The gut microbial metabolite formate exacerbates colorectal cancer progression. *Nature Metabolism*, 4(4):458–475.

Testerman, T. L., Vazquez-Torres, A., Xu, Y., Jones-Carson, J., Libby, S. J., and Fang, F. C. (2002). The alternative sigma factor sigmaE controls antioxidant

defences required for Salmonella virulence and stationary-phase survival. *Mol Microbiol*, 43(3):771–782.

Thompson, K. M., Rhodius, V. A., and Gottesman, S. (2007). SigmaE regulates and is regulated by a small RNA in Escherichia coli. *J Bacteriol*, 189(11):4243–4256.

Thorvaldsdottir, H., Robinson, J. T., and Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, 14(2):178–192.

Todor, H., Osadnik, H., Campbell, E. A., Myers, K. S., Li, H., Donohue, T. J., and Gross, C. A. (2020). Rewiring the specificity of extracytoplasmic function sigma factors. *Proceedings of the National Academy of Sciences*, 117(52):33496–33506.

Todor, H., Silvis, M. R., Osadnik, H., and Gross, C. A. (2021). Bacterial CRISPR screens for gene function.

Tomasini, A., Moreau, K., Chicher, J., Geissmann, T., Vandenesch, F., Romby, P., Marzi, S., and Caldelari, I. (2017). The RNA targetome of Staphylococcus aureus non-coding RNA RsaA: Impact on cell surface properties and defense mechanisms. *Nucleic Acids Research*, 45(11):6746–6760.

Tong, Y., Jørgensen, T. S., Whitford, C. M., Weber, T., and Lee, S. Y. (2021). A versatile genetic engineering toolkit for E. coli based on CRISPR-prime editing. *Nature Communications*, 12(1):1–11.

Toussi, D. N., Liu, X., and Massari, P. (2012). The FomA porin from Fusobacterium nucleatum is a toll-like receptor 2 agonist with immune adjuvant activity. *Clinical and Vaccine Immunology*, 19(7):1093–1101.

Tran, N. T., Huang, X., Hong, H. J., Bush, M. J., Chandra, G., Pinto, D., Bibb, M. J., Hutchings, M. I., Mascher, T., and Buttner, M. J. (2019). Defining the regulon of genes controlled by σE, a key regulator of the cell envelope stress response in Streptomyces coelicolor. *Molecular Microbiology*, 112(2):461–481.

Travers, A. A. and Burgess, R. R. (1969). Cyclic Re-use of the RNA polymerase sigma factor. *Nature*, 222(5193):537–540.

Trunk, T., Casasanta, M. A., Yoo, C. C., Slade, D. J., and Leo, J. C. (2019). Comparison of type 5d autotransporter phospholipases demonstrates a correlation between high activity and intracellular pathogenic lifestyle. *Biochemical Journal*, 476(18):2657–2676.

Tsien, R. Y. (1998). The green fluorescent protein.

Tsirigotaki, A., De Geyter, J., Šoštaric, N., Economou, A., and Karamanou, S. (2017). Protein export through the bacterial Sec pathway. *Nat Rev Microbiol*, 15(1):21–36.

Udekwu, K. I., Darfeuille, F., Vogel, J., Reimegård, J., Holmqvist, E., and Wagner,

E. G. (2005). Hfq-dependent regulation of OmpA synthesis is mediated by an antisense RNA. *Genes Dev*, 19(19):2355–2366.

Umaña, A., Nguyen, T. T., Sanders, B. E., Williams, K. J., Wozniak, B., and Slade, D. J. (2022). DNA methyltransferase enhanced Fusobacterium nucleatum genetics. *bioRxiv*, page 2022.07.20.500824.

Umaña, A., Sanders, B. E., Yoo, C. C., Casasanta, M. A., Udayasuryan, B., Verbridge, S. S., and Slade, D. J. (2019). Utilizing Whole Fusobacterium Genomes To Identify, Correct, and Characterize Potential Virulence Protein Families. *Journal of Bacteriology*, 201(23).

Urban, J. H. and Vogel, J. (2007). Translational control and target recognition by Escherichia coli small RNAs in vivo. *Nucleic Acids Res*, 35(3):1018–1037.

Vakulskas, C. A., Pannuri, A., Cortés-Selva, D., Zere, T. R., Ahmer, B. M., Babitzke, P., and Romeo, T. (2014). Global effects of the DEAD-box RNA helicase DeaD (CsdA) on gene expression over a broad range of temperatures. *Molecular Microbiology*, 92(5):945–958.

Valdivia, R. H. and Falkow, S. (1996). Bacterial genetics by flow cytometry: Rapid isolation of Salmonella typhimurium acid-inducible promoters by differential fluorescence induction. *Molecular Microbiology*, 22(2):367–378.

Van Zyl, W. F., Dicks, L. M. T., and Deane, S. M. (2019). Development of a novel selection/counter-selection system for chromosomal gene integrations and deletions in lactic acid bacteria. *BMC Molecular Biology*, 20(1):10.

Vander Haar, E. L., So, J., Gyamfi-Bannerman, C., and Han, Y. W. (2018). Fusobacterium nucleatum and adverse pregnancy outcomes: Epidemiological and mechanistic evidence.

Vigouroux, A. and Bikard, D. (2020). CRISPR Tools To Control Gene Expression in Bacteria. *Microbiology and Molecular Biology Reviews*, 84(2).

Vitreschak, A. G., Rodionov, D. A., Mironov, A. A., and Gelfand, M. S. (2003). Regulation of the vitamin B12 metabolism and transport in bacteria by a conserved RNA structural element. *RNA*, 9(9):1084–1097.

Vogel, J. (2020). An RNA biology perspective on species-specific programmable RNA antibiotics.

Vogel, J. and Papenfort, K. (2006). Small non-coding RNAs and the bacterial outer membrane. *Current Opinion in Microbiology*, 9(6):605–611.

Walsh, N. P., Alba, B. M., Bose, B., Gross, C. A., and Sauer, R. T. (2003). OMP peptide signals initiate the envelope-stress response by activating DegS protease via relief of inhibition mediated by its PDZ domain. *Cell*, 113(1):61–71.

Warman, E. A., Forrest, D., Guest, T., Haycocks, J. J., Wade, J. T., and Grainger, D. C. (2021). Widespread divergent transcription from bacterial and archaeal

promoters is a consequence of DNA-sequence symmetry. *Nature Microbiology*, 6(6):746–756.

Wassarman, K. M. (2018). 6S RNA, a Global Regulator of Transcription. *Microbiol Spectr*, 6(3).

Wassarman, K. M. and Saecker, R. M. (2006). Synthesis-mediated release of a small RNA inhibitor of RNA polymerase. *Science*, 314(5805):1601–1603.

Wassarman, K. M. and Storz, G. (2000). 6S RNA regulates E. coli RNA polymerase activity. *Cell*, 101(6):613–623.

Wehner, S., Damm, K., Hartmann, R. K., and Marz, M. (2014). Dissemination of 6S RNA among bacteria. *RNA Biol*, 11(11):1467–1478.

Welch, J. L., Rossetti, B. J., Rieken, C. W., Dewhirst, F. E., and Borisy, G. G. (2016). Biogeography of a human oral microbiome at the micron scale. *Proceedings of the National Academy of Sciences of the United States of America*, 113(6):E791–E800.

Westermann, A. J., Barquist, L., and Vogel, J. (2017). Resolving host–pathogen interactions by dual RNA-seq.

Westermann, A. J., Forstner, K. U., Amman, F., Barquist, L., Chao, Y., Schulte, L. N., Muller, L., Reinhardt, R., Stadler, P. F., and Vogel, J. (2016). Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions. *Nature*, 529(7587):496–501.

Westermann, A. J. and Vogel, J. (2021). Cross-species RNA-seq for deciphering host–microbe interactions.

Whitaker, W. R., Shepherd, E. S., and Sonnenburg, J. L. (2017). Tunable Expression Tools Enable Single-Cell Strain Distinction in the Gut Microbiome. *Cell*, 169(3):538–546.e12.

White, M. L. and Eswara, P. J. (2021). ylm has more than a (Z Anchor) ring to it!

Wicke, L., Ponath, F., Coppens, L., Gerovac, M., Lavigne, R., and Vogel, J. (2021). Introducing differential RNA-seq mapping to track the early infection phase for Pseudomonas phage phiKZ. *RNA Biology*, 18(8):1099–1110.

Williams, M. D., Kerber, C. A., and Tergin, H. F. (2003). Unusual presentation of lemierre's syndrome due to Fusobacterium nucleatum. *Journal of Clinical Microbiology*, 41(7):3445–3448.

Williams, M. L., Crowley, P. J., Hasona, A., and Brady, L. J. (2014). YlxM is a newly identified accessory protein that influences the function of signal recognition particle pathway components in Streptococcus mutans. *Journal of Bacteriology*, 196(11):2043–2052.

Winkler, W. C., Nahvi, A., Roth, A., Collins, J. A., and Breaker, R. R. (2004).

Control of gene expression by a natural metabolite-responsive ribozyme. *Nature*, 428(6980):281–286.

Wirbel, J., Pyl, P. T., Kartal, E., Zych, K., Kashani, A., Milanese, A., Fleck, J. S., Voigt, A. Y., Palleja, A., Ponnudurai, R., Sunagawa, S., Coelho, L. P., Schrotz-King, P., Vogtmann, E., Habermann, N., Niméus, E., Thomas, A. M., Manghi, P., Gandini, S., Serrano, D., Mizutani, S., Shiroma, H., Shiba, S., Shibata, T., Yachida, S., Yamada, T., Waldron, L., Naccarati, A., Segata, N., Sinha, R., Ulrich, C. M., Brenner, H., Arumugam, M., Bork, P., and Zeller, G. (2019). Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nature Medicine*, 25(4):679–689.

Wu, C., Al Mamun, A. A. M., Luong, T. T., Hu, B., Gu, J., Lee, J. H., D'Amore, M., Das, A., and Ton-That, H. (2018). Forward Genetic Dissection of Biofilm Development by Fusobacterium nucleatum: Novel Functions of Cell Division Proteins FtsX and EnvC. *MBio*, 9(2).

Wu, C., Chen, Y. W., Scheible, M., Chang, C., Wittchen, M., Lee, J. H., Luong, T. T., Tiner, B. L., Tauch, A., Das, A., and Ton-That, H. (2021). Genetic and molecular determinants of polymicrobial interactions in Fusobacterium nucleatum. *Proc Natl Acad Sci U S A*, 118(23).

Wu, T., Cen, L., Kaplan, C., Zhou, X., Lux, R., Shi, W., and He, X. (2015). Cellular components mediating coadherence of Candida albicans and Fusobacterium nucleatum. *Journal of Dental Research*, 94(10):1432–1438.

Xiao, J., Peng, B., Su, Z., Liu, A., Hu, Y., Nomura, C. T., Chen, S., and Wang, Q. (2020). Facilitating Protein Expression with Portable 5'-UTR Secondary Structures in Bacillus licheniformis. *ACS Synth Biol*, 9(5):1051–1058.

Yakhnin, H., Yakhnin, A. V., and Babitzke, P. (2015). Ribosomal protein L10(L12)4 autoregulates expression of the Bacillus subtilis rplJL operon by a transcription attenuation mechanism. *Nucleic Acids Res*, 43(14):7032–7043.

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., and Baba, H. (2016). Human Microbiome Fusobacterium Nucleatum in Esophageal Cancer Tissue Is Associated with Prognosis. *Clinical Cancer Research*, 22(22):5574–5581.

Yan, B., Boitano, M., Clark, T. A., and Ettwiller, L. (2018). SMRT-Cappable-seq reveals complex operon variants in bacteria. *Nature Communications*, 9(1).

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., and 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 Upregulating Expression of MicroRNA-21. *Gastroenterology*, 152(4):851–866.e24.

Yanofsky, C. (1981). Attenuation in the control of expression of bacterial operons. *Nature*, 289(5800):751–758.

Yu, G., Lam, T. T., Zhu, H., and Guan, Y. (2018a). Two Methods for Mapping and Visualizing Associated Data on Phylogeny Using Ggtree. *Mol Biol Evol*, 35(12):3041–3043.

Yu, S. H., Vogel, J., and Forstner, K. U. (2018b). ANNOgesic: a Swiss army knife for the RNA-seq based annotation of bacterial/archaeal genomes. *Gigascience*, 7(9).

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., and Fang, J. Y. (2017). Fusobacterium nucleatum Promotes Chemoresistance to Colorectal Cancer by Modulating Autophagy. *Cell*, 170(3):548–563.e16.

Zhang, J., Fei, J., Leslie, B. J., Han, K. Y., Kuhlman, T. E., and Ha, T. (2015). Tandem spinach array for mRNA imaging in living bacterial cells. *Scientific Reports*, 5(1):1–9.

Zhang, S. and Voigt, C. A. (2018). Engineered dCas9 with reduced toxicity in bacteria: Implications for genetic circuit design. *Nucleic Acids Research*, 46(20):11115–11125.

Zhang, X. Z., Yan, X., Cui, Z. L., Hong, Q., and Li, S. P. (2006). mazF, a novel counter-selectable marker for unmarked chromosomal manipulation in Bacillus subtilis. *Nucleic Acids Res*, 34(9):e71.

Zhang, Y. F., Han, K., Chandler, C. E., Tjaden, B., Ernst, R. K., and Lory, S. (2017). Probing the sRNA regulatory landscape of P. aeruginosa: post-transcriptional control of determinants of pathogenicity and antibiotic susceptibility. *Mol Microbiol*, 106(6):919–937.

Zheng, J. J., Perez, A. J., Tsui, H. T., Massidda, O., and Winkler, M. E. (2017). Absence of the KhpA and KhpB (JAG/EloR) RNA-binding proteins suppresses the requirement for PBP2b by overproduction of FtsA in Streptococcus pneumoniae D39. *Mol Microbiol*, 106(5):793–814.

Zheng, K., Wang, Y., Li, N., Jiang, F. F., Wu, C. X., Liu, F., Chen, H. C., and Liu, Z. F. (2018). Highly efficient base editing in bacteria using a Cas9-cytidine deaminase fusion. *Communications Biology*, 1(1).

Zheng, X., Hu, G. Q., She, Z. S., and Zhu, H. (2011). Leaderless genes in bacteria: Clue to the evolution of translation initiation mechanisms in prokaryotes. *BMC Genomics*, 12.

Zschiedrich, C. P., Keidel, V., and Szurmant, H. (2016). Molecular Mechanisms of Two-Component Signal Transduction. *Journal of Molecular Biology*, 428(19):3752–3775.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization

prediction. *Nucleic Acids Research*, 31(13):3406–3415.

8 Supplementary Data

The Excel files for all below mentioned data can be found on the enclosed USB drive.

8.1 Supplementary Data 8.1

Overview of detected TSSs.

8.2 Supplementary Data 8.2

Results of differential gene expression analysis for the dRNA-seq data.

8.3 Supplementary Data 8.3

Overview of predicted 5'UTRs.

8.4 Supplementary Data 8.4

Comparison of putative virulence factors to pTSSs in *Fnn*.

8.5 Supplementary Data 8.5

Reannotated CDSs in *Fnn*.

8.6 Supplementary Data 8.6

Predicted operons in *Fnn*.

8.7 Supplementary Data 8.7

Overview of predicted riboregulatory elements.

8.8 Supplementary Data 8.8

Overview of predicted sRNAs.

8.9 Supplementary Data 8.9

DEseq results for *rpoE* induction.

8.10 Supplementary Data 8.10

DEseq results for exposure to oxygen and polymyxin B.

8.11 Supplementary Data 8.11

DEseq results for FoxI pulse expression.

8.12 Supplementary Data 8.12

IntaRNA target prediction.

9 Abbreviations

Table 9.1 | List of abbreviations.

Abbreviation	Full expansion
aa	amino acid
ATP	adenosine triphosphate
aTSS	antisense TSS
bp	base pair
cDNA	complementary DNA
CDS	coding sequence
CFU	colony forming unit
CRC	colorectal cancer
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DNA	deoxyribonuleic acid
DNase	deoxyribonuclease
DTT	dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
FACS	Fluorescence-activated cell sorting
FC	fold-change
FDR	false discovery rate
FMN	flavin mononucleotide
Fna	*Fusobacterium nucleatum* subspecies *animalis*
Fnn	*Fusobacterium nucleatum* subspecies *nucleatum*
Fnp	*Fusobacterium nucleatum* subspecies *polymorphum*
Fnv	*Fusobacterium nucleatum* subspecies *vencentii*
FuH	*Fusobacterium Hwasookii*
FuP	*Fusobacterium Periodonticum*
gDNA	genomic DNA
GFP	green fluorescent protein
Grad-seq	gradient profiling by sequencing
h	hour(s)

Table 9.1 *continued from previous page*

Abbreviation	Full expansion
IM	inner membrane
iTSS	internal TSS
LB	Lennox broth
LC-MS/MS	Liquid Chromatography with tandem mass spectrometry
min	minute(s)
mRNA	messenger RNA
MSA	multiple sequence alignment
ncRNA	non-coding RNA
nt	nucleotide
OD_{600nm}	optical density measured at 600nm
OM	outer membrane
ORF	open reading frame
oTSS	oprhan TSS
P/C/I	Phenol/Chloroform/Isopropanol
PAA	polyacrylamide
PAGE	Polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
pTSS	primary TSS
RBP	RNA-binding protein
RBS	ribosome binding site
RNA	ribonucleic acid
RNAP	RNA polymerase
RNase	ribonuclease
RNA-seq	RNA sequencing
RNP	ribonucleoprotein complex
rRNA	ribosomal RNA
SD	Standard Deviation
SDS	sodium dodecyl sulfate
SEC	general secretory pathway
sORF	small open reading frame
sRNA	small regulatory RNA
SRP	signal recognition particle
sTSS	secondary TSS
TCS	two-component system

Table 9.1 *continued from previous page*

Abbreviation	**Full expansion**
TF	transcription factor
TSS	transcription start site
UTR	untranslated region
V/V	volume/volume
w/V	weight/volume

10 Curriculum Vitae

11 List of publications

First author manuscripts

Ponath, F., Zhu, Y., Cosi, V., & Vogel, J. (2022). Expanding the genetic toolkit helps dissect a global stress response in the early-branching species *Fusobacterium nucleatum*. *Proceedings of the National Academy of Sciences of the United States of America*, 119(40), e2201460119. https://doi.org/10.1073/pnas.2201460119

Ponath, F., Hör, J., & Vogel, J. (2022). An overview of gene regulation in bacteria by small RNAs derived from mRNA 3' ends. *FEMS Microbiology Reviews*, 46(5), fuac017. https://doi.org/10.1093/femsre/fuac017

Ponath, F., Tawk, C., Zhu, Y., Barquist, L., Faber, F., & Vogel, J. (2021). RNA landscape of the emerging cancer-associated microbe *Fusobacterium nucleatum*. *Nature Microbiology*, 6(8), 1007–1020. https://doi.org/10.1038/s41564-021-00927-7

Other manuscripts

El Mouali, Y., **Ponath, F.**, Scharrer, V., Wenner, N., Hinton, J., & Vogel, J. (2021). Scanning mutagenesis of RNA-binding protein ProQ reveals a quality control role for the Lon protease. *RNA* (New York, N.Y.), 27(12), 1512–1527. https://doi.org/10.1261/rna.078954.121

Hoffmann, T., Rahfeld, J. U., Schenk, M., **Ponath, F.**, Makioka, K., Hutter-Paier, B., Lues, I., Lemere, C. A., & Schilling, S. (2021). Combination of the Glutaminyl Cyclase Inhibitor PQ912 (Varoglutamstat) and the Murine Monoclonal Antibody PBD-C06 (m6) Shows Additive Effects on Brain Aβ Pathology

in Transgenic Mice. *International Journal of Molecular Sciences*, 22(21), 11791. https://doi.org/10.3390/ijms222111791

Fuchs, M., Lamm-Schmidt, V., Sulzer, J., **Ponath, F.**, Jenniches, L., Kirk, J. A., Fagan, R. P., Barquist, L., Vogel, J., & Faber, F. (2021). An RNA-centric global view of *Clostridioides difficile* reveals broad activity of Hfq in a clinically important gram-positive bacterium. *Proceedings of the National Academy of Sciences of the United States of America*, 118(25), e2103579118. https://doi.org/10.1073/pnas.21035791180

Wicke, L., **Ponath, F.**, Coppens, L., Gerovac, M., Lavigne, R., & Vogel, J. (2021). Introducing differential RNA-seq mapping to track the early infection phase for Pseudomonas phage ϕKZ. *RNA Biology*, 18(8), 1099–1110. https://doi.org/10.1080/15476286.2020.1827785

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. *Nature Communications*, 11(1), 3259. https://doi.org/10.1038/s41467-020-16967-2

12 Acknowledgments

All good things must come to an end and so does this work and my PhD time. The last few years would not have been possible nor as enjoyable without the people that surrounded me and so I would like to thank you all for your support:

Jörg Vogel, thank you for welcoming me into your lab and entrusting me with the *Fusobacterium* project. I appreciate all the support and guidance you have given me throughout this time.

My thesis committee Ana Rita Brochado, Chase Beisel and Till Strowig, I appreciate the time and effort you put in to help improve my project.

All my past and present colleagues in the Vogel lab which enriched my PhD time both in terms of science and fun.

Both points also apply to all the friends I have made throughout the past years in Würzburg: Austin Mottola, Caroline Tawk, Daniel Ryan, Elisa Venturini, Elise Bornet, Franziska Faber, Gianluca Matera, Gianluca Prezza, Kotaro Chihara, Manuela Fuchs, Svetlana Durica-Mitic, Tobias Kerrinnes, Yan Zhu, Youssef el Moali Benomar and more. Thank you for being there and enjoying life and work together. You have made this time not only more enjoyable but also inspired me.

And lastly, my family and friends from home & Halle times. You played a big part in shaping me into the person I am today. Thank you for your unwavering support and understanding - it has not always been easy with me.

Again, thank you all - it has been a blast!

13 Affidavit/Eidesstattliche Erklärung

Affidavit

I hereby confirm that my thesis entitled "Investigating the molecular biology of *Fusobacterium nucleatum*", is the result of my own work. I did not receive any help or support from commercial consultants. All sources and / or materials applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

Würzburg,
Place, Date Signature

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, die Dissertation "Untersuchung der Molekularbiologie von *Fusobacterium nucleatum*", eigenständig, d.h. insbesondere selbständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Würzburg,
Ort, Datum Unterschrift