Perturbed human sub-networks by \textit{Fusobacterium nucleatum} candidate virulence proteins

Andreas Zanzoni$^{1,*}$, Lionel Spinelli$^1$, Shérazade Braham$^1$ and Christine Brun$^{1,2}$

**Abstract**

**Background:** \textit{Fusobacterium nucleatum} is a gram-negative anaerobic species residing in the oral cavity and implicated in several inflammatory processes in the human body. Although \textit{F. nucleatum} abundance is increased in inflammatory bowel disease subjects and is prevalent in colorectal cancer patients, the causal role of the bacterium in gastrointestinal disorders and the mechanistic details of host cell functions subversion are not fully understood.

**Results:** We devised a computational strategy to identify putative secreted \textit{F. nucleatum} proteins (\textit{Fuso}Secretome) and to infer their interactions with human proteins based on the presence of host molecular mimicry elements. \textit{Fuso}Secretome proteins share similar features with known bacterial virulence factors thereby highlighting their pathogenic potential. We show that they interact with human proteins that participate in infection-related cellular processes and localize in established cellular districts of the host–pathogen interface. Our network-based analysis identified 31 functional modules in the human interactome preferentially targeted by 138 \textit{Fuso}Secretome proteins, among which we selected 26 as main candidate virulence proteins, representing both putative and known virulence proteins. Finally, six of the preferentially targeted functional modules are implicated in the onset and progression of inflammatory bowel diseases and colorectal cancer.

**Conclusions:** Overall, our computational analysis identified candidate virulence proteins potentially involved in the \textit{F. nucleatum}—human cross-talk in the context of gastrointestinal diseases.

**Keywords:** Fusobacterium nucleatum, Secretome, Molecular mimicry, Short linear motifs, Bioinformatics, Interaction network, Colorectal cancer, Inflammatory bowel diseases, Virulence proteins

**Background**

\textit{Fusobacterium nucleatum} is a gram-negative anaerobic bacterium best known as a component of the oral plaque and a key pathogen in gingivitis and periodontitis [1]. It has also been isolated in several inflammatory processes in distinct body sites (e.g., endocarditis, septic arthritis, liver and brain abscesses) and implicated in adverse pregnancy outcomes (reviewed in [2]). Moreover, it has been demonstrated that \textit{F. nucleatum} can adhere to and invade a variety of cell types, thereby inducing a pro-inflammatory response [3–8]. Recent work showed that (i) \textit{F. nucleatum} is prevalent in colorectal cancer (CRC) patients [9–11] and (ii) its abundance is increased in new-onset Crohn’s disease (CD) subjects [12]. Interestingly, follow-up studies suggested a potential role of this bacterium in CRC tumorigenesis and tumor-immune evasion [13–16].

Despite these findings, a large fraction of \textit{F. nucleatum} gene products are still uncharacterized. Moreover, to date, only a handful of pathogenic factors has been experimentally identified [17, 18] and protein interaction data between these factors and human proteins, which could inform on the molecular details underlying host-cell subversion mechanisms, are sparse [4, 16, 19]. Altogether, this underlines that a comprehensive view of the molecular details of the \textit{F. nucleatum}—human cross-talk is currently missing.

How could \textit{F. nucleatum} hijack human cells? Pathogens employ a variety of molecular strategies to reach an advantageous niche for survival. One of them consists of

---

$^*$ Correspondence: andreas.zanzoni@univ-amu.fr

$^1$Aix-Marseille Université, Inserm, TAGC UMR_S1090, Marseille, France

Full list of author information is available at the end of the article

© The Author(s). 2017 Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
subverting host protein interaction networks. Indeed, they secrete and deliver factors such as toxic compounds, small peptides, and even proteins to target the host molecular networks. To achieve this, virulence factors often display structures resembling host components in form and function [20–22] to interact with host proteins, thus providing a benefit to the pathogen [23]. Such “molecular mimics” (e.g., targeting motifs, enzymatic activities, and protein–protein interaction elements) allow pathogens to enter the host cell and perturb cell pathways (e.g., [24–26]).

Over the years, several experimental approaches have been applied to identify protein–protein interactions (PPIs) between pathogens and their hosts providing new insights on the pathogen’s molecular invasion strategies. However, the vast majority of these systematic studies focused on viruses (e.g., [27–29]) and, to a lesser extent, on bacteria [30–33] and eukaryotic parasites [33, 34]. Indeed, as cellular pathogens have large genomes and complex life cycles, the experimental identification of virulence proteins and the large-scale mapping of host-pathogen PPIs require a lot of effort and time [35, 36]. In this context, computational approaches have proved to be instrumental for the identification of putative pathogenic proteins (e.g., [37, 38]), the characterization of molecular mimics [23, 39, 40], and the inference of their interactions with host proteins (for a review see [41]).

Here, in order to gain new insights on the molecular cross-talk between \( F. \) *nucleatum* and the human host, we devised a computational strategy combining secretion prediction, protein–protein interaction inference, and protein interaction network analyses (Fig. 1). Doing so, we defined a secretome of the bacterium and the human proteins with which they interact based on the presence of mimicry elements. We identified the host cellular pathways that are likely perturbed by \( F. \) *nucleatum* including immune and infection response, homeostasis, cytoskeleton organization, and gene expression regulation. Interestingly, our results identify candidate virulence proteins, including the established Fap2 adhesin, and provide new insights underlying the putative causative role of \( F. \) *nucleatum* in colorectal cancer and inflammatory bowel diseases.

### Results

**Prediction of \( F. \) *nucleatum* secreted proteome**

Previous computational analyses highlighted that \( F. \) *nucleatum* has a reduced repertoire of secretion machinery [42, 43] meaning that it might exploit alternative “non-classical” translocation mechanisms to unleash virulence proteins. Thus, we sought to identify putative \( F. \) *nucleatum* secreted proteins by analyzing the 2046 protein sequences of the type species \( F. \) *nucleatum* subsp. *nucleatum* (strain ATCC 25586) proteome using

![Flow strategy of our computational approach](image-url)
two distinct algorithms: SignalP [44] for peptide-triggered secretion and SecretomeP [45] for leaderless protein secretion. While the SignalP algorithm predicted 61 F. nucleatum sequences being secreted via classical/regular secretion pathways, SecretomeP found 176 proteins as possibly secreted through non-classical routes. In total, we identified 237 putative secreted proteins in the F. nucleatum proteome (herein called “Fusosecretome”) (see Additional file 1: Table S1). Notably, we were able to correctly predict as secreted all the F. nucleatum virulence proteins known so far, namely FadA (FN0264), Fap2 (FN1449), RadD (FN1526), and the recently identified Aid1 adhesin (FN1253) [46]. This result underlines the relevance of secretion prediction to identify novel putative virulence proteins in the F. nucleatum proteome.

It has been shown that disorder propensity is an emerging hallmark of pathogenicity [47, 48]. As SecretomeP exploits protein disorder as a predicting feature, we analyzed the intrinsic disorder content of the Fusosecretome proteins identified by the SignalP algorithm only. We indeed observed a significantly higher disorder propensity of these proteins compared to the non-secreted proteins ($P$ value $= 1.9 \times 10^{-4}$, Kolgomorov–Smirnov test, two-sided) (Fig. 2; Additional file 2: Figure S1; Additional file 3: Table S2), further reinforcing the possible role of the Fusosecretome in the infection/invasion process.

To detect functional elements that can further contribute to F. nucleatum pathogenicity, we sought for the presence of globular domains in the Fusosecretome. We observed an enrichment of domains mainly belonging to the outer membrane beta-barrel protein superfamily (Table 1). Six out of the eight over-represented domains among the Fusosecretome proteins are also found in known virulence proteins of gram-negative bacteria [49] and are involved in adhesion, secretion, transport, and invasion.

Altogether, these findings suggest that Fusosecretome proteins display features of known virulence proteins and can likely be involved in the cross-talk with the human host.

**Inference of the Fusosecretome—human interaction network**

Generally, pathogens employ a variety of molecular strategies to interfere with host-cell networks, controlling key functions such as plasma membrane and cytoskeleton dynamics, immune response, and cell death/survival. In particular, their proteins often carry a range of mimics, which resemble structures of the host at the molecular level, to “sneak” into host cells [20–22, 50].

Here, we focused on putative molecular mimicry events that can mediate the interaction with host proteins: (i) globular domains that occur in both Fusosecretome and the human proteome and (ii) known eukaryotic short linear motifs (SLiMs) found in Fusosecretome proteins. SLiMs are short stretches of 3–10 contiguous amino acids residues that often mediate transient PPIs and tend to bind with low affinity [51].

We first scanned the sequences of the Fusosecretome and human proteins for the presence of domains as defined by Pfam [52]. We identified 55 “host-like” domains in 50 Fusosecretome proteins out of 237, including several domains related to ribosomal proteins, aminopeptidases, and tetratricopeptide repeats (TPR) (Additional file 4: Table S3). Interestingly, 29 of these domains are also found in known bacterial binders of human proteins [30].

We next detected the occurrence of experimentally identified SLiMs gathered from the Eukaryotic Linear Motif (ELM) database [53]. As linear motifs are short and degenerate in sequence, SLiM detection is prone to over-prediction [54]. To reduce the number of false positives, we kept occurrences falling in conserved and disordered protein sequences (see the “Methods” section). Indeed, known functional SLiMs show a higher degree of conservation compared to surrounding residues [51] and are located in unstructured regions [55, 56]. In this way, we identified at least one putative mimicry SLiM in 139 Fusosecretome proteins. Most of the 57 different detected SLiMs represents binding sites such as motifs recognized by PDZ, SH3, and SH2 domains (Additional file 4: Table S3).

We exploited these putative mimicry events to infer the interaction with human proteins by using templates of domain–domain and SLiM–domain interactions (see the “Methods” section for further details). Doing so, we obtained 3744 interactions (1544 domain- and 2201 SLiM-mediated interactions, respectively) between 144 Fusosecretome, which we designated as “candidate virulence proteins,” and 934 human proteins (Additional file 5: Table S4 and Additional file 6: Table S5) designated as “human inferred interactors.”
In order to assess the reliability of the inferences, we evaluated the biological relevance of the putative human interactors by performing enrichment analyses of different orthogonal datasets using as a reference background all the proteins encoded by the human genome.

First, human proteins experimentally identified as binders or targets of bacterial and viral proteins are over-represented among the 934 inferred human interactors of the FusoSecretome proteins (415 proteins, 1.3-fold, \( P \text{ value} = 1.61 \times 10^{-11} \)). Notably, the over-representation holds when bacterial and viral binders are considered separately (176 bacterial interactors, 1.1-fold, \( P \text{ value} = 3.5 \times 10^{-3} \) and 338 viral interactors, 1.5-fold, \( P \text{ value} < 2.2 \times 10^{-16} \)). This result is consistent with current knowledge on convergent targeting of host proteins by distinct pathogens [30, 33, 57, 58].

Second, according to the Human Proteins Atlas (see the “Methods” section), the vast majority of the inferred human interactors has been detected either in small intestine (652, 70%) or colorectal (671 proteins, 72%) tissues as well as in the saliva (673, 72%), confirming their presence in human body sites hosting F. nucleatum. Third, we assessed whether the inferred human interactors are implicated in gastrointestinal disorders by seeking for an over-representation of genes associated to such diseases (see the “Methods” section). Indeed, the human interactors of the FusoSecretome are enriched in (i) proteins identified in the human colon secretomes of colorectal cancer (CRC) tissue samples (3.5-fold, \( P \text{ value} < 2.2 \times 10^{-16} \)), (ii) proteins encoded by genes whose expression correlates with F. nucleatum abundance in CRC patients [13] (twofold, \( P \text{ value} = 4 \times 10^{-4} \)), and (iii) genes associated with inflammatory bowel diseases (IBDs) (twofold, \( P \text{ value} = 8 \times 10^{-4} \)). We obtained very similar enrichments by using a reduced statistical background corresponding to the interaction inference space (see the “Methods” section and Additional file 7: Supplementary Results).

Altogether, the results of these analyses highlight the relevance of the inferred human interactors as putative binders of FusoSecretome proteins and their potential implication in gut diseases, therefore validating the undertaken inference approach.

### Functional role of the human proteins targeted by F. nucleatum

Globally, the inferred FusoSecretome human interactors are involved in several processes related to pathogen infection...
such as immune response and inflammation, response to stress, endocytosis as shown by the 137 significantly enriched Biological Processes Gene Ontology (GO) terms among their annotations (Table 2, Additional file 8: Table S6A). Similarly, the targeted human proteins are over-represented in 125 pathways (54 from KEGG and 71 from Reactome databases [59, 60]) involved in cell adhesion and signaling, extracellular matrix remodeling, immunity, response to infection, and cancer-related pathways (Table 2, Additional file 8: Table S6A). These human proteins are mainly localized in the extracellular space, plasma membrane, and at cell-cell junctions that represent the main districts involved in the initial encounter between a pathogen and the host, as indicated by the over-representation of 30 Cellular Component GO terms (Table 2, Additional file 8: Table S6A). A substantial fraction of these enriched functional categories is significantly over-represented when using the reduced statistical background as well (see the “Methods” section and Additional file 8: Table S6B). Overall, this indicates that our inferred interactions can participate in the F. nucleatum—human cross-talk.

**F. nucleatum targets topologically important proteins in the host network**

To gain a broader picture of the inferred interactions in the cellular context, we mapped the FusosSecretome human interactors on a binary human interactome built by gathering protein interactions data from both small-scale experiments and systematic screens reported in the literature (see the “Methods” section and Additional file 9: Table S7). Around 70% of the inferred human interactors (i.e., 663 proteins) are present in the human binary interactome. Interestingly, the human targeted proteins occupy topologically important positions in the interactome as shown by their significantly higher number of interactions and higher values of betweenness centrality compared to other network proteins (number of interactions: mean = 23 vs. 11, \( P \) value = \( 1.9 \times 10^{-12} \); betweenness centrality: mean = 0.00078 vs. 0.00018, \( P \) value = \( 6.2 \times 10^{-12} \); two-sided Mann–Whitney test) (Fig. 3).

The human interactome is composed of functional network modules, defined as groups of proteins densely connected through their interactions and involved in the same biological process [61] (see the “Methods” section). We thus next investigated the 855 functional modules that we previously detected [62] using the OCG algorithm that decomposes a network into overlapping modules, based on modularity optimization [63] (Additional file 10: Table S8). A significant number of interactors participate in 2 or more of these functional units (259 proteins, 1.3-fold enrichment, \( P \) value = \( 1.4 \times 10^{-7} \)), indicating that the FusosSecretome tends to target multifunctional proteins in the human interactome [63]. Moreover, among the multifunctional inferred human interactors, we found an enrichment of extreme multifunctional proteins (52 interactors, twofold enrichment, \( P \) value = \( 1.0 \times 10^{-5} \)), which are defined as proteins involved in unrelated cellular functions and may represent candidate moonlighting proteins [64]. This suggests that the FusosSecretome might perturb multiple cellular pathways simultaneously by targeting preferentially a whole range of multifunctional proteins.

**Functional subnetworks of the human interactome perturbed by F. nucleatum and identification of the main candidate virulence proteins**

Based on their enrichment in inferred human interactors, 31 network modules (~4% of the 855 detected modules) are preferentially targeted by 138 distinct proteins of the FusosSecretome (Table 3). Targeted modules are involved in relevant processes such as immune response, cytoskeleton organization, cancer, and infection-related pathways (Table 3 and Additional file 11: Table S9). Moreover, proteins belonging to these modules are mainly localized in the extracellular space or in membranous structures (Table 3 and Additional file 11: Table S9), which represent important districts of the microbe-host interface. Interestingly, the enrichment of functional categories related to gene expression regulation (Additional file 11: Table S9) in several modules suggests novel potential host subversion mechanisms by *F. nucleatum*.

These modules are targeted on average by 50 FusosSecretome proteins (ranging from 2 to 104 per module) and the number of inferred host–pathogen interactions for each module varies considerably (Table 3). What are the main network perturbators among the FusosSecretome proteins? To quantify their impact on network modules based on the number of interactions, they have with each of them, we computed a Z score (see the “Methods” section, Additional file 12: Table S10). We considered the 26 FusosSecretome proteins having a perturbation Z score >2 in at least one module as main candidate virulence proteins. They consist in outer membrane proteins, enzymes, iron-binding proteins, and protein involved in transport (Table 4). Ten of them (38%) can perturb at least two distinct modules (Fig. 4a). Notably, we identified among the candidates, the known virulence protein Fap2 (FN1449) (Fig. 4b) that targets 4 modules, and a protein containing the MORN_2 domain (FN2118) (Fig. 4c) recently identified as a key element in actively invading *F. nucleatum* species [65], which perturbs 6 modules. On the other hand, 25 preferentially targeted modules are perturbed by at least two candidate virulence proteins, Module 78 involved in immune response being the most potentially subverted (Fig. 4a).

**F. nucleatum and gut diseases from a network perspective**

Among the 855 network modules detected in the human interactome, 38 are enriched in genes involved in at least
### Table 2 Significant Gene Ontology and pathways annotations among FusoSecretome inferred human interactors

| Annotation source | Annotation ID | Annotation name                                                                 | Corrected P-value         |
|-------------------|---------------|----------------------------------------------------------------------------------|----------------------------|
| **Biological process** |               |                                                                                  |                            |
| GO:0006415        | Translational termination | 7.57 × 10^{-41}                                                                  |                            |
| GO:0006414        | Translational elongation   | 2.65 × 10^{-34}                                                                  |                            |
| GO:0006457        | Protein folding           | 2.51 × 10^{-29}                                                                  |                            |
| GO:0006413        | Translational initiation  | 6.63 × 10^{-27}                                                                  |                            |
| GO:0072376        | Protein activation cascade| 1.17 × 10^{-25}                                                                  |                            |
| GO:0051604        | Protein maturation        | 5.45 × 10^{-23}                                                                  |                            |
| GO:0006614        | SRP-dependent cotranslational protein targeting to membrane | 1.93 × 10^{-22} |                            |
| GO:0006956        | Complement activation     | 3.62 × 10^{-21}                                                                  |                            |
| GO:0002697        | Regulation of immune effector process | 4.03 × 10^{-18} |                            |
| GO:0030449        | Regulation of complement activation | 4.10 × 10^{-18} |                            |
| **Cellular component** |               |                                                                                  |                            |
| GO:0005840        | Ribosome                 | 9.81 × 10^{-37}                                                                  |                            |
| GO:0022626        | Cytosolic ribosome        | 2.76 × 10^{-33}                                                                  |                            |
| GO:0005912        | Adherens junction         | 4.15 × 10^{-22}                                                                  |                            |
| GO:0098552        | Side of membrane         | 2.46 × 10^{-18}                                                                  |                            |
| GO:0030055        | Cell-substrate junction   | 2.85 × 10^{-17}                                                                  |                            |
| GO:0072562        | Blood microparticle       | 9.20 × 10^{-12}                                                                  |                            |
| GO:0005761        | Mitochondrial ribosome    | 4.28 × 10^{-12}                                                                  |                            |
| GO:0019887        | Extrinsic component of plasma membrane | 2.35 × 10^{-09} |                            |
| GO:0005911        | Cell-cell junction        | 4.54 × 10^{-08}                                                                  |                            |
| GO:0031012        | Extracellular matrix      | 3.91 × 10^{-07}                                                                  |                            |
| **KEGG**          |               |                                                                                  |                            |
| KEGG03010         | Ribosome                 | 4.34 × 10^{-39}                                                                  |                            |
| KEGG04610         | Complement and coagulation cascades | 2.28 × 10^{-32} |                            |
| KEGG04514         | Cell adhesion molecules (CAMs) | 2.88 × 10^{-30} |                            |
| KEGG04141         | Protein processing in endoplasmic reticulum | 2.03 × 10^{-28} |                            |
| KEGG04660         | T cell receptor signaling pathway | 1.33 × 10^{-27} |                            |
| KEGG05150         | Staphylococcus aureus infection | 1.78 × 10^{-26} |                            |
| KEGG04380         | Osteoclast differentiation | 7.25 × 10^{-26}                                                                  |                            |
| KEGG05203         | Viral carcinogenesis      | 8.21 × 10^{-26}                                                                  |                            |
| KEGG05169         | Epstein-Barr virus infection | 1.11 × 10^{-25} |                            |
| KEGG05164         | Influenza A               | 1.77 × 10^{-25}                                                                  |                            |
| **Reactome**      |               |                                                                                  |                            |
| REAC:1,592,389    | Activation of Matrix Metalloproteinases | 4.77 × 10^{-28} |                            |
| REAC:192,823      | Viral mRNA Translation    | 2.14 × 10^{-19}                                                                  |                            |
| REAC:156,902      | Peptide chain elongation  | 2.14 × 10^{-19}                                                                  |                            |
| REAC:975,956      | Nonsense Mediated Decay independent of the Exon Junction Complex  | 1.83 × 10^{-17} |                            |
| REAC:977,606      | Regulation of Complement cascade | 5.13 × 10^{-16} |                            |
| REAC:202,733      | Cell surface interactions at the vascular wall | 7.22 × 10^{-16} |                            |
| REAC:5,368,287    | Mitochondrial translation | 3.61 × 10^{-14}                                                                  |                            |
| REAC3,371,453     | Regulation of HSF1-mediated heat shock response | 3.92 × 10^{-14} |                            |
| REAC:3,371,599    | Defective HLCS causes multiple carboxylase deficiency | 1.33 × 10^{-08} |                            |
| REAC4,320,597     | Nectin/Necl trans heterodimerization | 1.33 × 10^{-08} |                            |

For each annotation source, the ten most significant terms are reported. The full list of annotation enrichments is available in Additional file 8: Table S6.
one gut disease (i.e., CRC and IBDs, see the “Methods” section). Interestingly, 27 of them (i.e., 71%) are targeted by at least one FusoSecretome protein, among which 3 contain a statistically significant fraction of inferred human interactors (Fig. 5). Notably, Module 78, involved in immune response, is enriched in genes associated to inflammatory bowel diseases (IBDs) (28 proteins, 5.2-fold enrichment, $P = 4.78 \times 10^{-4}$) as well as in CD-specific (9 proteins, 13.4-fold enrichment, $P = 2.52 \times 10^{-3}$) and CRC-mutated (11 proteins, fourfold enrichment, $P = 1.46 \times 10^{-2}$) genes. Moreover, it is enriched in genes whose expression correlates with $F. nucleatum$ abundance in CRC patients (24 proteins, 3.7-fold enrichment, $P = 3.35 \times 10^{-4}$). This module is targeted by several main candidate virulence proteins, including a hemolysin (FN0291), an outer membrane protein (FN1554) and the MORN_2 domain containing protein (FN21118) (Fig. 4a), which therefore, may play critical roles in these diseases. IBD genes are also enriched in Module 702 (5 proteins, 11-fold enrichment, $P = 2.13 \times 10^{-2}$), whose proteins participate in Jak-STAT signaling, whereas CD-specific genes are over-represented in Module 9 implicated in immunity (5 proteins, 28-fold enrichment, $P = 2.52 \times 10^{-3}$). Interestingly, Module 9 is specifically perturbed by Fap2 (FN1449) (Fig. 4b), which is known to modulate the host immune response.

Three other modules enriched in inferred human interactors show a significant dysregulation of the expression of their constituent proteins during CRC progression [66] and are implicated in infection-response pathways and cytoskeleton organization (Fig. 5). In particular, two of these modules (Modules 138 and 216) show significant and specific upregulation in stage II, whereas the third (Module 371) is significantly upregulated in normal and stage II samples. Overall, these results indicate that $F. nucleatum$ could contribute to the onset and progression of IBDs and CRC by perturbing some of the underlying network modules.

Comparison with additional bacterial strains

We applied our computational approach on the recently released proteomes of 6 actively invading Fusobacteria strains isolated from biopsy tissues [8, 65] (i.e., 4 $F. nucleatum$ subspecies and 2 $F. periodonticum$ strains), and the proteome of $E. coli$ K-12 as a “control strain” (see Additional file 7: Supplementary Results, Table S12). We found that the secretomes of these 7 bacteria share common features (i.e., disorder propensity, enriched domains, host-like domain and mimicry SLiM content) with the FusoSecretome (Additional file 7: Table S13–S15 and Figure. S2–S8). However, we observed a moderate overlap in terms of inferred interactors, enriched functions and preferentially targeted network modules (Additional File 7: Table S16–S18), and a modest concordance in term of network module perturbators (Additional File 7: Table S19).

The results of these analyses suggest that, on the one hand, actively invading Fusobacteria species share common mechanisms to interact with host cell and, on the other hand, are consistent with the fact that $F. nucleatum$ is an unusual heterogeneous species both at the genotypic and phenotypic level [8, 65, 67]. Finally, the commonalities between the FusoSecretome and the $E. coli$ K-12 secreted proteins are not surprising, since previous work showed that $E. coli$ K-12 carries cryptic genes coding for virulence factors [68], whose expression is activated by mutations in the histone-like protein HU,
| Module | Module proteins | Interactors | Inferred interactions | Fusotecreteome proteins | Annotations |
|--------|-----------------|-------------|----------------------|------------------------|-------------|
| 9      | 74              | 13          | 188                  | 64                     | Immune response-regulating cell surface receptor signaling pathway (GO:0002768), cell-cell junction (GO:0005911) |
| 16     | 67              | 12          | 122                  | 65                     | Metal ion homeostasis (GO:0055065), cell surface (GO:0009986) |
| 19     | 80              | 15          | 260                  | 55                     | Cellular response to organonitrogen compound (GO:0071417), membrane raft (GO:0045121) |
| 42     | 140             | 19          | 312                  | 56                     | Endocytosis (GO:0006897), membrane raft (GO:0045121) |
| 74     | 113             | 26          | 78                   | 26                     | Extracellular structure organization (GO:0043062), cell surface (GO:0009986) |
| 78     | 292             | 54          | 752                  | 100                    | Immune response-regulating cell surface receptor signaling pathway (GO:0002768), cell surface (GO:0009986) |
| 89     | 45              | 10          | 169                  | 48                     | Immune response-activating cell surface receptor signaling pathway (GO:0002429), nucleolar ribonuclease P complex (GO:0005655) |
| 90     | 51              | 10          | 28                   | 12                     | I-kappaB kinase/NF-kappaB cascade (GO:0007249), inclusion body (GO:0016234) |
| 138    | 126             | 22          | 57                   | 26                     | I-kappaB kinase/NF-kappaB cascade (GO:0007249), perinuclear region of cytoplasm (GO:0048471) |
| 165    | 81              | 17          | 327                  | 71                     | Neuron projection guidance (GO:0097485), synapase (GO:0045202) |
| 194    | 38              | 12          | 119                  | 48                     | G1/S transition of mitotic cell cycle (GO:0000082), cyclin-dependent protein kinase holoenzyme complex (GO:0000307) |
| 216    | 47              | 11          | 261                  | 80                     | Blood coagulation (GO:0007596), membrane raft (GO:0045121) |
| 246    | 50              | 15          | 83                   | 32                     | T cell activation (GO:0042110), Golgi membrane (GO:0000139) |
| 277    | 25              | 7           | 7                    | 2                      | Collagen catabolic process (GO:0030574), extracellular matrix (GO:0031012) |
| 298    | 106             | 16          | 242                  | 75                     | Actin cytoskeleton organization (GO:0030036), Arp2/3 protein complex (GO:0005885) |
| 300    | 37              | 9           | 164                  | 49                     | Stress-activated MAPK cascade (GO:0015103), nuclear speck (GO:0016607) |
| 371    | 25              | 7           | 79                   | 48                     | Actin filament organization (GO:0007015), lamellipodium (GO:0030027) |
| 433    | 38              | 9           | 121                  | 49                     | Positive regulation of intracellular protein kinase cascade (GO:0010740), spindle (GO:0005819) |
| 451    | 40              | 9           | 55                   | 34                     | Mitotic cell cycle phase transition (GO:0044772), heterochromatin (GO:0000792) |
| 456    | 36              | 9           | 142                  | 59                     | Regulation of system process (GO:0044057), dendrite (GO:0030425) |
| 563    | 26              | 8           | 43                   | 29                     | Regulation of sequence-specific DNA binding transcription factor activity (GO:00051090), external side of plasma membrane (GO:0009987) |
| 571    | 42              | 11          | 206                  | 53                     | Cell cycle phase transition (GO:0044770), transcription factor complex (GO:0005667) |
| 577    | 33              | 10          | 66                   | 48                     | Complement activation (GO:0006956), ER membrane insertion complex (GO:0072379) |
| 587    | 17              | 6           | 109                  | 46                     | Axonogenesis (GO:0007409), signalosome (GO:0008180) |
| 615    | 36              | 11          | 26                   | 10                     | Response to unfolded protein (GO:0006986), perinuclear region of cytoplasm (GO:0048471) |
| 625    | 25              | 8           | 226                  | 104                    | Regulation of sequence-specific DNA binding transcription factor activity (GO:00051090), chromatin (GO:0000785) |
| 689    | 38              | 14          | 151                  | 72                     | Blood coagulation (GO:0007596), apical junction complex (GO:0043296) |
| 702    | 23              | 8           | 157                  | 46                     | Peptidyl-tyrosine phosphorylation (GO:0018108), nucleolar ribonuclease P complex (GO:0005655) |
| 745    | 18              | 6           | 113                  | 46                     | Axon guidance (GO:0007411), cell leading edge (GO:0031252) |
| 794    | 22              | 7           | 151                  | 46                     | Gamma-aminobutyric acid signaling pathway (GO:0007214), postsynaptic membrane (GO:0045211) |
| 831    | 15              | 6           | 129                  | 45                     | Fc-gamma receptor signaling pathway involved in phagocytosis (GO:0038096), cell leading edge (GO:0031252) |

For each module the following information is reported: identifier, number of constituent proteins, number of inferred human interactors in the module, number of inferred interactions between proteins in the module and Fusotecreteome proteins, number of interacting Fusotecreteome proteins, representative annotations (Biological Process and Cellular Component) selected as the most frequent and significantly enriched annotations for the given module (for the complete list of functional annotations see Additional file 11: Table S9)
### Table 4 List of the main candidate virulence proteins in the Fusosecretome

| UniprotKB AC | Protein name | Gene symbol | Domains | Interacting domains | Interacting SLiMs |
|--------------|--------------|-------------|---------|---------------------|-------------------|
| Q8RIM1       | Fusobacterium outer membrane protein family | FN1554      | Autotransporter\(^a\) | – | LIG_FHA\(_1\), LIG_FHA\(_2\), LIG_PP\(_1\), LIG_SH2\(_2,\_SRC\), LIG_SH2\(_2,\_STATS\), LIG_SH2\(_3\), LIG_SUMO\_SBM\(_1\), MOD_N\_GLC\(_1\), TRG\_ENDOCYTIC\(_2\) |
| Q8RGK2       | Hemolysin    | FN0291      | Fil\_haemagg\(_2\)\(^b\) | – | LIG_FHA\(_1\), LIG_FHA\(_2\), LIG_Rb\_pABgroove\(_1\), LIG_SH2\_GRB2, LIG_SH2\_SRC, LIG_SH2\_STATS, LIG_SUMO\_SBM\(_1\), MOD_CK1\(_1\), MOD_CK2\(_1\), MOD_GSK3\(_1\), MOD_PKA\(_1\), MOD_PKA\(_2\), TRG\_ENDOCYTIC\(_2\) |
| Q8RG79       | Peptide methionine sulfoxide reductase MsrA | msrA        | PMSR, SelR | PMSR, SelR | CLV_PCSK\_PC1ET2\(_1\), LIG_FHA\(_1\), LIG_SH2\_GRB2, LIG_SH2\_SRC, LIG_SH2\_STATS, MOD_Cter\_Amidation, MOD_PKA\(_1\), MOD_PKA\(_2\), TRG\_ENDOCYTIC\(_2\) |
| Q8RHB9       | Hypothetical exported 24-amino acid repeat protein | FN2118      | MORN\(_2\) | – | LIG_SH2\_GRB2, LIG_SH2\_SRC |
| Q8R609       | Pyruvate-flavodoxin oxidoreductase | FN1421      | POR\(_N,\_POR\), EKR, Fer4\(_7\), TPP\_enzyme\(_C\) | | LIG_CYCLIN\(_1\), LIG_SH2\_GRB2, LIG_SH2\_STATS, LIG_SH2\(_3\), LIG_SUMO\_SBM\(_1\), LIG_WW\_Pin1\(_4\), MOD_CK1\(_1\), MOD_CK2\(_1\), MOD_GSK3\(_1\), MOD_PKA\(_1\), MOD_PKA\(_2\), TRG\_ENDOCYTIC\(_2\) |
| Q8RH03       | Chaperone protein DnaJ | dnaJ        | DnaJ, DnaJ\_CXXC\_CXXG, CTD\_II | | CLV_NDR\_ND1\(_1\), CLV_PCSK\_SK1\(_1\), LIG_CYCLIN\(_1\), LIG_FHA\(_2\), LIG_SH2\_STATS, LIG_SH2\(_3\), LIG_SUMO\_SBM\(_1\), LIG_WW\_Pin1\(_4\), MOD_CK1\(_1\), MOD_CK2\(_1\), MOD_GSK3\(_1\), MOD_PKA\(_1\), MOD_PKA\(_2\), TRG\_ENDOCYTIC\(_2\), MOD_PKA\(_2\) |
| Q8R643       | Pyruvate-flavodoxin oxidoreductase | FN1170      | POR\(_N,\_POR\), EKR, Fer4\(_7\), TPP\_enzyme\(_C\) | | LIG_BRCT\_BCRA1\(_1\), LIG_SH2\_GRB2, LIG_SH2\_STATS, LIG_SH2\(_3\), LIG_SUMO\_SBM\(_1\), LIG_WW\_Pin1\(_4\), MOD_CK1\(_1\), MOD_CK2\(_1\), MOD_GSK3\(_1\), MOD_PKA\(_1\), MOD_PKA\(_2\), TRG\_ENDOCYTIC\(_2\) |
| Q8R0D9       | Fusobacterium outer membrane protein family | FN1449      | – | – | CLV_PCSK\_SK1\(_1\), LIG_FHA\(_2\), LIG_FHA\(_2,\_Class\_2\), LIG_SH2\_SRC, LIG_SH2\_STATS, LIG_SUMO\_SBM\(_1\), MOD_GSK3\(_1\), MOD_GSK3\(_1,\_2\), MOD_PKA\(_1\), MOD_PKA\(_2\) |
| Q8R08       | Serine protease | FN1426      | Peptidase\(_S8\), Autotransporter\(^a\) | Peptidase\(_S8\) | LIG_FHA\(_2\), LIG_FHA\(_2,\_Class\_2\), LIG_SH2\_STATS, LIG_SH2\(_3\), LIG_SUMO\_SBM\(_1\), MOD_CK1\(_1\), MOD_CK2\(_1,\_2\), MOD_GSK3\(_1\), MOD_PKA\(_1\), MOD_PKA\(_2\), TRG\_ENDOCYTIC\(_2\) |
| Q8RFV3       | Hypothetical cytosolic protein | FN0579      | MG1, A2M\(_N\), A2M\(_N\), A2M\(_N\), A2M\(_N\), A2M\(_N\), A2M\(_N\) | | LIG_CYCLIN\(_1\), LIG_FHA\(_2\), LIG_SH2\_STATS, LIG_SH2\(_3\), LIG_SUMO\_SBM\(_1\), MOD_CK1\(_1\), MOD_CK2\(_1\), MOD_GSK3\(_1\), MOD_PKA\(_1\), MOD_PKA\(_2\) |
| Q8R5P1       | DNAse I homologous protein DHP2 | FN0891      | Exo\_endo\_phos | Exo\_endo\_phos | LIG_SH2\_GRB2, TRG\_ENDOCYTIC\(_2\) |
| Q8R5Y8       | Biotin carboxyl carrier protein of glutamate- COA decarboxylase | FN0200      | Biotin\_lipoyl | Biotin\_lipoyl | LIG_SUMO\_SBM\(_1\), LIG_WW\_Pin1\(_4\), MOD_PKA\(_1\), MOD_PKA\(_2\) |
| Q8R6D6       | Serine protease | FN1950      | Peptidase\(_S8\), Autotransporter\(^a\) | Peptidase\(_S8\) | LIG_FHA\(_2\), LIG_SH2\_STATS, LIG_SH2\(_3\), LIG_SUMO\_SBM\(_1\), MOD_CK1\(_1\), MOD_CK2\(_1,\_2\), MOD_GSK3\(_1\), MOD_PKA\(_1\), MOD_PKA\(_2\) |
| Q8R26        | Single-stranded DNA-binding protein | ssb         | SSB\(^a\) | SSB | LIG_BRCT\_BCRA1\(_1\), LIG_FHA\(_1\), LIG_FHA\(_2\), LIG_PDZ\_Class\_2, LIG_SH2\_STATS, LIG_SH2\(_3\), LIG_SUMO\_SBM\(_1\), MOD_CK1\(_1\), MOD_CK2\(_1\), MOD_GSK3\(_1\), MOD_PKA\(_1\) |
| Q8R7J1       | Dipeptide-binding protein | FN1111      | SBP\_bac\(_5\) | – | CLV_PCSK\_SK1\(_1\), LIG_BRCT\_BCRA1\(_1\), LIG_CYCLIN\(_1\), LIG_SH2\_GRB2, LIG_SH2\_STATS, LIG_SH2\(_3\), LIG_SUMO\_SBM\(_1\), LIG_WW\_Pin1\(_4\), MOD_CK1\(_1\), MOD_PKA\(_1\) |
| Q8R20        | Hemin receptor | FN0499      | Plug\(^a\) | – | CLV_PCSK\_PC1ET2\(_1\), LIG_CYCLIN\(_1\), LIG_FHA\(_1\), LIG_MAPP\(_1\), LIG_PDZ\_Class\_2, LIG_SH2\_STATS, LIG_SH2\(_3\), LIG_SUMO\_SBM\(_1\), MOD_GSK3\(_1\), MOD_N\_GLC\(_1\) |
which convert this established commensal strain to an invasive species in intestinal cells [69].

Discussion

Over the years, it has been shown that *F. nucleatum* can adhere and invade human cells triggering a pro-inflammatory response. Nevertheless, the current knowledge on the molecular players underlying the *F. nucleatum*—human cross-talk is still limited. For this reason, we carried out a computational study to identify *F. nucleatum* putative secreted factors (*Fusosecretome*) that can interact with human proteins.

The originality of our study is manifold compared to previous work. First, we used secretion prediction to identify potential *F. nucleatum* proteins that can be present at the microbe–host interface. Second, we exploited both domain–domain and domain–motif templates to infer interactions with human proteins. Earlier works, including one on *F. nucleatum*, chiefly applied homology-based methods for interaction inference with host proteins (e.g., [70–73]). To our knowledge, domain–motif templates have been only exploited so far to infer or to resolve human–virus protein interaction networks [39, 74]. Indeed, SLiM mimicry is widespread among viruses [21, 75], but increasing evidence shows that it can be an effective subversion strategy in bacteria as well [22]. Third, we performed a network-based analysis on the human interactome to identify the main candidate *F. nucleatum* virulence proteins and the sub-networks they likely perturb.

Our approach relies on two prediction steps: (i) the definition of the *Fusosecretome* based either on the presence of a signal peptide or several protein features such as disorder content, and (ii) the detection of host mimicry elements involved in the interaction with the host. It could be argued that the SecretomeP algorithm may incorrectly predict some proteins as secreted because of their high disorder content. For instance, a previous study considered as erroneous the secretion prediction of ribosomal proteins [76]. We assigned 20

| Table 4 | List of the main candidate virulence proteins in the Fusosecretome (Continued) |
|---------|-----------------------------------------------------------------------------|
| Q8RG959 | Iron(III)-binding protein | FN0375 | SBP_bac_8 | – | LIG_PP1\(^a\), LIG_SH2_GRB2, LIG_SUMO_SBM_1\(^b\), LIG_WW_Pin1_4, MOD_CK1\(^c\), MOD_ProDKin_1 |
| Q8RZ4   | Hemolysin activator protein | FN0131 | POTRA_2\(^a\), ShlB\(^b\) | – | CLV_PCSK_PCE1_2, LIG_SH2_STATS, LIG_SUMO_SBM_1\(^b\), LIG_TRAF2\(^a\) |
| Q8RHO5  | Chaperone protein DnaK | dnaK | HSP70 | HSP70 | CLV_PCSK_PCE1_2, LIG_SH2_STATS, LIG_SH3_3\(^c\), LIG_SUMO_SBM_1\(^b\), LIG_WW_Pin1_4, MOD_CK1\(^c\), MOD_CK2\(^b\), MOD_GSK3_1, MOD_PPIK_1, MOD_PLK, MOD_ProDKin_1, TRG_ENDOCYTIC_2\(^b\), TRG_LysEnd_APsAcL_1\(^b\) |
| Q8RHM4  | Tetratricopeptide repeat protein | FN1990 | DnaJ | DnaJ | – |
| Q8RHL4  | Hypothetical lipoprotein | FN1899 | DUF3798 | – | CLV_C14_Caspase3–7, LIG_FHA_1, LIG_FHA_2, LIG_SH2_STATS, LIG_SUMO_SBM_1\(^b\), LIG_WW_Pin1_4, MOD_CK2\(^b\), MOD_GSK3_1, MOD_PPIK_1, MOD_PLK, MOD_ProDKin_1 |
| Q8RJ9   | Hemolysin | FN1817 | Fil_haemagg_2\(^a\) | – | LIG_CYCLIN_1\(^b\), LIG_FHA_1, LIG_FHA_2, LIG_SH2_STATS, LIG_SUMO_SBM_1\(^b\), LIG_WW_Pin1_4, MOD_CK1\(^c\), MOD_CK2\(^b\), MOD_GSK3_1, MOD_N-GLC_1\(^b\), MOD_PPIK_1, MOD_PLK, MOD_ProDKin_1, TRG_ENDOCYTIC_2\(^b\) |
| Q8RJ47  | Tetratricopeptide repeat family protein | FN1787 | TPR_11\(^a\), TPR_1 | TPR_11, TPR_1 | – |
| Q8RJ75  | S05 ribosomal protein L2 | rplB | Ribosomal_L2, Ribosomal_L2_C | Ribosomal_L2, Ribosomal_L2_C | CLV_PCSK_SK1_1, LIG_BIR_L_1, LIG_PP1\(^a\), LIG_SH3_3\(^c\), LIG_SUMO_SBM_1\(^b\), LIG_USP7\(^b\), LIG_WW_Pin1_4, MOD_PPIK_1, MOD_PLK, MOD_ProDKin_1, TRG_ENDOCYTIC_2\(^b\) |
| Q8RJ5   | Competence protein | FN1611 | HHH_3 | – | LIG_TRAF2\(^a\), MOD_CK2\(^b\) |
| Q8RIK0  | Hypothetical lipoprotein | FN1590 | DUF3798 | – | CLV_C14_Caspase3–7, LIG_FHA_1, LIG_FHA_2, LIG_SH2_STATS, LIG_SUMO_SBM_1\(^b\), LIG_WW_Pin1_4, MOD_CK2\(^b\), MOD_GSK3_1, MOD_PPIK_1, MOD_PLK, MOD_ProDKin_1, TRG_ENDOCYTIC_2\(^b\) |

For every protein, the detected Pfam domains are reported along with the list of domains and SLiMs for which at least one interaction has been inferred.

\(^a\)Pfam entry detected in at least one protein sequence stored in the database of known bacterial virulence factors

\(^b\)Motif for which it was experimentally identified at least one instance in a viral or bacterial protein
ribosomal proteins to the FusoSecretome. Although we cannot exclude a misprediction, ribosomal proteins can be secreted in some bacteria and be involved in host interaction [77, 78]. Furthermore, increasing evidence shows that ribosomal proteins are moonlighting proteins with extra-ribosomal functions such as the E. coli ribosomal L2 protein that moonlights by affecting the activity of replication proteins [79]. Among the 337 inferred interactions between the 20 FusoSecretome ribosomal and 183 human proteins, only a third of latter belong to ribosomal protein families. Interestingly, only 3 of the 41 human interactors inferred for F. nucleatum

Fig. 4 Interaction network between FusoSecretome candidate virulence proteins and preferentially targeted modules. a Candidate virulence proteins are depicted as green rectangular nodes labeled with respective gene symbol, whereas network modules as orange circles, whose size is proportional to the number of proteins belonging to each module and are labeled with the corresponding identifier. Edge width is proportional to the number of inferred interactions of a virulence protein with a given module. Network modules enriched in gut-related disease gene sets are labeled with symbols of different colors (i.e., light blue star: Crohn’s disease, CD; dark blue star: Inflammatory bowel disease, IBD; violet star: genes whose expression correlates with F. nucleatum abundance in colorectal cancer patients, FusoExpr; rose star: genes mutated in colorectal cancer, CRC-mutated; rose zig-zag arrow: dysregulated expression during colorectal cancer progression, CRC-dysregulated). b The protein Fap2 (FN1449) interacts with 9 proteins (nodes with a green border) of Module 9 and c the MORN2 domain containing protein (FN2118) interacts with 8 proteins in Module 89.
L2 are ribosomal proteins, and we identified the L2 protein as candidate virulence protein preferentially targeting Module 451. As this module is mainly involved in cell cycle and DNA repair, this result is consistent with the ability of L2 in E. coli to interfere with DNA processing factors [79] and further reinforces the confidence in the secretome prediction. Moreover, we have here underlined the value of the proposed approach: the interactome provides, on the one hand, the proper biological context to filter out potential false positive inferred interactions and, on the other, pinpoints candidate proteins that can be involved in the F. nucleatum—host interface.

Concerning the host mimicry elements, SLiM detection is notorious for over-prediction [54, given their relative short length and degeneracy (i.e., few fixed amino acid positions). Our strategy to control for false positives was to consider only conserved SLiM occurrences in the Fusosecretome protein regions predicted as disordered. Indeed, the vast majority of known functional SLiMs falls in unstructured regions [54, 56] and shows higher levels of conservation compared to neighboring sequences. Conversely, we might also have missed some “true” mimicry instances in the Fusosecretome by using too stringent parameters for domains and SLiMs identification and our interaction inferences may well be incomplete due to the limited number of available interaction templates. However, their functional significance fortifies our confidence in the predictive approach. Indeed, the Fusosecretome shares similar features with known virulence proteins highlighting its pathogenic potential. In addition, interactors are implicated in established biological processes and cellular districts of the host–pathogen interface and significantly overlap with known pathogen protein binders. Furthermore, more than 70% of interactors are expressed in either the saliva or intestinal tissues. This suggests that most of the inferred interactions can occur in known F. nucleatum niches in the human body. Finally, we found among the human interactors an over-representation of genes whose expression correlates with F. nucleatum abundance in CRC patients; CRC: genes mutated in colorectal cancer samples; IBD: genes associated to inflammatory bowel disease; CD and UC: genes specifically associated to Crohn’s disease and ulcerative colitis respectively.

![Fig. 5 Enrichment of Fusosecretome inferred human interactors and gut disease related proteins in network modules. Each column of the heatmap represents a module. The color of the cells corresponds to the log-transformed enrichment ratio. Pink circles indicate enriched sets. Modules showing a significant dysregulation in CRC progression are highlighted by an empty circle with green border. For the six modules showing an enrichment in inferred interactions and at least one of gut disease related proteins, the most representative functions are reported. Fusosexpr genes whose expression correlates with F. nucleatum abundance in CRC patients; CRC: genes mutated in colorectal cancer samples; IBD: genes associated to inflammatory bowel disease; CD and UC: genes specifically associated to Crohn’s disease and ulcerative colitis respectively.](image-url)
interactions, several modules involved in chromatin modification and transcription regulation (Modules 246, 451, 571, and 625), and localized in compartments such as perinuclear region of the cytoplasm (Modules 90, 138, and 615). Intriguingly, this is reminiscent of the fact that invading \textit{F. nucleatum} strains localize in perinuclear district of colorectal adenocarcinoma cells [8] and that bacteria can tune host-cell response by interfering directly—or indirectly—with the chromatin organization and the regulation of gene expression [84].

We propose 26 \textit{Fusobacterium} candidate virulence proteins as major network perturbators. They are the predominant interactors of preferentially targeted modules. Among the candidates, we identified the known virulence protein Fap2, which was recently shown to promote immune system evasion by interacting with the immunoreceptor TIGIT [19]. Interestingly, Fap2 interacts specifically with Module 9, which is involved in immune response, thus suggesting novel potential binders mediating Fap2 subversion.

A recent report found that abundance of \textit{F. nucleatum} is associated with high microsatellite instability tumors and shorter survival [14]. Notably, three preferentially targeted network modules (i.e., Modules 138, 216, and 371) show a significant upregulation in a stage associated to high microsatellite instability during CRC progression (stage II) [66, 85] and poor prognosis [86, 87]. This suggests that these modules may be important for CRC progression and outcome and that the inferred interactions targeting these modules can mediate the cross-talk between \textit{F. nucleatum} and the host in this particular subtype of CRC.

Overall, our functional and network-based analysis shows that the proposed interactions can occur in vivo and be biologically relevant for the \textit{F. nucleatum}—human host dialog.

**Conclusions**

Over the last years, many microbes have been identified as key players in chronic disease onset and progression. However, untangling these complex microbe–disease associations requires lots of effort and time, especially in the case of emerging pathogens that are often difficult to manipulate genetically. By detecting the presence of host mimicry elements, we have inferred the protein interactions between the putative secretome of \textit{F. nucleatum} and human proteins, and ultimately provided a list of candidate virulence proteins and their human interactors that can be experimentally exploited to test new hypotheses on the \textit{F. nucleatum}—host cross-talk. Our computational strategy can be helpful in guiding and speeding-up wet lab research in microbes–host interactions.

**Methods**

**Protein sequence data**

The reference proteomes of \textit{Fusobacterium nucleatum subsp. nucleatum} strain ATCC 25586 (Proteome ID: UP000002521) and \textit{Homo sapiens} (Proteome ID: UP000005640) were downloaded from the UniProtKB proteomes portal [88] (April 2013). The protein sequences of known gram-negative bacteria virulence factors were taken from the Virulence Factors DataBase [49] (January 2014).

**Secretome prediction**

We identified putative secreted proteins among the \textit{F. nucleatum} proteins by applying two algorithms: SignalP 4.1 [44] that detects the presence of a signal peptide and SecretomeP 2.0 [45] that identifies non-classical secreted proteins (i.e., not triggered by a signal peptide) using a set of protein features such as amino acid composition and intrinsic disorder content.

**Disorder propensity**

To evaluate the intrinsic disorder propensity of \textit{F. nucleatum} proteins predicted as secreted, we used the stand-alone programs of the following algorithms: DISOPRED (version 2.0) [89], IUPred (both long and short predictions) [90] and DisEMBL (COILS and HOTLOOPS predictions, version 1.4) [91]. We compared the disorder propensity distribution of SignalP-predicted secreted proteins to non-secreted proteins using the Kolmogorov–Smirnov test (two-sided, alpha = 0.05).

**Detection of functional domains**

We ran the pfamscan program [92] on \textit{F. nucleatum}, \textit{H. sapiens}, and virulence factors protein sequences to detect the presence of Pfam domains [52] (release 26). We kept only Pfam-A matches with an E value $<10^{-5}$.

**Identification of short linear motifs**

We used the SLiMSearch 2.0 tool from the SLiMSuite [93] to identify occurrences of known short linear motifs from the ELM database [53] (downloaded in May 2013) in the \textit{F. nucleatum} proteome. To select putative mimicry motifs, we applied two SLiMSearch context filters: (i) the motif must be in a disordered region (average motif disorder score >0.2, calculated by IUPred) and (ii) must be conserved in at least one putative ortholog detected in a database of 694 proteomes of commensal/pathogen bacteria in Mammalia downloaded from UniprotKB (March 2014). Sequence alignments and conservation assessment were performed using the GOPHER program from the SLiMSuite using standard parameters [94].
Protein interaction inference
We built an interaction network between *F. nucleatum* putative secretome and human proteins by using interaction templates from the 3did database [95], which stores 6290 high-resolution three-dimensional templates for domain–domain interactions, and the iELM resource [96, 97] that lists 578 high-confidence motif-mediated interfaces between 191 ELM motifs and 402 human proteins. Both datasets were downloaded in August 2013. The domain-based interaction inference works as follow: given a pair of known interacting domains *A* and *B*, if domain *A* is detected in the *F. nucleatum* protein *a* and domain *B* in the human protein *b*, then an interaction between *a* and *b* is inferred. Analogously, for the SLiM-mediated interaction inference: for a given known ELM motif *m* interacting with the domain *C* in the human protein *c*, if the motif *m* occurs in the *F. nucleatum* protein *a*, then *a* is inferred to interact with *c*.

Human proteins targeted by bacteria and viruses
We gathered a list of 3428 human proteins that were experimentally identified as interaction partners of three bacterial pathogen proteins (*Bacillus anthracis*, *Francisella tularensis*, and *Yersinia pestis*) in a large-scale yeast two-hybrid screen [30]. We downloaded interaction data with viruses for 4897 human proteins from the VirHostNet database [98].

Human expression data
RNA-seq expression data for 20,345 protein coding genes in normal colorectal, salivary gland and small intestine (i.e., jejunum and ileum) tissues was downloaded from the Human Protein Atlas (version 13), a compendium of gene and protein expression profiles in 32 tissues [99]. We considered as expressed those protein-coding genes with a FPKM >1, that is 13,640 for colorectal, 13,742 for salivary gland and 13,220 for small intestine.

Functional enrichment analyses
We have compiled several gut-related disease gene sets gathering data from the literature and public repositories. Patient secretome profiling (2566 proteins) for tumor colorectal tissue samples were taken from [100]. We retrieve 152 colorectal cancer genes from the Network of Cancer Genes database (version 4.0, [101]). The list of human genes whose expression correlates with *F. nucleatum* abundance in colorectal cancer patients [13] was kindly provided by Aleksandar Kostic (Broad Institute, USA). The compendium of 163 loci associated with inflammatory bowel diseases was taken from a large meta-analysis of Crohn’s disease and ulcerative colitis genome-wide association studies [80]. The enrichment of these gut-related disease gene sets among inferred interactors was tested using a one-sided Fisher’s exact test.

We assessed the over-representation of cellular functions by performing a enrichment analysis on the list of inferred human interactors using the gProfileR webserver [102] (version: r1488_e83_eg30, build date: December 2015). We analyzed the following annotations: Biological Process and Cellular Component from the Gene Ontology [103]; biological pathways from KEGG [59], and Reactome [60]. Functional categories containing less than 5 and more than 500 genes were discarded.

We used two different reference backgrounds for these statistical analyses. The first background consists of the protein-coding genes in the human genome (i.e., 20 254 genes, UniprotKB, February 2013), whereas the second includes 11 284 protein-coding genes for which we could infer an interaction based on the available domain–domain and motif–domain interaction templates. In both cases, *P* values were corrected for multiple testing with the Benjamini–Hochberg procedure applying a significance threshold equal to 0.025.

Human interactome building, network module detection and annotation
We use the human interactome that we assembled and used in [62, 66]. Briefly, protein interaction data were gathered from several databases (e.g., BioGRID, InnateDB, Intact, MatrixDB, MINT, Reactome) through the PSICQUIC query interface [104] and from large-scale interaction mapping experiments (e.g., [105]). We kept only likely direct (i.e., binary) interactions according to the experimental detection method [106] and mapped protein identifiers to UniprotKB IDs. Given the redundancy among SwissProt and TrEMBL entries, protein sequences were clustered using the CD-HIT algorithm [107], SwissProt/TrEMBL pairs at 95% identity were considered as the same protein: interactions of TrEMBL protein were assigned to the SwissProt protein. As a result, we obtained a human binary interactome containing 74,388 interactions between 12,865 proteins (February 2013).

We detected 855 network modules detected using the Overlapping Cluster Generator algorithm [63]. Modules were functionally annotated by assessing the enrichment of Gene Ontology (GO) biological process and cellular component terms [103], and cellular pathways from KEGG [59] and Reactome [60]. Enrichment *P* values were computed using the R package gProfileR [102] and corrected for multiple testing with the Benjamini–Hochberg procedure (significance threshold = 0.025) and annotated proteins in the human interactome were used as statistical background. Similarly, the over-representation of inferred human interactors and gut disease gene sets in network modules of the human interactome was assessed using a one-sided Fisher’s exact test followed by Benjamini–Hochberg multiple testing correction (significance threshold = 0.025).
Network module perturbation Z score
We devised a score to quantify the contribution of *F. nucleatum* secreted proteins to the perturbation of a network module through their inferred interactions. We defined the perturbation Z score for each *F. nucleatum* protein *f* interacting with at least one protein in module *m* as follows:

\[
Z_{f,m} = \frac{x_{f,m} - \mu_m}{\sigma_m}
\]

Where \(x_{f,m}\) is the number of inferred interactions of the protein *f* with module *m*, \(Z_{f,m}\) is the perturbation Z score of the protein *f* in the module *m*, \(\mu_m\) and \(\sigma_m\) are the mean of the inferred interaction values and their standard deviation in the module *m*, respectively.

Network modules significantly dysregulated during CRC progression
The 77 network modules showing a significant dysregulation during CRC progression were taken from our previous work [66], in which we devised a computational method that combines quantitative proteomic profiling of TCGA CRC samples, protein interaction network, and statistical analysis to identify significantly dysregulated cellular functions during cancer progression.

Additional files

| Additional file 1: Table S1. List of the *F. nucleatum* proteins predicted as secreted (FusoSecretome). (XLSX 61 kb) |
| Additional file 2: Figure S1. Assessment of the disorder propensity of the FusoSecretome proteins (SignalP prediction) with additional prediction algorithms. (PDF 120 kb) |
| Additional file 3: Table S2. Disorder content of FusoSecretome proteins (SignalP prediction) as determined by different algorithms. (XLSX 52 kb) |
| Additional file 4: Table S3. List of detected mimicry elements used for protein interaction inference between FusoSecretome and human proteins. (XLSX 53 kb) |
| Additional file 5: Table S4. Inferred interactions between FusoSecretome and human proteins. (XLSX 116 kb) |
| Additional file 6: Table S5. List of the FusoSecretome inferred human interactors and their annotations. (XLSX 77 kb) |
| Additional file 7: Supplementary Information. Supplementary results, tables and figures. (PDF 1670 kb) |
| Additional file 8: Table S6. Functional annotations significantly enriched among inferred interactions. (XLSX 51 kb) |
| Additional file 9: Table S7. The human binary interactome used in this study. (XLSX 1174 kb) |
| Additional file 10: Table S8. List of network modules detected by the OCG algorithm. (XLSX 471 kb) |
| Additional file 11: Table S9. Functional annotations for the 31 network modules preferentially targeted by the FusoSecretome. (XLSX 140 kb) |
| Additional file 12: Table S10. Perturbation scores of the FusoSecretome proteins for the 31 preferentially targeted modules. (XLSX 65 kb) |

Acknowledgements
The authors would like to thank the members of the TAGC laboratory for fruitful discussion, Anaïs Baudot (I2M, CNRS, France) for critically reading the first draft of the manuscript, Aleksandar Kostic (Broad Institute, USA) for kindly providing the list of human genes whose expression correlates with *F. nucleatum* abundance in colorectal cancer patients, and Henrik Nielsen (DTU Bioinformatics, Denmark) for assistance in running SecretomeP predictions. AZ is grateful to Coralie, Olivia, and Claire for their constant support.

Funding
The project leading to this publication has received funding from Excellence Initiative of Aix-Marseille University—A*MIDEX, a French “Investissements d’Avenir” program, to CB, and was partially supported by the French ‘Plan Cancer 2009–2013’ program (Systems Biology call, A12171AS). The funding organizations had no role in the design of the study and collection, analysis, interpretation of data, and in writing the manuscript.

Availability of data and materials
All data generated or analyzed on the ATCC 25586 strain are included in this published article (and its Additional files). All other data are available from the corresponding author on reasonable request.

Authors’ contributions
AZ conceived the study, designed and performed the experiments, analyzed the data, and wrote the manuscript. LS performed the experiments and analyzed the data. SB performed the experiments. CB designed the experiments, analyzed the data, and wrote the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate
This study is based on publicly available datasets only. Thus, no ethical approval is needed/applicable nor is consent from any participants.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
1Aix-Marseille Université, Inserm, TAGC UMR_S1090, Marseille, France. 2CNRS, Marseille, France.

Received: 28 October 2016 Accepted: 13 July 2017
Published online: 10 August 2017

References
1. Moore WE, Moore LV. The bacteria of periodontal diseases. Periodontol. 1994;5:66–77.
2. Han YW. Fusobacterium nucleatum: a commensal-turned pathogen. Curr Opin Microbiol. 2015;23:141–7.
3. Dharmani P, Strauss J, Ambrose C, Allen-Vereco E, Chadee K. Fusobacterium nucleatum infection of colonic cells stimulates MUC2 mucin and tumor necrosis factor alpha. Infect Immun. 2011;79:2597–607.
4. Fardini Y, Wang X, Témoin S, Nithianantham S, Lee D, Shoham M, et al. Fusobacterium nucleatum adhesin FadA binds vascular endothelial cadherin and alters endothelial integrity. Mol Microbiol. 2011;82:1468–80.
5. Gunsøy UK, Königén E, Uitto V-J. Intracellular replication of fusobacteria requires new actin filament formation of epithelial cells. APMS Acta Pathol Microbiol Immunol Scand. 2008;116:1063–70.
6. Han YW, Shi W, Huang GT, Kinder Haake S, Park NH, Kuramitsu H, et al. Interactions between periodontal bacteria and human oral epithelial cells: Fusobacterium nucleatum adheres to and invades epithelial cells. Infect Immun. 2000;68:3140–6.
7. Qah H, Bergenholtz G, Tan KS. Fusobacterium nucleatum induces cytokine production through Toll-like-receptor-independent mechanism. Int Endod J. 2014;47:550–9.
8. Strauss J, Kaplan GG, Beck PL, Roux K, Panaccione R, Devinsney R, et al. Invasive potential of gut mucosa-derived Fusobacterium nucleatum positively correlates with IBD status of the host. Inflamm Bowel Dis. 2011;17:1971–8.
9. Castellaín M, Warren RL, Freeman JD, Dreoni L, Koziwinski M, Strauss J, et al. Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. Genome Res. 2012;22:299–306.
10. Kostic AD, Gevers D, Peddamsulu CS, Michaud M, Duke F, Earl AM, et al. Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. Genome Res. 2012;22:292–8.
11. McCoy AN, Araújo-Pérez F, Azcárate-Peril A, Yeh JJ, Sandler RS, Keku TO. Fusobacterium is associated with colorectal adenomas. PLoS One. 2013;8:e53653.
12. Gevers D, Kugathasan S, Denson LA, Vázquez-Baeza Y, Van Treuren W, Ren B, et al. The treatment-naive microbiome in new-onset Crohn’s disease. Cell Host Microbe. 2014;15:382–92.
13. Castellarin M, Warren RL, Freeman JD, Dreolini L, Krzywinski M, Strauss J, et al. Zanzoni et al. Microbiome
14. Kostic AD, Chun E, Robertson L, Glickman JN, Gallini CA, Michaud M, et al. Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. Cell Host Microbe. 2013;14:207–15.
15. Mima K, Nishihara R, Qian ZR, Yaku Y, Nowak JA, et al. Fusobacterium nucleatum in colorectal carcinoma tissue and patient prognosis. Gut. 2016;65(12):1973–1980.
16. Mima K, Suka Y, Nishihara R, Qian ZR, Yamauchi M, Inamura K, et al. Fusobacterium nucleatum and T cells in colorectal carcinoma. JIMA Oncol. 2015;1563–61.
17. Rubinstein MR, Wang X, Liu W, Hao Y, Cai G, Han YW. Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating E-cadherin/E-catenin signaling via its FadA adhesin. Cell Host Microbe. 2013;14:195–206.
18. Han YW, Ikekami A, Rajanna C, Kawasaki H, Zhou Y, Li M, et al. Identification and characterization of a novel adhesin unique to oral fusobacteria. J Bacteriol. 2015;197:3330–40.
19. Kaplan CW, Ma X, Paranjpe A, Jewett A, Lux R, Kinder-Haake S, et al. Fusobacterium nucleatum outer membrane proteins Fap2 and Radl induce cell death in human lymphocytes. Infect Immun. 2010;78:4773–8.
20. Davey NE, Travé G, Gibson TJ. How viruses hijack cell regulation. Trends Biochem Sci. 2011;36:159–69.
21. Via A, Uyar B, Brun C, Zanazzi A. How pathogens use linear motifs to perturb host cell networks. Trends Biochem Sci. 2015;40:316–8.
22. Luddin P, Nilsson D, Mäser P. Genome-wide identification of molecular mimicry candidates in parasites. PLoS One. 2011;6:e17546.
23. Baxt LA, Navratil V, Tafforeau L, Hiet MS, Aublin-Gex A, Agaugué S, et al. Hepatitis C virus infection protein networks and their interaction with the human proteome. Science. 2011;332:3135–7.
24. Rudel T, Kepp O, Kozjak-Pavlovic V. Interactions between bacterial pathogens and mitochondrial cell death pathways. Nat Rev Microbiol. 2010;8:693–701.
25. Elde NC, Malik HS. The evolutionary conundrum of pathogen mimicry. Nat Commun. 2015;6:7412.
26. Haglund CM, Welch MD. Pathogens and polymers: microbe-host interactions targeting of a common host protein-network by pathogen effectors from three kingdoms of life. Cell Host Microbe. 2014;16:364–75.
27. Elde NC, Malik HS. The evolutionary conundrum of pathogen mimicry. Nat Commun. 2015;6:7412.
28. Baxt LA, Garza-Mayers AC, Goldberg MB. Bacterial subversion of host innate immune response by C-terminal features. BMC Genomics. 2014;15:653.
29. Jäger S, Cimermancic P, Gulbahce N, Johnson JR, McGovern KE, Clarke SC, et al. Global mapping of the human-proteome interactions reveals that retromer is a multifunctional protein identified from a human protein interaction network. Nat Commun. 2015;6:7412.
63. Becker E, Robisson B, Chapple CE, Guinoche A, Brun C. Multifunctional proteins revealed by overlapping clustering in protein interaction network. Bioinformatics. 2012;28:940.

64. Chapple CE, Brun C. Redefining protein moonlighting. Oncotarget. 2015;6:16812–3.

65. Manson McGuire A, Cochrane K, Griggs AD, Haas BJ, Abee T, Zeng Q, et al. Evolution of invasion in a diverse set of Fusobacterium species. mBio. 2014;5:e01864.

66. Zanzoni A, Brun C. Integration of quantitative proteomics data and interaction networks: Identification of dysregulated cellular functions during cancer progression. Methods. 2016;93:103–9.

67. Strauss J, White A, Ambrose C, McDonald J, Allen-Venice E. Phenotypic and genotypic analyses of clinical Fusobacterium nucleatum and Fusobacterium periodonticum isolates from the human gut. Anaerobe. 2008;14:301–9.

68. Kar S, Edgar R, Adhya S. Nucleoid remodeling by an altered HU protein. Reorganization of the transcription program. Proc Natl Acad Sci U S A. 2005;102:16397–402.

69. Koli P, Sudan S, Fitzgerald D, Adhya S, Kar S. Conversion of commensal Escherichia coli K-12 to an invasive form via expression of a mutant histone-like protein. MBio. 2011;2(5). doi:10.1128/mBio.00182-11.

70. Tyagi N, Krishnadev O, Srinivasan N. Prediction of protein-protein interactions between Helicobacter pylori and a human host. Mol Biol Symp. 2009;1635:30–5.

71. Doolittle JM, Gomez SM. Mapping protein interactions between Dengue virus and its human and insect hosts. PLoS Negl Trop Dis. 2011;5:e954.

72. Schleier S, Garcia-Garcia J, Klein-Seetharaman J, Oliva B. Prediction and comparison of Salmonella-human and Salmonella-Madtopsis interactions. Chem Biodivers. 2012;9991–1018.

73. Kumar A, Thotakura PL, Tiwary BK, Krishna R. Target identification in Fusobacterium nucleatum by subtractive genomics approach and enrichment analysis of host-pathogen protein-protein interactions. BMC Microbiol. 2016:1684.

74. Evans P, Dampier W, Ungar L, Tozeren A. Prediction of HIV-1 virus-host protein interactions using virus and host sequence motifs. BMC Med Genet. 2009;2:27.

75. Harigat S, Asa A, Babu MM, Andino R. Use of host-like peptide motifs in viral proteomes is a prevalent strategy in host-virus interactions. Cell Rep. 2014;7:1729–39.

76. Christie-Olizaj JA, Piña-Villalonga JM, Bosch R, Nagolves B, Armandoa J. Comparative proteogenomics of eleven Roseobacter exproteomes reveals different adaptive strategies among these marine bacteria. Mol Cell Proteomics. 2012;11(M111.013110).

77. Tjalma H, Lambrouy L, Hermens PW, Swinkels DW. Shedding & sharing: disclosure of proteomic expressions on a bacterial face. Proteomics. 2008;14:15–28.

78. Pérez-Cruz C, Delgado L, López-Iglesias C, Mercade E. Outer-inner membrane vesicles naturally secreted by gram-negative pathogenic bacteria. PLoS One. 2015;10:e0161896.

79. Chodavarapu S, Felczak MM, Kauraj MJ. Two forms of ribosomal protein L2 of Escherichia coli that inhibit DnaK in DNA replication. Nucleic Acids Res. 2011;39:14803–91.

80. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-genetic analysis of inflammatory bowel disease. Nature. 2012;491:119–26.

81. Coupon F, Brinkman F, Charloteaux B, Pevzner SJ, Zhong Q, Sahni N, et al. A web server to explore context for short linear motifs in proteins. Bioinformatics. 2012;28:16812–3.

82. Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB. Protein disorder prediction: implications for structural proteomics. Structure. 2003;11:1453–9.

83. Li W, Cowley A, Ulludag M, Gur T, McWilliam H, Squizzato S, et al. The EMBL-EBI bioinformatics web and programmatic tools framework. Nucleic Acids Res. 2015;43:W580–4.

84. Davey NE, Haslam NJ, Shields DC, Edwards RJ. SlimSearch 2.0: biological context for short linear motifs in proteins. Nucleic Acids Res. 2011;39:W56–60.

85. Davey NE, Shields DC, Edwards RJ. SlimSdc: short, linear motif discovery, correcting for common evolutionary descent. Nucleic Acids Res. 2006;34:546–54.

86. Stein A, Cao A, Aloy P. 3dId: identification and classification of domain-based interactions of known three-dimensional structure. Nucleic Acids Res. 2011;39:D718–23.

87. Weisheft RJ, Luck K, Petralaki E, Davey NE, Gibson TJ. The identification of short linear motif-mediated interfaces within the human interactome. Bioinforma. 2012;28:9976–82.

88. Weisheft RJ, Jelí P, Dinkelh G, Gibson TJ. ELM—a web server to explore short linear motif-mediated interactions. Nucleic Acids Res. 2012;40:W364–9.

89. Navratil V, de Chassey B, Meynien L, Delmotte S, Gautier C, André P, et al. VirHostNet: a knowledge base for the management and the analysis of proteome-wide virus-host interaction networks. Nucleic Acids Res. 2009;37:D661–8.

90. Uhlem M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. Science. 2015;347:1260419.

91. de Witt M, Kant H, Perisna SR, Pham TV, Mongera S, van Berk MP, et al. Colorectal cancer candidate biomarkers identified by tissue secretome proteome profiling. J Proteome. 2014;69:26–39.

92. An O, Pendino V, D’Antonio M, Ratti E, Gentilini M, Ciccaelli FD. NGC 40: the network of cancer genes in the era of massive mutational screenings of cancer genomes. Database J Biol Databases Curation. 2014;2014:bau015.

93. Reimand J, Arak T, Adler P, Reisberg S, Peterson H, Vilo J, g:Profiler-a web server for functional interpretation of gene lists (2016 update). Nucleic Acids Res. 2016;44(W1):W83–9.

94. The Gene Ontology Consortium. The gene ontology in 2010: extensions and refinements. Nucleic Acids Res. 2010;38:355–56.

95. Aranda B, Blankenburg H, Kerrien S, Brinkman FSL, Ceol A, Chautard E, et al. The STRING database in 2011: protein interactions and functional outcome predictions from integration of different interaction sources. Nucleic Acids Res. 2011;39:D300–6.

96. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. Science. 2015;347:1260419.

97. Aranda B, Blankenburg H, Kerrien S, Brinkman FSL, Ceol A, Chautard E, et al. The STRING database in 2011: protein interactions and functional outcome predictions from integration of different interaction sources. Nucleic Acids Res. 2011;39:D300–6.

98. Navratil V, de Chassey B, Meynien L, Delmotte S, Gautier C, André P, et al. VirHostNet: a knowledge base for the management and the analysis of proteome-wide virus-host interaction networks. Nucleic Acids Res. 2009;37:D661–8.

99. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. Science. 2015;347:1260419.

100. de Witt M, Hamb K, Pernis SR, Pham TV, Mongera S, van Berk MP, et al. Colorectal cancer candidate biomarkers identified by tissue secretome proteome profiling. J Proteome. 2014;69:26–39.

101. Rolland T, Tajan M, Charleaux B, Pavnjier SJ, Zhong Q, Sahni N, et al. A proteome-scale map of the human interactome network. Cell. 2014;159:1212–26.

102. Rual J-F, Venkatesan K, Hao T, Hirozane-Kishiiawa T, Dricot A, Li N, et al. Towards a proteome-scale map of the human protein-protein interaction network. Nature. 2005;437:1173–8.

103. Li W, Liu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for the next generation sequencing data. Bioinformatics. 2012;28:3150–2.