

Review

# Omics in Aid of Host-pathogen Interaction Studies: A Bacterial Perspective

Ursula Fels<sup>1,2</sup>, Kris Gevaert<sup>1,2</sup> and Petra Van Damme<sup>1,2,\*</sup>

<sup>1</sup> VIB-UGent Center for Medical Biotechnology, Ghent 9000, Belgium

<sup>2</sup> Department of Biochemistry

\* Correspondence: petra.vandamme@vib-ugent.be; Tel.: +32 (0)9 264 9279

**Abstract:** By providing useful tools to study host-pathogen interactions, next-generation omics has recently enabled the study of gene expression changes in both pathogen and infected host simultaneously. However, since great discriminative power is required to study pathogen and host simultaneously throughout the infection process, the depth of quantitative gene expression profiling has proven to be unsatisfactory when focusing on bacterial pathogens, thus preferentially requiring specific strategies or the development of novel methodologies based on complementary omics approaches. In this review, we focus on the difficulties encountered when making use of omics approaches to study bacterial pathogenesis. Besides, we review different omics strategies (i.e. transcriptomics, proteomics and secretomics) and their applications for studying interactions of pathogens with their host.

**Keywords:** host-pathogen interaction; bacterial pathogens; infection biology; omics

---

## 1. General introduction

Over the years, efforts have been made to classify microorganisms as either pathogenic or non-pathogenic, and to understand the virulence of the former. At the very beginning, virulence was thought to be an intrinsic capacity of only certain microorganisms [1]. While such a strict division is embodied in Koch postulates [2], widely used in medical microbiology since 1880, advances in the research field of infectious diseases led to a progressive flexibility of the threshold to classify a microorganism as pathogenic or not (reviewed by Pierre-Olivier Méthot and co-workers [1]). Accumulating evidence suggests the involvement of host innate and adaptive immune responses and environmental conditions as key factors in determining a pathogen's ability to infect a host. Altogether this brings us to the current definition of virulence as 'the outcome of a host-pathogen interaction', thus not considering intrinsic properties of the microbe or host separately [3]. Despite significant efforts and the knowledge gained when studying microorganisms as separate entities, one of the remaining challenges of the host-pathogen interaction research field is to understand the behavior of pathogens in the context of their hosts and *vice versa*. Overall, this stresses the need for integrative studies of host-pathogen interactions.

More specifically, unravelling host-pathogen interactions is of central importance to understand the host response to infection and the mechanisms employed by microorganisms to subvert host defenses, and for the discovery of novel antimicrobial targets. Omics approaches have the advantage

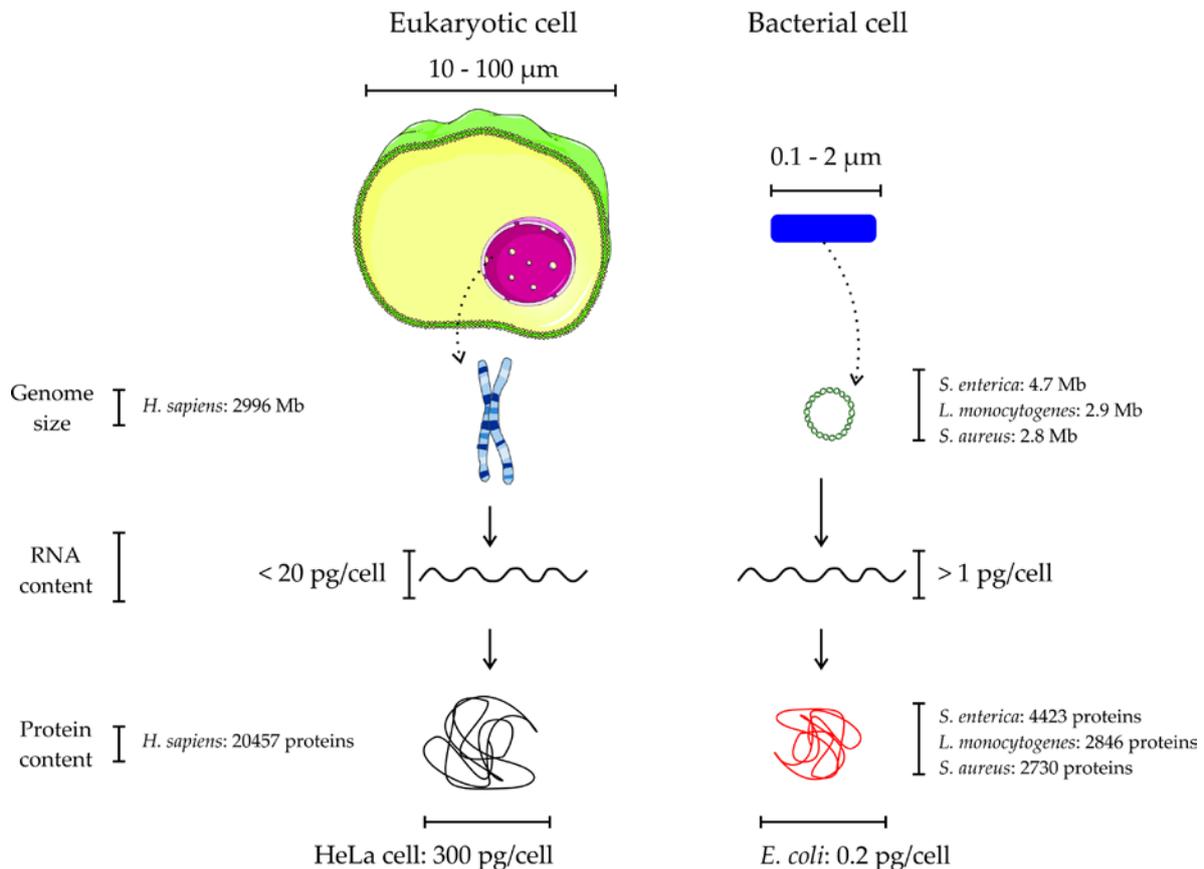
of enabling integrative studies of system alterations at different levels, including gene transcription, protein synthesis and alterations of the proteome amongst others. Therefore, the applications of omics-based strategies are key for the discovery of mechanisms steering bacterial virulence and pathogenesis. However, despite major advancements, omics approaches still have significant limitations when applied to study pathogens in the context of the host. A shortcoming which is mainly due to the macromolecular content of eukaryotic cells differing by several orders of magnitude from that of prokaryotes, making such studies extremely challenging. Further, the lengthy procedures frequently applied to *a priori* enrich for bacteria might influence the data outcome. Another limitation comes from the fact that reproduction and interpretation of reported findings is often complicated or even precluded due to the scarcity of details provided regarding the experimental setups used (Table 1). Moreover, the high variability of the infection capacity of the pathogen and the overall infection efficiency also contribute to the observed high variability in infected cell (sub-) populations. Thus, ideally host-pathogen interactions are studied on the single cell level, however, such studies are not yet routinely performed. In this review, we focus on the application of various omics approaches that permit the study of host-pathogen interactions using bacterial pathogens as infection models and specifically focus on studies enabling detection of changes in the bacterial pathogen within infected cells while not on studies reporting host cell changes, since omics analysis of the latter can essentially be performed routinely. For proteomic tools to study bacterial pathogen protein/ host protein interactions, we would like to refer the reader to a recent review discussing such studies in more detail [4].

## 2. Current limitations for the application of omics in host-pathogen interaction studies

When applying omics approaches for studying host-pathogen interactions it is important to consider the differences concerning the host and pathogen molecular contents (Figure 1). If the focus is on the host, studies can largely neglect the pathogen content [5-7]. Nonetheless, if the focus is on the pathogen's response, its molecular content becomes limiting, making it highly desirable to implement strategies to enrich the molecular content of the pathogen, strategies which are however, not always available or easy to implement. As such, a frequently used alternative approach to study pathogenic responses is to grow the pathogen in conditions mimicking the host environment [8-11].

To address the differences between the host and bacterial pathogen, one of the first things to consider is their difference in size. The diameter of eukaryotic cells typically ranges from 10 to 100  $\mu\text{m}$ , whereas that of bacteria ranges from 0.1 to 2  $\mu\text{m}$  [12]. Consequently, when considering bacterial and eukaryotic cells as spheres, their volumes range between  $\sim 10^5$  to  $\sim 10^9$   $\mu\text{m}^3$ , for bacteria and eukaryotic cells respectively. In terms of genomic content, eukaryotes possess larger and more complex genomes than prokaryotes; the genome of bacterial pathogens varies between 0.8 to 8 Mb, whereas the genome of higher eukaryotic organisms such as human (*Homo sapiens*) and mouse (*Mus musculus*) is about 2996 and 2671 Mb respectively [13], thus about several hundred times larger. A bacterial pathogen contains approximatively 5 to 10 times less genes than their eukaryotic hosts. For

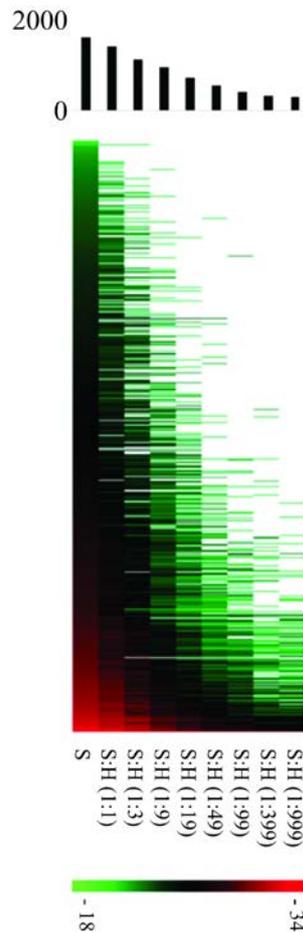
example, *Homo sapiens* possesses about 20,457 protein-coding genes, while *Salmonella enterica* subsp. *enterica* serovar Typhimurium strain LT2 contains 4423 protein-coding genes [14]. As such, without considering all possible proteoforms (e.g. variant of proteins raised upon alternative splicing and alternative translation initiation amongst others) [15], eukaryotic proteomes are about one order of magnitude larger than their prokaryotic counterparts (Figure 1).



**Figure 1 | Macromolecular differences between cells of eukaryotic and bacterial origin.** Eukaryotic and bacterial cells differ greatly in many aspects. First, in terms of size, these organisms are considerably different. Eukaryotic cells have a size that ranges between 10 to 100  $\mu\text{m}$ , whereas bacteria are typically 0.1 to 2  $\mu\text{m}$ . Macromolecules are the building blocks of life, present in all known living organisms. Eukaryotic cells and bacteria also have great differences in terms of their macromolecule content. Genomes differ in aspects such as size and organization, whereas the RNA content is about 20 times higher in eukaryotic cells. Finally, at the protein level, these organisms display differences not only in absolute numbers of proteins [14] and proteoforms expressed, but also in terms of the total amount of proteins per cell.

Moreover, when studying the proteome of an infected host, the mass spectrometry-based identification of bacterial proteins is highly challenging given their overall much lower levels and the fact [16] that despite recent improvements in accuracy, sensitivity and speed of mass spectrometers and the performance of liquid chromatography, ionization suppression and an overall low loading capacity still result in under-sampling. This limits the coverage of complex samples which can only be partially overcome by increasing the overall time spent on analyzing samples [17]. A representative example of how the number of *Salmonella* proteins identified using LC-MS/MS

decreases when analyzing complex peptide mixtures, being a *Salmonella* total protein dilution series in lysates of human HeLa cells, is shown in Figure 2. Here, the number of identified *Salmonella* proteins when comparing a pure *Salmonella* versus a 1:999 pathogen/host peptide mixture drops already ~5 fold (i.e. from 1600 to 295 protein identifications).



**Figure 2 | Heat map representation of identified *Salmonella* proteins and corresponding label-free quantification (LFQ) intensities when analyzing complex host/pathogen peptide mixtures.** Complex peptide mixtures obtained from a trypsin digested total *Salmonella* protein lysate (S) dilution series (as indicated in the figure) in protein lysates of human HeLa cells (H) were analyzed by LC-MS/MS and the protein abundance quantified using label-free quantification (LFQ). In the heat map, the red and green color indicates high and low LFQ intensities ( $\log_2(\text{LFQ})$ ) as shown by the color scale bar in  $\log_2(\text{LFQ})$ , respectively and white, indicates that the protein was not identified in the corresponding experimental setup analyzed. In the upper part of the figure, the bar chart shows the number of proteins identified in every condition. The precise number of proteins identified was 1600 (S), 1391 (S:H 1:1), 1120 (S:H 1:3), 950 (S:H 1:9), 736 (S:H 1:19), 553 (S:H 1:49), 427 (S:H 1:99), 328 (S:H 1:399), 295 (S:H 1:999).

Besides the increased genome complexity of most eukaryotes, the DNA of eukaryotic cells is also organized differently from that of prokaryotic cells. Eukaryotic genetic material is organized as a macromolecular structure of highly packed DNA and proteins known as chromatin. Bacteria usually possess a unique circular chromosome and, in some cases, extrachromosomal replicons known as

plasmids, which for several bacteria such as *Shigella* spp. are relevant for their virulence [18]. The RNA content of eukaryotic and prokaryotic organisms is typically about two orders of magnitude different, with about 20 pg and 0.1 pg of total RNA per eukaryotic and prokaryotic cell, respectively [19]. When considering their average sizes, the protein content of eukaryotic and prokaryotic organisms differs typically over ~1000 fold (i.e. 300 pg and 0.2 pg in case of a HeLa and *E.coli* cell, respectively) [12] (Figure 1). Altogether, this highlights that when applying omics to understand pathogenic processes involved in host pathogen interactions, these substantial differences in molecular content cannot be neglected.

Another level of complexity is imposed by the heterogeneity among infected cells, and not all cells will be infected in a cell culture. More specifically, the number of bacteria per infected cell and stage of infection can (greatly) vary among infected cells (Supplementary Table 1). As such, when setting up an infection model, several considerations have to be made. Some of these are the multiplicity of infection (MOI) and internalization conditions in the case of intracellular bacteria, host cell-type dependency of intracellular pathogen replication in infected cells, and viability of the infected cells and of the bacteria. Illustrative of this, upon infection of the host, *Salmonella* colonizes the gut where it promotes its own internalization in epithelial cells [20]. After passing through the epithelial barrier, *Salmonella* Typhimurium can be phagocytized by macrophages and dendritic cells (DCs), explaining the widespread use of epithelial, macrophage and dendritic cell lines next to primary cell cultures (Supplementary Table 1). Since a bacterial pathogen targets only (a) particular cell type(s) during infection, it is thus extremely important to carefully consider which representative infection model to use. As illustrated in Supplementary Table 1, in case of *Francisella tularensis* many studies used myeloid primary cells or cell lines since these are the primary target cells and replicative niche of this bacterial pathogen. Many other different cell lines can be used as host models such as epithelial cell lines (HeLa or Caco-2), phagocytic cell lines (J774A.1 murine macrophages, THP-1 human monocytes) or primary cells such as monocyte or bone marrow derived macrophages or dendritic cells from human or murine origin, as well as neutrophils.

Most studies on intracellular pathogens use antibiotics to kill extracellular bacteria following their addition to host cell cultures enabling the focus on internalized bacteria. Generally, host cells are first seeded after which bacteria are added to the culture medium, which is followed by an incubation with the host cells for a given period (typically ranging from 30 to 60 minutes) to allow bacterial internalization. Then, non-internalized bacteria are washed out, and antibiotics are added to ensure killing of bacteria that adhered to the cells but were not internalized. This procedure is referred to as an antibiotic protection assay and most commonly the antibiotic gentamicin is used. With short incubation times, gentamicin does not enter host cells, hence it does not affect intracellular bacteria [21]. Following an antibiotic protection assay, intracellular bacteria are quantified by colony forming unit (CFU) enumeration or, alternatively, the number of infected cells can be estimated following fluorescence microscopy or using flow cytometry. Using microscopy, the number and the distribution or location of bacteria within host cells or cell-associated bacteria can be determined in a

quantitative manner as for example shown by Malik-Kale *et al.* [22] who made use of mCherry expressing *Salmonella*. By measuring mean fluorescence intensities, flow cytometry enables semi-quantitative analyses as demonstrated by Raybourne *et al.* [23] who compared the intracellular *L. monocytogenes* content of human monocytes and granulocytes by pre-labeling bacteria with the lipophilic fluorescent dye PKH-2. Isolation of infected cells using FACS typically relies on the exogenous expression of fluorescent proteins, like green fluorescent protein (GFP), by pathogens [24-27]. Not surprisingly, several of these techniques are often used in parallel to answer complex biological questions. Moreover, it is important to highlight that expression of exogenous fluorescent proteins from plasmids can affect *Salmonella's* virulence [28]. Evidence for this was provided by Clark *et al.* [29] who observed a reduction in the expression of *Salmonella* pathogenic island 1 (SPI-1) genes and a reduced infectivity either in bacteria containing the GFP-expressing plasmid or the empty control plasmid. On the other hand, to avoid the expression of GFP protein from plasmids, the *gfp* gene can be directly cloned in the bacterial chromosome allowing for its expression from a constitutive promoter [30].

Another important parameter to consider is the multiplicity of infection (MOI), in bacteriology defined as the ratio of bacteria to eukaryotic host cell. This parameter should be newly established for different pathogens in relation to their hosts and their cellular states. For instance, when *Y. pestis* is used to infect macrophages, the latter may undergo apoptosis [31], which shows the importance of the appropriate MOI selection dependent of the research question. However, often it is not easy to find a consensus MOI between different studies (Supplementary Table 1) as for example, *L. monocytogenes* internalization was shown to be significantly different when comparing MOIs of 1 and 10 amongst different studies [32]. Further, using the same approach, it was reported that the percentage of infected cells at 4 and 24 hpi depends on the MOI used to infect cells [30]. The authors reported that more than 75% of HeLa cells infected with *S. Typhimurium* are infected at 4 hpi when a MOI of 100 is used, whereas a MOI of 10 only resulted in 5% of the cells infected. As such caution is needed when extrapolating the results from different studies. Generally, increasing the MOI will increase the percentage of infected cells, until a certain point where it reaches a plateau [33]. The first MOI that allows reaching this plateau is defined as the optimal MOI (as much as possible infected cells). After this point it does not make sense to use higher MOIs. As described for *Yersinia*, cell viability should also be considered when defining MOI values. Therefore, a compromise must be made between an optimal MOI to obtain sufficiently high number of infected cells and one that will not compromise the viability of cells. Dependent of the research question, appropriate MOIs typically range from 5 to 100 (Supplementary Table 1). Even at an optimal MOI, not all cells will be infected; Jantsch *et al.* [34] reported that upon infection of bone marrow derived DCs (BMDCs), in which *Salmonella* is not replicative, only 40% of the BMDCs were infected as determined by flow cytometry. The absence of replication in these cells makes omics studies more challenging given the limited amount of pathogenic material available however, it resembles the early time points of *Salmonella* infection in case of macrophages and epithelial cells when bacteria are not yet replicative. In another

study performed in HeLa cells using the same *Salmonella* strain [22], the number of intracellular bacteria was found to differ among infected cells throughout time. Using mCherry-labeled bacteria the number of intracellular replicative *Salmonella* bacteria was counted by fluorescence microscopy and it was found that 2 hpi, infected cells contained about 20 bacteria, whereas at 8 and 16 hpi this number ranged between 20 and 100. Overall, these examples clearly illustrate that not all cells become infected and, and when considering infected cells, the number of intracellular bacteria per cell can vary greatly.

Finally, the growth phase and composition of the media used during *in vitro* infections must be considered. For example, when studying *Salmonella* invasion of epithelial cell lines, the late exponential phase was shown to be the phase during which *Salmonella* pathogenesis island-1 (SPI-1) is highly induced, and therefore the more suitable growth phase for infection [35]. Many studies make use of opsonizing antibodies when phagocytic cells are used. This is the case for *Francisella tularensis* infecting macrophages, where opsonized bacteria are more efficiently phagocytized than non-opsonized bacteria as shown by Geier *et al.* [36] during infection of bone marrow derived macrophages. Moreover, when studying the Gram-positive pathogen *L. monocytogenes*, Kolb-Maurer *et al.* [37] showed that upon replacing fetal calf serum (FCS) in the culture medium by human plasma, the bacteria were more actively taken up by the cells (Supplementary Table 1). It is also important to consider, that the switch from bacterial to host cell growth media when performing the infection might introduce (substantial) changes in the gene expression profile of bacteria, necessitating the need of the inclusion of appropriate controls.

### 3. Transcriptome analyses during infection to study gene expression profiles of host and pathogen simultaneously: RNA-seq supplanting microarrays

Transcriptomics refers to the study of an organism's transcriptome, which is the totality of its RNA transcripts. Hence, transcriptomics allows detection of the expressed gene pool. Transcriptomes have classically been studied using DNA microarrays [38]. Briefly, RNA is extracted from the cells, converted to cDNA via reverse transcription, fluorescently labeled by PCR amplification [39] and hybridized with immobilized gene-specific DNA probes. Upon hybridization, the fluorescence signals allow detecting the transcripts and quantification of their levels. Moreover, by using different fluorescent labels, multiple transcriptomes can be studied simultaneously [40]. Even though microarrays allowed the study of pathogen transcriptomes in infection models [41], with the advent of next-generation sequencing approaches enabling high-throughput sequencing, massively parallel cDNA sequencing or RNA-seq quickly took over. An extensive description of RNA-seq approaches together with their technical advantages over microarray techniques has recently been reviewed [42, 43].

RNA-seq holds some interesting advantages over microarrays that are important when studying host-pathogen interactions. More specifically, RNA-seq has a broader dynamic range than microarrays ( $10^5$  and  $10^3$ , respectively [44]), allowing to address larger fold-changes in gene

expression. RNA-seq also enables detection of low abundance transcripts and allows discriminating between different transcript isoforms and identification of genetic variants. Important for host-pathogen studies and due to its high sensitivity [45], simultaneous gene expression profiling of both pathogen and host became possible. Moreover, as RNA-seq is sequencing-based and read mapping onto the respective genomes is performed *in silico*, cross-reactivity when studying different species, an important drawback of microarrays, does not any longer occur. More specifically, in recent years several studies made use of dual RNA-seq, a term coined by Westermann *et al.* [43], when studying host-pathogen interactions using *in vitro* infection models (reviewed in [19]). When assuming 10 bacteria per infected cell, sequencing depth requirements to obtain correct reflections of both host and pathogen expression profiles were estimated to be 200 and 2,000 million total reads for rRNA depleted and total RNA samples, respectively [43]. Of note however, while these sequencing depths suffice to determine changes in gene expression, sequence coverage of the transcriptomes obtained this way is still too low to provide detailed information regarding transcript structure, delineation of transcript boundaries etc. Using dual RNA-seq, some of the studied pathogens include *Salmonella Typhimurium* [30], *Staphylococcus aureus* [46], Uropathogenic *Escherichia coli* (UPEC) [47], *Chlamydia trachomatis* [48], *Mycobacterium bovis* [49] and *Haemophilus influenzae* [50]. More recently, dual RNA-seq was also used to simultaneously study host and pathogen gene expression profiles in *in vivo* infection models (reviewed in [19]). Briefly, gene expression was analyzed in mice infected with *Pseudomonas aeruginosa*, after lung tissue resection [51]. In another study, Peyer's patches from mice infected with *Yersinia pseudotuberculosis* were analyzed [52]. In both examples, infected tissues were isolated and homogenized prior to total RNA extraction for sequencing. Since multiple types of cells might be involved, this approach has also been generally referred to as multi RNA-seq [19].

While both *in vitro* [30] and *in vivo* approaches ideally require *a priori* isolation of infected cells and even though *in vivo* models of host-pathogen interactions represent optimal conditions to understand bacterial pathogenesis, the composite gene expression profiles originating from a variety of host cell types observed in case of the latter complicate data interpretation even further and thus optimally require separation of the different types of infected cells. Moreover, increased heterogeneity in bacterial pathogen populations might add another level of complexity to multi RNA-seq studies, making it for example also necessary to differentiate between the intra- or extracellular origin of bacteria without *a priori* isolation of infected cells as for instance, *Salmonella* alternates between a proliferative intracellular stage and an extracellular disseminative stage [53]. Since efficient and cost-effective separation methods are still lacking, the use of multi RNA-seq to study host-pathogen interactions remains somewhat limited [19].

#### 4. Proteomics of bacterial pathogens in the context of the host - overcoming the overwhelming host proteome contribution

To obtain sufficient quantities of a pathogen's proteome from the infected host, enrichment and isolation protocols have been devised. Overall, the main objective of these is to reduce the

contribution of host proteins to the isolated bacterial proteome content, thereby increasing the overall sensitivity of mass spectrometry (MS) based detection of bacterial proteins. Intact bacteria can be isolated from lysates of infected host cells using different methods such as differential centrifugation, fluorescence activated cell sorting (FACS) based cell sorting and immune (magnetic) selection, all of which rely on intrinsic or acquired properties that enable their selection. Next to these strategies, various approaches have been reported that enable specific bacterial protein labeling with non-natural amino acids and subsequent enrichment of labeled proteins.

#### 4.1 Physical separation based approaches for isolating intact bacterial pathogens upon infection

Centrifugation-based approaches make use of infected host cells that are lysed by specific detergents. Triton-X100 is typically used as it enables selective lysis of the membranes from eukaryotic cells, leaving the membranes of bacterial cells intact. Thus, detergent-based lysis of infected host cells allows the isolation of intact and viable bacteria upon centrifugation. Using differential centrifugation, Yanhua Liu and coworkers studied the proteome of intracellular *Salmonella* infected HeLa cells [54]. The authors recovered a large amount of bacteria (i.e.  $\sim 10^8$  bacteria). After MS analysis, 1675 *Salmonella* proteins and 251 host contaminants proteins (i.e. about 13% of all proteins identified) were identified. Differential centrifugation was also used to study the proteome of several bacterial pathogens in both *in vitro* and *in vivo* infection models. Examples include *Shigella* [55,56], *Listeria* [57-59], *Yersinia* [60] and *Brucella* [61].

Upon internalization, certain pathogens end up in specific subcellular vacuoles; so-called bacteria containing vacuoles (BCVs). Some of these pathogens are *Salmonella*, *Legionella*, *Shigella*, *Mycobacterium*, *Yersinia* and *Francisella*. Pathogens inside BCVs can manipulate host cells pathways by interacting with particular targets through secretion of effector molecules. For instance, *Salmonella* induces its own engulfment in epithelial cells and is found in a membrane-bound compartment, known as the *Salmonella*-containing vacuole (SCV). This compartment is specially adapted by the bacteria to allow its own survival and replication [62]. By means of subcellular fractionation of infected cells using sucrose-density centrifugation, SCVs can be fairly purely isolated with subsequent recovery of viable bacteria. In the work of Santos *et al.* [63], the authors report a recovery of about 30 million bacterial cells when infecting 60 million HeLa cells at a MOI of 100. Inclusion of an immunoprecipitation (IP) step, such as the SseF IP performed in [64] and targeting a *Salmonella* secreted effector residing in the SCV membrane, might also be considered as a bacterial enrichment method as it yields purer SCV fractions. However, viewing the losses associated with such a workflow, the input material required is high and the sample manipulations costly and labor intensive making it not very suitable for downstream proteomics analysis [64]. Finally, in the particular case of *Salmonella* which can survive and replicate in the cytosol of infected cells [65] [22], a drawback of this approach is that it only enriches the *Salmonella* population inside SCVs, thus neglecting the (possible) contribution of cytosolic bacteria to the overall infection process.

Using FACS based isolation, Becker and coworkers purified GFP expressing *Salmonella* from *ex vivo* infected mouse spleen and caecum cells [66]. In their study, 370 *Salmonella* proteins could be identified from spleen isolates and 835 proteins from caecum. Nevertheless, host contaminants were frequently observed in both samples as from all identified proteins in the spleen and caecum samples 70% and 25% originated from host proteins respectively.

Another strategy uses immune (magnetic) selection (I(M)S) of bacteria. Again, this strategy relies on the isolation of the microorganisms after selective lysis of host cells by detergents, followed by the use of specific antibodies raised against a particular microorganism [67]. These antibodies are obtained by immunization of animals and collection of antisera against accessible epitopes presented by the pathogen. Twine *et al.* used IMS [67] to isolate *Francisella* bacteria. In their study, spleen homogenates from infected mice were incubated with magnetic beads coupled with anti-*Francisella* rabbit sera which allowed for the isolation of up to ~49% of all bacteria (80 million bacterial cells) present in the spleen homogenate [67].

Even though physical isolation-based methods such as FACS and I(M)S, allow separating bacteria from the host upon infection, their application is biased and restricted by several aspects, such as the ability to genetically manipulate the microorganisms, the capacity to obtain specific antibodies and the amount of sample needed. For instance, genetic manipulation of the Gram-positive pathogen *Staphylococcus aureus* to express fluorescent proteins cannot be easily achieved [68]. Moreover, the scarce availability of antibodies targeting specific pathogens limits the wide applicability of immunoselection. An alternative approach is based on labeling bacteria with gold (Au) or ferric oxide-core (FeOx) poly (vinyl alcohol) coated fluorescence-labeled nanoparticles (NP) [69]. These NPs are internalized by *S. aureus* prior to infection and allow their subsequent isolation by FACS or magnetic separation. Noteworthy however, NP engulfment has been shown to affect the normal growth of the bacteria as a longer lag phase was observed due to bacterial adaptation to the NPs. To further characterize this, the authors used quantitative proteome analysis to analyze the protein expression profile of bacteria upon incubation with or without NPs. Several proteins with different expression profiles were observed this way. Moreover, upregulation of proteins involved in the anti-oxidative stress response indicates that *S. aureus* is exposed to mild oxidative stress when grown in the presence of FeOx-NP. To conclude, all these strategies are powerful methods and have been successfully applied to study host-pathogen interactions. Nevertheless as reported in several studies, optimization is still needed to increase the purity of the bacterial proteins obtained.

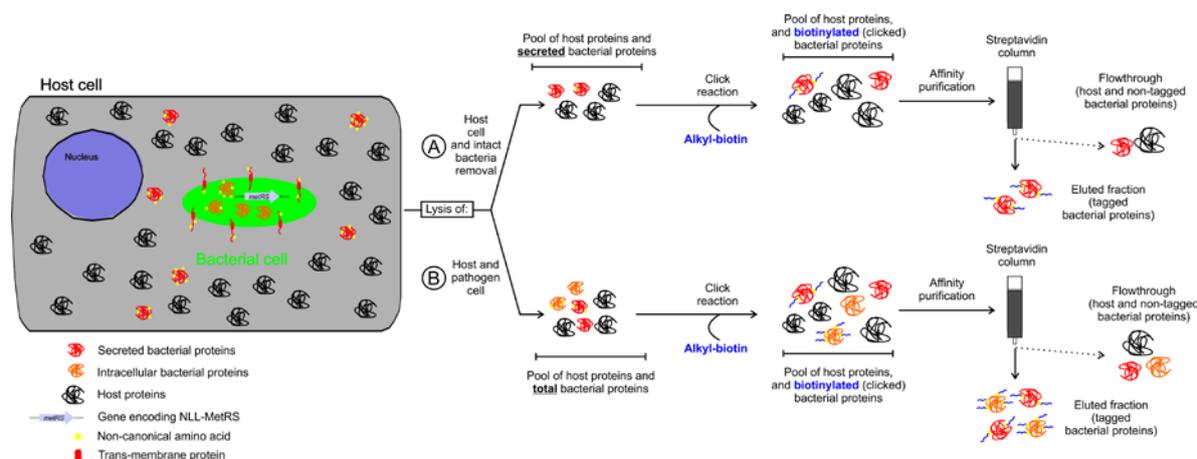
#### 4.2 Protein labeling-based isolation of the bacterial proteome in infection models

*In cellulo* bacterial proteome labeling is typically achieved using non-natural amino acids that are exclusively incorporated into the bacterial but not the eukaryotic proteome (Figure 3A). Given that selective labeling of bacterial proteins occurs before lysis of infected cells, such approaches do not require *a priori* physical separation of bacteria from host cells in contrast to the strategies described above [70, 71]. For a comprehensive review of labeling approaches, we refer to Tirrell *et al.* [72].

Typically, functional groups of non-natural amino acids incorporated into bacterial proteins by the translation machinery enable specific derivatization of labeled proteins and their subsequent isolation [72]. Some of the non-natural amino acids used such as the azide-bearing methionine surrogate azido homoalanine (AHA), can be loaded to methionyl-tRNA following wild-type aminoacyl-tRNA synthetase (aaRS) activation of AHA. Unlike AHA however, trans-crotylglycine (Tcg) or azidonorleucine (ANL), other methionine surrogates, are not activated (efficiently) by wild-type aminoacyl-tRNA synthetases and respectively require overexpression of methionyl-tRNA synthetase, or the expression of a mutant aaRS [72], enabling the (almost) exclusive incorporation in bacterial proteins. One area of research where these approaches have been applied is the selective isolation and identification of the newly synthesized proteome [73, 74]. More specifically, when AHA is added to the media for a (short) pulse, it will first be incorporated into newly synthesized proteins, an approach referred to as biorthogonal noncanonical amino acid tagging (BONCAT) [73]. Since AHA contains a biorthogonal azido group, a chemical group that does not interfere with native biochemical processes nor has any biological activity, upon extraction of the total proteome content, specific labeling of proteins containing AHA can be obtained by azide-alkyne cycloaddition also known as click chemistry [75]. Here, the alkyne group can be conjugated to different molecules, like fluorophores or biotin, which allows fluorescent based inspection (e.g. microscopy) or isolation by affinity purification respectively. Nevertheless, when studying host-pathogen interactions, labeling with AHA is not applicable because both the pathogen and host proteomes may incorporate this surrogate, as explained above.

A way to overcome this issue came with the discovery of variants of aaRS that can activate specific non-natural amino acids [76]. Since such aaRS mutants can be selectively expressed in bacteria but not in the host, which allows for selective labeling of bacterial proteomes in the context of the infected host [70, 71]. More specifically, Ngo *et al.* [70] used an *Escherichia coli* strain constitutively expressing mutant methionyl-tRNA synthetase (i.e. NLL-MetRS) [76] to infect macrophages in medium supplemented with ANL. After lysis of the infected cells, proteins were clicked with alkyne-functionalized biotin. Biotin-clicked bacterial proteins were subsequently enriched by streptavidin-aided affinity purification. In another study, Grammel and coworkers [71] ectopically expressed *E. coli* NLL-MetRS in *Salmonella*, which also enabled ANL incorporation into *Salmonella* proteins. Additionally, the authors used the alkynyl isostere of ANL, 2-amino-octynoic (AOA) instead of ANL [71], and showed that, in their experimental conditions, AOA resulted in a lower background signal typically resulting from non-specific cross-reactions. In this study, different AOA pulses were used to study the alteration of the *Salmonella* proteome in infected macrophages. Despite the fact that 85% of all protein identifications originated from *Salmonella*, only 195 *Salmonella* proteins were identified [71]. Even though biorthogonal labelling represents an alternative approach to avoid physical isolation methods, its implementation has the limitation that only proteins containing methionines will be labelled. In the case of *Salmonella*, 96% of its proteome contains one or more Met residues in addition to the N-terminal initiator Met residue. Moreover, the incorporation

of non-natural amino acids into the proteins can have a metabolic effect as shown for the intracellular parasite *Toxoplasma gondii* [77], and therefore might affect the infection process.



**Fig. 3 | Bio-orthogonal labeling and purification of labeled bacterial (secreted) proteins from infected host cells.** The scheme depicts the workflow for bio-orthogonal tagging of bacterial proteins with non-natural amino acids followed by click chemistry, and isolation of labeled pathogen proteins. Host cells are infected with bacteria, non-natural amino acids are added to the medium prior to infection (pre-labeled bacteria), or upon infection. Bacteria express NLL-MetRS aminoacyl-tRNA synthetase, which can incorporate non-natural Met analogues exclusively into bacterial proteins. Newly synthesized bacterial proteins will thus incorporate this non-natural Met. After infection, host cells are lysed either in (A) a selective manner keeping bacterial cells intact (e.g. to study the bacterial secretome) or (B) by complete lysis enabling to study the proteome of the pathogen and host simultaneously during infection in a more global manner. Particularly, in case of (A), bacterial cells are typically removed by centrifugation before downstream processing of the obtained lysate. Subsequently, proteins are clicked using alkyl-biotin (blue). In this step, only proteins that have incorporated non-natural amino acid will be conjugated to biotin. Finally, a purification step using streptavidin affinity chromatography is performed. Here, non-tagged proteins do not bind the streptavidin resin and are thus easily removed, while the biotinylated (clicked) proteins can be recovered. Thus, bacterial (secreted) proteins are highly enriched in the sample and can be studied to understand changes in the secretome (A) or in the bacterial proteome (B) of a bacterial pathogen during infection.

### 5. Secretomics for identifying bacterial extracellular effectors involved in infection

The term bacterial secretome was coined in studies of the eubacteria *Bacillus subtilis* by Tjalsma *et al.* in 2000 [78] and Antelmann *et al.* in 2001 [79]. It was initially used to refer to both the secreted proteins and the actual secretion machinery. However, the current paradigm in the field has shifted to define the secretome in a broader way as not only the subset of secreted proteins or so-called exoproteome [80], but also including bacterial proteins that interact with the extracellular environment such as the membrane-anchored adhesins which facilitate adherence to the surface or to host cells [81].

Bacterial secretomes have been studied by different approaches including *in silico* identification of secretion signals, isolation of the secretomes of *in vitro* cultured microorganisms under different growth conditions and screening of proteins by phage display. *In silico* based approaches rely on algorithms to find consensus secretion signals characteristic of secreted proteins [82]. Secreted proteins are usually targeted to the secretory pathway by a signal either encoded in the N-terminal

region of the protein or by the 5' end of the mRNA sequence [83]. While *in silico* secretome predictions enable the rapid discovery of potential secreted proteins, empiric validation of candidates is required to show that these proteins indeed belong to the secretome. Such validation is not always straightforward, amongst others because proteins predicted to be secreted might be poorly expressed. Further, this approach relies on proper genome annotation which is not always available. In addition, *in vitro* techniques typically identify secreted proteins in culture supernatants after physical separation from the bacteria [84-89]. However, in such studies, bacteria are grown in conditions that only mimic the host environment. Hence, the actual set of proteins secreted upon interaction with the host might be different from the protein set obtained *in vitro*. Phage display is based on the generation of a library containing up to  $10^{12}$  phage particles displaying different bacterial protein(s) (parts) [90]. This strategy for studying the secretome can be used when proteins interacting with host cells, the extracellular matrix, serum proteins such as antibodies and the complement system are used as baits to fish out those bacterial proteins that are potentially secreted or membrane associated [91]. Therefore, the use of phage display for the identification of secretome proteins is limited to those proteins interacting with host surfaces (reviewed by Gagic *et al.* [90]), thus missing the intracellular targets of secreted proteins.

Recently however, enrichment of bacterial secretomes using *in vitro* infection models was reported [92, 93]. Both studies used biorthogonal labeling of bacterial proteins with ANL to isolate the secretome (Figure 3B). Mahdavi and coworkers [92] studied the secretome of *Yersinia enterocolitica* infected HeLa cells. In this work, effectors secreted by extracellular or intracellular *Yersinia* were identified. At the early stage of infection, *Yersinia* is extracellular and uses a type III secretion system (T3SS) to inject effectors into the host. Interestingly, infection can be detected due to the changed morphology of HeLa cells after injection of the effectors. In their experimental setup, *Yersinia* constitutively expressing NLL-MetRS was used to infect HeLa cells in media supplemented with ANL. Upon selective lysis of the host and removal of intact bacteria, effectors injected in host cells were enriched and subsequently identified by mass spectrometry. A similar strategy, but now including a step to eliminate extracellular bacteria by a gentamicin protection assay, was performed in [21]. This way, upon lysis of the host cells, protein secreted by intracellular bacteria could be identified. The outcome of these two strategies, was the identification of two different groups of secreted proteins; namely 8 T3SS-effectors (3 secreted by extracellular and 5 by intracellular *Yersinia*) and 5 effectors not secreted by T3SS as demonstrated when making use of a mutant strain deficient in type III secretion. Finally, pulses with ANL allowed for a time-lapse study of the secreted effectors during the process of *Yersinia* infection and the order of the different injected proteins could be established. Another study carried out by Chande and coworkers [93] also used ANL incorporation into mycobacterial proteins to identify secreted effectors by different virulent *Mycobacterium* strains infecting THP-1 monocytes at different time points. Here, a *Mycobacterium* codon-adapted NLL-MetRS was used and the bacteria labeled 12 h prior to infection. Presumably, pre-labeling is required due to the slow growth rate of *Mycobacterium* [94]. By using label-free proteomics, the authors could

identify 15 to 30 secreted proteins in three different strains, and quantified differences in their abundance at different time points. Of note, the time points performed in this study differed 12 h in time, while the time points of *Yersinia* study described above were collected 30 min apart [92], again highly likely in line with the different growth rates exhibited by these pathogens.

6. Ribosome profiling as the next big step of future endeavors for studying bacterial translomes upon host cell infection?

Despite the recent advances made in studying bacterial protein synthesis by means of proteomics, both the difficulty of distinguishing newly translated and nascent proteins from the pre-existing protein pools (e.g. when identifying newly translated bacterial proteins upon infection) next to the increasing capabilities of NGS in terms of throughput and sensitivity, spurred an awareness to start analyzing bacterial protein translation at the level of the transcriptome/genome.

Only recently, ribosome profiling (or Ribo-seq) revolutionized the study of translation by deep sequencing of ribosome-protected mRNA fragments [95], thereby expanding our current understanding of the genomic architecture and coding potential for a wide variety of organisms to an unprecedented extent [96]. Moreover, recent ribosome profiling studies performed by us and others enabled translome studies of bacterial cells, not only revealing translation of numerous previously unidentified (small) open reading frames, so-called pseudogenes and translation from numerous alternative translation initiation events including near-cognate start sites [97-100], but also empowering differential gene expression studies under infection relevant conditions [101].

Unlike the multitude of viral translation studies reported in the context of infection [102], and likely complicated due to the significant differences between bacterial and eukaryotic translation, to date, no bacterial translome-oriented studies in the context of infection have been reported. Interestingly however, unlike strategies relying on physical methods to separate cell types and cell enrichment strategies, genetically encoded systems were reported studying mRNA translation in different cell (sub-)populations [103, 104]. More specifically, by expressing an epitope tagged ribosomal protein in a cell-dependent manner or under a tissue-specific promoter, translating ribosome affinity purification (TRAP) and RiboTag approaches enabled tissue or cell-type specific profiling of translomes by the selective recovery (i.e. immunopurification) of tagged polysome-associated mRNAs from cells expressing a tagged ribosomal subunit [105, 106].

In principle, similar strategies making use of ribosome tagged bacterial strains [107] would permit selective isolation of bacterial ribosomes when studying the bacterial translome in a host cell specific context. Moreover, as in the case of the multi RNA-seq strategies described above, it is still highly preferable to enrich for the desired subset of bacterial transcripts, especially in the context of newly translated proteins. Again, *in silico* discrimination between host and pathogen can simply be performed by mapping the data onto respective reference genomes.

Since for most bacterial pathogens, the complete repertoire of effectors employed to establish a successful infection remains unknown, by performing integrative analysis of gene and protein expression data, a more qualitative and refined set of true protein coding potential of the pathogen in its host context can be envisaged. Furthermore, by providing an unprecedented understanding of interactions between pathogen and host, the identification of potentially new pathogen virulence factors will be facilitated.

### Acknowledgments

The authors would like to thank Ignacio Santeccchia for valuable discussions and help with preparation of the figures.

### Conflict of interest

The authors declare no conflict of interest.

### References

1. Methot, P.O. and S. Alizon, *What is a pathogen? Toward a process view of host-parasite interactions. Virulence*, 2014. **5**(8): p. 775-85.
2. Falkow, S., *Molecular Koch's postulates applied to bacterial pathogenicity [mdash] a personal recollection 15 years later. Nat Rev Micro*, 2004. **2**(1): p. 67-72.
3. Casadevall, A. and L.-a. Pirofski, *The damage-response framework of microbial pathogenesis. Nat Rev Micro*, 2003. **1**(1): p. 17-24.
4. Jean Beltran, P.M., et al., *Proteomics and integrative omic approaches for understanding host-pathogen interactions and infectious diseases. 2017. 13*(3): p. 922.
5. Qi, L., et al., *Quantitative proteomic analysis of host epithelial cells infected by Salmonella enterica serovar Typhimurium. Proteomics*, 2017.
6. Chromy, B.A., et al., *Proteomic characterization of host response to Yersinia pestis and near neighbors. Biochem Biophys Res Commun*, 2004. **320**(2): p. 474-9.
7. Li, P., et al., *Comparative Proteomics Analysis of Human Macrophages Infected with Virulent Mycobacterium bovis. Front Cell Infect Microbiol*, 2017. **7**: p. 65.
8. Zhu, L., et al., *The proteome of Shigella flexneri 2a 2457T grown at 30 and 37 degrees C. Mol Cell Proteomics*, 2010. **9**(6): p. 1209-20.
9. Stancik, L.M., et al., *pH-dependent expression of periplasmic proteins and amino acid catabolism in Escherichia coli. J Bacteriol*, 2002. **184**(15): p. 4246-58.
10. Weber, A., S.A. Kogl, and K. Jung, *Time-dependent proteome alterations under osmotic stress during aerobic and anaerobic growth in Escherichia coli. J Bacteriol*, 2006. **188**(20): p. 7165-75.
11. Kim, K., et al., *Mass spectrometry-based quantitative proteomic analysis of Salmonella enterica serovar Enteritidis protein expression upon exposure to hydrogen peroxide. BMC Microbiology*, 2010. **10**: p. 166-166.
12. Alberts, B., *Molecular biology of the cell*. 2015.
13. Karlin, S., et al., *Heterogeneity of Genome and Proteome Content in Bacteria, Archaea, and Eukaryotes. Theoretical Population Biology*, 2002. **61**(4): p. 367-390.

14. Wang, M., et al., *PaxDb, a database of protein abundance averages across all three domains of life*. *Mol Cell Proteomics*, 2012. **11**(8): p. 492-500.
15. Smith, L.M. and N.L. Kelleher, *Proteoform: a single term describing protein complexity*. *Nat Methods*, 2013. **10**(3): p. 186-7.
16. Plasman, K., et al., *Perfecting Progress*. Same dan Ltd Pharmaceutical Publishers, 2017: p. 70- 72.
17. Bekker-Jensen, D.B., et al., *An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes*. *Cell Syst*, 2017. **4**(6): p. 587-599.e4.
18. Buchrieser, C., et al., *The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of Shigella flexneri*. *Mol Microbiol*, 2000. **38**(4): p. 760-71.
19. Westermann, A.J. and L. Barquist, *Resolving host-pathogen interactions by dual RNA-seq*. 2017. **13**(2): p. e1006033.
20. Mastroeni, P., et al., *A dynamic view of the spread and intracellular distribution of Salmonella enterica*. *Nat Rev Microbiol*, 2009. **7**(1): p. 73-80.
21. Kaneko, M., Y. Emoto, and M. Emoto, *A Simple, Reproducible, Inexpensive, Yet Old-Fashioned Method for Determining Phagocytic and Bactericidal Activities of Macrophages*. *Yonsei Medical Journal*, 2016. **57**(2): p. 283-290.
22. Malik-Kale, P., S. Winfree, and O. Steele-Mortimer, *The bimodal lifestyle of intracellular Salmonella in epithelial cells: replication in the cytosol obscures defects in vacuolar replication*. *PLoS One*, 2012. **7**(6): p. e38732.
23. Raybourne, R.B., et al., *Uptake and killing of Listeria monocytogenes by normal human peripheral blood granulocytes and monocytes as measured by flow cytometry and cell sorting*. *FEMS Immunol Med Microbiol*, 2001. **31**(3): p. 219-25.
24. Chalfie, M., et al., *Green fluorescent protein as a marker for gene expression*. *Science*, 1994. **263**(5148): p. 802-5.
25. Surmann, K., et al., *Comparative proteome analysis reveals conserved and specific adaptation patterns of Staphylococcus aureus after internalization by different types of human non-professional phagocytic host cells*. *Frontiers in Microbiology*, 2014. **5**: p. 392.
26. Schmidt, F., et al., *Time-resolved quantitative proteome profiling of host-pathogen interactions: the response of Staphylococcus aureus RN1HG to internalisation by human airway epithelial cells*. *Proteomics*, 2010. **10**(15): p. 2801-11.
27. Burton, N.A., et al., *Disparate impact of oxidative host defenses determines the fate of Salmonella during systemic infection in mice*. *Cell Host Microbe*, 2014. **15**(1): p. 72-83.
28. Knodler, L.A., et al., *Cloning Vectors and Fluorescent Proteins Can Significantly Inhibit Salmonella enterica Virulence in Both Epithelial Cells and Macrophages: Implications for Bacterial Pathogenesis Studies*. *Infection and Immunity*, 2005. **73**(10): p. 7027-7031.
29. Clark, L., et al., *GFP plasmid-induced defects in Salmonella invasion depend on plasmid architecture, not protein expression*. *Microbiology*, 2009. **155**(Pt 2): p. 461-7.
30. Westermann, A.J., et al., *Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions*. *Nature*, 2016. **529**(7587): p. 496-501.
31. Bi, Y., et al., *Yersinia pestis versus Yersinia pseudotuberculosis: effects on host macrophages*. *Scand J Immunol*, 2012. **76**(6): p. 541-51.
32. Arnett, E., et al., *The pore-forming toxin listeriolysin O is degraded by neutrophil metalloproteinase-8 and fails to mediate Listeria monocytogenes intracellular survival in neutrophils*. *J Immunol*, 2014. **192**(1): p. 234-44.

33. Gog, J.R., et al., *Dynamics of Salmonella infection of macrophages at the single cell level*. Journal of The Royal Society Interface, 2012.
34. Jantsch, J., et al., *Intracellular activities of Salmonella enterica in murine dendritic cells*. Cell Microbiol, 2003. 5(12): p. 933-45.
35. Ibarra, J.A., et al., *Induction of Salmonella pathogenicity island 1 under different growth conditions can affect Salmonella–host cell interactions in vitro*. Microbiology, 2010. 156(Pt 4): p. 1120-1133.
36. Geier, H. and J. Celli, *Phagocytic Receptors Dictate Phagosomal Escape and Intracellular Proliferation of Francisella tularensis*. Infection and Immunity, 2011. 79(6): p. 2204-2214.
37. Kolb-Maurer, A., et al., *Listeria monocytogenes-infected human dendritic cells: uptake and host cell response*. Infect Immun, 2000. 68(6): p. 3680-8.
38. Fodor, S.P., et al., *Multiplexed biochemical assays with biological chips*. Nature, 1993. 364(6437): p. 555-6.
39. Veres, G., et al., *The molecular basis of the sparse fur mouse mutation*. Science, 1987. 237(4813): p. 415-7.
40. Rotter, A., et al., *Finding Differentially Expressed Genes in Two-Channel DNA Microarray Datasets: How to Increase Reliability of Data Preprocessing*. OMICS : a Journal of Integrative Biology, 2008. 12(3): p. 171-182.
41. Hossain, H., S. Tchatalbachev, and T. Chakraborty, *Host gene expression profiling in pathogen-host interactions*. Curr Opin Immunol, 2006. 18(4): p. 422-9.
42. Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics*. Nature reviews. Genetics, 2009. 10(1): p. 57-63.
43. Westermann, A.J., S.A. Gorski, and J. Vogel, *Dual RNA-seq of pathogen and host*. Nat Rev Microbiol, 2012. 10(9): p. 618-30.
44. Zhao, S., et al., *Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells*. PLoS ONE, 2014. 9(1): p. e78644.
45. Bloom, J.S., et al., *Measuring differential gene expression by short read sequencing: quantitative comparison to 2-channel gene expression microarrays*. BMC Genomics, 2009. 10: p. 221.
46. Szafranska, A.K., et al., *High-resolution transcriptomic analysis of the adaptive response of Staphylococcus aureus during acute and chronic phases of osteomyelitis*. MBio, 2014. 5(6).
47. Mavromatis, C.H., et al., *The co-transcriptome of uropathogenic Escherichia coli-infected mouse macrophages reveals new insights into host-pathogen interactions*. Cell Microbiol, 2015. 17(5): p. 730-46.
48. Humphrys, M.S., et al., *Simultaneous transcriptional profiling of bacteria and their host cells*. PLoS One, 2013. 8(12): p. e80597.
49. Rienksma, R.A., et al., *Comprehensive insights into transcriptional adaptation of intracellular mycobacteria by microbe-enriched dual RNA sequencing*. BMC Genomics, 2015. 16: p. 34.
50. Baddal, B., et al., *Dual RNA-seq of Nontypeable Haemophilus influenzae and Host Cell Transcriptomes Reveals Novel Insights into Host-Pathogen Cross Talk*. MBio, 2015. 6(6): p. e01765-15.
51. Damron, F.H., et al., *Dual-seq transcriptomics reveals the battle for iron during Pseudomonas aeruginosa acute murine pneumonia*. Scientific Reports, 2016. 6: p. 39172.
52. Nuss, A.M., et al., *Tissue dual RNA-seq allows fast discovery of infection-specific functions and riboregulators shaping host–pathogen transcriptomes*. Proceedings of the National Academy of Sciences, 2017. 114(5): p. E791-E800.
53. Silva, M., *Classical Labeling of Bacterial Pathogens According to Their Lifestyle in the Host: Inconsistencies and Alternatives*. Frontiers in Microbiology, 2012. 3(71).
54. Liu, Y., et al., *Proteomic Analyses of Intracellular Salmonella enterica Serovar Typhimurium Reveal Extensive Bacterial Adaptations to Infected Host Epithelial Cells*. Infection and Immunity, 2015. 83(7): p. 2897-2906.

55. Pieper, R., et al., *The Shigella dysenteriae serotype 1 proteome, profiled in the host intestinal environment, reveals major metabolic modifications and increased expression of invasive proteins*. *Proteomics*, 2009. **9**(22): p. 5029-45.
56. Pieper, R., et al., *Analysis of the proteome of intracellular Shigella flexneri reveals pathways important for intracellular growth*. *Infect Immun*, 2013. **81**(12): p. 4635-48.
57. Van de Velde, S., et al., *Isolation and 2-D-DIGE proteomic analysis of intracellular and extracellular forms of Listeria monocytogenes*. *Proteomics*, 2009. **9**(24): p. 5484-96.
58. Garcia-del Portillo, F., et al., *Association of ActA to peptidoglycan revealed by cell wall proteomics of intracellular Listeria monocytogenes*. *J Biol Chem*, 2011. **286**(40): p. 34675-89.
59. Donaldson, J.R., et al., *Proteomic expression profiles of virulent and avirulent strains of Listeria monocytogenes isolated from macrophages*. *J Proteomics*, 2011. **74**(10): p. 1906-17.
60. Ponnusamy, D., S.D. Hartson, and K.D. Clinkenbeard, *Intracellular Yersinia pestis expresses general stress response and tellurite resistance proteins in mouse macrophages*. *Vet Microbiol*, 2011. **150**(1-2): p. 146-51.
61. Al Dahouk, S., et al., *Quantitative analysis of the Brucella suis proteome reveals metabolic adaptation to long-term nutrient starvation*. *BMC Microbiology*, 2013. **13**: p. 199-199.
62. Steele-Mortimer, O., *The Salmonella-containing Vacuole – Moving with the Times*. *Current opinion in microbiology*, 2008. **11**(1): p. 38-45.
63. Santos, J.C., et al., *The COPII complex and lysosomal VAMP7 determine intracellular Salmonella localization and growth*. *Cell Microbiol*, 2015. **17**(12): p. 1699-720.
64. Vorwerk, S., et al., *Proteomes of host cell membranes modified by intracellular activities of Salmonella enterica*. *Mol Cell Proteomics*, 2015. **14**(1): p. 81-92.
65. Knodler, L.A., et al., *Dissemination of invasive Salmonella via bacterial-induced extrusion of mucosal epithelia*. *Proceedings of the National Academy of Sciences of the United States of America*, 2010. **107**(41): p. 17733-17738.
66. Becker, D., et al., *Robust Salmonella metabolism limits possibilities for new antimicrobials*. *Nature*, 2006. **440**(7082): p. 303-307.
67. Twine, S.M., et al., *In vivo proteomic analysis of the intracellular bacterial pathogen, Francisella tularensis, isolated from mouse spleen*. *Biochem Biophys Res Commun*, 2006. **345**(4): p. 1621-33.
68. Monk, I.R. and T.J. Foster, *Genetic manipulation of Staphylococci—breaking through the barrier*. *Frontiers in Cellular and Infection Microbiology*, 2012. **2**: p. 49.
69. Depke, M., et al., *Labeling of the pathogenic bacterium Staphylococcus aureus with gold or ferric oxide-core nanoparticles highlights new capabilities for investigation of host-pathogen interactions*. *Cytometry A*, 2014. **85**(2): p. 140-50.
70. Ngo, J.T., et al., *Cell-selective metabolic labeling of proteins*. *Nat Chem Biol*, 2009. **5**(10): p. 715-717.
71. Grammel, M., M.M. Zhang, and H.C. Hang, *Orthogonal alkynyl-amino acid reporter for selective labeling of bacterial proteomes during infection*. *Angewandte Chemie (International ed. in English)*, 2010. **49**(34): p. 5970-5974.
72. Ngo, J.T. and D.A. Tirrell, *Noncanonical amino acids in the interrogation of cellular protein synthesis*. *Acc Chem Res*, 2011. **44**(9): p. 677-85.
73. Dieterich, D.C., et al., *Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT)*. *Proc Natl Acad Sci US A*, 2006. **103**(25): p. 9482-7.

74. Dieterich, D.C., et al., *Labeling, detection and identification of newly synthesized proteomes with bioorthogonal non-canonical amino-acid tagging*. *Nat Protoc*, 2007. **2**(3): p. 532-40.
75. Hong, V., et al., *Analysis and optimization of copper-catalyzed azide-alkyne cycloaddition for bioconjugation*. *Angew Chem Int Ed Engl*, 2009. **48**(52): p. 9879-83.
76. Tanrikulu, I.C., et al., *Discovery of Escherichia coli methionyl-tRNA synthetase mutants for efficient labeling of proteins with azidonorleucine in vivo*. *Proc Natl Acad Sci U S A*, 2009. **106**(36): p. 15285-90.
77. Wier, G.M., et al., *New method for the orthogonal labeling and purification of Toxoplasma gondii proteins while inside the host cell*. *MBio*, 2015. **6**(2): p. e01628.
78. Tjalsma, H., et al., *Signal Peptide-Dependent Protein Transport in Bacillus subtilis: a Genome-Based Survey of the Secretome*. *Microbiology and Molecular Biology Reviews*, 2000. **64**(3): p. 515-547.
79. Antelmann, H., et al., *A proteomic view on genome-based signal peptide predictions*. *Genome Res*, 2001. **11**.
80. Tjalsma, H., *Feature-based reappraisal of the Bacillus subtilis exoproteome*. *Proteomics*, 2007. **7**(1): p. 73-81.
81. Desvaux, M., et al., *Secretion and subcellular localizations of bacterial proteins: a semantic awareness issue*. *Trends Microbiol*, 2009. **17**(4): p. 139-45.
82. Cornejo-Granados, F., et al., *Secretome Prediction of Two M. tuberculosis Clinical Isolates Reveals Their High Antigenic Density and Potential Drug Targets*. *Frontiers in Microbiology*, 2017. **8**: p. 128.
83. Samudrala, R., F. Heffron, and J.E. McDermott, *Accurate Prediction of Secreted Substrates and Identification of a Conserved Putative Secretion Signal for Type III Secretion Systems*. *PLoS Pathogens*, 2009. **5**(4): p. e1000375.
84. Zheng, J., et al., *Analysis of the Secretome and Identification of Novel Constituents from Culture Filtrate of Bacillus Calmette-Guérin Using High-resolution Mass Spectrometry*. *Molecular & Cellular Proteomics : MCP*, 2013. **12**(8): p. 2081-2095.
85. Trost, M., et al., *Comparative proteome analysis of secretory proteins from pathogenic and nonpathogenic Listeria species*. *Proteomics*, 2005. **5**(6): p. 1544-57.
86. Niemann, G.S., et al., *Discovery of Novel Secreted Virulence Factors from Salmonella enterica Serovar Typhimurium by Proteomic Analysis of Culture Supernatants*. *Infection and Immunity*, 2011. **79**(1): p. 33-43.
87. Deng, W., et al., *Quantitative Proteomic Analysis of Type III Secretome of Enteropathogenic Escherichia coli Reveals an Expanded Effector Repertoire for Attaching/Effacing Bacterial Pathogens*. *Molecular & Cellular Proteomics : MCP*, 2012. **11**(9): p. 692-709.
88. Chinison, J.J., et al., *Identification of Mycobacterium avium subsp. hominissuis secreted proteins using an in vitro system mimicking the phagosomal environment*. *BMC Microbiology*, 2016. **16**: p. 270.
89. Dreisbach, A., et al., *Profiling the surfacome of Staphylococcus aureus*. *Proteomics*, 2010. **10**(17): p. 3082-96.
90. Gagic, D., et al., *Exploring the Secretomes of Microbes and Microbial Communities Using Filamentous Phage Display*. *Frontiers in Microbiology*, 2016. **7**: p. 429.
91. Jankovic, D., et al., *Direct selection and phage display of a Gram-positive secretome*. *Genome Biology*, 2007. **8**(12): p. R266-R266.
92. Mahdavi, A., et al., *Identification of secreted bacterial proteins by noncanonical amino acid tagging*. *Proceedings of the National Academy of Sciences of the United States of America*, 2014. **111**(1): p.433-438.
93. Chande, A.G., et al., *Selective enrichment of mycobacterial proteins from infected host macrophages*. *Scientific Reports*, 2015. **5**: p. 13430.

94. Colston, M.J. and R.A. Cox, *Mycobacterial Growth and Dormancy*, in *Mycobacteria*. 2009, Blackwell Publishing Ltd. p. 198-219.
95. Ingolia, N.T., et al., *Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling*. *Science*, 2009. **324**(5924): p. 218-23.
96. Gawron, D., K. Gevaert, and P. Van Damme, *The proteome under translational control*. *Proteomics*, 2014. **14**(23-24): p. 2647-62.
97. Li, G.W., E. Oh, and J.S. Weissman, *The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria*. *Nature*, 2012. **484**(7395): p. 538-41.
98. Oh, E., et al., *Selective ribosome profiling reveals the co-translational chaperone action of trigger factor in vivo*. *Cell*, 2011. **147**(6): p. 1295-1308.
99. Ndah, E., et al., *REPARATION: Ribosome Profiling Assisted (Re-)Annotation of Bacterial genomes*. bioRxiv, 2017.
100. Giess, A., et al., *A Unique Ribosome Signature Reveals Bacterial Translation Initiation Sites*. bioRxiv, 2016.
101. Baek, J., et al., *Identification of Unannotated Small Genes in Salmonella*. *G3 (Bethesda)*, 2017. **7**(3): p. 983-989.
102. Stern-Ginossar, N. and N.T. Ingolia, *Ribosome Profiling as a Tool to Decipher Viral Complexity*. *Annual Review of Virology*, 2015. **2**(1): p. 335-349.
103. Sanz, E., et al., *Cell-type-specific isolation of ribosome-associated mRNA from complex tissues*. *Proceedings of the National Academy of Sciences*, 2009. **106**(33): p. 13939-13944.
104. Heiman, M., et al., *A translational profiling approach for the molecular characterization of CNS cell types*. *Cell*, 2008. **135**(4): p. 738-48.
105. Hornstein, N., et al., *Ligation-free ribosome profiling of cell type-specific translation in the brain*. *Genome Biol*, 2016. **17**(1): p. 149.
106. Gao, X., et al., *Quantitative profiling of initiating ribosomes in vivo*. 2015. **12**(2): p. 147-53.
107. Ederth, J., et al., *A single-step method for purification of active His-tagged ribosomes from a genetically engineered Escherichia coli*. *Nucleic Acids Research*, 2009. **37**(2): p. e15-e15.