## TITLE

# RNA landscape of the emerging cancer-associated microbe *Fusobacterium nucleatum*

## AUTHOR LIST

Falk Ponath[1], Caroline Tawk[2], Yan Zhu[2], Lars Barquist[1,3],

Franziska Faber[2], Jörg Vogel[1,2,3]*

## AFFILIATIONS

[1] Helmholtz Institute for RNA-based Infection Research (HIRI), Helmholtz Centre for Infection Research (HZI), D-97080 Würzburg, Germany

[2] Institute for Molecular Infection Biology (IMIB), University of Würzburg, D-97080 Würzburg, Germany

[3] Faculty of Medicine, University of Würzburg, Würzburg, Germany.

*Correspondence: joerg.vogel@uni-wuerzburg.de

Running title: RNA landscape of *Fusobacterium*

Keywords: Transcriptional start site, differential RNA-seq, noncoding RNA, antisense RNA, small RNA, CRISPR/Cas, post-transcriptional control

## ABSTRACT

*Fusobacterium nucleatum*, long known as a constituent of the oral microflora, has recently garnered renewed attention for its association with several different human cancers. The growing interest in this emerging cancer-associated bacterium contrasts with a paucity of knowledge about its basic gene expression features and physiological responses. As fusobacterial lack all established small RNA-associated proteins, post-transcriptional networks in these bacteria are also unknown. Here, using differential RNA-seq (dRNA-seq), we generate high-resolution global RNA maps for five clinically relevant fusobacterial strains - *F. nucleatum* subspecies *nucleatum*, *animalis*, *polymorphum* and *vincentii* as well as *F. periodonticum* – for early, mid-exponential growth and early stationary phase. These data are made available in an online browser, and we use these to uncover fundamental aspects of fusobacterial gene expression architecture and a suite of noncoding RNAs. Developing a vector for functional analysis of fusobacterial genes, we discover a conserved fusobacterial oxygen-induced small RNA, FoxI, which serves as a post-transcriptional repressor of the major outer membrane porin FomA. Our findings provide a crucial step towards delineating the regulatory networks enabling *F. nucleatum* adaptation to different environments, which may elucidate how these bacteria colonize different compartments of the human body.

## MAIN TEXT

## INTRODUCTION

Bacteria associated with cancer are of interest for diverse clinical reasons, which include their potential oncogenic properties, use in diagnostics, and influence on treatment outcome[1]. A prominent recent example is the association of *Fusobacterium nucleatum* with colorectal and breast cancers[2-4]. Originally observed through genomics studies of colorectal cancer (CRC) samples[3,4], *F. nucleatum* is now established as a constituent of microbial communities in human colorectal carcinomas[5,6] and metastatic lesions[7], where it contributes to disease progression[7,8] and resistance to chemotherapy[9,10]. Links to esophageal and pancreatic cancers have also been reported[11-13].

These phenomenological observations were underpinned by the discovery of several molecular factors such as the adhesin FadA[14] or the lectin Fap2[15], which enable *F. nucleatum* to interact with epithelial cells and recognize Gal-GalNAc (a sugar abundantly displayed by CRC and breast cancer tissue), respectively. On the host side, the molecular response of mammalian cells to *F. nucleatum* is also beginning to be better understood, revealing which pathways allow the bacterium to adapt to host-imposed milieus[2,14-21]. Yet, in stark contrast with this growing interest in *F. nucleatum*, our molecular understanding of the bacterium itself is lagging much behind other disease-associated species.

A gram-negative, non-motile, obligate anaerobic bacterium with a small genome in the 2-2.5 Mb range, *F. nucleatum* is an abundant oral commensal of humans[6,22]. Possessing broad adhesion capabilities, the bacterium bridges early and late colonizers in the oral plaque[6]. *F. nucleatum* belongs to the invasive branch of the phylum *Fusobacteria* that evolved in an adaptive radiation, separating it from non-invasive fusobacteria[23]. Complete genome sequences are available for several members of the Fusobacterium genus[24,25]. However, some pioneering gene expression studies notwithstanding[26,27], little remains known about the overall transcriptome structure, RNA output and stress responses of *F. nucleatum*. To illustrate this, a single transcription start site (TSS) has been determined in the entire phylum[28], and RNA-based annotations of gene structure including untranslated regions (UTRs) and operons are generally lacking for any fusobacterial species.

This paucity extends to post-transcriptional control: while *cis* and *trans* acting regulatory RNAs, including small noncoding RNAs (sRNAs), have been comprehensively described in many other bacteria[29-31], they have remained unreported in *F. nucleatum*. Intriguingly, fusobacteria do not seem to encode any of the three common sRNA-associated proteins, CsrA, Hfq, and ProQ[32], which precludes straightforward sRNA searches via RNA co-immunopreciptation[30]. Instead, high-resolution transcriptome maps covering both coding and noncoding genomic regions will be needed for sRNA discovery in fusobacteria. Importantly, sRNA genes tend to be tightly regulated, making them excellent markers of the physiological state of a cell; yet, this also means that for experimental sRNA discovery to be comprehensive, global RNA maps must be obtained under different growth conditions[33]. This global information could also help to overcome the poor genetic tractability of fusobacteria. For example, there is a predicted CRISPR-Cas system in *F. nucleatum*[34] but whether it is active and repurposable for chromosomal modifications, as done in other bacteria[35,36], remains unknown.

Here, we have determined the primary transcriptomes[37] and noncoding RNA suites of five different fusobacterial strains associated with humans: *F. nucleatum* subspecies *nucleatum*, *animalis*, *polymorphum* and *vincentii* as well as *F. periodonticum* (included for its repeatedly reported association with oral squamous cell carcinoma[38]), henceforth referred to as *Fnn*, *Fna*, *Fnp*, *Fnv* and *Fup* respectively. Our genome-wide annotations of TSS, 5'UTRs, and operons identify the major expression units in different stages of growth. We reveal a rich layer of putative noncoding regulators and amongst these identify an oxygen-induced sRNA repressor of the synthesis of *F. nucleatum*'s most abundant outer membrane porin FomA. To this end, we develop an expression vector that should be broadly applicable to the functional analysis of fusobacterial genes, noncoding and coding ones alike.

## RESULTS

### *The primary transcriptome of Fusobacterium nucleatum*

We obtained high-resolution RNA maps of the above five fusobacterial strains using the dRNA-seq method[37,39] for global annotation of 5' ends of transcripts. dRNA-seq determines TSS by distinguishing primary (5'PPP) transcript ends from processed (5'P,

5'OH) RNA by a specific nuclease treatment. We generally analyzed RNA from three different growth phases (Fig. 1a), i.e., early-exponential (E), mid-exponential (M) and early stationary phase (S). Since dRNA-seq gave largely similar results for those genes that are shared by the five fusobacterial strains analyzed here (Extended Data Fig. 1-4, Supplementary Dataset 1), data presented hereafter refer to *Fnn* unless stated otherwise. All transcriptomic data can be viewed in an online browser at https://helmholtz-hiri.de/en/datasets/fusobacterium.

As to benchmarking, dRNA-seq readily detected the annotated major TSS of the *tnaAB* tryptophanase operon[28] (Fig. 1a). Using ANNOgesic[10] and manual curation, we assigned TSS to five different classes based on location and relative expression (Fig. 1b): pTSS (primary or main TSS of a gene or operon), sTSS (secondary TSS, less expressed than pTSS), iTSS (internal TSS within an open reading frame (ORF)), aTSS (antisense to a gene or within 100 nt distance), and oTSS (orphan TSS, no associated gene).

The primary transcriptome of *Fnn* determined here comprises 930 TSS (Fig. 1b), showing the common multiple associations with different TSS classes[39]. For example, 37 pTSS (~5%) lie within upstream genes, thus also classifying as iTSS. While most TSS seem constitutively active (Supplementary Dataset 1), we also observed many growth-phase dependent TSS, as exemplified by the fructose uptake genes FN1438-FN1441 whose transcription diminishes in stationary phase (Supplementary Fig. 1, Supplementary Dataset 1). However, different from the frequently observed pervasive antisense transcription in other bacteria[40], aTSS only make up ~13%. These aTSS were detected primarily in the early or mid-exponential growth phases (Supplementary Dataset 1), as exemplified by the aTSS near the start codon of *glpK* (FN1839, glycerol kinase). Intriguingly, the *glpK* mRNA only accumulates in stationary phase (Supplementary Dataset 2), which suggests a growth-phase dependent antisense control of glycerol uptake and metabolism.

*Promoter and 5'UTR structure in F. nucleatum*

To initiate transcription, bacterial RNA polymerase (RNAP) associates with a sigma factor (σ factor) to recognize promoter regions[41,42]. *Fnn* encodes four σ factors: $\sigma^{70}$/RpoD, $\sigma^{E}$/RpoE, one unclassified ECF family σ-factor, and one with similarity to the sporulation

SigH factor of phylogenetically distant *Clostridiales*[43,44]. Scanning the -50 to +1 regions upstream of pTSS with MEME[45] identified an extended -10 box as well as a -35 box for recognition by housekeeping $\sigma^{70}$ in ~93% of the pTSS (Fig. 1c). These two boxes flank an AT-rich region, an arrangement known from ε-proteobacteria[37,46]. Beyond that, we find no other obvious promoter motifs under the growth conditions tested here. This suggests that fusobacteria, in order to adapt to different environments, must control their genes by additional mechanisms including post-transcriptional control by regulatory RNA species.

The 5' UTR of mRNA is the primary region for post-transcriptional control in bacteria. Using pTSS and sTSS, we define a median 5' UTR length of 36 nts, while we also detect unusually long ones, e.g. the 324-nt 5'UTR of FN0328 (Fig. 1d, Supplementary Dataset 3). A 5'-AGGAGG-3' motif resembling a Shine Dalgarno (SD) sequence is present in three-quarters (76%) of all 5'UTRs. Leaderless mRNAs seem surprisingly rare in *Fnn*: only 4 genes have a 5'UTR of <10 nt in length, and encode an integrase/recombinase, a pseudouridine synthetase, a putative transcriptional regulator and a predicted drug efflux pump, respectively (Supplementary Dataset 3).

*Transcription structure of virulence-associated genes*

Although *Fnn* lacks typical secretion systems for export of effector proteins, there is a growing list of *Fnn* genes with proven roles in virulence[23,24,47], including genes encoding Type V autotransporters[44,48-50]. Our RNA maps not only show that these genes are constitutively expressed but also provide expression context for many of them (Fig. 2a), e.g., revealing monocistronic expression of adhesin FadA, which recognizes host cells and triggers β-catenin signalling[14,16,51]. By contrast, the important Gal-GalNac lectin Fap2 is part of a dicistron. Likewise, the predicted FadA paralog FN1529 is co-expressed with RadD (FN1526), a putative type Va autotransporter important for inter-species adherence and biofilm formation[52,53].

The serine protease fusolisin (FN1426)[54] is a more complex case: the ORF is monocistronic but the 3' region of this gene is independently transcribed into the FunR47 sRNA (Fig. 2a). Altogether, 65 out of 208 proposed virulence-associated genes[55] are transcribed from a pTSS (Supplementary Dataset 4). Now having their 5'UTRs defined,

these genes lend themselves to investigation of potential post-transcriptional control of fusobacterial virulence.

*RNA-based annotations of ORFs and operons*

RNA maps are a powerful tool to assign and correct diverse gene expression features based on experimental data, which includes the global verification of ORF annotations in fully sequenced genomes[37]. To correct *Fnn* ORFs where necessary, we double-checked all coding sequences (CDS) lacking a canonical starting codon or RBS, or with an extendable ORF. These reannotations are in excellent agreement with the latest genome sequence update at Fusoportal[24] (Supplementary Dataset 5).

Small proteins represent an overlooked class of bacterial gene products whose functional importance is just beginning to unfold[56]. *Fnn* has 22 annotated small proteins <50aa in length, many of which are associated with transposase genes or are ambiguous, lacking an AUG start codon or RBS, respectively. Using our TSS maps, we enrich the genome annotation with three previously overlooked high-confidence candidates of small ORFs, naming them *fspC1* to *fspC3*. For example, we propose *fspC*3 to encode a conserved ~48 aa hydrophobic peptide and to be co-transcribed with an operon for nucleotide metabolism and tRNA maturation/repair functions (Fig. 2b). The *fspC*1 ORF encodes a 41-aa peptide with a predicted transmembrane domain and lies close to the gene of anti-termination factor NusB (Supplementary Fig. 2a); whereas *fspC*2 lies upstream of a predicted glutamate carboxypeptidase and encodes a 33-aa hydrophobic peptide (Supplementary Fig. 2b). Their strong sequence conservation among *Fusobacterium* species suggests that the FspC1, FspC2 and FspC3 peptides play important roles in fusobacterial physiology.

Previous operon annotations in *Fnn* have much relied upon computational inference from other bacteria[57]. Our TSS-based annotation predicts a total of 428 operons (Supplementary Dataset 6), which includes previously unknown polycistrons such as that of phospholipase A1 type Vd autotransporter FplA (FN1704)[58] with the FN1706-FN1707 genes (Supplementary Fig. 7a). The longest operon predicted spans 23 genes, encoding multiple ribosomal proteins and preprotein translocase SecY (Supplementary Dataset 6). TSS inside primary operons predict the presence of 53 sub-operons. Of note, in ~40% (21

sub-operons) of such cases the iTSS uncouples the last or the last two genes from the full operon (Supplementary Dataset 6). This is well illustrated with the FN1326-FN1320 operon, wherein gene FN1321 encoding an orphan response regulator seems to be conditionally uncoupled from an upstream biosynthesis gene cluster (Supplementary Fig. 7b).

*Riboregulatory elements in F. nucleatum*

Typical 5'UTR borne *cis*-regulatory RNA elements are riboswitches and RNA thermometers, which are RNA structures used by many pathogenic bacteria for location-dependent post-transcriptional control of virulence genes[59,60]. The Rfam database[61] and literature searches led us to annotate 12 putative riboswitches in *Fnn* (Supplementary Dataset 7), three of which may sense cobalamin and thereby control genes involved in uptake of vitamin $B_{12}$, iron, or both (FN0300, Fe2+/B12-binding protein; FN1971, TonB-dependent receptor; FN1381, putative autotransporter) (Extended Data Fig. 5). Two riboswitches may sense flavin mononucleotide (FMN), one of which is associated with *ribH* (FN1505) encoding the riboflavin synthase β-subunit. The remaining ones belong to the families of glycine, lysine, purine, SAM, and TPP riboswitches, and are generally found upstream of biosynthesis and/or metabolism genes of these ligands (Supplementary Dataset 7). Ligand-responsive 5'UTRs in *F. nucleatum* further include a glucoseamine-6-phosphate (GlcN6P) sensing ribozyme[62]; as in *Bacillus subtilis*, this ribozyme may feedback-control GlcN6P synthase (FN0452) levels, and thereby cell wall production.

Of other *cis* elements in 5'UTRs (Supplementary Dataset 7), we found an RNA leader known to auto-regulate the synthesis of ribosomal protein L10[63,64]. Additionally, a putative binding site of the regulatory PyrR protein in the 5' UTR of the *pyr* operon indicates transcriptional attenuation[65]. By contrast, we hardly found candidates for RNA thermometers. None of the five predicted candidates lie in a 5' UTR (Supplementary Dataset 7). This currently leaves *F. nucleatum* without standard RNA thermometers, perhaps owing to the bacterium's stable environmental temperature in the human body.

*Core noncoding RNAs and an active CRISPR-Cas system*

Bacteria possess several ubiquitous stable RNAs other than rRNA and tRNA, all of which lacked annotation in the *Fnn* genome. Guided by Rfam database predictions, we successfully probed 4.5S RNA, M1 RNA (RNase P) and tmRNA on northern blots and observed constitutive expression over the course of growth (Fig. 3a). The 105-nt 4.5S RNA shows the conserved apical GGAA tetraloop on which the signal recognition particle (SRP) for co-translational delivery of inner membrane proteins assembles[66] (Extended Data Fig. 6). As in many other bacteria[67], the ~330-nt M1 RNA is expressed independently of the RNase P protein (FN0002); it is processed off a dicistronic transcript with hypothetical ORF FN1315. The *Fnn* tmRNA is 363 nts long and its internal small ORF (14 aa) for *trans*-peptide tagging shows an interesting sequence dichotomy between oral and non-oral isolates of fusobacteria (Extended Data Fig. 7). Importantly, however, its overall conservation argues that tmRNA-SmpB (FN0609) protein-associated rescue of stalled ribosomes on damaged mRNAs is a core process in fusobacteria.

6S RNA is a bacterial riboregulator shown to bind RNAP and modulate transcription in *E. coli* and *B. subtilis*[68]. 6S RNA genes (*ssrS*) are hard to predict for a general lack of conserved primary sequence. Here, gene synteny searches combined with previous RNA structure predictions by others[69] identified an *ssrS* gene between serine protease FN0508 and arginyl-tRNA-synthetase ArgS (FN0506), antisense to hypothetical ORF FN0507 (Fig. 3a-b). Several observations argue for this to be a functional 6S RNA: its overall abundance and accumulation towards stationary phase; its conserved structure as a long hairpin with an internal bulge mimicking a DNA open promoter complex; and our successful detection of tiny (26-nt long) pRNAs whose synthesis marks 6S RNA-related RNAP activity[68] (Fig. 3b-c, Extended Data Fig. 8).

CRISPR-Cas systems protect against unwanted foreign DNA in many bacteria. In the *Fnn* genome we identified a putative type I-B system with 17 repeats (Fig. 3d), one of which (repeat #11) displays a full match for the fusobacterial phage ΦFunu2 (ref. [70]). This CRISPR-Cas locus must be constitutively active, given that both dRNA-seq and northern blot analysis showed pre-crRNA processing into individual crRNA under all conditions tested. As previously seen with type I-B systems in *Clostridium thermocellum* and *Methanococcus maripaludis*[71], the processing patterns are complex, exhibiting stable intermediates of double or even multiple repeat-spacer pairs. Interestingly, both the Cas genes and the crRNAs are upregulated in stationary phase cells, indicating a possible

trade-off between active anti-phage defense and growth optimization (Supplementary Dataset 2). Further investigation across a greater number of strains is warranted to see whether these expression and processing patterns are conserved. A preliminary analysis of the four additional strains suggests that *F. periodonticum* also upregulates its CRISPR-Cas defense towards the stationary phase (Supplementary Dataset 2).

*The full noncoding RNA suite*

To comprehensively annotate additional noncoding transcripts, we combined ANNOgesic predictions[10] with manual inspection of the dRNA-seq data. This yielded 43 sRNA candidates from all over the genome (Fig. 4a-b), which were either transcribed from independent genes in intergenic regions (IGRs) or processed off mRNAs. For nomenclature, we will refer to them as FunR sRNAs until a function is assigned. Extensive northern blot validation experiments using RNA samples from three growth phases confirmed 24 of 43 tested candidates, ranging from 56 to 345 nts in length (Fig. 4c). Nine of these validated sRNAs are expressed from stand-alone sRNA genes in IGRs (Supplementary Dataset 8). Additionally, 12 ncRNAs possess promoters that exhibit the same $\sigma^{70}$ motif seen in mRNA genes (Supplementary Fig. 4).

Most of the candidate sRNAs are highly specific to *Fnn*, at least in primary sequence (Fig. 4b). BLASTN analysis (75% nucleotide-identity cut-off) across 36 fusobacterial species and subspecies suggests that there are some broadly-conserved fusobacterial core sRNAs, e.g., FunR12 or FunR23. On the other end of the spectrum, FunR19 is found in the reference strain *F. nucleatum* ATCC 25586 and only 2 additional strains. The FunR19 sRNA is located in the 5' UTR of a putative reverse transcriptase gene, which is reminiscent of bacterial retrons and indicates a function in specialized anti-phage defence (Fig. 4c)[72]. In addition, we note intriguing similarities in genomic location to sRNAs from other bacterial genera that may provide functional clues. For example, FunR7 being located in the 5' region of stress chaperone gene *clpB* shares its location with the abundant ProQ-binding sRNA RyfD of *E. coli* and *Salmonella*[73,74] (Fig. 4c). In *E. coli*, RyfD can inhibit biofilm formation[75], which is an important physiological trait for fusobacteria as well. FunR27 is a 5'UTR-derived sRNA that comprises a predicted SAM-I riboswitch region of the *metK* mRNA (S-adenosylmethionine synthase). Such riboswitch-derived

sRNAs were proposed to act *in trans* as gene expression regulators in *Listeria monocytogenes*[76]. An interesting example of a *cis*-antisense sRNA is FunR43; it overlaps with the 3' end of a SIR2-domain protein (FN1185) in a putative genomic "defence island", indicating an anti-phage function[77].

*FoxI is a conserved oxygen-induced sRNA*

To gain initial insight into fusobacterial sRNA functions, we selected FunR22 for further analysis. As shown in Fig. 5a, FunR22 sequences are present in all *F. nucleatum* strains as well as in related *F. periodonticum* and *F. hwasookii*, usually in close vicinity to the *rsmB* (rRNA methyltransferase) or *trpB* (tryptophan synthase) genes. The dRNA-seq data argue that FunR22 is transcribed from a stand-alone sRNA gene. Northern blotting validated both, the predicted length (87 nt) and growth-phase dependent expression of FunR22 (Fig. 5b). Sequence alignment of FunR22 sequences and *in silico* RNA folding suggest two major sRNA regions, a 33-nt single-stranded region with a putative conserved seed sequence for mRNA recognition and a long ρ-independent terminator hairpin (Fig. 5c-d). Additionally, the alignment shows a highly conserved promoter region that differs from the $\sigma^{70}$ consensus (Fig. 5c), indicating that expression of FunR22 was regulated under specific conditions. To identify such conditions, we exposed *F. nucleatum* to oxidative or oxygen stress ($H_2O_2$ or $O_2$), heat shock (42°C), membrane stress (bile, acidic pH, lysozyme) and iron limitation (depletion of $Fe^{2+}$). Within this panel, we observed a specific upregulation of FunR22 upon a 20-min oxygen shock, based on which we renamed it FoxI for 'Fusobacterial oxygen-induced sRNA' (Fig. 5e).

*Plasmid pEcoFus-1 as a shuttle vector for functional analysis*

Functional characterization of bacterial sRNAs requires gene disruption and overexpression methods as well as knowledge of sRNA-associated proteins. To advance functional RNA analysis in fusobacteria, we developed a genetic system for sRNA expression by building a new shuttle vector, pEcoFus (Fig. 6a), on a pRPF185 chassis[78], in which the cloned (sRNA) gene of interest is expressed from the constitutive fusobacterial 4.5S RNA promoter mapped above. Northern blot analysis showed that we

achieved stable expression of several sRNAs from this vector (Fig. 6a). Importantly, none of these sRNAs was toxic, neither in the cloning vehicle *E. coli* nor in target organism, *F. nucleatum*.

*FoxI represses the major outer membrane protein FomA*

To identify putative mRNA targets of FoxI, we overexpressed it in *F. nucleatum* from pEcoFus (Fig. 6a). While the overexpression caused no apparent morphological changes, it prevented *F. nucleatum* from reaching full cell density (Extended Data Fig. 9a), showing that the cloned sRNA was physiologically active. Comparison of protein profiles revealed depletion of a very abundant protein in the 35-55 kDa range in the FoxI expressing strain, compared to empty pEcoFus (Fig. 6b). Mass spectrometry of the excised band predicted this depleted protein to be the ~42 kDa outer membrane (OM) porin FomA (Fig. 6c and Extended Data Fig. 9c-f). The OM localization of this putative FoxI target was fully supported by cell fractionation experiments as well as western blot analysis with a FomA antiserum (Fig. 6b).

Next, we used the IntaRNA algorithm[79] to predict possible base-pairing interactions between FoxI and the *fomA* mRNA. The top prediction was an 8-bp RNA helix with a bulged A, sequestering the RBS of this target and so repressing translation initiation of the *fomA* mRNA (Fig. 6d). Most importantly, this interaction would fully engage a conserved region of the *fomA* mRNA (Fig. 6e) and one of the two candidate seed regions in FoxI, i.e. the conserved cytidine-rich stretch upstream of the long 3' hairpin (Fig. 5c-d). Indeed, mutation of three consecutive cytidines to adenines (sRNA variant FoxI-3C) abrogated both, the growth phenotype and the downregulation of FomA (Figs. 6b-c and Extended Data Fig. 9a). Expression of the sRNA itself, however, was largely unaffected by this mutation (Extended Data Fig. 9b). Thus, the almost complete depletion of one of the most abundant *Fnn* proteins by the FoxI sRNA is very likely to occur on the post-transcriptional level by a mechanism of translational repression.

## DISCUSSION

The phylum Fusobacteria, despite its importance for human and veterinary medicine, is understudied with respect to molecular mechanisms of gene expression and RNA biology. Our global RNA maps obtained for five fusobacterial strains provide an important resource in the quest to understand how gene regulation enables this group of microbes to dwell and proliferate in diverse animals[6]. In addition, while high-throughput screening for fusobacterial gene function is in its infancy[80,81], our single-nucleotide expression maps will be invaluable for scoring effects of transposon insertions outside reading frames. Our study increases by a factor of thousand the number of mapped 5' ends, effectively assigning a primary TSS or polycistronic transcription to the majority of *F. nucleatum* genes. Specifically, we have assigned TSS to 706 genes or operons, which include the two major virulence factors FadA and Fap2 and ~200 other putative virulence factors[55] (Figs. 1b and 2a, Supplementary Dataset 1). Their observed growth-independent expression starkly contrasts with the common condition-dependent induction of virulence genes in many other human pathogens. This supports *F. nucleatum*'s role as an opportunistic pathogen and generalist[82], thus in part explaining why the bacterium can colonize additional sites in the human body.

Regarding general transcription signals, our analysis of different growth stages shows that the majority of fusobacterial promoters possess an extended -10 box; a -35 box is less prevalent in *Fnp* and *Fnv*. Despite these differences, our results indicate a shared recognition sequence for the fusobacterial housekeeping $\sigma^{70}$ factor. Intriguingly, the fusobacterial promoters differ from those of closely related species (e.g., *Bacteroides fragilis* [83], *B. thetaiotaomicron* [84] or *Porphyromonas gingivalis* [85]), and seem more similar to promoters of proteobacteria [37,41,46] or Firmicutes [86].

We not only correct the *in silico* annotation for ~22% of all genes of *F. nucleatum*, but also add hundreds of functional elements. In addition to three previously overlooked conserved small ORFs, we provide evidence for a rich layer of noncoding RNAs of diverse origin including a dozen sRNAs from "empty" intergenic regions. Except for four ubiquitous RNAs (6S RNA, 4.5S RNA, tmRNA, M1 of RNase P), the fusobacterial sRNAs seem unique, showing no obvious sequence homology outside this phylum. Yet, our discovery of FoxI as a repressor of the major porin FomA establishes proof-of-principle that fusobacteria share with many other bacteria the use of sRNAs to regulate envelope composition[29]. Mechanistically, we consider it likely that *in vivo*, stable formation of the

predicted 8-bp FoxI-*fomA* RNA duplex is mediated by an RNA-binding protein (RBP). Since fusobacteria lack CsrA, Hfq, and ProQ[87,88], the use of FoxI as an RNA bait in experimental RBP discovery[30,74] promises to expand the currently known set of sRNA-related RBPs. Candidates include the KhpA/B proteins, which have been predicted to associate with sRNAs in gram-positive species[89-91].

Work in other bacteria has shown that conserved sRNAs are often the most stringently regulated genes within a given regulon[92-95]. While the reported colonization of different body sites[2-4,96-101] implies that *F. nucleatum* possesses multiple environmental sensing and stress response pathways, the responsible TFs are unknown. Here, we observe strong and selective upregulation of the FoxI sRNA after oxygen exposure, paired with exceptional sequence conservation of the *foxI* promoter, indicating the presence of key elements for its oxygen-dependent activation (or derepression). These observations provide invaluable starting points to find a TF that enables *F. nucleatum* to respond to elevated oxygen levels. Physiologically, the FoxI-mediated repression of FomA synthesis when *F. nucleatum* senses an oxygen-rich environment, i.e., after leaving its stable niches in oral biofilms or cancer tissue, may protect from the host's immune system, for FomA is known to be recognized by both adapted and innate immune pathways[102].

In conclusion, high-resolution RNA maps are essential prerequisites for the study of host-pathogen interactions by advanced transcriptomics methods such as dual RNA-seq[103] or bacterial single-cell RNA-seq[104,105], in particular, when trying to track *F. nucleatum* gene activity in cancer tissue. Improving the poor genetic tractability of this organism, our new shuttle vector (Fig. 6a) should help to accelerate the discovery of gene function in niche adaptation and survival. Furthermore, this expression system may facilitate genome-wide antisense knockdown of chromosomal genes, as pioneered in *Staphylococcus aureus* 19 years ago[106]. Alternatively, it may be used to repurpose CRISPR-Cas locus of *F. nucleatum*, which is here shown to be functional (Fig. 3d), for intrinsic gene regulation. All in all, the RNA-centric approach taken here opens up new avenues for molecular microbiology excursions into an understudied phylum of great medical importance.

## METHODS

*Bacterial strains and growth conditions*

All oligonucleotides, plasmids or strains used in this study can be found in Supplementary Dataset 9. This study utilized four different subspecies of *Fusobacterium nucleatum*: *Fusobacterium nucleatum* subsp*. nucleatum* (ATCC 25586 and ATCC 23726), subsp. *polymorphum* (ATCC 10953), subsp. *vincentii* 3_1_36A2 and subsp. *animalis* 7_1, as well as a strain of *Fusobacterium periodonticum*. ATCC 25586 and ATCC 10953 were obtained from the German Collection of Microorganisms and Cell Culture (DSMZ, Germany) and ATCC 23726 was obtained from the American Type Culture Collection (ATCC, USA). The *Fnv*, *Fna* and *Fup* strains were a kind gift from Emma Allen-Vercoe (University of Guelph, Canada). *F. nucleatum* was routinely grown at 37 °C in 80:10:10 ($N_2$:$H_2$:$CO_2$) on Brain-Heart-Infusion (BHI) 2% agar plates supplemented with 1% (w/vol) yeast-extract, 1% (w/vol) glucose, 5µg/ml hemin and 1% (vol/vol) FBS (BHI-C). Liquid cultures for all experiments were grown in Columbia Broth medium without agitation (see below). Pre-cultures were prepared 24h prior to inoculating working cultures at a 1:50 dilution. When using *F. nucleatum* subsp. *nucleatum* ATCC 23726 carrying a plasmid, agar plates were supplemented with 5 µg/ml Thiamphenicol and liquid cultures contained 2.5 µg/ml of the antibiotic.

*Sample collection for dRNA-seq analysis and RNA-extraction*

Samples for dRNA-seq analysis were collected from bacterial cultures corresponding to the early logarithmic, mid-logarithmic and early stationary growth phase. Three biological replicates were collected for each time point. Samples were fixed by the addition of STOP Mix [95% (vol/vol) EtOH, 5% (vol/vol)] phenol prior to snap-freezing in liquid nitrogen. All samples were subsequently stored at −80 °C until RNA extraction. RNA extraction was performed as previously reported [37]. In short, frozen bacterial cultures were thawed on ice, centrifuged and cell pellets were resuspended in lysis solution (600 µl of 0.5 mg/ml lysozyme in TE buffer (pH 8.0) with 60 µl 10% SDS). Bacterial cells were lysed by placing the samples for 2 minutes at 65°C in a water bath and the reaction stopped by addition of 65 µl of 3M Sodium acetate (pH 5.2). Total RNA was extracted from the lysates using the hot phenol method[107].

*Generation and sequencing of cDNA libraries for dRNA-seq and RNA-seq*

cDNA libraries for dRNA-seq were generated by Vertis Biotechnology AG, Munich, Germany. This was performed as described in *Berezikov et al. (2010)*[108] while omitting the RNA size-fractionation step prior to cDNA-synthesis. To summarize, two libraries were created for each total RNA sample for which one included all transcripts (TEX-) and the second one was enriched for primary transcripts by treatment with terminator exonuclease (TEX+). This was followed by addition of a poly(A)-tail to equal amounts of RNA samples using a poly(A) polymerase. Next, the 5'triphosphate residues were removed with tobacco acid pyrophosphatase (TAP) prior to ligation of 5'RNA adapter. This was used for the generation of first-strand cDNA using oligo(dT)-adapter primer and MMLV transcriptase. The cDNA concentration was further increased to 20-30 ng/µl through PCR-amplification utilizing a high-fidelity DNA-polymerase. Library specific barcodes were included for the 5'sequencing adapters to allow for multiplexing. cDNA libraries for the untreated samples were performed in a similar fashion except that the TEX treatment was omitted, the final cDNA concentration were 10-20ng/µl and barcode sequences were included in the 5' and 3' TrueSeq sequencing adapters. The different cDNA libraries were pooled in approximately equimolar amounts before being sequenced on an Illumina NextSeq 500 system (75bp single-end read length). The cDNA libraries were sequenced with Vertis Biotechnology AG, Munich, Germany (for *F. n.* subsp. *nucleatum* and *polymorphum*) or by the Core Unit SysMed (University Würzburg) (for *F. n.* subsp. *animalis*, *vincentii* and *F. periodonticum*)

*Read mapping of dRNA-seq data*

The RAW data is available at GEO under accession GSE161360. Adapters clipping and quality trimming of the Illumina reads in FASTQ format was performed using the fastq_quality_trimmer function of the FASTX toolkit version 0.10.1 (http://hannonlab.cshl.edu/fastx_toolkit/). The READemption 0.4.3 tool was used to perform the following steps with the usage of its subcommands "create", "align" and "coverage"[109]: The poly(A)-tail sequences were computationally removed prior to size-filtering step which excluded all sequences <12 nt. The following reference genomes for each isolate were downloaded from the NCBI ftp server and used for mapping via

Segemehl version 0.2.0 (ref. [110]): *F. nucleatum* subsp. *nucleatum* (NC_003454.1), *F. nucleatum* subsp. *polymorphum* (NZ_CM000440.1), *F. nucleatum* subsp. *animalis* (NZ_CP007062.1), *F. nucleatum* subsp. *vincentii* (NC_022196.1), *F. periodonticum* (NZ_CP028108.1).

*Prediction of TSS, transcripts, UTRs, operons, sORFs, sRNAs, CRISPR locus, regulatory RNA-elements and terminators*

For the following predictions, the ANNOgesic tool was used[10]: All parameters were kept at the default setting, if not otherwise specified. TSS were predicted using the default setting of ANNOgesic, which further categorizes the TSS into 5 different classes: primary, secondary, internal, antisense and orphan TSS. For this, all parameters had to be met within all replicates of each condition in order to be annotated as a TSS. Additionally, predicted secondary TSSs were excluded if a primary TSS was predicted to be <7 nt away. The TSS prediction was further improved by manual curation. The annotation of UTRs was conducted by adjusting the default settings to allow for a maximum length of 300 nt and by extending 5'UTRs up to 25 nt to be connected with detected transcripts. All sRNAs were predicted through the input of detected promoter sequences (see below) while further allowing for the detection of 5'UTR-derived sRNAs. Additionally, ANNOgesic takes TSS, coverage and the presence of terminators into account when predicting the length of sRNA candidates. For predictions of transcripts, operon structures, CRISPR locus and regulatory RNA-elements the default settings of ANNOgesic were used. Rho-independent terminator prediction was carried out using ANNOgesic which combines two heuristic prediction algorithms: TransTermHP[111] analysis and a detection of sharp coverage decreases around the predicted terminator sequence.

*Promoter, Shine-Dalgarno and prophage sequence detection*

To detect putative promoter motifs, 50nt upstream of each detected TSS including the TSS position were extracted using BEDtools[112] and further analysed with MEME version 4.12.0[45]. The same procedure was performed for all genes lacking a primary TSS. For this, 100 nt long sequences upstream of the start codon were extracted in order to account for

unaccounted 5'UTRs. To identify SD sequences, all 5'UTR sequences were extracted (see above) and used as input for MEME analysis.

*Northern blot analysis*

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

*Re-annotation of coding sequences*

All annotated CDS were checked for the presence of an AUG start codon. Additionally, 20 nt upstream of the start codon were extracted to analyse if a Shine Dalgarno sequence was present. In the absence of either, the ORFs were manually inspected. In the case of a missing AUG start codon, the sequences were surveyed for an in-frame start codon up- or downstream of the annotated start. A similar analysis was conducted in case of an absent Shine Dalgarno sequence. CDS for which the corrected start was supported by both the presence of a start codon (AUG, GUG or UUG) and a Shine Dalgarno sequence as well as by sequence conservation in other *F. n. susbp. nucleatum* strains were included for re-annotation.

*Construction of shuttle vector pEcoFus*

The *E. coli – C. difficile* shuttle vector pRPF185[78] was used as a backbone. In the first step, the backbone was opened by inverse PCR using primers JVO-17251/17252 to remove *repA* and *orfB* while introducing PvuI and NotI restriction sites. The origin of replication for *F. nucleatum* was amplified from pORI92 (a gift from G. Bachrach) using primers JVO-17207/17248 to introduce PvuI and NotI restriction sites, cut and ligated with the backbone resulting in pFP14. Next, pFP14 was digested with KpnI and BamHI before being assembled together with an amplified fragment of the 100nt promoter region of

the 4.5S RNA using the NEBuilder® Hifi Assembly Cloning kit and transformed into *E. coli* topF'. This resulted in the vector pEcoFus for constitutive overexpression for any gene of interest.

*Construction of sRNA expression plasmids*

Overexpression plasmids for different sRNAs were assembled using Gibson Assembly protocol. Full-length sRNA sequences for FoxI (JVO-17259/17302), FunR47 (JVO-17306/17307), FunR7 (JVO-17543/17544), FunR7+extended 3'end (JVO-17543/17545), FunR12 (JVO-17549/17550) and FunR16 (JVO-17558/17559) were amplified from chromosomal DNA of *F. nucleatum subsp. nucleatum* ATCC 25586 and assembled with BamHI-digested pEcoFus using NEBuilder® Hifi Assembly Cloning kit and transformed into *E. coli* topF'. Correct sequences were verified by sequencing.

Electroporation of *F. nucleatum subsp. nucleatum* ATCC 23726

Electro-competent cells of the genetically tractable F. nucleatum subsp. nucleatum ATCC 23726 were generated in a similar fashion as described in *Haake et al. (2000)*[113]. In short, cells were harvested from mid-exponential phase at 4000xg for 10min at 4°C. Subsequently the cells were washed in ice-cold and pre-reduced 10% (V/V) glycerol solution for a total of 5 washes. Finally, the cells were resuspended to 60 OD/ml in 10% glycerol solution, aliquoted and stored at -80°C until use. 80µl of competent cells were transformed with 5µg of dialyzed plasmid (2.0 kV, 1 mm gap) and recovered in Columbia broth containing 1 mM $MgCl_2$ for 2h before plating the bacteria on BHI-C plates containing 5 µg/ml Thiamphenicol. This routinely yielded >10 CFU/µg plasmid DNA. Transformants were re-streaked onto fresh BHI-C before verification of carrying the correct insert and preparation of glycerol stocks or downstream analyses.

*Stress exposure of F. nucleatum*

*Fnn* was grown to mid-exponential phase before being exposed to different stress conditions for 20min. The following were conducted inside the anaerobic chamber: $H_2O_2$

(400µM; AppliChem), bile (0.05 % (w/v); Sigma Aldrich (70168)), iron depletion (300µM; 2,2-Bipyridyl; Sigma Aldrich (D216305-10G)) and lysozyme (50µg/ml; Roth (8259.2)). All treated cultures and the untreated control were kept in the incubator at 37°C for the duration of the treatment. For acidic pH treatment the bacteria were harvested (37°C; 5min at 4000xg) and resuspended in pH-adjusted (pH 5.5 or pH 7 for the control) pre-warmed and reduced media. Similarly, for the $O_2$ shock, the samples were spread onto petri dishes and incubated at 37°C under atmospheric conditions. The heat shock was conducted at 42.5°C in a water bath outside the chamber. Prior to removing samples from the chamber, the falcons were wrapped with parafilm to avoid oxygen exposure. At the end of each treatment, cells were fixed in STOP mix and RNA was isolated as described above.

*Subcellular fractionation*

Subcellular fractionation was performed as described by *Knoke et al. (2020)*[114] for two biological replicates. In short, 50 $OD_{600}$ units were harvested from mid-exponential phase cultures at 4000xg for 10min at 4°C. The cell pellet was resuspended in lysis buffer (10 mM Tris-HCl (pH 7.5), 0.2 mM PMSF, 1 mM $MgCl_2$) and total protein samples were collected. The cells were lysed via sonication and centrifuged at 4000xg for 10min at 4°C. A lysate sample was collected prior to performing another centrifugation step at 15,000xg for 1h at 4°C. Samples for the cytosolic fraction were collected and the pellet was incubated with 0.5 % N-lauroylsarcosine in lysis buffer overnight at 4°C to separate inner and outer membrane. Another high-speed centrifugation step was performed (100,000xg for 2h at 4°C) collecting the inner membrane fraction and dissolving the outer membrane fraction in lysis buffer containing 0.5% Triton X-100. For denaturing SDS-page analysis, 0.1 $OD_{600}$ were loaded for total cell lysate and the cytosolic fraction while 1 $OD_{600}$ was loaded for both inner and outer membrane fraction. For western blot analysis of FomA, a polyclonal antibody was generated in rabbits against a synthetic peptide derived from FomA (amino acid sequence: KKFATYNKGDKKSQF). Peptide synthesis, validation via ELISA and subsequent antigen-purification was performed by Eurogentec (Belgium). Afterwards an unstained gel was transferred to a PVDF membrane (GE Healthcare Life

Sciences) and incubated with anti-FomA antibody before being detected using an anti-rabbit secondary antibody (Thermo Fisher Scientific, cat# 31460).

*nanoLC-MS/MS analysis of protein samples*

Quantification of FomA from *F. nucleatum* samples was conducted similarly as described in *Hör et al. (2020)*[91]. Denatured samples for total protein (empty vector control, FoxI or FoxI-3C overexpression) or the outer membrane fraction (empty vector control) were separated on a denaturing 15% SDS-gel and for each sample the region of interest was excised from the gels for two replicates (Fig. 6b). The samples were destained in 100 mM ammonium bicarbonate containing 30% acetonitrile and shrunk in 100% acetonitrile prior to trypsin digestion (0.1 µg in 100 mM ammonium bicarbonate, overnight at 37°C). The samples were then dissolved in 5% formic acid for subsequent nanoLC-MS/MS analysis. For nanoLC-MS/MS analysis an Orbitrap Fusion (Thermo Fisher Scientific) combined with a PicoView Ion source (New Objective) and an Easy-nLC 1000 (Thermo Fisher Scientifc) was used. The dissolved samples were loaded on PicoFrit capillary columns (30cm x 150 µM ID, New Objective) packed with ReproSil-Pur 120 C18-AQ (1.9 µM, Dr. Maisch) and separated using a 140-min linear gradient (3-30% acetonitrile, 0.1% formic acid) at a flow rate of 500nl/min. Both MS (60,000 scans; target value for AGC: $2x10^5$) and MS/MS (7,500 scans; target value for AGC: $5x10^4$) analysis were conducted using the Orbitrap with the HCD fragmentation set to 35% normalized collision energy. A fixed cycle time of 3 s was used for the Top Speed data-dependent MS/MS method while a further exclusion of 1 repeat count per minute was applied. Precursor selection was conducted at a minimum threshold of 50,000 while excluding single charged ones. Internal calibration with Easy-IC was used to improve mass-to-charge ratio assignment. Subsequent analysis of the data was conducted using MaxQuant (version 1.5.7.4)[115] with integrated Andromeda comparing it against the Uniprot database for *F. nucleatum subsp. nucleatum* ATCC 25586 (FUSNN).

*Data visualization*

Coverage plots were generated with the R packages Gviz (v1.32.0). Gene cartoons were generated using the R packages gggenes (v0.4.0). The web browser was build using jbrowse (v1.16.2).

*Statistics and Reproducibility*

The dRNA-seq was performed in biological triplicates. Northern blot analysis for growth condition was performed in biological triplicates and for the sRNA overexpression in biological duplicates. The subcellular fractionation was conducted for two biological replicates. The samples size was determined based upon experience with previous studies.

## Data availability

RNA-seq data can be accessed at NCBI Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) under the accession number GSE161360.

MS data can be accessed at the Proteomics Identification Database PRIDE (https://www.ebi.ac.uk/pride) under the accession number PXD022474.

The RFAM database can accessed at http://rfam.xfam.org/.

The Uniprot database can be accessed at https://www.uniprot.org/uniprot/?query=taxonomy:190304.

## ACKNOWLEDGEMENT

We are grateful to Youssef el Mouali Benomar, Jens Hör, Daniel Ryan and Gilad Bachrach for fruitful discussions; Elisa Venturini for support with the subcellular fractionation; Gilad Bachrach (The Hebrew University of Jerusalem, Israel) for providing pORI92; Emma Allen-Vercoe (University Guelph, Canada) for providing *Fusobacterium* strains; Laura Jenniches for help with launching the web browser. We thank Stephanie Lamer and Andreas Schlosser for mass spectrometry analysis. We thank the Vogel Stiftung Dr. Eckernkamp for supporting FP with a Dr. Eckernkamp Fellowship. This work was funded by a DFG Gottfried Wilhelm Leibniz Award to JV (DFG Vo875-18). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

## REFERENCES

1 Dzutsev, A. *et al.* Microbes and Cancer. *Annual review of immunology* **35**, 199-228, doi:10.1146/annurev-immunol-051116-052133 (2017).

2 Parhi, L. *et al.* Breast cancer colonization by Fusobacterium nucleatum accelerates tumor growth and metastatic progression. *Nat Commun* **11**, 3259, doi:10.1038/s41467-020-16967-2 (2020).

3 Kostic, A. D. *et al.* Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. *Genome Res* **22**, 292-298, doi:10.1101/gr.126573.111 (2012).

4 Castellarin, M. *et al.* Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. *Genome Res* **22**, 299-306, doi:10.1101/gr.126516.111 (2012).

5 Hussan, H., Clinton, S. K., Roberts, K. & Bailey, M. T. Fusobacterium's link to colorectal neoplasia sequenced: A systematic review and future insights. *World journal of gastroenterology* **23**, 8626-8650, doi:10.3748/wjg.v23.i48.8626 (2017).

6 Brennan, C. A. & Garrett, W. S. Fusobacterium nucleatum - symbiont, opportunist and oncobacterium. *Nature reviews. Microbiology* **17**, 156-166, doi:10.1038/s41579-018-0129-6 (2019).

7 Bullman, S. *et al.* Analysis of Fusobacterium persistence and antibiotic response in colorectal cancer. *Science (New York, N.Y.)* **358**, 1443-1448, doi:10.1126/science.aal5240 (2017).

8 Amitay, E. L. *et al.* Fusobacterium and colorectal cancer: causal factor or passenger? Results from a large colorectal cancer screening study. *Carcinogenesis* **38**, 781-788, doi:10.1093/carcin/bgx053 (2017).

9 Yu, T. *et al.* Fusobacterium nucleatum Promotes Chemoresistance to Colorectal Cancer by Modulating Autophagy. *Cell* **170**, 548-563.e516, doi:10.1016/j.cell.2017.07.008 (2017).

10 Yu, S. H., Vogel, J. & Forstner, K. U. ANNOgesic: a Swiss army knife for the RNA-seq based annotation of bacterial/archaeal genomes. *GigaScience* **7**, doi:10.1093/gigascience/giy096 (2018).

11 Gaiser, R. A. *et al.* Enrichment of oral microbiota in early cystic precursors to invasive pancreatic cancer. *Gut* **68**, 2186-2194, doi:10.1136/gutjnl-2018-317458 (2019).

12 Yamamura, K. *et al.* Fusobacterium nucleatum in gastroenterological cancer: Evaluation of measurement methods using quantitative polymerase chain reaction and a literature review. *Oncology letters* **14**, 6373-6378, doi:10.3892/ol.2017.7001 (2017).

13 Yamamura, K. *et al.* Human Microbiome Fusobacterium Nucleatum in Esophageal Cancer Tissue Is Associated with Prognosis. *Clinical cancer research : an official journal of the American Association for Cancer Research* **22**, 5574-5581, doi:10.1158/1078-0432.Ccr-16-1786 (2016).

14 Rubinstein, M. R. *et al.* Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating E-cadherin/beta-catenin signaling via its FadA adhesin. *Cell Host Microbe* **14**, 195-206, doi:10.1016/j.chom.2013.07.012 (2013).

15 Abed, J. *et al.* Fap2 Mediates Fusobacterium nucleatum Colorectal Adenocarcinoma Enrichment by Binding to Tumor-Expressed Gal-GalNAc. *Cell Host Microbe* **20**, 215-225, doi:10.1016/j.chom.2016.07.006 (2016).

16 Han, Y. W. *et al.* Identification and characterization of a novel adhesin unique to oral fusobacteria. *J Bacteriol* **187**, 5330-5340, doi:10.1128/JB.187.15.5330-5340.2005 (2005).

17 Holt, R. A. & Cochrane, K. Tumor Potentiating Mechanisms of Fusobacterium nucleatum, A Multifaceted Microbe. *Gastroenterology* **152**, 694-696, doi:10.1053/j.gastro.2017.01.024 (2017).

18 Kaplan, C. W. *et al.* Fusobacterium nucleatum outer membrane proteins Fap2 and RadD induce cell death in human lymphocytes. *Infect Immun* **78**, 4773-4778, doi:10.1128/IAI.00567-10 (2010).

19 Gur, C. *et al.* Binding of the Fap2 protein of Fusobacterium nucleatum to human inhibitory receptor TIGIT protects tumors from immune cell attack. *Immunity* **42**, 344-355, doi:10.1016/j.immuni.2015.01.010 (2015).

20 Cochrane, K., Robinson, A. V., Holt, R. A. & Allen-Vercoe, E. A survey of Fusobacterium nucleatum genes modulated by host cell infection. *Microb Genom* **6**, doi:10.1099/mgen.0.000300 (2020).

21 Abed, J. *et al.* Colon Cancer-Associated Fusobacterium nucleatum May Originate From the Oral Cavity and Reach Colon Tumors via the Circulatory System. *Front Cell Infect Microbiol* **10**, 400, doi:10.3389/fcimb.2020.00400 (2020).

22 Han, Y. W. Fusobacterium nucleatum: a commensal-turned pathogen. *Current opinion in microbiology* **23**, 141-147, doi:10.1016/j.mib.2014.11.013 (2015).

23 Manson McGuire, A. *et al.* Evolution of invasion in a diverse set of Fusobacterium species. *MBio* **5**, e01864, doi:10.1128/mBio.01864-14 (2014).

24 Sanders, B. E., Umana, A., Lemkul, J. A. & Slade, D. J. FusoPortal: an Interactive Repository of Hybrid MinION-Sequenced Fusobacterium Genomes Improves Gene Identification and Characterization. *mSphere* **3**, doi:10.1128/mSphere.00228-18 (2018).

25 Kook, J. K. *et al.* Genome-Based Reclassification of Fusobacterium nucleatum Subspecies at the Species Level. *Curr Microbiol* **74**, 1137-1147, doi:10.1007/s00284-017-1296-9 (2017).

26 Merritt, J., Niu, G., Okinaga, T. & Qi, F. Autoaggregation response of Fusobacterium nucleatum. *Appl Environ Microbiol* **75**, 7725-7733, doi:10.1128/AEM.00916-09 (2009).

27 Mutha, N. V. R. *et al.* Transcriptional responses of Streptococcus gordonii and Fusobacterium nucleatum to coaggregation. *Mol Oral Microbiol* **33**, 450-464, doi:10.1111/omi.12248 (2018).

28 Sasaki-Imamura, T., Yano, A. & Yoshida, Y. Production of indole from L-tryptophan and effects of these compounds on biofilm formation by Fusobacterium nucleatum ATCC 25586. *Appl Environ Microbiol* **76**, 4260-4268, doi:10.1128/aem.00166-10 (2010).

29 Wagner, E. G. & Romby, P. Small RNAs in bacteria and archaea: who they are, what they do, and how they do it. *Advances in genetics* **90**, 133-208, doi:10.1016/bs.adgen.2015.05.001 (2015).

30 Hör, J., Gorski, S. A. & Vogel, J. Bacterial RNA Biology on a Genome Scale. *Mol Cell* **70**, 785-799, doi:10.1016/j.molcel.2017.12.023 (2018).

31 Hör, J., Matera, G., Vogel, J., Gottesman, S. & Storz, G. Trans-Acting Small RNAs and Their Effects on Gene Expression in Escherichia coli and Salmonella enterica. *EcoSal Plus* **9**, doi:10.1128/ecosalplus.ESP-0030-2019 (2020).

32 Babitzke, P., Lai, Y. J., Renda, A. J. & Romeo, T. Posttranscription Initiation Control of Gene Expression Mediated by Bacterial RNA-Binding Proteins. *Annu Rev Microbiol* **73**, 43-67, doi:10.1146/annurev-micro-020518-115907 (2019).

33 Barquist, L. & Vogel, J. Accelerating Discovery and Functional Analysis of Small RNAs with New Technologies. *Annual review of genetics* **49**, 367-394, doi:10.1146/annurev-genet-112414-054804 (2015).

34 Ang, M. Y. *et al.* Comparative Genome Analysis of Fusobacterium nucleatum. *Genome Biol Evol* **8**, 2928-2938, doi:10.1093/gbe/evw199 (2016).

35 Pyne, M. E., Bruder, M. R., Moo-Young, M., Chung, D. A. & Chou, C. P. Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in Clostridium. *Scientific reports* **6**, 25666, doi:10.1038/srep25666 (2016).

36 Li, Y. *et al.* Harnessing Type I and Type III CRISPR-Cas systems for genome editing. *Nucleic acids research* **44**, e34, doi:10.1093/nar/gkv1044 (2016).

37 Sharma, C. M. *et al.* The primary transcriptome of the major human pathogen Helicobacter pylori. *Nature* **464**, 250-255, doi:10.1038/nature08756 (2010).

38 Minarovits, J. Anaerobic bacterial communities associated with oral carcinoma: Intratumoral, surface-biofilm and salivary microbiota. *Anaerobe* **68**, 102300, doi:10.1016/j.anaerobe.2020.102300 (2021).

39 Sharma, C. M. & Vogel, J. Differential RNA-seq: the approach behind and the biological insight gained. *Current opinion in microbiology* **19**, 97-105, doi:10.1016/j.mib.2014.06.010 (2014).

40 Georg, J. & Hess, W. R. Widespread Antisense Transcription in Prokaryotes. *Microbiology spectrum* **6**, doi:10.1128/microbiolspec.RWR-0029-2018 (2018).

41 Feklistov, A., Sharon, B. D., Darst, S. A. & Gross, C. A. Bacterial sigma factors: a historical, structural, and genomic perspective. *Annu Rev Microbiol* **68**, 357-376, doi:10.1146/annurev-micro-092412-155737 (2014).

42 Browning, D. F. & Busby, S. J. W. The regulation of bacterial transcription initiation. *Nature Reviews Microbiology* **2**, 57, doi:10.1038/nrmicro787 (2004).

43 Karpathy, S. E. *et al.* Genome sequence of Fusobacterium nucleatum subspecies polymorphum - a genetically tractable fusobacterium. *PLoS One* **2**, e659, doi:10.1371/journal.pone.0000659 (2007).

44 Kapatral, V. *et al.* Genome Sequence and Analysis of the Oral Bacterium Fusobacterium nucleatum Strain ATCC 25586. *Journal of Bacteriology* **184**, 2005-2018, doi:10.1128/jb.184.7.2005-2018.2002 (2002).

45 Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME Suite. *Nucleic acids research* **43**, W39-49, doi:10.1093/nar/gkv416 (2015).

46 Dugar, G. *et al.* High-resolution transcriptome maps reveal strain-specific regulatory features of multiple Campylobacter jejuni isolates. *PLoS Genet* **9**, e1003495, doi:10.1371/journal.pgen.1003495 (2013).

47 Zanzoni, A., Spinelli, L., Braham, S. & Brun, C. Perturbed human sub-networks by Fusobacterium nucleatum candidate virulence proteins. *Microbiome* **5**, 89, doi:10.1186/s40168-017-0307-1 (2017).

48 Kapatral, V. *et al.* Genome analysis of F. nucleatum sub spp vincentii and its comparison with the genome of F. nucleatum ATCC 25586. *Genome Res* **13**, 1180-1189, doi:10.1101/gr.566003 (2003).

49 Desvaux, M., Khan, A., Beatson, S. A., Scott-Tucker, A. & Henderson, I. R. Protein secretion systems in Fusobacterium nucleatum: genomic identification of Type 4 piliation and complete Type V pathways brings new insight into mechanisms of pathogenesis. *Biochimica et biophysica acta* **1713**, 92-112, doi:10.1016/j.bbamem.2005.05.002 (2005).

50 Umana, A. *et al.* Utilizing Whole Fusobacterium Genomes To Identify, Correct, and Characterize Potential Virulence Protein Families. *J Bacteriol* **201**, doi:10.1128/JB.00273-19 (2019).

51 Xu, M. *et al.* FadA from Fusobacterium nucleatum utilizes both secreted and nonsecreted forms for functional oligomerization for attachment and invasion of host cells. *The Journal of biological chemistry* **282**, 25000-25009, doi:10.1074/jbc.M611567200 (2007).

52 Kaplan, C. W., Lux, R., Haake, S. K. & Shi, W. The Fusobacterium nucleatum outer membrane protein RadD is an arginine-inhibitable adhesin required for inter-species adherence and the structured architecture of multispecies biofilm. *Molecular microbiology* **71**, 35-47, doi:10.1111/j.1365-2958.2008.06503.x (2009).

53 Engevik, M. *et al.* Fusobacterium nucleatum adheres to Clostridioides difficile via the RadD adhesin to enhance biofilm formation in intestinal mucus. *Gastroenterology*, doi:10.1053/j.gastro.2020.11.034 (2020).

54 Doron, L. *et al.* Identification and characterization of fusolisin, the Fusobacterium nucleatum autotransporter serine protease. *PLoS One* **9**, e111329, doi:10.1371/journal.pone.0111329 (2014).

55 Kumar, A., Thotakura, P. L., Tiwary, B. K. & Krishna, R. Target identification in Fusobacterium nucleatum by subtractive genomics approach and enrichment analysis of host-pathogen protein-protein interactions. *BMC Microbiol* **16**, 84, doi:10.1186/s12866-016-0700-0 (2016).

56 Storz, G., Wolf, Y. I. & Ramamurthi, K. S. Small proteins can no longer be ignored. *Annu Rev Biochem* **83**, 753-777, doi:10.1146/annurev-biochem-070611-102400 (2014).

57 Cao, H., Ma, Q., Chen, X. & Xu, Y. DOOR: a prokaryotic operon database for genome analyses and functional inference. *Brief Bioinform* **20**, 1568-1577, doi:10.1093/bib/bbx088 (2019).

58 Casasanta, M. A. *et al.* A chemical and biological toolbox for Type Vd secretion: Characterization of the phospholipase A1 autotransporter FplA from Fusobacterium nucleatum. *The Journal of biological chemistry* **292**, 20240-20254, doi:10.1074/jbc.M117.819144 (2017).

59 Loh, E., Righetti, F., Eichner, H., Twittenhoff, C. & Narberhaus, F. RNA Thermometers in Bacterial Pathogens. *Microbiology spectrum* **6**, doi:10.1128/microbiolspec.RWR-0012-2017 (2018).

60 McCown, P. J., Corbino, K. A., Stav, S., Sherlock, M. E. & Breaker, R. R. Riboswitch diversity and distribution. *RNA (New York, N.Y.)* **23**, 995-1011, doi:10.1261/rna.061234.117 (2017).

61 Kalvari, I. *et al.* Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic acids research* **46**, D335-D342, doi:10.1093/nar/gkx1038 (2018).

62 Winkler, W. C., Nahvi, A., Roth, A., Collins, J. A. & Breaker, R. R. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428**, 281-286, doi:10.1038/nature02362 (2004).

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

64 Yakhnin, H., Yakhnin, A. V. & Babitzke, P. Ribosomal protein L10(L12)4 autoregulates expression of the Bacillus subtilis rplJL operon by a transcription attenuation mechanism. *Nucleic acids research* **43**, 7032-7043, doi:10.1093/nar/gkv628 (2015).

65 Bonner, E. R., D'Elia, J. N., Billips, B. K. & Switzer, R. L. Molecular recognition of pyr mRNA by the Bacillus subtilis attenuation regulatory protein PyrR. *Nucleic acids research* **29**, 4851-4865, doi:10.1093/nar/29.23.4851 (2001).

66 Akopian, D., Shen, K., Zhang, X. & Shan, S. O. Signal recognition particle: an essential protein-targeting machine. *Annu Rev Biochem* **82**, 693-721, doi:10.1146/annurev-biochem-072711-164732 (2013).

67 Hartmann, E. & Hartmann, R. K. The enigma of ribonuclease P evolution. *Trends Genet* **19**, 561-569, doi:10.1016/j.tig.2003.08.007 (2003).

68 Wassarman, K. M. 6S RNA, a Global Regulator of Transcription. *Microbiology spectrum* **6**, doi:10.1128/microbiolspec.RWR-0019-2018 (2018).

69 Wehner, S., Damm, K., Hartmann, R. K. & Marz, M. Dissemination of 6S RNA among bacteria. *RNA biology* **11**, 1467-1478, doi:10.4161/rna.29894 (2014).

70 Cochrane, K. *et al.* Complete genome sequences and analysis of the Fusobacterium nucleatum subspecies animalis 7-1 bacteriophage ϕFunu1 and ϕFunu2. *Anaerobe* **38**, 125-129, doi:10.1016/j.anaerobe.2015.10.013 (2016).

71 Richter, H. *et al.* Characterization of CRISPR RNA processing in Clostridium thermocellum and Methanococcus maripaludis. *Nucleic acids research* **40**, 9887-9896, doi:10.1093/nar/gks737 (2012).

72 Millman, A. *et al.* Bacterial Retrons Function In Anti-Phage Defense. *Cell*, doi:10.1016/j.cell.2020.09.065 (2020).

73 Melamed, S., Adams, P. P., Zhang, A., Zhang, H. & Storz, G. RNA-RNA Interactomes of ProQ and Hfq Reveal Overlapping and Competing Roles. *Mol Cell* **77**, 411-425 e417, doi:10.1016/j.molcel.2019.10.022 (2020).

74 Smirnov, A. *et al.* Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. *Proc Natl Acad Sci U S A* **113**, 11591-11596, doi:10.1073/pnas.1609981113 (2016).

75 Bak, G. *et al.* Identification of novel sRNAs involved in biofilm formation, motility, and fimbriae formation in Escherichia coli. *Scientific reports* **5**, 15287, doi:10.1038/srep15287 (2015).

76 Loh, E. *et al.* A trans-acting riboswitch controls expression of the virulence regulator PrfA in Listeria monocytogenes. *Cell* **139**, 770-779, doi:10.1016/j.cell.2009.08.046 (2009).

77 Koonin, E. V., Makarova, K. S. & Wolf, Y. I. Evolutionary Genomics of Defense Systems in Archaea and Bacteria. *Annu Rev Microbiol* **71**, 233-261, doi:10.1146/annurev-micro-090816-093830 (2017).

78 Fagan, R. P. & Fairweather, N. F. Clostridium difficile has two parallel and essential Sec secretion systems. *The Journal of biological chemistry* **286**, 27483-27493, doi:10.1074/jbc.M111.263889 (2011).

79 Mann, M., Wright, P. R. & Backofen, R. IntaRNA 2.0: enhanced and customizable prediction of RNA-RNA interactions. *Nucleic acids research* **45**, W435-W439, doi:10.1093/nar/gkx279 (2017).

80 Wu, C. *et al.* Forward Genetic Dissection of Biofilm Development by Fusobacterium nucleatum: Novel Functions of Cell Division Proteins FtsX and EnvC. *mBio* **9**, doi:10.1128/mBio.00360-18 (2018).

81 Coppenhagen-Glazer, S. *et al.* Fap2 of Fusobacterium nucleatum is a galactose-inhibitable adhesin involved in coaggregation, cell adhesion, and preterm birth. *Infect Immun* **83**, 1104-1113, doi:10.1128/IAI.02838-14 (2015).

82 Brown, S. P., Cornforth, D. M. & Mideo, N. Evolution of virulence in opportunistic pathogens: generalism, plasticity, and control. *Trends in microbiology* **20**, 336-342, doi:10.1016/j.tim.2012.04.005 (2012).

83 Bayley, D. P., Rocha, E. R. & Smith, C. J. Analysis of cepA and other Bacteroides fragilis genes reveals a unique promoter structure. *FEMS microbiology letters* **193**, 149-154, doi:10.1111/j.1574-6968.2000.tb09417.x (2000).

84 Ryan, D., Jenniches, L., Reichardt, S., Barquist, L. & Westermann, A. J. A high-resolution transcriptome map identifies small RNA regulation of metabolism in the gut microbe Bacteroides thetaiotaomicron. *Nat Commun* **11**, 3557, doi:10.1038/s41467-020-17348-5 (2020).

85 Jackson, C. A. *et al.* A consensus Porphyromonas gingivalis promoter sequence. *FEMS microbiology letters* **186**, 133-138, doi:10.1111/j.1574-6968.2000.tb09094.x (2000).

86 Michaux, C. *et al.* Single-Nucleotide RNA Maps for the Two Major Nosocomial Pathogens Enterococcus faecalis and Enterococcus faecium. *Front Cell Infect Microbiol* **10**, 600325, doi:10.3389/fcimb.2020.600325 (2020).

87 Olejniczak, M. & Storz, G. ProQ/FinO-domain proteins: another ubiquitous family of RNA matchmakers? *Molecular microbiology* **104**, 905-915, doi:10.1111/mmi.13679 (2017).

88 Holmqvist, E. & Vogel, J. RNA-binding proteins in bacteria. *Nature reviews. Microbiology* **16**, 601-615, doi:10.1038/s41579-018-0049-5 (2018).

89 Zheng, J. J., Perez, A. J., Tsui, H. T., Massidda, O. & Winkler, M. E. Absence of the KhpA and KhpB (JAG/EloR) RNA-binding proteins suppresses the requirement for PBP2b by overproduction of FtsA in Streptococcus pneumoniae D39. *Molecular microbiology* **106**, 793-814, doi:10.1111/mmi.13847 (2017).

90 Lamm-Schmidt, V. *et al.* Grad-seq identifies KhpB as a global RNA-binding protein in Clostridioides difficile that regulates toxin production. *microLife* **2**, doi:10.1093/femsml/uqab004 (2021).

91 Hör, J. *et al.* Grad-seq in a Gram-positive bacterium reveals exonucleolytic sRNA activation in competence control. *Embo j* **39**, e103852, doi:10.15252/embj.2019103852 (2020).

92 Altuvia, S., Weinstein-Fischer, D., Zhang, A., Postow, L. & Storz, G. A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator. *Cell* **90**, 43-53, doi:10.1016/s0092-8674(00)80312-8 (1997).

93 Pfeiffer, V. *et al.* A small non-coding RNA of the invasion gene island (SPI-1) represses outer membrane protein synthesis from the Salmonella core genome. *Molecular microbiology* **66**, 1174-1191, doi:10.1111/j.1365-2958.2007.05991.x (2007).

94 Guisbert, E., Rhodius, V. A., Ahuja, N., Witkin, E. & Gross, C. A. Hfq modulates the sigmaE-mediated envelope stress response and the sigma32-mediated cytoplasmic stress response in Escherichia coli. *J Bacteriol* **189**, 1963-1973, doi:10.1128/JB.01243-06 (2007).

95 Vanderpool, C. K. & Gottesman, S. The novel transcription factor SgrR coordinates the response to glucose-phosphate stress. *J Bacteriol* **189**, 2238-2248, doi:10.1128/JB.01689-06 (2007).

96 Nejman, D. *et al.* The human tumor microbiome is composed of tumor type-specific intracellular bacteria. *Science (New York, N.Y.)* **368**, 973-980, doi:10.1126/science.aay9189 (2020).

97 Swidsinski, A. *et al.* Acute appendicitis is characterised by local invasion with Fusobacterium nucleatum/necrophorum. *Gut* **60**, 34-40, doi:10.1136/gut.2009.191320 (2011).

98 Han, X. Y. *et al.* Fusobacterial brain abscess: a review of five cases and an analysis of possible pathogenesis. *J Neurosurg* **99**, 693-700, doi:10.3171/jns.2003.99.4.0693 (2003).

99 Gregory, S. W., Boyce, T. G., Larson, A. N., Patel, R. & Jackson, M. A. Fusobacterium nucleatum Osteomyelitis in 3 Previously Healthy Children: A Case Series and Review of the Literature. *J Pediatric Infect Dis Soc* **4**, e155-159, doi:10.1093/jpids/piv052 (2015).

100 Truant, A. L., Menge, S., Milliorn, K., Lairscey, R. & Kelly, M. T. Fusobacterium nucleatum pericarditis. *J Clin Microbiol* **17**, 349-351, doi:10.1128/JCM.17.2.349-351.1983 (1983).

101 Altshuler, G. & Hyde, S. Clinicopathologic considerations of fusobacteria chorioamnionitis. *Acta Obstet Gynecol Scand* **67**, 513-517, doi:10.3109/00016348809029862 (1988).

102 Toussi, D. N., Liu, X. & Massari, P. The FomA porin from Fusobacterium nucleatum is a Toll-like receptor 2 agonist with immune adjuvant activity. *Clin Vaccine Immunol* **19**, 1093-1101, doi:10.1128/CVI.00236-12 (2012).

103 Westermann, A. J. *et al.* Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions. *Nature* **529**, 496-501, doi:10.1038/nature16547 (2016).

104 Imdahl, F., Vafadarnejad, E., Homberger, C., Saliba, A. E. & Vogel, J. Single-cell RNA-sequencing reports growth-condition-specific global transcriptomes of individual bacteria. *Nat Microbiol*, doi:10.1038/s41564-020-0774-1 (2020).

105 Blattman, S. B., Jiang, W., Oikonomou, P. & Tavazoie, S. Prokaryotic single-cell RNA sequencing by in situ combinatorial indexing. *Nat Microbiol* **5**, 1192-1201, doi:10.1038/s41564-020-0729-6 (2020).

106 Ji, Y. *et al.* Identification of critical staphylococcal genes using conditional phenotypes generated by antisense RNA. *Science (New York, N.Y.)* **293**, 2266-2269, doi:10.1126/science.1063566 (2001).

107 Blomberg, P., Wagner, E. G. & Nordstrom, K. Control of replication of plasmid R1: the duplex between the antisense RNA, CopA, and its target, CopT, is processed specifically in vivo and in vitro by RNase III. *Embo j* **9**, 2331-2340 (1990).

108 Berezikov, E. *et al.* Diversity of microRNAs in human and chimpanzee brain. *Nature genetics* **38**, 1375-1377, doi:10.1038/ng1914 (2006).

109 Förstner, K. U., Vogel, J. & Sharma, C. M. READemption-a tool for the computational analysis of deep-sequencing-based transcriptome data. *Bioinformatics* **30**, 3421-3423, doi:10.1093/bioinformatics/btu533 (2014).

110 Otto, C., Stadler, P. F. & Hoffmann, S. Fast and sensitive mapping of bisulfite-treated sequencing data. *Bioinformatics* **28**, 1698-1704, doi:10.1093/bioinformatics/bts254 (2012).

111 Kingsford, C. L., Ayanbule, K. & Salzberg, S. L. Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome biology* **8**, R22, doi:10.1186/gb-2007-8-2-r22 (2007).

112 Quinlan, A. R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Current protocols in bioinformatics* **47**, 11.12.11-34, doi:10.1002/0471250953.bi1112s47 (2014).

113 Haake, S. K., Yoder, S. C., Attarian, G. & Podkaminer, K. Native plasmids of Fusobacterium nucleatum: characterization and use in development of genetic systems. *J Bacteriol* **182**, 1176-1180, doi:10.1128/jb.182.4.1176-1180.2000 (2000).

114 Knoke, L. R. *et al.* Agrobacterium tumefaciens Small Lipoprotein Atu8019 Is Involved in Selective Outer Membrane Vesicle (OMV) Docking to Bacterial Cells. *Front Microbiol* **11**, 1228, doi:10.3389/fmicb.2020.01228 (2020).

115 Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-1372, doi:10.1038/nbt.1511 (2008).

116 Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. JPred4: a protein secondary structure prediction server. *Nucleic acids research* **43**, W389-394, doi:10.1093/nar/gkv332 (2015).

117 Almagro Armenteros, J. J. *et al.* SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* **37**, 420-423, doi:10.1038/s41587-019-0036-z (2019).

118 Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms for molecular biology : AMB* **6**, 26, doi:10.1186/1748-7188-6-26 (2011).

## Author Contribution Statement

F.P., C.T. and Y.Z. conducted experiments. F.P. performed data analysis. F.P. and J.V. designed research. L.B. set up the web browser. J.V. directed research. F.P., F.F. and J.V. wrote the manuscript.

## Competing Interest Statement

The authors declare no competing interests.

## Figure Legends

**Figure 1: dRNA-seq for *F. nucleatum* subsp. *nucleatum***
**a**, Overview of analyzed growth conditions (early (E), mid-exponential (M) and early stationary (S) phase) and experimental workflow for transcriptome analysis via dRNA-seq. Genome-wide read distribution for *F. nucleatum* subsp. *nucleatum* ATCC 25586 are shown followed by the validation of the previously annotated TSS for *tnaA*. Growth data are represented by the mean (+/- standard deviation) from three biological replicates. **b**, Venn diagram showing the number of detected TSS for each class. The lower panel shows TSS classification based upon expression strength and genomic location. **c**, Analysis of promoter regions associated with detected TSSs using the MEME[45] suite identified a promoter motif in ~93% of analyzed pTSS. An extended -10 box and the -35 box that are separated by an AT-rich region are indicated. **d**, Length distribution and corresponding occurrences of all 5'UTRs associated with pTSS (black) and sTSS

(red). A consensus Shine Dalgarno sequence was predicted using MEME analysis. The average distance from the start codon is indicated.

**Figure 2: Transcriptome features of known virulence factors and prediction of small proteins in *F. nucleatum* subsp. nucleatum.**

**a**, Genomic context and transcriptome features for representative examples of known virulence-associated genes. Rho-independent terminators were predicted with ANNOgesic. **b**, Conservation analysis combined with TSS prediction uncovers the small protein candidate FspC3 in the intergenic region upstream of FN0239. Genomic alignment for representative strains of different fusobacteria harbouring a *fspC3* sequence (FNN = *F. n.* subsp. *nucleatum*; FNA = *F. n.* subsp. *animalis*; FNP = *F. n.* subsp. *polymorphum*; FNV = *F. n.* subsp. *vincentii*; FuH = *F. hwasookii*; FuP = *F. periodonticum*). For protein alignment, protein sequences from all available annotated genomes of the above fusobacteria were used. The web-application of jpred 4[116] and SignalP (5.0)[117] was used to predict the secondary structure and signal peptide respectively.

**Figure 3: Identification of core noncoding RNAs and an active CRISPR-Cas system in *F. nucleatum* subsp. *nucleatum***

**a**, Expression profiling of 6S RNA, 4.5S RNA, tmRNA and M1 RNA of RNase P across the different growth stages by northern blot. **b**, Read distribution for the 6S RNA detects an antisense TSS supporting the transcription of the pRNA. **c**, Secondary structure prediction of the putative 6S RNA shows a two-handle stem-loop structure with an internal bulge region commonly found in known 6S RNAs. The TSS associated with the identified pRNA (blue) is indicated with an arrow. **d**, Northern blot validation for the processing of the CRISPR array by probing for the overlapping region between individual spacer and repeat pairs. The results show generation of single spacer-repeat pairs of ~67nt increasing towards stationary phase, and also larger fragments indicating a more complex processing of the array. Rho-independent terminator was predicted via ANNOgesic.

**Figure 4: The small noncoding RNA landscape in *Fusobacterium* species**

**a**, Distribution of all predicted TSS and sRNAs across the Fnn genome. **b**, Shown is the conservation analysis for all validated candidates within 36 complete genomes of *F. nucleatum* strains and related species in the Fusobacterium genus. For each strain, sRNAs are considered to be conserved if they display ≧75% nucleotide identity and were not shorter than the 75% of the sRNAs length in the reference strain. An asterisk (*) marks sRNA candidate which overlap with riboswitch predictions but detection of a stable fragment by northern blot analysis indicate potential dual functions. Grey boxes indicate that the ncRNAs did not meet the required criteria but were predicted via RFAM. Purple ncRNAs were only identified when comparing genomic synteny in addition to the secondary structure of these regions. **c**, Northern blot validation of predicted sRNAs in three different growth phases. The sRNA candidates are classified into established classes of intergenic, 5'UTR or 3'UTR-derived or antisense sRNAs. Rho-independent terminators were predicted by ANNOgesic.

**Figure 5: FoxI is an oxygen-induced conserved sRNA**

**a**, Comparison of genome synteny for FoxI in *F. nucleatum* susbp. *nucleatum* (FNN), subsp. *polymorphum* (FNP), subsp. *vincentii* (FNV), subsp. *animalis* (FNA) as well as *Fusobacterium periodonticum* (FuP) and *F. hwasookii* (FuH) reveals the conserved

genomic localization of the sRNA between the tryptophanase subunit B (*trpB*) and the ribosomal small subunit methyltransferase B (*rsmB*). The *trpB* gene is absent in all FNV strains. **b**, Expression profiling of FoxI by Northern blot in three different growth phases (E – early exponential, M – mid exponential, S - stationary). **c**, Alignment of FoxI and its upstream region highlights the strong sequence conservation for the sRNA as well as the promoter region. A predicted terminator loop is indicated. **d**, Predicted secondary structure of FoxI using RNAfold[118] . The nucleotide exchanges generating FoxI-3C are shown. **e**, Expression profiling of FoxI by northern blot analysis under different stress conditions shows FoxI to be highly upregulated upon 20 min. of exposure to oxygen.

**Figure 6: Overexpression of FoxI identifies the OMP FomA as a potential target of the sRNA**

**a**, Schematic representation of the shuttle vector pEcoFus used for sRNA overexpression. The vector is derived from pRPF185[78] and carries a plasmid origin of replication functioning in *F. nucleatum* (OriFn). Expression of the sRNA is driven by the 4.5S RNA promoter ($P_{4.5S\ RNA}$). Northern blot validation for the overexpression of FoxI and four additional sRNAs compared to the wildtype (WT) and the empty vector control (ctrl.) which carries a short nonsense stretch after the promoter. From top to bottom, the panel shows probing for: FoxI, FunR7, FunR12, FunR16 and FunR47. **b**, Denaturing SDS page and western blot analysis of subcellular fractionation of *Fnn* carrying the empty vector control, the overexpression vector for FoxI or FoxI-3C. Equal OD units for the total protein and cytosolic fraction were collected from mid-exponential phase cultures and 10x the amount was loaded for the inner membrane (IM) and outer membrane (OM) fraction to ensure proper visualization after coomassie staining. Western blot analysis was carried out loading half the amounts prior to detection of FomA using an anti-FomA antibody. **c**, Results of LC-MS/MS analysis comparing the ratios for label free quantification (LFQ) of the overexpression of FoxI or FoxI-3C to the empty vector control. Shown are only the proteins detected across all samples in the size range of 35-55 kDa. **d**, Base-pairing between the 5'region of fomA and the linear region FoxI as predicted by IntaRNA . The start codon of fomA (red) as well as the mutated region in FoxI-3C (grey) are indicated. **e**, Alignment of the 5'UTR and proximal coding region of fomA from representative strains of the different fusobacterial strains (FNN = *F. n.* subsp. *nucleatum*; FNA = *F. n.* subsp. *animalis*; FNP = *F. n.* subsp. *polymorphum*; FNV = *F. n.* subsp. *vincentii*; FuH = *F. hwasookii*; FuP = *F. periodonticum*). The predicted binding site of FoxI as well as the TSS and AUG start codon are indicated.