Cross-species RNA-seq for deciphering host–microbe interactions

Alexander J. Westermann [1,2] ✉ *and Jörg Vogel* [1,2] ✉

Abstract | The human body is constantly exposed to microorganisms, which entails manifold interactions between human cells and diverse commensal or pathogenic bacteria. The cellular states of the interacting cells are decisive for the outcome of these encounters such as whether bacterial virulence programmes and host defence or tolerance mechanisms are induced. This Review summarizes how next-generation RNA sequencing (RNA-seq) has become a primary technology to study host–microbe interactions with high resolution, improving our understanding of the physiological consequences and the mechanisms at play. We illustrate how the discriminatory power and sensitivity of RNA-seq helps to dissect increasingly complex cellular interactions in time and space down to the single-cell level. We also outline how future transcriptomics may answer currently open questions in host–microbe interactions and inform treatment schemes for microbial disorders.

Interactions between host and bacterial cells form an integral part of human physiology. Colonizing bacteria may attach to mucosal surfaces or bind specifically to host receptors at epithelial linings and must adapt their growth and metabolism to the respective microenvironment. Conversely, to respond adequately to colonizing microorganisms, our immune system must continuously distinguish beneficial bacteria from harmful pathogens. In addition, it is increasingly appreciated that the potentially harmful encounter with a bacterial pathogen is not only determined by the presence or absence of virulence traits in that bacterium but also by when, where and how it interacts with host cells. Furthermore, many infectious diseases are polymicrobial in nature and disorders in the composition of the myriads of bacteria, fungi and viruses that constitute the host-associated microbiota substantially influence the susceptibility or resistance to infection.

Complexity extends to the host whose tissues and organs comprise multiple different cell types that engage in extensive crosstalk. Furthermore, spatial aspects determine the outcome of host–microbe encounters. Pathogens occupy specific niches within their host body, wherein they persist or wherefrom they disseminate, creating vastly different microenvironments. However, even within a defined host niche, phenotypic heterogeneity between genetically identical cell populations can lead to divergent results with sometimes severe effects for the host. The resultant variability in host–microbe interactions might contribute to therapeutic failures and the establishment of chronic, recurring infections[1]. Overall, to fully understand the major principles of host–microbe interactions in health and disease, high-resolution approaches are needed that can capture their full complexity at different scales, from the level of the whole organism down to its individual single cells.

Experimental infection systems — both animal models[2] and advanced 3D in vitro tissues[3] — are improving in their capacity to recapitulate the in vivo situation within human niches. In parallel, the technologies to measure the biological activities of the involved organisms keep improving. For instance, infection biologists can currently draw on different global approaches to study host–microbe encounters at a systems level (BOX 1). Among these approaches, transcriptomic methods record steady-state levels of transcripts and can thereby provide a snapshot of the cellular physiology.

Thanks to their sensitivity, cost-efficiency and generic nature, transcriptomics by next-generation RNA sequencing (RNA-seq) has been popular in studying host–microbe interactions[4]. Since the introduction of RNA-seq to the field of infection biology about a decade ago, there have been three major phases with respect to the analysis of host–microbe interactions (FIG. 1). In the first phase, host and bacterial cells were physically separated from one another and their transcriptomes were analysed individually. The second phase begun with the realization that its high discriminatory power and intrinsic single-nucleotide resolution render RNA-seq an exceptionally powerful technique for the simultaneous detection and quantification of transcripts from different interacting organisms[5]. This was accompanied by a rethink that non-coding transcripts — long regarded as mere by-products of RNA-seq profiling — reveal crucial

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

[2]*Institute for Molecular Infection Biology (IMIB), University of Würzburg, Würzburg, Germany.*

✉*e-mail: alexander.westermann@uni-wuerzburg.de; joerg.vogel@uni-wuerzburg.de*

https://doi.org/10.1038/s41576-021-00326-y

Box 1 | **Orthogonal 'omics' technologies and their use for host–microbe interaction studies**

Genomics and epigenomics

Omics techniques aim at the global detection of biomolecules in an organism to infer biological insights (see the Table). Genomics refers to the determination of the genetic content of an organism by sequencing its genomic DNA. Comparative genomics was harnessed to unravel the similarities and differences within the (virulence) gene content of related bacterial species[198], to study human host adaptation of bacterial pathogens[199], to trace the spread of infectious diseases[200], or to uncover SNPs and insertions/deletions (indels) that render human subjects or whole populations more susceptible to certain infections[201]. However, genomics is also moving away from single-sided studies and striving to set host and pathogen in context by profiling host genetic susceptibility and pathogen genetic variation in conjunction[202].

Whereas the genetic blueprint defines the molecular processes that a given organism is theoretically capable of, it fails to determine the genes that are currently activated under the given condition. Chemical chromatin modifications are marks of the transcriptional activity at a given locus. Technological breakthroughs in deep-sequencing technologies (BOX 2) now enable the detection of modified DNA bases and gave rise to the field of host–pathogen epigenomics[203,204]. More commonly, RNA serves as a direct readout of transcriptional activity; however, it should be noted that the transcriptome of a cell always reflects the sum of de novo transcription and transcript decay. Although mRNA levels are typically used as a proxy for protein expression, there may be cases where this assumption is violated (see main text). Rather, ribosome profiling (Ribo-seq) or mass spectrometry (MS) provide insight into the composition of a cellular proteome at any given point in time.

Translatomics and proteomics

Ribo-seq has long been restricted to cultivatable species and has traditionally been used on pure bacterial cultures[205]; however, technological improvements, particularly with respect to input material constraints, now enable 'MetaRibo-seq' studies of complex bacterial consortia such as the intestinal microbiota[206]. Ribo-seq can also be applied to mammalian cells and even dual Ribo-seq studies have recently been reported, for example, for diverse virus-infected cells[207] and for human fibroblasts infected with a eukaryotic parasite[208]. However, divergence in the technical details, foremost the different compounds needed to stall the prokaryotic and eukaryotic translation apparatus, has so far prevented cross-kingdom Ribo-seq of bacteria-infected host cells.

Whereas Ribo-seq measures global translation rates, label-free and label-based MS methods detect translation products directly, providing an accurate reflection of the composition of a cellular proteome at any given point in time[206]. However, while RNA-seq-based approaches are capable of discovering novel, previously unannotated transcripts, MS generally relies on reference data, limiting proteomics to the detection of known proteins and, owing to technical limitations, conventional proteomics biases heavily against the detection of membrane proteins[209]. Additionally, owing to codon degeneracy, the resolving power of MS is inferior to RNA-seq and, lacking the possibility to amplify peptides prior to MS, current proteomics cannot reach the sensitivity of genomics, transcriptomics or Ribo-seq for which the input nucleic acids may be amplified at will. For example, metaproteomics that profile protein levels in complex bacterial consortia (such as the human faecal microbiome) typically detect only a minimal fraction of the total number of proteins present in that sample (only the most abundant 0.1–0.0001%)[210]. Consequently, host–microbe proteomics — measuring bacterial and eukaryotic protein levels simultaneously — is uncommon. Nevertheless, one-sided infection studies exist that profiled, for example, proteome changes in *Salmonella enterica* subsp. *enterica* serovar Typhimurium during host infection[211] or determined the host factors required for intra-epithelial *Salmonella* replication[212] or cell death of *Salmonella*-infected macrophages[213]. Proteomics methods also lend themselves to mapping protein–protein interaction networks during infection, for example, interactions of secreted *Salmonella* effectors with host target proteins[214]. Thus, proteomic approaches to host–microbe interactions can complement transcriptomics[215] and both RNA-seq and MS are being increasingly used in the clinics, for example, to accelerate the diagnosis of bacterial infections[216,217].

Metabolomics

Not all proteins lingering in a cell may be functioning. Proteomics has the potential to detect post-translational modifications during host–pathogen interactions[182] and so-called activity-based protein profiling detects active enzymes in biological samples and has been successfully applied to shed light on the enzymatic crosstalk of the host and pathogen during enteric infection[218]. Additionally, metabolomics gained attraction as it offers an even more direct readout of a cell's phenotype. Mirroring infection proteomics, host–pathogen metabolomics is a blooming field[219]. One-sided metabolomics was performed to measure host[220] or — upon physical separation — pathogen metabolic flux changes during an infection[221]. Analysing host and microbial metabolism in parallel is challenging given that many core metabolites are shared across kingdoms and cannot be unequivocally assigned to their source organism. This difficulty may be partially overcome by intelligent experimental design that involves the use of axenic cultures or spike-ins, or by complementing host–pathogen metabolomics with orthogonal information. For example, a recent study combined host–pathogen metabolomics with dual RNA-seq during the early infection of human macrophages by *Mycobacterium tuberculosis* and uncovered the robustness of this obligate intracellular pathogen in the face of metabolic interventions[97]. Similarly, metabolomics and metagenomics were measured in parallel in lung samples of patients with cystic fibrosis to reconstruct the metabolic flow during disease[222].

Multi-omics

More generally, the progress made in the individual omics techniques fosters such integrative multi-omics approaches to dissect host–microbe interactions from multiple angles. For instance, in a *Salmonella* infection study of mice, metagenomics, proteomics and metabolomics of faecal samples were combined to link infection with gut microbiota compositional and metabolic changes[223]. Furthermore, in an extremely comprehensive survey, microbial metagenomics, transcriptomics, proteomics and metabolomics were integrated with host expression profiling to unravel the molecular aspects underlying dysbiosis in the human gut microbiome during inflammatory bowel disease[224]. However, integrative omics are demanding new bioinformatic tools to take full advantage of the enormous amount of information buried within the massive datasets[225].

Omics method	Target biomolecules	Key technologies	Sensitivity	Resolution	From genotype to phenotype	Cost per sample
Genomics	Genomic DNA	DNA-seq	High	High	Genetic blueprint	Low
Transcriptomics	mRNA, ncRNA	RNA-seq	High	High	Gene expression (RNA level)	Low
Proteomics	Protein	LC-MS/MS (label based or label free)	Low	Intermediate	Gene expression (protein level)	Intermediate
Metabolomics	Metabolites	MS, NMR (targeted or untargeted)	Low	Low	Direct phenotypic readout	High

Side-by-side comparison of the properties, strengths and limitations of the different omics methods. DNA-seq, DNA sequencing; LC-MS/MS, liquid chromatography–tandem mass spectrometry; mRNA, messenger RNA; MS, mass spectrometry; ncRNA, non-coding RNA; RNA-seq, RNA sequencing.

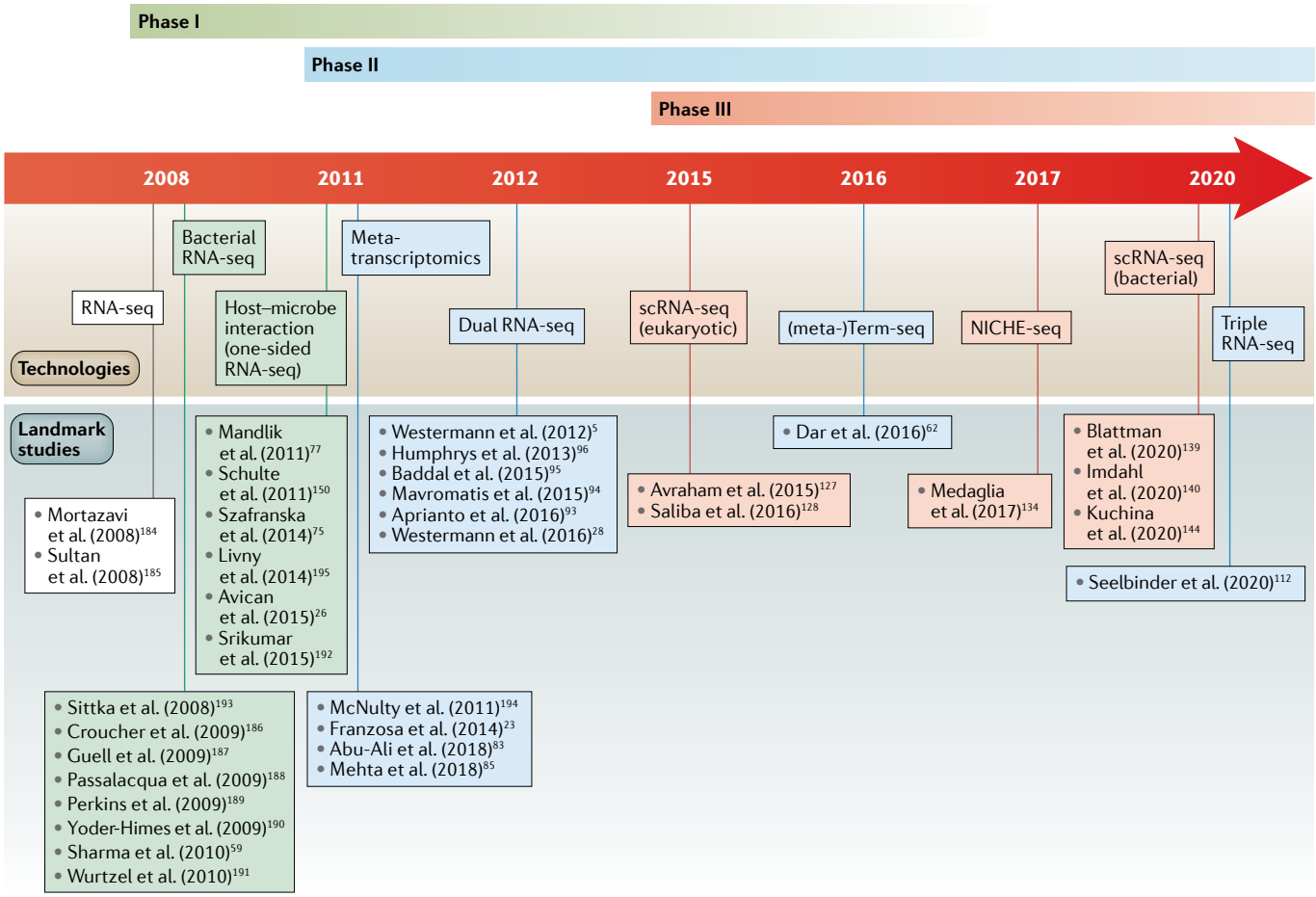

Fig. 1 | **The history of RNA-seq-based infection research.** The timeline can be broadly subdivided into three phases: phase I is characterized by one-sided RNA sequencing (RNA-seq) studies of either host or microorganism, phase II is associated with the concept of sequencing multiple transcriptomes together and phase III was heralded with the introduction of single-cell RNA-seq (scRNA-seq) technology into infection biology. For each phase and each technology, select milestone studies are cited, without claim to completeness. REFS[5,23,26,28,59,62,75,77,83,85,93–96,112,127,128,134,139,140,144,150,184–195].

insights into host–microbe interplay[6–8]. The present third phase is marked by the increasing use of single-cell RNA-seq (scRNA-seq) to dissect cellular heterogeneity in host–microbe encounters[9,10].

Here, we review cross-species transcriptomics, focusing on the interaction of bacterial pathogens or commensals with mammalian (mostly human) host cells. We first summarize the similarities and differences between bacterial and mammalian transcriptomes and transcriptomics, followed by an overview of current cross-species approaches. We bring the technologies to life with examples of the biological insights that they have already provided, including the realization that host–microbe interactions are both context dependent and highly heterogeneous. We also discuss the current limitations of cross-species transcriptomics and how to overcome them, especially against the backdrop of the current scRNA-seq revolution.

Eukaryotic and bacterial transcriptome features

An organism's transcriptome may provide a snapshot of its physiological state. At the cellular level, this transcriptome is an RNA potpourri of myriads of different coding and non-coding transcripts from several major classes, some of which are being made, others are already present in the cell in their active form, and yet others are undergoing decay. Despite two billion years of separate evolution, the fundamental RNA classes are shared between prokaryotes and eukaryotes (FIG. 2). Ribosomal RNAs (rRNAs), serving scaffolding and enzymatic functions in the ribosome, are highly abundant, accounting for >80% (and >95% in fast growing cells) of the RNA content of bacterial and eukaryotic cells alike. Transfer RNA (tRNA) molecules translate RNA language (ribonucleobases) into protein information (amino acid chains) and generally contribute ~10% to the cellular RNA content.

Messenger RNA (mRNA) instructs the ribosome on which protein to synthesize. Making up ~5% of the total cellular transcriptome in both bacteria and eukaryotes, mRNAs are characterized by kingdom-specific features. Bacterial genes are on average 1 kb long; however, they are often transcribed as polycistrons, resulting in long mRNAs that span multiple genes. Primary transcripts typically carry a 5′ triphosphate group, while the 3′ ends of bacterial mRNAs lack an extended poly(A) stretch (3′ polyadenylation does occur, but is limited to a few nucleotides that enable rapid transcript decay[11]).

Axenic cultures
Describes cultures comprised of only a single, defined bacterial species or strain.

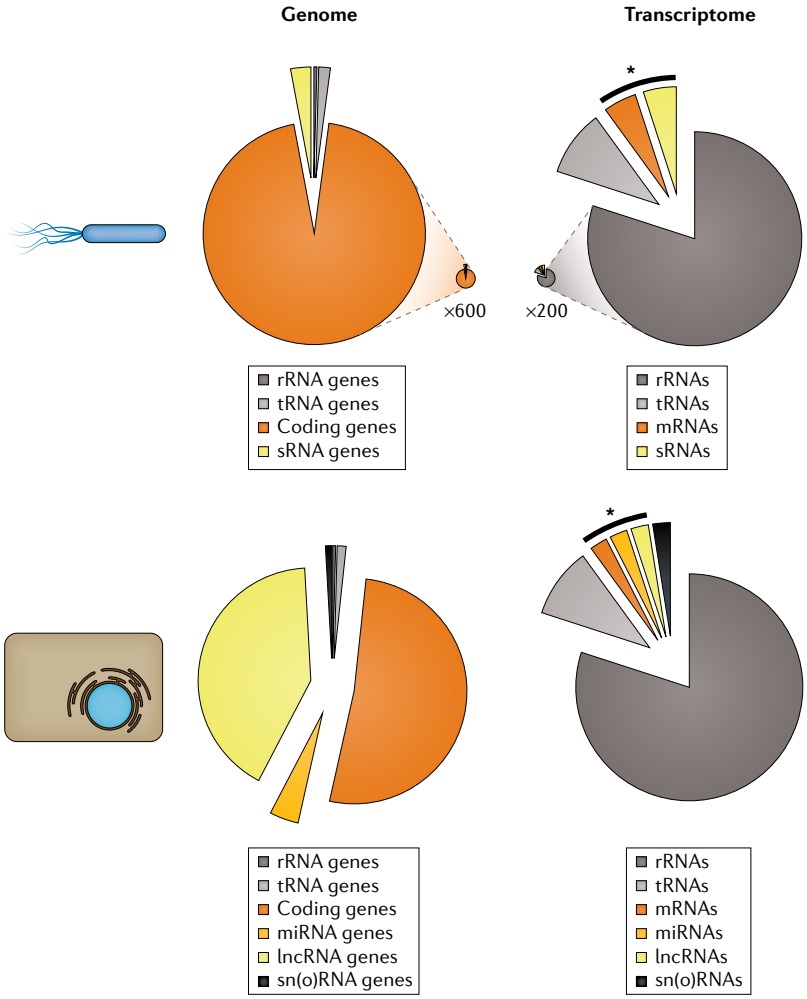

Fig. 2 | **Comparison of cellular RNA content between bacteria and mammals.** Pie charts of the fraction of different gene classes in the genome (left; from RefSeq) or RNA molecules in the transcriptome (from REF.[5]) in *Salmonella enterica* subsp. *enterica* serovar Typhimurium (upper) and humans (lower). The area of the bacterial pie charts is magnified by the indicated factors; the unmagnified pie charts reflect the relative size of the *Salmonella* genome/transcriptome compared with the human genome/transcriptome. Informative transcript classes (mRNAs and regulatory non-coding RNAs) to deduce cellular states are highlighted (*). lncRNAs, long non-coding RNAs; miRNAs, microRNAs; rRNAs, ribosomal RNAs; sRNAs, small RNAs; snRNAs, small nuclear RNAs; snoRNAs, small nucleolar RNAs; tRNAs, transfer RNAs.

By contrast, eukaryotic mRNAs carry a 7-methylguanosine cap and a poly(A) tail in the range of 70–250 nucleotides (nt). Unlike bacterial mRNAs, which are made in a translation-ready form, eukaryotic coding transcripts are synthesized as precursor mRNAs with introns and exons and undergo splicing (intron removal) in the nucleus prior to export to the cytoplasm. The average length of mature human mRNAs is ~3 kb; as only monocistrons are present in mammals, extremely long coding transcripts (>10 kb) are rare.

Certain additional, specialized RNA molecules exist in both kingdoms such as the Y-RNAs, which are stable non-coding transcripts that associate with the Ro60 autoantigen in mammals[12] and Ro60-related proteins in bacteria[13], or the RNA component of RNase P (M1 RNA)[14]. However, most non-coding RNA species are exclusively found in either bacteria or eukaryotes, where they carry out regulatory, enzymatic or scaffolding functions. In their sum, these heterogeneous transcripts typically contribute another ~5–10% to the bacterial or eukaryotic transcriptome, although the relative proportion of bacterial non-coding RNAs seems to strongly increase under certain conditions[15]. Most bacteria express dozens to hundreds of different small regulatory RNAs (sRNAs)[16]. An sRNA may be transcribed from an intergenic non-coding gene or cleaved off the end of an mRNA. Most sRNAs regulate target genes post-transcriptionally by short base pair interactions with the respective mRNAs.

Non-coding RNAs specific to eukaryotic transcriptomes include small nuclear RNAs (snRNAs) and small nucleolar RNAs (snoRNAs), which are involved in the splicing of precursor mRNAs or the chemical modification of rRNAs and tRNAs, respectively, and localize to nuclear compartments. PIWI-interacting RNAs (piRNAs) are best known for silencing transposable elements in the germ line, while non-coding RNAs in somatic cells are generally divided into microRNAs (miRNAs) and long non-coding RNAs (lncRNAs). The homogeneous class of miRNAs, with their mature form being ~22 nt long, localize to the cytosol, where they regulate target mRNAs in a sequence-dependent manner reminiscent of bacterial sRNAs. By contrast, lncRNAs are more heterogeneous and are operationally defined as non-coding transcripts >200 nt in length (note, however, that there is an ongoing debate as to how many of them encode short peptides[17]). Some lncRNAs carry a 3′ poly(A) tail while others are not polyadenylated. lncRNAs may be retained in the nucleus, where they can help to regulate the chromatin state, or localize to the cytoplasm and fulfil diverse functions, for example, maintaining subcellular compartments, serving as scaffolds for ribonucleoparticle assembly or sequestering miRNAs, to name but a few[18].

The differences between transcript repertoires are accompanied by a quantitative disparity between bacterial and eukaryotic cellular RNA contents (FIG. 2). The *Escherichia coli* genome has a size of ~5 Mbp, whereas the entire human genome is ~3,000 Mbp. The RNA content also differs dramatically: an average *E. coli* cell contains ~100 fg of RNA, whereas an average human cell contains up to 30 pg RNA, that is, several hundred times more[19]. Overall, these qualitative and quantitative transcriptome differences are reflected in the current standard RNA-seq protocols for bacterial and mammalian samples.

Bacterial versus eukaryotic RNA-seq workflows

The standard steps in an RNA-seq experiment are the purification of nucleic acids from a biological sample, enzymatic digestion of contaminating genomic DNA, depletion of abundant but less informative rRNA, conversion of the remaining transcripts into complementary DNA (cDNA), high-throughput sequencing of the cDNA fragments, alignment of the resulting sequencing reads to a reference genome sequence and quantification of the reads mapped to individual genetic features. RNA-seq provides various readouts, as follows. Counting all sequencing reads from the same transcript reveals the relative expression of the corresponding gene.

Additionally, read coverage distributions reveal general transcript features such as 5′ and 3′ boundaries, processing sites, and operon (for bacteria) or intron–exon (for eukaryotes) structures. Researchers can choose between different library preparation pipelines and RNA-seq platforms that all come with specific strengths and weaknesses (BOX 2). For more details of the sequencing technologies available, we refer the reader to recent specialized reviews[20,21]. Suffice to say here that the current standard for both bacterial and eukaryotic expression profiling is 'short-read' sequencing-by-synthesis using Illumina technology. Here, we restrict ourselves to discussing the commonalities and differences of the commonly used bacterial and eukaryotic Illumina-based RNA-seq protocols (FIG. 3a), as these are closest to being developed into multi-species RNA-seq approaches.

Transcriptome fixation, cellular lysis and RNA extraction. Sample processing may be time-consuming and can involve multiple handling steps, especially so in scenarios where specific subpopulations of bacterial or host cells are collected. Given that transcriptomes are notoriously unstable, in such cases, the RNA composition should be preserved with the help of fixatives that block de novo transcription and RNA decay. A variety of RNA-preserving reagents exist — each with specific strengths and weaknesses[22] — and are increasingly used for robust microbe–microbe and host–microbe transcriptomics[23–27]. A systematic evaluation of several standard fixation methods showed unexpected substantial fragmentation of RNA after isolation from infection samples, with consequences for the coverage of different RNA classes in subsequent RNA-seq[28]. As a general guideline, as both formaldehyde-based and alcohol-based fixation can have adverse effects on RNA integrity[24], ammonium sulfate-containing stabilizing reagents, such as the commercial RNA*later* (Sigma-Aldrich) or RNA*protect* (QIAGEN), are often preferred for whole-transcriptome studies. However, in experimental setups that rely on sort-enrichment, careful evaluation of these RNA-preserving reagents is recommended as they might quench the fluorescent signals of marker proteins.

The physicochemical properties of prokaryotic and eukaryotic cells differ vastly, as they do among bacterial species themselves (for example, Gram-negative versus Gram-positive bacteria). For instance, whereas Gram-negative bacterial cells can easily be cracked, also allowing joint lysis with mammalian host cells, the efficient disruption of Gram-positive bacterial cell walls often depends on enzymatic and/or mechanical treatment. It is thus advisable to carefully evaluate lysis efficiencies for cross-species transcriptomics. That is, lysis conditions need to be empirically established that are sufficiently harsh to effectively and homogeneously break up all cell types in the sample, while still being mild enough to not degrade cellular RNAs.

Once the lysates are prepared, downstream RNA isolation techniques are typically interchangeable between organisms but may bias towards individual RNA classes dependent on their size, secondary structure and degree of modification. Further information on these topics can be obtained from REF.[29].

Ribodepletion. Ribosomes are the most abundant ribonucleoprotein particles in any living cell and, although their number fluctuates with growth rate[30,31], the rRNA they contain will dominate the cellular RNA pool. Although rRNA reads in RNA-seq data are occasionally used to infer bacterial replication rates[32,33], this transcript class is commonly depleted in both bacterial and eukaryotic RNA-seq to indirectly increase coverage of mRNAs and other informative RNA classes. However, ribodepletion strategies partly differ between bacterial and eukaryotic studies.

In bacteria, rRNA is typically removed actively before or during cDNA library generation. Most commonly, it is pulled out from the pool of total RNA by sequence-specific 'capture oligonucleotides' that can be biotinylated (for example, riboPOOL from siTOOLs) or coupled to magnetic beads (for example, RiboCop from Lexogen), available as commercial or 'do-it-yourself' kits[34], reaching depletion efficiencies of >95%[35,36]. Alternatively, upon incubation with sequence-specific DNA oligonucleotides, the resulting rRNA–DNA hybrids are digested with RNase H (a principle incorporated in some NEBNext protocols and in Illumina's new Ribo-Zero Plus technology). Recently, we have introduced bacterial rRNA depletion at the cDNA level with programmed Cas9 nuclease cleavage[37]. This technique,

Box 2 | **Alternative cDNA library preparation and sequencing technologies**

RNAtag-seq

The primary aim of most RNA sequencing (RNA-seq) studies is to determine the set of differentially expressed genes in an organism in response to an experimental stimulus. However, for differential expression profiling, full transcript coverage is not needed, resulting in the fact that many RNA-seq studies accumulate more information than is actually required. In RNAtag-seq[226], a barcoded adapter is ligated to the 3′ end of input RNA fragments, thus allowing for multiplexing early during library construction. This represents a time-efficient and cost-efficient alternative to standard complementary DNA (cDNA) library generation protocols. RNAtag-seq works with both prokaryotic and eukaryotic input RNA and, because counting the resulting sequencing reads is sufficient to call differentially expressed genes, it may be incorporated into cross-kingdom RNA-seq approaches.

SEnd-seq

There are opposite cases when ensemble RNA levels are insufficient but when the detection of full-length transcripts is key to interpreting biological processes. Various methods can map transcription start sites in bacteria[227] or eukaryotes[228] and there are alternative methods that determine 3′ ends across bacterial[62] or eukaryotic transcriptomes[229]. Recently, simultaneous 5′ and 3′ end sequencing (SEnd-seq) was developed and maps both transcription start and end sites in parallel, in the same experiment[230]. Developed for *Escherichia coli*, the protocol is generic and should be transferable to metatranscriptomics. However, given that SEnd-seq determines only the extreme 5′ and 3′ ends of an input cDNA molecule, the protocol is less suited to deducing the full-length mRNA structure in an organism where splicing occurs.

Long-read sequencing

Illumina short-read sequencing technology relies on transcript fragmentation, thereby inevitably losing the complete nucleotide composition of individual transcripts. As alternatives to the Illumina-based methods above, long-read sequencing platforms — such as Pacific Biosciences (which sequences full-length, unfragmented cDNA molecules) or Oxford Nanopore (with the possibility to directly sequence RNA molecules, omitting the need for reverse transcription) — are entering the market[20,21]. Although the current read depth, error rate and cost of these platforms cannot compete with Illumina technology, they have the ability to read a cDNA/RNA molecule from one end to the other and bear great potential for future bacterial[231] and eukaryotic[232] (epi)transcriptomics.

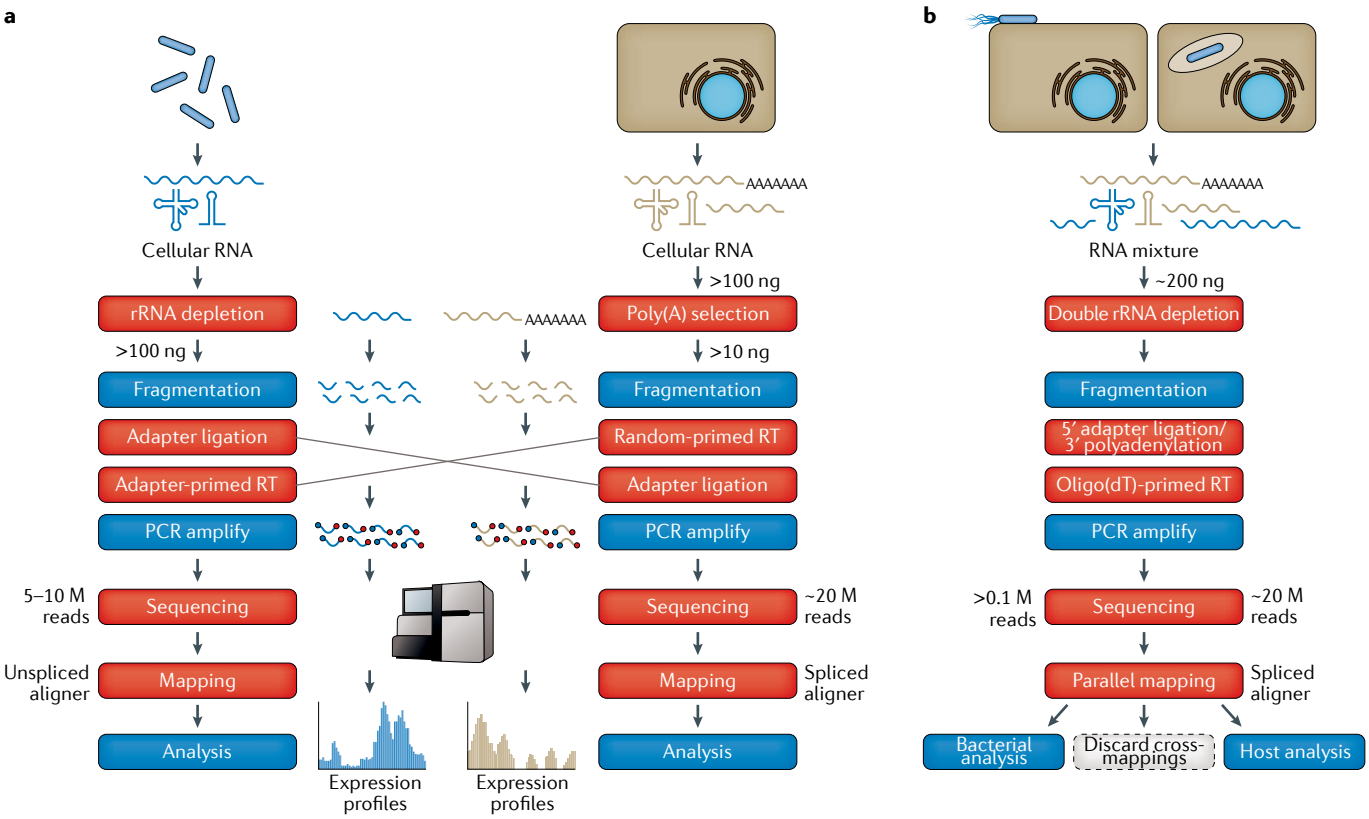

Fig. 3 | **The basic steps in commonly used protocols for strand-specific bacterial, mammalian or dual expression profiling. a** | Single-species RNA sequencing (RNA-seq). The bacterial RNA-seq protocol (left) is based on the NEBNext Small RNA Library Prep Set workflow (New England Biolabs) and the mammalian protocol (right) is based on the TruSeq Stranded mRNA workflow (Illumina). Commonalities are in blue boxes and differences in red boxes. The minimal amounts of total or enriched input RNA are those stated in the respective user manuals. **b** | Dual RNA-seq protocol based on REF.[24]. Sequencing depth estimates for conventional bacterial or mammalian RNA-seq are from REF.[42] and REFS[43,44], respectively, and those for dual RNA-seq are from REF.[29]. rRNA, ribosomal RNA; RT, reverse transcription.

which is generally known as DASH (depletion of abundant sequences by hybridization)[38], removes rRNA less efficiently but seems particularly suitable for low-input samples[37]. What remains after bacterial ribodepletion are mRNAs, sRNAs and tRNAs. As tRNAs require specialized protocols for efficient cDNA conversion[39–41], it is primarily mRNAs and sRNAs that will eventually make it into conventional sequencing libraries.

In eukaryotic RNA-seq, rRNA is generally depleted indirectly by selective cDNA priming on polyadenylated transcripts, that is, primarily mRNAs and lncRNAs (poly(A)-selection is an integral part of the TruSeq Stranded mRNA kit from Illumina). Alternatively, if non-polyadenylated transcripts (for example, poly(A)-negative lncRNAs) are also of interest, active rRNA removal strategies, such as the commercial riboPOOL, RiboCop or Ribo-Zero Plus technologies or customized DASH[38], may be used. These active rRNA removal methods will also be the top choice when analysing mixtures of bacteria and eukaryotic host cells by so-called dual RNA-seq[28], which will be covered further below.

cDNA library preparation and sequencing. The dominating Illumina technology offers a maximum read length of 300 bp, necessitating fragmentation of total RNA samples prior to cDNA conversion for bacteria and eukaryotes alike. In bacterial RNA-seq pipelines, RNA adapters are then often ligated to the 3′ end of the obtained RNA fragments and subsequently used as anchors for reverse transcription (RT) via adapter-specific primers (as in the popular NEBNext Small RNA Library Prep Set; New England Biolabs). Similar adapter-based library generation protocols can also be used in combination with size-selected, unfragmented eukaryotic short RNAs, for example, for miRNA profiling. By contrast, profiling of eukaryotic mRNA expression generally involves the enrichment of polyadenylated transcripts followed by RNA fragmentation and RT by random priming (for example, TruSeq Stranded mRNA kit; Illumina). Dependent on the sample type and the scope of the RNA-seq experiment, alternative, custom-tailored library preparation protocols exist, some of which are described in BOX 2.

The sequencing depth requirements also depend on the desired outcome of an RNA-seq experiment. Although the accurate mapping of transcript boundaries, operon, or alternative splicing structures and the quantification of low-abundance transcripts demand a high sequencing depth, most RNA-seq studies in bacteria and eukaryotes are, in essence, differential expression analyses and, as such, much less sequencing intensive.

Simulation studies proposed ballpark figures of 5–10 million non-rRNA reads for exhaustive expression profiling in bacteria[42] and ~20 million non-rRNA reads in mammals[43,44].

Read alignment, normalization, quantification and functional analyses. The concepts of read mapping and quantification are shared between bacterial and eukaryotic RNA-seq, with differences in the details[29]. As splicing is a molecular process that is exclusive to eukaryotes, standard read mappers for prokaryotic RNA-seq generally save both computing power and time by blanking it out, whereas eukaryotic mapping pipelines strictly depend on spliced aligners[45]. Normalization is similar for bacterial and eukaryotic RNA-seq data, with the most common methods assuming that the expression of a substantial fraction of genes is not altered between two distinct conditions[46]. However, this assumption fails for the pairwise comparison of extreme conditions, which is particularly likely in bacteria that can change a considerable part of their transcriptome and such phenomena have fostered the development of normalization tools tailored to bacterial RNA-seq data[47]. On the experimental side, the inclusion of artificial RNA spike-ins in the RNA-seq protocol provides an opportunity to normalize to cell numbers[48–50].

For differential expression analysis, the DESeq2 (REF.[51]), edgeR[52] and limma/voom[53] algorithms have been most popular. As these tools cannot analyse differential isoform usage, RNA-seq studies addressing alternative splicing in eukaryotes rely on different algorithms[54]. Once the differentially expressed genes (or isoforms) have been identified, these changes are usually subjected to global pathway analysis. Both bacterial and eukaryotic transcriptomics rely on well-curated databases such as Gene Ontology (GO)[55] or Kyoto Encyclopedia of Genes and Genomes (KEGG)[56]. Additionally, for the analysis of mouse or human RNA-seq data, more-refined databases exist that specialize in, for example, metabolic[57] or immune-related gene sets[58].

In summary, despite the basic steps being shared between conventional prokaryotic and eukaryotic RNA-seq protocols, experimental and analytical pipelines have diverged over the past decade to satisfy the specific needs of the respective target transcriptome. As a consequence, early RNA-seq profiling of bacterial gene expression in vivo ignored the host transcriptome and entailed the physical enrichment of bacteria from host tissue, as will be reviewed in the next section.

Bacterial RNA-seq for in vivo transcriptomics

In order to fully exploit global expression data of a host and microorganisms, high-resolution transcriptome maps are required. Whereas the mouse and human transcriptomes have been well annotated over the years, fine-grained transcriptome maps are not a given for most bacteria. Rather, bacterial transcriptomes are often annotated based on computational searches for potential open reading frames (ORFs) with a length of >100 amino acids but short ORFs, untranslated regions (UTRs) or non-coding RNA genes may be missing. Consequently, a truly comprehensive global expression analysis usually requires prior transcriptome refinement using RNA-seq-based techniques, such as differential RNA-seq (dRNA-seq; initially applied to Gram-negative species, such as *Helicobacter pylori*[59] and *Salmonella enterica*[60], and later to Gram-positive species such as *Streptococcus pneumoniae*[61]), to globally map transcriptional start sites or terminator-sequencing (Term-seq; established in *Listeria monocytogenes*[62]) or end-enriched RNA-seq (Rend-seq; established in *E. coli* and *Bacillus subtilis*[63]) to determine transcription termination sites. Such reference transcriptomes are now available for numerous aerobic[60,64–69] and anaerobic bacteria[70]. In addition, research on *Salmonella enterica* subsp. *enterica* serovar Typhimurium has greatly benefited from a gene expression atlas called SalCom, compiling RNA-seq profiles under 22 defined in vitro conditions, each mimicking specific phases of the *Salmonella* Typhimurium infection cycle[15]. Similarly, the recently established PATHOgenex database[71], which features host stress-related in vitro transcriptome signatures of 32 human pathogens, provides an excellent resource for infection biologists.

The emergence of the field of cellular microbiology three decades ago[72] brought about a need to study gene expression in bacterial pathogens within the context of their host organism. The first such approaches relied on the physical enrichment of bacteria from infection samples, followed by transcriptome analysis of the purified bacteria alone. For example, infected eukaryotic cells were selectively lysed with a detergent and the released intracellular bacteria were collected by differential centrifugation[73,74]. Strikingly, the gene expression profile of *S. enterica* thus recovered from macrophages revealed several activated genes that were silent in any of the 22 in vitro conditions in the SalCom atlas. This finding highlights the difficulty in reconstructing host-like conditions in a culture flask and underscores the importance of true cellular microbiology approaches. Similar studies have been performed in animal infection models, for example, using anti-*Staphylococcus aureus* antibodies coupled to magnetic beads to enrich these bacteria from a mouse model of osteomyelitis[75], engineering bioluminescent strains of *Citrobacter rodentium* that enabled their recovery from mouse colon tissue by bioluminescence imaging[27], and *Pseudomonas aeruginosa* in the sputum of patients with cystic fibrosis[76]. Collectively, these studies revealed new virulence strategies and metabolic adaptations of these pathogens within their host niches.

Even if bacteria are enriched, the resulting transcriptome data may still be dominated by host sequences. To minimize the host RNA/cDNA background, bacterial transcripts may be enriched using commercial kits (for example, Microb*Enrich*; Ambion) that deplete eukaryotic rRNA and polyadenylated transcripts. This was used for in vivo expression studies of *Vibrio cholerae* in an infant rabbit and mouse model[77], of *Yersinia pseudotuberculosis* in mouse caecal tissue biopsy samples[26], and of *C. rodentium* recovered upon infection of mouse colon tissue[27].

Bacterial transcripts can also be enriched directly. 'Hybrid-selection' refers to the incubation of complex cDNA samples with species-specific, biotinylated probes

Gnotobiotic
Pertaining to 'gnotobiosis', which is Greek for 'known life'. The term generally describes biological systems wherein all present organisms can be accounted for. In the present context, the term refers to ex-germ-free mice that were inoculated and colonized with a defined bacterial species or consortium.

Metatranscriptomics
Methods to detect and quantify steady-state transcript levels from multiple bacterial species within a community present in a given environmental or host-derived sample.

to capture and enrich the cDNAs of interest away from the host cDNA background. Hybrid-selection recently enabled niche-specific RNA-seq of the gut commensal *Bacteroides fragilis* within the proximal colon[78]. This revealed, for the first time, spatially distinct expression profiles of a microbiota member between its luminal, mucus-associated and epithelial niches. Although the capture probes comprised bait sequences for the complete *B. fragilis* genome, prior knowledge of a bacterium's transcriptome structure can refine probe design. For example, biotinylated probes specific for bacterial mRNAs and sRNAs (omitting rRNA, tRNA or non-transcribed loci) were employed in a technical study looking at intracellular expression profiles of *P. aeruginosa* or *Mycobacterium tuberculosis*[79]. Of note, given that enrichment occurs at the cDNA level — that is, after sample amplification — hybrid-selection is extremely sensitive and works down to the single-cell level[79]. So far, hybrid selection of bacterial cDNA has been applied to host cell cultures or gnotobiotic mice after monocolonization with the target bacterium. However, it should also permit the enrichment of target sequences from multi-species samples from conventional mice with a complex microbiota to reveal potentially protective effects of commensal species.

Cross-species RNA-seq approaches

Steady improvements in RNA-seq technology have allowed researchers to advance beyond one-sided RNA-seq studies and investigate microorganisms together with other microorganisms or with host cells in unprecedented detail and under physiological conditions (FIG. 4). The following discussions will focus on metatranscriptomics within host-associated bacterial consortia and on dual RNA-seq to simultaneously read out bacterial and host transcriptomes. We conclude the section with an outline of envisaged extensions of cross-species RNA-seq.

Metatranscriptomics. Sequencing-based analysis of the structure of bacterial communities began with metagenomics, that is, the analysis of 16S rDNA sequences (16S profiling), before the more sensitive shotgun sequencing of genomic DNA became feasible. However, the abundance of a certain microbial species or a certain gene does not necessarily correlate with the functional contribution of that species or gene in the consortium. In this regard, metatranscriptomics has been gaining popularity for functional insights into interactions within microbiomes[80].

To minimize the risk of RNA degradation during sample preparation, an RNA preserving reagent is often used to 'freeze' metatranscriptomes[23,81]. The standardized usage of such agents should facilitate cross-comparisons between different studies. As for conventional RNA-seq, metatranscriptomics typically involves rRNA depletion using the same techniques as used for single-species RNA-seq[35]. However, it needs to be stressed that, because most commercial rRNA removal kits target model bacteria, they might be less efficient for the many non-model species of, for example, the human gut microbiota. Cas9-based ribosomal cDNA removal[37] on metatranscriptomic samples is yet to be tested; again, it should be particularly suitable for low-input samples, for example, when wanting to profile microbiome activity in an insect gut at the RNA level[82].

Accurate quantification of gene expression in a microbiome requires the calibration of RNA for the

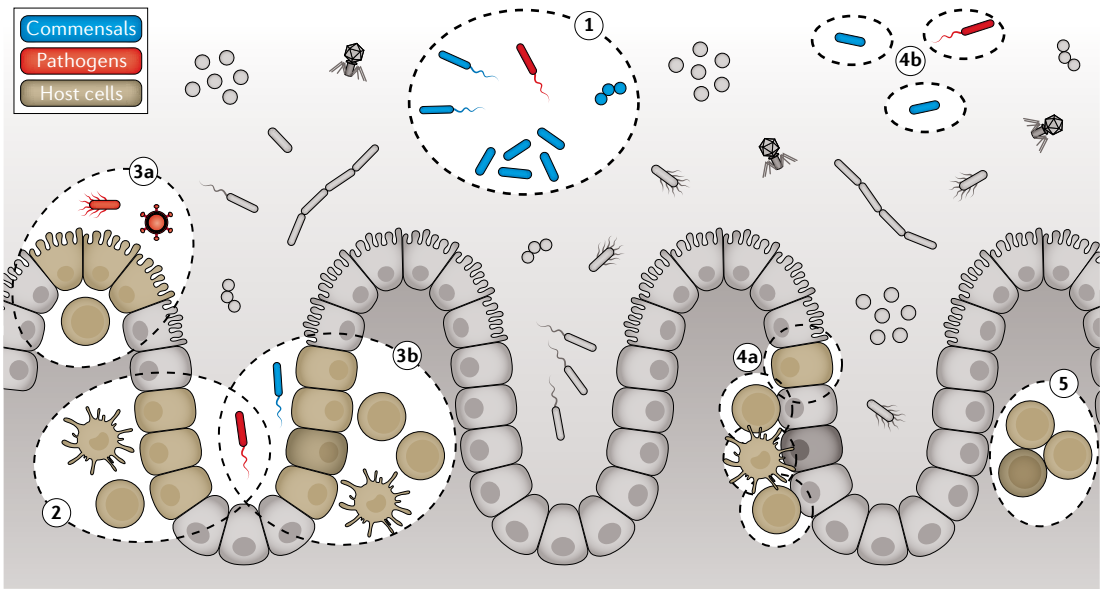

Fig. 4 | **Graphical overview of RNA-seq-based approaches to study inter-species interactions in the mammalian intestine.** (1) Metatranscriptomics[23,83,85,90,196]. (2) Dual RNA sequencing (RNA-seq) of an enteric pathogen and infected host tissue[102,103]. (3) Triple RNA-seq[112] of viral/bacterial co-infections (3a) or of a bacterial pathogen, a competing commensal and their host (3b). (4) Single-cell RNA-seq of individual host cells[127,128] (4a) or single bacteria[139,140,144] (4b). (5) NICHE-seq[134] to maintain spatial information about the local microenvironment at single-cell resolution. Adapted from REF.[197], CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

DNA content of the same sample, as done in hybrid DNA/RNA-seq analyses[23,83–85]. To this end, DNA and RNA samples are each sequenced to a depth of 20–40 million Illumina reads. Briefly, the resulting DNA and RNA sequencing data are processed separately, with the metagenomic data then being used to normalize the metatranscriptomic reads. Unsurprisingly, the computational steps are far from trivial and two orthogonal approaches have been developed to this end, namely taxonomic classification involving either reference-based or de novo metagenome-assembled genomes (MAGs). Reference-based classifiers, such as HUMAnN2 (REF.[86]), Kaiju[87] or Kraken2 (REF.[88]), work well when community member species are largely predetermined and only their relative abundances are unknown such as in the human intestinal microbiota. If bacterial species within a sampled community are not known a priori, de novo taxonomic classification that assembles MAGs from metagenome reads is preferred[89].

Comparative studies of healthy human gut microbiomes revealed metatranscriptomic profiles to be generally less individual than their metagenomic counterparts[23,83,85]. This argues for subject-specific, whole-community regulation, wherein a core set of housekeeping transcripts is universally expressed over individual subjects but by different microbial species. By contrast, longitudinal profiling showed the metatranscriptome of an individual to be more variable over time than that person's metagenome, suggesting the active regulation of microbial gene expression in response to environmental and nutritional fluctuations[83,85]. Compositional (metagenomics) and functional (metatranscriptomics) changes in the human intestinal microbiota have also been linked with the outcome of infectious diseases. In a human challenge model, a gut metatranscriptome dominated by mRNAs encoding antioxidants and metal ion homeostasis proteins was linked to a reduced risk of developing symptoms of typhoid upon infection with *Salmonella enterica* subsp. *enterica* serovar Typhi[90].

Dual RNA-seq. Transcriptomics is increasingly being harnessed to study interactions across kingdoms, foremost between bacterial pathogens and mammalian hosts[5,29]. Dual RNA-seq measures gene expression simultaneously in (intracellular or extracellular) bacteria and infected host cells or tissues. As eukaryotic and bacterial material is processed together without prior physical separation (FIG. 3b), dual RNA-seq promises the discovery of novel interdependencies in host–microbe interactions. Technically, dual RNA-seq benefits from commercial kits optimized for the parallel depletion of prokaryotic and eukaryotic rRNA; when it comes to the analysis of clinical samples that often offer only minute amounts of RNA, rRNA depletion at the cDNA level (by DASH, see above) should be an attractive strategy. The sequencing depth overall is not significantly different from conventional mammalian RNA-seq provided that the bacterial RNA proportion in the sample is within 0.5–10%[29].

Dual RNA-seq has been extensively used to unveil previously hidden aspects of *Salmonella* Typhimurium–host cell interactions, for example, how the activities of PinT sRNA, which is highly induced in intracellular *Salmonella*, alters epithelial STAT3 signalling[28]. Furthermore, dual RNA-seq linked the global activities of *Salmonella* RNA-binding protein ProQ with MAPK signalling in epithelial host cells[91] and showed that *Salmonella* persisters arising during macrophage infection maintain a metabolically active state that allows them to reprogramme their host cells[92] (FIG. 5a). Dual RNA-seq studies have also been performed with cell culture models of infection with *S. pneumoniae*[93], uropathogenic *E. coli*[94] and *Haemophilus influenza*[95]. Dual RNA-seq is particularly suitable to study obligate intracellular pathogens, as demonstrated with *Chlamydia trachomatis*[96], *M. tuberculosis*[97,98] (FIG. 5b) and *Orientia tsutsugamushi*[99] (FIG. 5c). Many of these studies were designed as temporal studies to follow the kinetics of bacterial and host gene expression during their interaction or as comparative studies comparing the expression patterns between infections with different bacterial strains. The power of these designs is well illustrated with the PinT sRNA, for which dual RNA-seq profiling over the course of infection with wild-type versus Δ*pinT* bacteria revealed that PinT functions as a post-transcriptional timer in the transition of *Salmonella*'s two major virulence programmes, with a massive impact on gene expression in infected host cells[28].

The application of dual RNA-seq is increasingly shifting from host cell monoculture towards in vivo-like infection models. For example, a new 3D intestinal tissue model recapitulates aspects of human gastroenteritis in response to *Salmonella* Typhimurium infection[100]. Dual RNA-seq of individual cell types isolated from this model helped to parse out the direct and indirect effects of bacterial infection in the epithelial and endothelial compartments. While it corroborated a previously observed connection between the type III secretion systems of *Salmonella* and the STAT3-dependent inflammatory response[101], it revealed that this response stays local in the epithelial lining[100].

Mouse infection models that can closely recapitulate human disease exist for many bacterial pathogens. Generally, bulk RNA-seq data from infected host tissues or organs are difficult to analyse as altered transcript abundances represent the sum of differentially expressed genes and alterations in host cell-type composition. This notwithstanding, dual RNA-seq has been successfully used with some of these infection models, analysing Peyer patches, kidneys or lungs of mice infected with the extracellular pathogens *Y. pseudotuberculosis* (FIG. 5d), *S. aureus* and *P. aeruginosa*, respectively[102–104]. Comparative dual RNA-seq of two different strains of *S. pneumoniae* in mouse lung tissue linked a SNP in the bacterial raffinose utilization pathway to differential neutrophil recruitment as a potential cause of the divergent disease outcomes[105]. Examples of dual RNA-seq revealing how bacteria adapt to and even exploit the human immune response include a human challenge model infected with the extracellular pathogen *Haemophilus ducreyi*[106] or samples from human patients with intracellular *Mycobacterium leprae*[32].

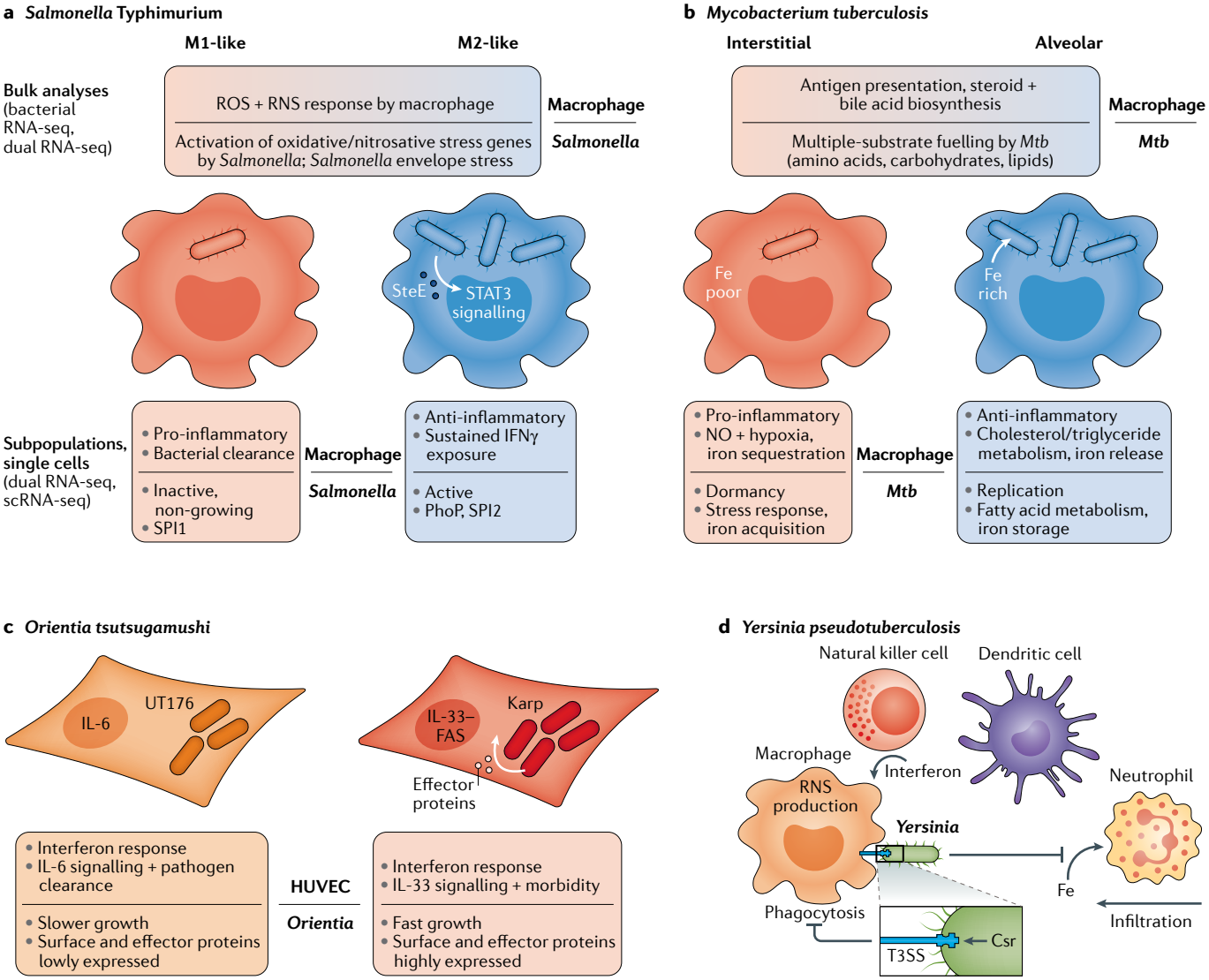

Fig. 5 | **Molecular aspects of host–pathogen interactions revealed by transcriptomics. a** | *Salmonella enterica* subsp. *enterica* serovar Typhimurium infection of macrophages represents a prime example for the strength of combining different, complementary transcriptomic approaches to dissect the molecular aspects of an infection process from various angles. The global picture that emerges from the combined application of conventional bacterial RNA sequencing (RNA-seq)[192], dual RNA-seq[28,92] and single-cell RNA sequencing (scRNA-seq)[127,128] argues that intra-macrophage *Salmonella* face oxidative, nitrosative and envelope stress. Moreover, the intracellular replication rates are variable and follow a gradient that correlates with the expression of *Salmonella* pathogenicity island 2 (SPI2) as well as PhoP-controlled genes and anti-correlates with SPI1 genes. The extreme populations along this gradient — persistent or fast-replicating bacteria — reside in pro-inflammatory M1-like or anti-inflammatory M2-like macrophages, respectively, and host macrophage polarization seems to be actively induced by secreted *Salmonella* effector proteins as exemplified by SteE[131]. **b** | A similar virulence strategy is applied by *Mycobacterium tuberculosis* (*Mtb*). Initial bulk dual RNA-seq of *Mycobacterium*-infected THP1 macrophages revealed the metabolic flexibility of this intracellular pathogen during infection[97]. Macrophage phenotype-specific dual RNA-seq dissected the mycobacterial–host interactions of interstitial (M1-like) and alveolar (M2-like) mouse macrophages[108]. **c** | Two isolates of the obligate intracellular pathogen *Orientia tsutsugamushi* — UT176 and Karp — were used to infect human endothelial cells and analysed by dual RNA-seq, revealing a common interferon-based host response in addition to strain-specific immune signalling cascades, likely to result from differential expression of virulence factors between the two bacterial strains[99]. **d** | Bacterial RNA-seq of *Yersinia pseudotuberculosis* in small caecal tissue biopsy samples revealed an early induction of type III secretion system (T3SS) genes during the acute phase and their repression when bacteria adopt a persistent state[26]. Tissue dual RNA-seq of this pathogen within ileal Peyer patches recapitulated the early induction of *Yersinia* T3SS — a mechanism dependent on the carbon storage regulator (Csr) system — and identified the host's immune response to infection to be dominated by infiltrating neutrophils[103]. RNS, reactive nitrogen species; ROS, reactive oxygen species.

Others used the technique on human skin biopsy samples to dissect the pathophysiology of monomicrobial and polymicrobial necrotizing soft-tissue infections[107], revealing in vivo virulence expression patterns, including an upregulation of invasion genes of *Streptococcus pyogenes* during mono-infection and a pronounced interferon response in the surrounding host tissue. By contrast, a metatranscriptomic-like analysis of the bacterial expression data from polymicrobial biopsy samples revealed an elevated expression

of lipopolysaccharide (LPS) biosynthetic genes and adhesion (but not invasion) factors as compared with mono-infections[107]. The data further indicated functional specialization between the co-occurring species, with key metabolic pathways only expressed by a fraction of the community members, thereby suggesting bacterial synergy at the heart of the pathogenicity of polymicrobial soft-tissue infections. The host response to polymicrobial infection was dominated by a strong pro-inflammatory profile (potentially as a result of elevated LPS levels) and genes for extracellular matrix components, indicative of activated fibroblasts. Eventually, these divergent host response patterns were exploited to pinpoint biomarkers specific for monomicrobial or polymicrobial infection for early therapeutic intervention.

When sequencing bulk tissue samples, cell type-specific gene expression may be lost in the average expression profiles. However, the resolution can be increased in two ways: before sequencing, by dissociating an infected tissue and isolating pre-defined cell types, or after sequencing, by utilizing cell type-specific markers to computationally deconvolute the heterogeneous data. A proof-of-principle study of mouse colon tissue infected with *Salmonella* Typhimurium demonstrated that cell-type enrichment is generally feasible[25]. The use of a fluorescent *Salmonella* strain (expressing GFP) in combination with selective antibody staining enabled the separation of infected and bystander colonocytes. Another dual RNA-seq study used cell type-specific antibody staining and enrichment of alveolar and interstitial macrophages from mouse lungs infected with *M. tuberculosis*[108]. However, as the mycobacteria were not fluorescent, the enriched macrophages were a mixture of infected and uninfected cells. To not out-dilute the bacterial reads, the sorted host macrophages were selectively lysed, followed by enrichment and mechanical lysis of the mycobacterial cells and subsequent pooling of the host and bacterial lysates. This strategy yielded sufficient numbers of bacterial reads to observe different mycobacterial expression profiles in alveolar versus interstitial macrophages, indicating that alveolar macrophages represent a conducive environment, whereas interstitial macrophages represent a more hostile environment (FIG. 5b). However, there are considerable technical challenges posed by lengthy tissue dissociation, antibody staining and cell sorting, all of which threaten transcriptome integrity. Therefore, samples should ideally be fixed as early in the procedure as technically feasible[25].

As a post-sequencing alternative to the physical enrichment of distinct cell types, in silico dissection of host expression signatures upon mouse infection with *Y. pseudotuberculosis* revealed extensive neutrophil infiltration into infected Peyer patches[103]. Likewise, a cell type-specific expression signature tool was employed to dissect alterations in the cellular composition of leprosy skin lesions[32]. Although not yet combined with infection studies, more sophisticated cell-type deconvolution analyses fed with scRNA-seq data[109–111] bear great potential to increase the resolution of in vivo dual RNA-seq analysis in the future.

Triple RNA-seq. Conceptually speaking, it is a small step from dual RNA-seq to any type of 'multi' RNA-seq to investigate polymicrobial infections, especially with different types of pathogen. The latter is illustrated by a recent triple RNA-seq study in an ex vivo dendritic cell model co-infected with human cytomegalovirus and the pathogenic fungus *Aspergillus fumigatus*[112]; this combination of pathogens is a serious medical threat in organ and stem cell transplantation. The very different viral, fungal and human genomes facilitated the unequivocal assignment of the triple RNA-seq reads with negligible cross-mapping, revealing synergistic pathogen strategies during co-infection and inter-species expression network analyses pinpointing host genes with biomarker potential. The same approach should facilitate a better understanding of co-infections with a bacterium and a virus, which are common and usually lead to a more severe outcome than the respective mono-infections. Promising models to understand how one infection makes the host more susceptible to another include non-typhoidal *Salmonella* and human immunodeficiency virus[113,114], *Streptococcus* spp. and influenza virus[115], or *Chlamydia* spp. and human herpesvirus[116].

Towards 'omni-RNA-seq' approaches. Increasing read lengths in combination with better computational algorithms to resolve RNA-seq data at the taxonomic level enable RNA-seq studies of eukaryotic host cells interacting with more than one bacterium and, ideally, with hundreds of different bacterial species as in the case of the gastrointestinal microbiota. Although not yet conducted in a single-step experiment, some pioneering work towards such 'omni' RNA-seq includes transcriptomics of the nasal epithelium of asthmatic children complemented with independently obtained metatranscriptomics of the nasal microbiome[117]. Similarly, the pulmonary microbiome was profiled at the RNA level and integrated with host lung transcriptome data in patients with chronic obstructive pulmonary disease (COPD)[118]. Despite donor variability, the number of patients (8 and 25) and control subjects (6 and 9) in these studies were sufficient to define the paediatric asthma-associated or COPD-associated microbiome signatures, correlating with a modulation of the host's immune system (IL-1α was associated with asthmatic patients; a T helper 17 cell response correlated with COPD exacerbation).

Overall, the past 5 years have seen an enormous increase in the popularity of multi-organism RNA-seq approaches (FIGS 1,4). While metatranscriptomics and dual RNA-seq are on the way to becoming routine techniques in modern infection biology, the development of cross-species RNA-seq will continue by considering ever more interaction partners within an infected 'holobiont'[119] and tracing their complex interaction networks. Appropriate analysis of the resulting data mass demands sophisticated bioinformatic tools and should be followed up with experiments seeking to probe causality and functional consequences of the predicted multi-species interaction nodes. In parallel to these efforts, the field of infection research is currently shaken up by the

introduction of single-cell transcriptomics, which will be the focus of the next section.

Single-cell RNA-seq

The recent advances in scRNA-seq have ushered in a new era of transcriptomics, which has already witnessed profound new discoveries, ranging from previously unknown cell types or physiological states to new principles of stochastic gene expression[120]. At the same time, cellular heterogeneity is increasingly understood to play important roles in host–pathogen interactions, for example, pathogens exploit pre-existing and/or induced cellular heterogeneity to establish an infection niche. Likewise, pathogens themselves can present with substantial phenotypic diversification, for example, *Salmonella* Typhimurium populations are characterized by a bistable expression of invasion genes already before host cell contact[121] and gene expression variability even intensifies over the course of the infection[122]. Physiologically, such a bet-hedging strategy primes a subpopulation of *Salmonella* for invasion, inducing epithelial inflammation that in turn benefits the luminal population[123,124]. New transcriptomic methods going beyond bulk analysis are needed to capture and understand how cellular heterogeneity determines the outcome of host–microbe interactions.

Eukaryotic scRNA-seq. At this point in time, eukaryotic scRNA-seq has become a routine technique, with a wide variety of experimental and computational pipelines to choose from[125] and several commercial platforms available. Although transcriptome fixation is less common than in bulk sequencing approaches, the first RNA-preserving techniques are now compatible with single-cell applications[126]. Unsurprisingly, scRNA-seq is also making its way into infection biology[10], with the pioneering work again using *Salmonella* Typhimurium to study heterogeneity in populations of infected host cells[127,128]. This bacterium had been known to display heterogeneous replication rates inside macrophages[129] and in vivo tissues[130].

Using a macrophage infection model, scRNA-seq uncovered a rapid polarization of *Salmonella*-infected host cells that would have been impossible to detect by bulk RNA-seq (FIG. 5a). Moreover, it also showed that actively replicating *Salmonella* reside predominantly in anti-inflammatory, infection-permissive M2-like macrophages, whereas non-replicating bacteria primarily dwell in pro-inflammatory M1 macrophages[128], with bacterial replication rates correlating with the magnitude of the host's interferon response[127]. Subsequent integration of the scRNA-seq data with corresponding (population level) dual RNA-seq established a link between the expression of a specific *Salmonella* virulence factor and macrophage polarization[92]. In other words, *Salmonella* actively manipulate their host cells to commit development towards a replication-permissive state[131], a virulence strategy that this pathogen seems to share with *M. tuberculosis*[108] (FIG. 5b). scRNA-seq data have also become available for human primary immune cells infected with either *Salmonella* Typhimurium or *M. tuberculosis*[132]. Moreover, there is a comprehensive atlas of cell type-specific gene expression signatures and cellular compositional changes within the gut epithelium of a mouse model after *Salmonella* Typhimurium infection[133]. These single-cell datasets provide a rich resource for the study of heterogeneity in diverse host cell types.

A severe limitation of most in vivo scRNA-seq protocols is the loss of spatial information in the process of tissue dissociation. However, spatial information is pivotal to interpreting scRNA-seq results from infected samples because many bacterial pathogens colonize vastly different local microenvironments. By combining photoactivatable reporters and scRNA-seq, NICHE-seq overcomes this limitation and reconstructs the spatial organization of infection niches[134]. Specifically, transgenic mice ubiquitously expressing a photoactivatable GFP variant were infected with lymphocytic choriomeningitis virus and, upon photoactivation of subregions in B cell follicles or the T cell area of inguinal lymph nodes with two-photon laser scanning microscopy, the tissue was dissociated and GFP-positive cells were collected by fluorescence-activated cell sorting and analysed by massively parallel scRNA-seq.

Another common limitation of scRNA-seq protocols is their dependence on oligo(dT)-primed RT, thereby losing information for many non-polyadenylated RNA classes, foremost, small eukaryotic RNAs. To overcome this limitation, a 3′ adapter ligation-based method termed Small-seq was developed, which allows for the profiling of miRNAs, tRNA fragments and snoRNAs in individual mammalian cells[135,136]. Another method, Holo-seq, captures both small RNA species and mRNAs from single cells by the addition of in vitro-transcribed carrier RNA, which allows conventional library construction (as for bulk RNA-seq) and involves enzymatic digestion at the cDNA level prior to sequencing[137]. These developments reflect previous advancements in bulk RNA-seq, emanating from mRNA-centric approaches towards a 'sequence-all' strategy. The expression of several miRNAs and lncRNAs responds remarkably rapidly to a pathogenic stimulus, proposing them as suitable biomarkers for diagnostics. Poly(A)-independent scRNA-seq may assess whether some of these non-coding RNAs are heterogeneously expressed and could serve as biomarkers also for specific cellular subpopulations during infection.

Bacterial scRNA-seq. While single-cell transcriptomics began to revolutionize eukaryotic biology, technical hurdles prevented its robust application to bacteria until recently. As bacterial cells contain only a femtogram amount of RNA[19], that is, >100 times less than the typical eukaryotic cell, a sensitive cDNA synthesis and amplification protocol is required. Whereas most current eukaryotic scRNA-seq protocols have a lower detection limit of 10 copies of a transcript per cell, bacterial scRNA-seq must consider a much lower average copy-number of mRNAs (0.4 copies per cell[138]). The intrinsic lability of bacterial mRNAs (with half-lives in the minute range, as compared with hours in eukaryotes), is another issue and requires perforation of the bacterial envelope, cell lysis and subsequent RNA stabilization

to be done rapidly. Above all, however, the absence of a poly(A) tail on functional bacterial transcripts precludes oligo(dT)-based RT priming.

With the recent publication of robust bacterial scRNA-seq protocols[139,140], single-bacterium transcriptomics have now become a reality for major human pathogens. One study used Multiple Annealing and Tailing-based Quantitative scRNA-seq (MATQ-seq) — originally developed for eukaryotic scRNA-seq[141] — to profile individual *Salmonella* Typhimurium and *P. aeruginosa* cells after their isolation by cell sorting[140]. Benchmarking with an established bulk RNA-seq compendium for *Salmonella* revealed these single-bacterium transcriptomes to faithfully capture growth-dependent gene expression patterns. The other study introduced the so-called Prokaryotic Expression profiling by Tagging RNA In situ and sequencing (PETRI-seq) to study individual Gram-negative (*E. coli*) and Gram-positive (*S. aureus*) bacteria[139]. PETRI-seq also builds on a protocol previously developed for eukaryotic scRNA-seq; it uses combinatorial indexing to barcode transcripts in situ, a method originally known as 'SPLiT-seq' for Split-Pool Ligation-based Transcriptome sequencing[142]. Both MATQ-seq and PETRI-seq capture approximately 200–300 different mRNAs per single bacterium, that is, close to ~5% of all mRNAs in a typical bacterial cell[143] but two orders of magnitude up from previous approaches to studying the heterogeneity of gene expression in single bacteria with the help of fluorescent reporter genes. As of the time of writing, a third study reported microbial scRNA-seq at near genome-wide scale[144]. The protocol, termed microSPLiT, is also based on split-pool barcoding (like PETRI-seq) and was applied to single *B. subtilis* and *E. coli* cells and detected, on average, >300 mRNA copies per cell. Each of the three methods has strengths and weaknesses: MATQ-seq seems to have a lower dropout rate than PETRI-seq and microSPLiT, whereas split-pool barcoding overcomes the need to isolate single cells and thus offers much higher throughput. On a general note, although there are estimates of how many cells (and sequencing reads per cell) are typically required for an informative eukaryotic scRNA-seq study[145,146], bacterial scRNA-seq is too much in its infancy for such general guidelines. In the MATQ-seq and PETRI-seq studies, 19–27 or 204–875 bacteria, respectively, per condition were sequenced to ~60 million or ~40 million reads per library (without rRNA removal), but it remains unclear whether this was sufficient to resolve cellular variability in these samples to saturation. The microSPLiT study analysed >25,000 individual bacteria, enough to detect rare subpopulations[144]. The rule of thumb is that the complexity of the sample under investigation and the scope of the experiment will ultimately dictate the required cell numbers and sequencing depths.

The ramifications of single-bacterium RNA-seq for infection biology are manifold. To give just one example, persisters are a dangerous threat in diverse infectious diseases as the respective bacteria withstand antibiotic exposure and cause infection relapse[147]. However, within their host niche, persisters seem far from being transcriptionally inert[92]. Bacterial scRNA-seq of pathogenic persisters isolated from infected patients could help to study their in vivo activities and provide functional insights into when, how and why they reactivate. We expect that the technical refinement and automation of steps in the MATQ-seq, PETRI-seq and microSPLiT protocols will help bacterial scRNA-seq to become a widely used method for the study of microbial pathogens in infection settings.

Conclusions and future perspectives

RNA-seq has become a central methodology in the quest to understand the gene expression changes that ensue from host–microbe interactions. As described above, there has been a steady increase in the scope and sensitivity of RNA-seq-based methods. As regards scope, it is fair to say that, while the protein-centric history of infection biology tends to focus RNA-seq studies on mRNA expression levels, there is an increasing appreciation of the many non-coding transcripts that may change in either infection partner, that is, the microorganism and the host. For example, in human cells infected with *Salmonella*, lncRNA profiles changed faster than those of mRNAs, while in the bacteria themselves, several highly conserved sRNAs can be used as proxy for important regulons, inferring which type of stress the pathogen is experiencing[28]. Beyond their role as putative biomarkers, there is accumulating evidence that individual non-coding RNAs impact the outcome of host–microbe interactions, again with *Salmonella* Typhimurium infection models taking the lead. Certain host lncRNAs, for example, decrease the susceptibility to *Salmonella* infection[148,149] and individual miRNAs contribute to host defence against this pathogen[150–153], while others are actively exploited by *Salmonella* to favour pathogenesis[154]. In turn, *Salmonella* itself dispatches an arsenal of its own non-coding RNA elements in the tug-of-war between infection and clearance[28,155–158]. We expect the non-coding branch of host–microbe interaction studies to further expand in the future, particularly in the context of obligate anaerobic pathogens and commensals that have so far escaped in-depth analysis.

RNA modification is another aspect for expanding the scope of RNA-seq analysis of interspecies interaction. Traditionally known in rRNA and tRNA molecules, RNA modifications are now reported in many other cellular transcript classes, including mRNAs and regulatory non-coding RNAs in both bacteria[159] and mammals[160]. Modifications influence the base pairing, conformation and protein-binding properties of RNA molecules and have been shown to impact bacterial virulence and host–microbe interactions[161,162]. There is no shortage of protocols to profile RNA modifications in a transcriptome-wide manner for single-sided bacterial or eukaryotic studies[163,164]. These workflows are typically based on the chemical reactions of specific base modifications in the substrate RNA samples, resulting in differential reverse-transcription efficiencies during cDNA synthesis. Additionally, direct RNA sequencing using nanopores[165–167] (BOX 2) has the capability to infer modified ribonucleotides from the kinetic variation of the electrostatic potential in the pore. It is thus generic and may offer a possibility to simultaneously detect the

Box 3 | **Functional screens for factors shaping host–microbe interactions**

Deep sequencing proved useful for high-throughput genetic perturbation screens to uncover bacterial or host factors contributing to infection that may eventually be exploited as therapeutic targets. For instance, random mutagenesis followed by fitness screening of the resultant mutant pool can be applied to identify factors that are functionally important under the given selection pressure. Transposon insertion sequencing[233] has been widely harnessed to uncover virulence factors of diverse bacterial pathogens, including *Salmonella enterica* subsp. *enterica* serovar Typhimurium, *Haemophilus influenza*, *Vibrio cholerae*, *Neisseria meningitidis*, *Streptococcus* spp., *Legionella pneumophila*, *Acinetobacter baumannii*, *Bordetella pertussis*, *Brucella abortus*, *Coxiella burnetii* and *Enterococcus faecium*[234–247].

Random mutagenesis has also been applied for functional genomics in mice and in human cells[248,249]. However, given their larger genomes, genetic screens in mammals are typically performed in a more targeted manner. Initially, RNA interference screens were employed to identify host factors that influence their infection by bacterial pathogens; however, such screens typically yielded a high number of false-positives, likely due to 'off-targeting' by the small interfering RNA pool[250]. This drawback is overcome by recent CRISPR-based screens that have since been intensely used to address infection-related questions[251]. For human host models, multiplexed pools of single-guide RNAs are available[252]. This facilitates genome-wide screening approaches, for example, to identify host factors conferring resistance against intoxication by the Shiga toxin of enterohaemorrhagic *Escherichia coli*[253], against the type III secretion system of *Vibrio parahaemolyticus*[254], the typhoid toxin of *Salmonella enterica* subsp. enterica serovar Typhi[255] or the *Staphylococcus aureus* toxin leukocidin[256], and to uncover the colon epithelial receptor of *Clostridium difficile* toxin B[257].

In parallel, CRISPR tools have been developed to deliberately perturb target gene sets in bacteria[258] and can be exploited for in vivo screens of bacterial pathogenesis[259,260]. We predict that CRISPR-based perturbation screens will become a key technology for the discovery of novel bacterial colonization or virulence factors, especially amongst all the short genes[261,262] that are less likely to be hit by random mutagenesis. Additionally, CRISPR technologies arise — in parallel with antisense oligonucleotides[263] — as novel species-specific antibiotic candidates and programmable tools for microbiota editing[264–266].

epitranscriptomes of both the host and microorganism during their interplay.

With respect to sensitivity, ongoing progress in the acquisition[78,79] and analysis[86] of bacterial in vivo transcriptomes paves the way for an improved appreciation of bacterial gene expression within complex, host-associated communities. Inasmuch as scRNA-seq may seem to be in its infancy, it has already provided important insights into cellular heterogeneity arising during infection and it now works for both eukaryotic and microbial cells. It is important to stress that the protocols available could also fill the gap between bulk RNA-seq protocols, which commonly require input RNA in the nanogram to microgram range, and true single-cell studies, enabling robust analysis of biopsy samples from an infected organ or the small populations of bacteria in the gut of an insect infection model. Another application might be the profiling of the RNA content of extracellular vesicles, which are known to be produced by eukaryotic and bacterial cells alike[168,169]. The Extracellular RNA Communication Consortium (ERCC) investigates the roles these RNAs play in intercellular communication and their potential as biomarkers and therapeutic targets[170]. However, there is also an active debate about the putative function of vesicular RNA in interspecies communication[171], which would greatly benefit from having experimentally determined RNA cargo profiles from single vesicles, under physiological conditions.

Temporal and spatial resolution is also bound to increase further. With regards to deciphering expression kinetics, methods using metabolic labelling allow for the discrimination of de novo-transcribed from pre-existing RNA molecules and, thus, a better understanding of the order of gene expression[172]. This type of tracking of de novo transcription has been combined with scRNA-seq to provide a fine-grained temporal picture of the early cellular response to a viral attack[173]; it should also be straightforward to apply it to cells infected with bacterial cells. In parallel, spatial transcriptomics[174], including near-genome-wide fluorescence in situ hybridization[175] and in situ RNA-seq approaches that sequence nucleic acids directly in preserved tissue[176,177], are on the rise. The introduction of NICHE-seq[134] represented a breakthrough for studying the spatial aspects of host–microbe interactions at single-cell resolution. Although NICHE-seq is limited to local microenvironments, integrating spatial/temporal expression profiling at single-cell resolution with imaging and bioinformatic tools for higher-order tissue reconstruction has been achieved (reviewed in REFS[178,179]). Currently, technical constraints and high costs limit such analyses to small organs from mouse models. However, the potential to reconstruct infection processes in four dimensions at the tissue or organ scale is fascinating and would promise new avenues for interdisciplinary infection research.

The ultimate goal for RNA-seq in the present context is to simultaneously profile and correlate gene expression changes in single infected host cells, together with the pathogen. Dual scRNA-seq is in the starting blocks[79,180] but needs to be brought to a genome-wide scale. This should be in reach given that the now available bacterial scRNA-seq protocols[139,140] are independent of poly(A) and should thus be able to capture the full complement of both eukaryotic and bacterial RNA as does bulk dual RNA-seq[28]. Improved by targeted cDNA removal[37] and enrichment[79], dual scRNA-seq promises the dissection of host–microbe interactions at an unprecedented resolution.

A crucial inherent caveat of any transcriptomic approach is the fact that mRNA levels may not necessarily correlate with the abundance of the respectively encoded proteins[138]. This problem became very apparent in dual RNA-seq of the obligate intracellular pathogen *O. tsutsugamushi* and human endothelial cells, for which complementary proteomics data were obtained[99]. Mathematical modelling concluded that the bacterial mRNA expression levels alone resulted in predictions of protein abundance that were slightly better than chance. *Orientia* may be an extreme case owing to its high content of transposable elements being associated with pervasive antisense transcription. Indeed, consideration of the ratio of sense-to-antisense reads enhanced the predictability of protein levels from RNA-seq data substantially.

Nevertheless, such examples are a good reminder of the importance of validating RNA-seq-derived findings at the protein level through methods for protein quantification that range from low-throughput (for selected key factors by western blot or flow cytometry) to mid-throughput (multiplex ELISA or 'immuno-PCR'

with oligonucleotide-conjugated antibodies[181]) to high-throughput (BOX 1). Yet, even the presence of a protein does not necessarily guarantee its functional importance under the given condition. Instead, protein activity might be regulated by post-translational modifications, particularly so for enzymes and signalling molecules[182]. In this respect, although outside the focus of this Review, it is worth mentioning that sequencing-based perturbation screens are gaining traction for high-throughput studies of gene functionality within host–microbe interaction settings[183] (BOX 3). Integrating host–microbe transcriptomics with these functional genomics data will eventually result in testable hypotheses and propose RNA molecules that could serve as future biomarkers or drug targets to combat infectious diseases and microbial disorders.

Published online 17 February 2021

1. Bumann, D. Heterogeneous host-pathogen encounters: act locally, think globally. *Cell Host Microbe* **17**, 13–19 (2015).
2. Sarkar, S. & Heise, M. T. Mouse models as resources for studying infectious diseases. *Clin. Ther.* **41**, 1912–1922 (2019).
3. Barrila, J. et al. Modeling host-pathogen interactions in the context of the microenvironment: three-dimensional cell culture comes of age. *Infect. Immun.* https://doi.org/10.1128/IAI.00282-18 (2018).
4. Colgan, A. M., Cameron, A. D. & Kroger, C. If it transcribes, we can sequence it: mining the complexities of host-pathogen-environment interactions using RNA-seq. *Curr. Opin. Microbiol.* **36**, 37–46 (2017).
5. Westermann, A. J., Gorski, S. A. & Vogel, J. Dual RNA-seq of pathogen and host. *Nat. Rev. Microbiol.* **10**, 618–630 (2012).
A study describing a thought experiment that explores the concept of multi-organismal RNA-seq and coining the term 'dual RNA-seq'.
6. Eulalio, A., Schulte, L. & Vogel, J. The mammalian microRNA response to bacterial infections. *RNA Biol.* **9**, 742–750 (2012).
7. Agliano, F., Rathinam, V. A., Medvedev, A. E., Vanaja, S. K. & Vella, A. T. Long noncoding RNAs in host-pathogen interactions. *Trends Immunol.* **40**, 492–510 (2019).
8. Westermann, A. J. Regulatory RNAs in virulence and host-microbe interactions. *Microbiol. Spectr.* https://doi.org/10.1128/microbiolspec.RWR-0002-2017 (2018).
9. Saliba, A. E., Westermann, A. J., Gorski, S. A. & Vogel, J. Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Res.* **42**, 8845–8860 (2014).
10. Penaranda, C. & Hung, D. T. Single-cell RNA sequencing to understand host-pathogen interactions. *ACS Infect. Dis.* **5**, 336–344 (2019).
11. Dreyfus, M. & Regnier, P. The poly(A) tail of mRNAs: bodyguard in eukaryotes, scavenger in bacteria. *Cell* **111**, 611–613 (2002).
12. Wolin, S. L. & Steitz, J. A. Genes for two small cytoplasmic Ro RNAs are adjacent and appear to be single-copy in the human genome. *Cell* **32**, 735–744 (1983).
13. Sim, S. & Wolin, S. L. Bacterial Y RNAs: gates, tethers, and tRNA mimics. *Microbiol. Spectr.* https://doi.org/10.1128/microbiolspec.RWR-0023-2018 (2018).
14. Lundblad, E. W. & Altman, S. Inhibition of gene expression by RNase P. *N. Biotechnol.* **27**, 212–221 (2010).
15. Kroger, C. et al. An infection-relevant transcriptomic compendium for Salmonella enterica Serovar Typhimurium. *Cell Host Microbe* **14**, 683–695 (2013).
16. Hor, J., Matera, G., Vogel, J., Gottesman, S. & Storz, G. Trans-acting small RNAs and their effects on gene expression in Escherichia coli and Salmonella enterica. *EcoSal Plus* https://doi.org/10.1128/ecosalplus.ESP-0030-2019 (2020).
17. Rion, N. & Ruegg, M. A. LncRNA-encoded peptides: more than translational noise? *Cell Res.* **27**, 604–605 (2017).
18. Yao, R. W., Wang, Y. & Chen, L. L. Cellular functions of long noncoding RNAs. *Nat. Cell Biol.* **21**, 542–551 (2019).
19. Milo, R. & Phillips, R. *Cell Biology by the Number* (Garland Science, 2015).
20. Stark, R., Grzelak, M. & Hadfield, J. RNA sequencing: the teenage years. *Nat. Rev. Genet.* **20**, 631–656 (2019).
21. van Dijk, E. L., Jaszczyszyn, Y., Naquin, D. & Thermes, C. The third revolution in sequencing technology. *Trends Genet.* **34**, 666–681 (2018).
22. Cox, M. L. et al. Investigating fixative-induced changes in RNA quality and utility by microarray analysis. *Exp. Mol. Pathol.* **84**, 156–172 (2008).
23. Franzosa, E. A. et al. Relating the metatranscriptome and metagenome of the human gut. *Proc. Natl Acad. Sci. USA* **111**, E2329–E2338 (2014).
24. Westermann, A. J. & Vogel, J. Host-pathogen transcriptomics by dual RNA-Seq. *Methods Mol. Biol.* **1737**, 59–75 (2018).
25. Fronicke, L. et al. Toward cell type-specific in vivo dual RNA-Seq. *Methods Enzymol.* **612**, 505–522 (2018).
26. Avican, K. et al. Reprogramming of Yersinia from virulent to persistent mode revealed by complex in vivo RNA-seq analysis. *PLoS Pathog.* **11**, e1004600 (2015).
27. Connolly, J. P. R. et al. Host-associated niche metabolism controls enteric infection through fine-tuning the regulation of type 3 secretion. *Nat. Commun.* **9**, 4187 (2018).
28. Westermann, A. J. et al. Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions. *Nature* **529**, 496–501 (2016).
Comprehensive evaluation of the dual RNA-seq technology for 14 different mammalian host cell types infected with *Salmonella* and identification of virulence-related non-coding RNAs in this bacterial pathogen.
29. Westermann, A. J., Barquist, L. & Vogel, J. Resolving host-pathogen interactions by dual RNA-seq. *PLoS Pathog.* **13**, e1006033 (2017).
30. Jin, D. J., Cagliero, C. & Zhou, Y. N. Growth rate regulation in Escherichia coli. *FEMS Microbiol. Rev.* **36**, 269–287 (2012).
31. Goodfellow, S. J. & Zomerdijk, J. C. Basic mechanisms in RNA polymerase I transcription of the ribosomal RNA genes. *Subcell. Biochem.* **61**, 211–236 (2013).
32. Montoya, D. J. et al. Dual RNA-Seq of human leprosy lesions identifies bacterial determinants linked to host immune response. *Cell Rep.* **26**, 3574–3585.e3 (2019).
A dual RNA-seq study that measured gene expression of host and pathogen in human leprosy skin lesions, revealing that the human immune response is not primarily shaped by the bacterial dose but by the virulence programmes active in the infecting mycobacteria.
33. Sintsova, A. et al. Genetically diverse uropathogenic Escherichia coli adopt a common transcriptional program in patients with UTIs. *eLife* **8**, e49748 (2019).
34. Culviner, P. H., Guegler, C. K. & Laub, M. T. A simple, cost-effective, and robust method for rRNA depletion in RNA-sequencing studies. *mBio* **11**, e00010–20 (2020).
35. Giannoukos, G. et al. Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. *Genome Biol.* **13**, R23 (2012).
36. Petrova, O. E., Garcia-Alcalde, F., Zampaloni, C. & Sauer, K. Comparative evaluation of rRNA depletion procedures for the improved analysis of bacterial biofilm and mixed pathogen culture transcriptomes. *Sci Rep.* **7**, 41114 (2017).
37. Prezza, G. et al. Improved bacterial RNA-seq by Cas9-based depletion of ribosomal RNA reads. *RNA* **26**, 1069–1078 (2020).
The study reports the sensitive CRISPR-based depletion of unwanted bacterial sequences to improve targeted next-generation sequencing.
38. Gu, W. et al. Depletion of abundant sequences by hybridization (DASH): using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. *Genome Biol.* **17**, 41 (2016).
39. Zheng, G. et al. Efficient and quantitative high-throughput tRNA sequencing. *Nat. Methods* **12**, 835–837 (2015).
40. Cozen, A. E. et al. ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments. *Nat. Methods* **12**, 879–884 (2015).
41. Shigematsu, M. et al. YAMAT-seq: an efficient method for high-throughput sequencing of mature transfer RNAs. *Nucleic Acids Res.* **45**, e70 (2017).
42. Haas, B. J., Chin, M., Nusbaum, C., Birren, B. W. & Livny, J. How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes? *BMC Genomics* **13**, 734 (2012).
43. Liu, Y., Zhou, J. & White, K. P. RNA-seq differential expression studies: more sequence or more replication? *Bioinformatics* **30**, 301–304 (2014).
44. Ching, T., Huang, S. & Garmire, L. X. Power analysis and sample size estimation for RNA-Seq differential expression. *RNA* **20**, 1684–1696 (2014).
45. Engstrom, P. G. et al. Systematic evaluation of spliced alignment programs for RNA-seq data. *Nat. Methods* **10**, 1185–1191 (2013).
46. Dillies, M. A. et al. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief. Bioinforma.* **14**, 671–683 (2013).
47. Berghoff, B. A., Karlsson, T., Kallman, T., Wagner, E. G. H. & Grabherr, M. G. RNA-sequence data normalization through in silico prediction of reference genes: the bacterial response to DNA damage as case study. *BioData Min.* **10**, 30 (2017).
48. Munro, S. A. et al. Assessing technical performance in differential gene expression experiments with external spike-in RNA control ratio mixtures. *Nat. Commun.* **5**, 5125 (2014).
49. Consortium, S. M.-I. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nat. Biotechnol.* **32**, 903–914 (2014).
50. Loven, J. et al. Revisiting global gene expression analysis. *Cell* **151**, 476–482 (2012).
51. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
52. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
53. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
54. Garber, M., Grabherr, M. G., Guttman, M. & Trapnell, C. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat. Methods* **8**, 469–477 (2011).
55. Gene Ontology Consortium. Gene Ontology Consortium: going forward. *Nucleic Acids Res.* **43**, D1049–D1056 (2015).
56. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457–D462 (2016).
57. Caspi, R. et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* **44**, D471–D480 (2016).
58. Breuer, K. et al. InnateDB: systems biology of innate immunity and beyond–recent updates and continuing curation. *Nucleic Acids Res.* **41**, D1228–D1233 (2013).
59. Sharma, C. M. et al. The primary transcriptome of the major human pathogen Helicobacter pylori. *Nature* **464**, 250–255 (2010).
60. Kroger, C. et al. The transcriptional landscape and small RNAs of Salmonella enterica serovar Typhimurium. *Proc. Natl Acad. Sci. USA* **109**, E1277–E1286 (2012).
61. Slager, J., Aprianto, R. & Veening, J. W. Deep genome annotation of the opportunistic human pathogen Streptococcus pneumoniae D39. *Nucleic Acids Res.* **46**, 9971–9989 (2018).

62. Dar, D. et al. Term-seq reveals abundant ribo-regulation of antibiotics resistance in bacteria. *Science* **352**, aad9822 (2016).
The authors established Term-seq to map prokaryotic RNA 3′ ends at the genome-wide scale and, among other uses, applied the method to human oral microbiome samples.
63. Lalanne, J. B. et al. Evolutionary convergence of pathway-specific enzyme expression stoichiometry. *Cell* **173**, 749–761.e38 (2018).
64. Kroger, C. et al. The primary transcriptome, small RNAs and regulation of antimicrobial resistance in Acinetobacter baumannii ATCC 17978. *Nucleic Acids Res.* **46**, 9684–9698 (2018).
65. Dugar, G. et al. High-resolution transcriptome maps reveal strain-specific regulatory features of multiple Campylobacter jejuni isolates. *PLoS Genet.* **9**, e1003495 (2013).
66. Albrecht, M. et al. The transcriptional landscape of Chlamydia pneumoniae. *Genome Biol.* **12**, R98 (2011).
67. Thomason, M. K. et al. A rhlI 5′ UTR-derived sRNA regulates RhlR-dependent quorum sensing in Pseudomonas aeruginosa. *mBio* **10**, e02253-19 (2019).
68. Heidrich, N. et al. The primary transcriptome of Neisseria meningitidis and its interaction with the RNA chaperone Hfq. *Nucleic Acids Res.* **45**, 6147–6167 (2017).
69. Thomason, M. K. et al. Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in Escherichia coli. *J. Bacteriol.* **197**, 18–28 (2015).
70. Ryan, D., Jenniches, L., Reichardt, S., Barquist, L. & Westermann, A. J. A high-resolution transcriptome map identifies small RNA regulation of metabolism in the gut microbe Bacteroides thetaiotaomicron. *Nat. Commun.* **11**, 3557 (2020).
71. Avican, K. et al. RNA atlas of human bacterial pathogens uncovers stress dynamics linked to infection. Preprint at *bioRxiv* https://doi.org/10.1101/2020.06.29.177147 (2020).
72. Cossart, P., Boquet, P., Normark, S. & Rappuoli, R. Cellular microbiology emerging. *Science* **271**, 315–316 (1996).
73. Raynaud, S., Le Pabic, H. & Felden, B. Selective recovery of RNAs from bacterial pathogens after their internalization by human host cells. *Methods* **143**, 4–11 (2018).
74. Hinton, J. C., Hautefort, I., Eriksson, S., Thompson, A. & Rhen, M. Benefits and pitfalls of using microarrays to monitor bacterial gene expression during infection. *Curr. Opin. Microbiol.* **7**, 277–282 (2004).
75. Szafranska, A. K. et al. High-resolution transcriptomic analysis of the adaptive response of Staphylococcus aureus during acute and chronic phases of osteomyelitis. *mBio* **5**, e01775–14 (2014).
76. Rossi, E., Falcone, M., Molin, S. & Johansen, H. K. High-resolution in situ transcriptomics of Pseudomonas aeruginosa unveils genotype independent patho-phenotypes in cystic fibrosis lungs. *Nat. Commun.* **9**, 3459 (2018).
77. Mandlik, A. et al. RNA-Seq-based monitoring of infection-linked changes in Vibrio cholerae gene expression. *Cell Host Microbe* **10**, 165–174 (2011).
78. Donaldson, G. P. et al. Spatially distinct physiology of Bacteroides fragilis within the proximal colon of gnotobiotic mice. *Nat. Microbiol.* **5**, 746–756 (2020).
This study employed hybrid-selection RNA-seq to enrich *Bacteroides fragilis* transcripts from host–bacterial RNA mixtures to compare in vivo transcriptome signatures of this gut commensal between different host niches.
79. Betin, V. et al. Hybridization-based capture of pathogen mRNA enables paired host-pathogen transcriptional analysis. *Sci. Rep.* **9**, 19244 (2019).
80. Bashiardes, S., Zilberman-Schapira, G. & Elinav, E. Use of metatranscriptomics in microbiome research. *Bioinforma. Biol. Insights* **10**, 19–25 (2016).
81. Reck, M. et al. Stool metatranscriptomics: a technical guideline for mRNA stabilisation and isolation. *BMC Genomics* **16**, 494 (2015).
82. Engel, P. & Moran, N. A. The gut microbiota of insects - diversity in structure and function. *FEMS Microbiol. Rev.* **37**, 699–735 (2013).
83. Abu-Ali, G. S. et al. Metatranscriptome of human faecal microbial communities in a cohort of adult men. *Nat. Microbiol.* **3**, 356–366 (2018).
Comprehensive analysis of the human gut microbiome with a focus on ecological and molecular aspects and defining a 'core' and 'variable' metatranscriptome across participants.
84. Yan, Y. et al. Structure of the mucosal and stool microbiome in Lynch syndrome. *Cell Host Microbe* **27**, 585–600.e4 (2020).
85. Mehta, R. S. et al. Stability of the human faecal microbiome in a cohort of adult men. *Nat. Microbiol.* **3**, 347–355 (2018).
Analysis of the human gut microbiome from several hundred subjects revealed the stability of the metatranscriptome over time, highlighting its potential for diagnostics.
86. Franzosa, E. A. et al. Species-level functional profiling of metagenomes and metatranscriptomes. *Nat. Methods* **15**, 962–968 (2018).
This paper describes the launch of HUMAnN2, an algorithm to functionally analyse metatranscriptomic data from host-associated and environmental bacterial communities.
87. Menzel, P., Ng, K. L. & Krogh, A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat. Commun.* **7**, 11257 (2016).
88. Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**, R46 (2014).
89. Forouzan, E., Shariati, P., Mousavi Maleki, M. S., Karkhane, A. A. & Yakhchali, B. Practical evaluation of 11 de novo assemblers in metagenome assembly. *J. Microbiol. Methods* **151**, 99–105 (2018).
90. Zhang, Y. et al. Compositional and functional differences in the human gut microbiome correlate with clinical outcome following infection with wild-type salmonella enterica serovar typhi. *mBio* **9**, e00686–18 (2018).
91. Westermann, A. J. et al. The major RNA-binding protein ProQ impacts virulence gene expression in salmonella enterica serovar typhimurium. *mBio* **10**, e02504–18 (2019).
92. Stapels, D. A. C. et al. Salmonella persisters undermine host immune defenses during antibiotic treatment. *Science* **362**, 1156–1160 (2018).
Dual RNA-seq of *Salmonella*-infected mouse macrophages linked the polarization towards the M2-like macrophage phenotype with the secreted bacterial effector protein SteE.
93. Aprianto, R., Slager, J., Holsappel, S. & Veening, J. W. Time-resolved dual RNA-seq reveals extensive rewiring of lung epithelial and pneumococcal transcriptomes during early infection. *Genome Biol.* **17**, 198 (2016).
94. Mavromatis, C. H. et al. The co-transcriptome of uropathogenic Escherichia coli-infected mouse macrophages reveals new insights into host-pathogen interactions. *Cell Microbiol.* **17**, 730–746 (2015).
95. Baddal, B. et al. Dual RNA-seq of nontypeable haemophilus influenzae and host cell transcriptomes reveals novel insights into host-pathogen cross talk. *mBio* **6**, e01765-15 (2015).
96. Humphrys, M. S. et al. Simultaneous transcriptional profiling of bacteria and their host cells. *PLoS ONE* **8**, e80597 (2013).
97. Zimmermann, M. et al. Integration of metabolomics and transcriptomics reveals a complex diet of mycobacterium tuberculosis during early macrophage infection. *mSystems* **2**, e00057-17 (2017).
The study combined dual RNA-seq with host–pathogen metabolomics for a cell-culture model of *M. tuberculosis* infection and inferred a system-wide host–pathogen metabolic network.
98. Rienksma, R. A. et al. Comprehensive insights into transcriptional adaptation of intracellular mycobacteria by microbe-enriched dual RNA sequencing. *BMC Genomics* **16**, 34 (2015).
99. Mika-Gospodorz, B. et al. Dual RNA-seq of Orientia tsutsugamushi informs on host-pathogen interactions for this neglected intracellular human pathogen. *Nat. Commun.* **11**, 3363 (2020).
Dual RNA-seq unveiled differences in the pathogenicity of two strains of the obligate intracellular pathogen *O. tsutsugamushi* in an infected cell-culture model.
100. Schulte, L. N. et al. An advanced human intestinal coculture model reveals compartmentalized host and pathogen strategies during salmonella infection. *mBio* **11**, e03348–19 (2020).
101. Hannemann, S., Gao, B. & Galan, J. E. Salmonella modulation of host cell gene expression promotes its intracellular growth. *PLoS Pathog.* **9**, e1003668 (2013).
102. Thanert, R., Goldmann, O., Beineke, A. & Medina, E. Host-inherent variability influences the transcriptional response of Staphylococcus aureus during in vivo infection. *Nat. Commun.* **8**, 14268 (2017).
103. Nuss, A. M. et al. Tissue dual RNA-seq allows fast discovery of infection-specific functions and riboregulators shaping host-pathogen transcriptomes. *Proc. Natl Acad. Sci. USA* **114**, E791–E800 (2017).
One of the first applications of dual RNA-seq to an in vivo setting, analysing host and bacterial gene expression during the infection of mouse Peyer's patches with *Y. pseudotuberculosis*.
104. Damron, F. H., Oglesby-Sherrouse, A. G., Wilks, A. & Barbier, M. Dual-seq transcriptomics reveals the battle for iron during Pseudomonas aeruginosa acute murine pneumonia. *Sci. Rep.* **6**, 39172 (2016).
105. Minhas, V. et al. In vivo dual RNA-seq reveals that neutrophil recruitment underlies differential tissue tropism of Streptococcus pneumoniae. *Commun. Biol.* **3**, 293 (2020).
The authors devised a comparative dual RNA-seq approach to identify the impact of a SNP in a *Streptococcus raffinose* utilization regulator gene on infection outcome in a mouse model.
106. Griesenauer, B. et al. Determination of an interaction network between an extracellular bacterial pathogen and the human host. *mBio* **10**, e01193–19 (2019).
Combined dual RNA-seq and metabolomics study of punch biopsy samples from human volunteers inoculated with the bacterial skin pathogen *H. ducreyi*.
107. Thanert, R. et al. Molecular profiling of tissue biopsies reveals unique signatures associated with streptococcal necrotizing soft tissue infections. *Nat. Commun.* **10**, 3846 (2019).
Comprehensive host–bacterial transcriptomic analysis of necrotizing soft-tissue biopsy samples from human patients with either *S. pyogenes* monomicrobial or polymicrobial infections.
108. Pisu, D., Huang, L., Grenier, J. K. & Russell, D. G. Dual RNA-Seq of Mtb-infected macrophages in vivo reveals ontologically distinct host-pathogen interactions. *Cell Rep.* **30**, 335–350.e4 (2020).
In vivo dual RNA-seq of mouse lungs infected with *M. tuberculosis* that revealed the segregation of host–pathogen transcriptomes dependent on macrophage phenotype.
109. Jew, B. et al. Accurate estimation of cell composition in bulk expression through robust integration of single-cell information. *Nat. Commun.* **11**, 1971 (2020).
110. Zaitsev, K., Bambouskova, M., Swain, A. & Artyomov, M. N. Complete deconvolution of cellular mixtures based on linearity of transcriptional signatures. *Nat. Commun.* **10**, 2209 (2019).
111. Qiu, J. et al. Mixed-species RNA-seq for elucidation of non-cell-autonomous control of gene transcription. *Nat. Protoc.* **13**, 2176–2199 (2018).
112. Seelbinder, B. et al. Triple RNA-seq reveals synergy in a human virus-fungus co-infection model. *Cell Rep.* **33**, 108389 (2020).
First (and thus far only) triple RNA-seq study. It revealed the molecular basis of synergistic virulence strategies of two frequently co-occurring pulmonary pathogens.
113. Le-Bury, G. & Niedergang, F. Defective phagocytic properties of HIV-infected macrophages: how might they be implicated in the development of invasive salmonella typhimurium? *Front. Immunol.* **9**, 531 (2018).
114. Gordon, M. A. Salmonella infections in immunocompromised adults. *J. Infect.* **56**, 413–422 (2008).
115. Chertow, D. S. & Memoli, M. J. Bacterial coinfection in influenza a grand rounds review. *J. Am. Med. Assoc.* **309**, 275–282 (2013).
116. Steben, M. & Duarte-Franco, E. Human papillomavirus infection: epidemiology and pathophysiology. *Gynecol. Oncol.* **107** (Suppl. 1), S2–S5 (2007).
117. Perez-Losada, M., Castro-Nallar, E., Bendall, M. L., Freishtat, R. J. & Crandall, K. A. Dual transcriptomic profiling of host and microbiota during health and disease in pediatric asthma. *PLoS ONE* **10**, e0131819 (2015).
118. Ren, L. et al. Transcriptionally active lung microbiome and its association with bacterial biomass and host inflammatory status. *mSystems* **3**, e00199–18 (2018).
119. Bordenstein, S. R. & Theis, K. R. Host biology in light of the microbiome: ten principles of holobionts and hologenomes. *PLoS Biol.* **13**, e1002226 (2015).
120. Birnbaum, K. D. Power in numbers: single-cell RNA-Seq strategies to dissect complex tissues. *Annu. Rev. Genet.* **52**, 203–221 (2018).
121. Hautefort, I., Proenca, M. J. & Hinton, J. C. Single-copy green fluorescent protein gene fusions allow accurate measurement of Salmonella gene expression in vitro and during infection of mammalian cells. *Appl. Env. Microbiol.* **69**, 7480–7491 (2003).

122. Tsai, C. N. & Coombes, B. K. The role of the host in driving phenotypic heterogeneity in Salmonella. *Trends Microbiol.* **27**, 508–523 (2019).
123. Ackermann, M. et al. Self-destructive cooperation mediated by phenotypic noise. *Nature* **454**, 987–990 (2008).
124. Rivera-Chavez, F. et al. Salmonella uses energy taxis to benefit from intestinal inflammation. *PLoS Pathog.* **9**, e1003267 (2013).
125. Vieth, B., Parekh, S., Ziegenhain, C., Enard, W. & Hellmann, I. A systematic evaluation of single cell RNA-seq analysis pipelines. *Nat. Commun.* **10**, 4667 (2019).
126. Attar, M. et al. A practical solution for preserving single cells for RNA sequencing. *Sci. Rep.* **8**, 2151 (2018).
127. Avraham, R. et al. Pathogen cell-to-cell variability drives heterogeneity in host immune responses. *Cell* **162**, 1309–1321 (2015).
Combined eukaryotic single-cell and dual RNA-seq study that associated heterogeneity in *Salmonella* PhoP activity with interferon signalling in infected mouse macrophages.
128. Saliba, A. E. et al. Single-cell RNA-seq ties macrophage polarization to growth rate of intracellular Salmonella. *Nat. Microbiol.* **2**, 16206 (2016).
One of the first applications of eukaryotic single-cell RNA-seq to an infection model; revealed *Salmonella*-infected macrophages to differentiate into divergently polarized phenotypes with consequences for the intracellular behaviour of this pathogen.
129. Helaine, S. et al. Internalization of Salmonella by macrophages induces formation of nonreplicating persisters. *Science* **343**, 204–208 (2014).
130. Claudi, B. et al. Phenotypic variation of salmonella in host tissues delays eradication by antimicrobial chemotherapy. *Cell* **158**, 722–733 (2014).
131. Panagi, I. et al. Salmonella effector SteE converts the mammalian serine/threonine kinase GSK3 into a tyrosine kinase to direct macrophage polarization. *Cell Host Microbe* **27**, 41-53.e6 (2020).
132. Ben-Moshe, N. B. et al. Predicting bacterial infection outcomes using single cell RNA-sequencing analysis of human immune cells. *Nat. Commun.* **10**, 3266 (2019).
133. Haber, A. L. et al. A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333 (2017).
134. Medaglia, C. et al. Spatial reconstruction of immune niches by combining photoactivatable reporters and scRNA-seq. *Science* **358**, 1622–1626 (2017).
This study introduced NICHE-seq, which combines eukaryotic gene expression analysis at single-cell resolution with spatial information to dissect cellular and molecular aspects of infection niches.
135. Hagemann-Jensen, M., Abdullayev, I., Sandberg, R. & Faridani, O. R. Small-seq for single-cell small-RNA sequencing. *Nat. Protoc.* **13**, 2407–2424 (2018).
136. Faridani, O. R. et al. Single-cell sequencing of the small-RNA transcriptome. *Nat. Biotechnol.* **34**, 1264–1266 (2016).
137. Xiao, Z. et al. Holo-Seq: single-cell sequencing of holo-transcriptome. *Genome Biol.* **19**, 163 (2018).
138. Taniguchi, Y. et al. Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533–538 (2010).
139. Blattman, S. B., Jiang, W., Oikonomou, P. & Tavazoie, S. Prokaryotic single-cell RNA sequencing by in situ combinatorial indexing. *Nat. Microbiol.* **5**, 1192–1201 (2020).
Pioneering study of bacterial scRNA-seq using a SPLiT-seq-related protocol termed 'PETRI-seq' for individual *E. coli* and *S. aureus* cells.
140. Imdahl, F., Vafadarnejad, E., Homberger, C., Saliba, A.-E. & Vogel, J. Single-cell RNA-sequencing reports growth-condition-specific global transcriptomes of individual bacteria. *Nat. Microbiol.* **5**, 1202–1206 (2020).
Pioneering study of bacterial scRNA-seq that applied MATQ-seq to individual *S. typhimurium* and *P. aeruginosa* cells.
141. Sheng, K. et al. Effective detection of variation in single-cell transcriptomes using MATQ-seq. *Nat. Methods* **14**, 267–270 (2017).
142. Rosenberg, A. B. et al. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360**, 176–182 (2018).
143. Bartholomaus, A. et al. Bacteria differently regulate mRNA abundance to specifically respond to various stresses. *Philos. Trans. A Math. Phys. Eng. Sci.* **374**, 20150069 (2016).
144. Kuchina, A. et al. Microbial single-cell RNA sequencing by split-pool barcoding. *Science* https://doi.org/10.1126/science.aba5257 (2020).
Pioneering study of bacterial scRNA-seq via 'micro-SPLiT' that was used to detect the transcriptome of >25,000 individual *B. subtilis* cells.
145. Haque, A., Engel, J., Teichmann, S. A. & Lonnberg, T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med.* **9**, 75 (2017).
146. Luecken, M. D. & Theis, F. J. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol. Syst. Biol.* **15**, e8746 (2019).
147. Fisher, R. A., Gollan, B. & Helaine, S. Persistent bacterial infections and persister cells. *Nat. Rev. Microbiol.* **15**, 453–464 (2017).
148. Gomez, J. A. et al. The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon-gamma locus. *Cell* **152**, 743–754 (2013).
149. Imamura, K. et al. Diminished nuclear RNA decay upon Salmonella infection upregulates antibacterial noncoding RNAs. *EMBO J.* **37**, e97723 (2018).
150. Schulte, L. N., Eulalio, A., Mollenkopf, H. J., Reinhardt, R. & Vogel, J. Analysis of the host microRNA response to Salmonella uncovers the control of major cytokines by the let-7 family. *EMBO J.* **30**, 1977–1989 (2011).
151. Schulte, L. N., Westermann, A. J. & Vogel, J. Differential activation and functional specialization of miR-146 and miR-155 in innate immune sensing. *Nucleic Acids Res.* **41**, 542–553 (2013).
152. Maudet et al. Functional high-throughput screening identifies the miR-15 microRNA family as cellular restriction factors for Salmonella infection. *Nat. Commun.* **5**, 4718 (2014).
153. Rodriguez, A. et al. Requirement of bic/microRNA-155 for normal immune function. *Science* **316**, 608–611 (2007).
154. Aguilar, C. et al. Functional screenings reveal different requirements for host microRNAs in Salmonella and Shigella infection. *Nat. Microbiol.* **5**, 192–205 (2020).
155. Gong, H. et al. A Salmonella small non-coding RNA facilitates bacterial invasion and intracellular replication by modulating the expression of virulence factors. *PLoS Pathog.* **7**, e1002120 (2011).
156. Lee, E. J. & Groisman, E. A. An antisense RNA that governs the expression kinetics of a multifunctional virulence gene. *Mol. Microbiol.* **76**, 1020–1033 (2010).
157. Padalon-Brauch, G. et al. Small RNAs encoded within genetic islands of Salmonella typhimurium show host-induced expression and role in virulence. *Nucleic Acids Res.* **36**, 1913–1927 (2008).
158. Ellis, M. J. et al. Silent but deadly: IS200 promotes pathogenicity in Salmonella Typhimurium. *RNA Biol.* **15**, 176–181 (2018).
159. Hofer, K. & Jaschke, A. Epitranscriptomics: RNA modifications in bacteria and archaea. *Microbiol. Spectr.* https://doi.org/10.1128/microbiolspec.RWR-0015-2017 (2018).
160. Frye, M., Jaffrey, S. R., Pan, T., Rechavi, G. & Suzuki, T. RNA modifications: what have we learned and where are we headed? *Nat. Rev. Genet.* **17**, 365–372 (2016).
161. Rimbach, K., Kaiser, S., Helm, M., Dalpke, A. H. & Eigenbrod, T. 2'-O-methylation within bacterial RNA acts as suppressor of TLR7/TLR8 activation in human innate immune cells. *J. Innate Immun.* **7**, 482–493 (2015).
162. Shippy, D. C. & Fadl, A. A. tRNA modification enzymes GidA and MnmE: potential role in virulence of bacterial pathogens. *Int. J. Mol. Sci.* **15**, 18267–18280 (2014).
163. Helm, M. & Motorin, Y. Detecting RNA modifications in the epitranscriptome: predict and validate. *Nat. Rev. Genet.* **18**, 275–291 (2017).
164. Marbaniang, C. N. & Vogel, J. Emerging roles of RNA modifications in bacteria. *Curr. Opin. Microbiol.* **30**, 50–57 (2016).
165. Soneson, C. et al. A comprehensive examination of nanopore native RNA sequencing for characterization of complex transcriptomes. *Nat. Commun.* **10**, 3359 (2019).
166. Garalde, D. R. et al. Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods* **15**, 201–206 (2018).
167. Workman, R. E. et al. Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat. Methods* **16**, 1297–1305 (2019).
168. Tsatsaronis, J. A., Franch-Arroyo, S., Resch, U. & Charpentier, E. Extracellular vesicle RNA: a universal mediator of microbial communication? *Trends Microbiol.* **26**, 401–410 (2018).
169. Brody, H. Extracellular RNA. *Nature* https://doi.org/10.1038/d41586-020-01762-2 (2020).
170. Das, S. et al. The extracellular RNA communication consortium: establishing foundational knowledge and technologies for extracellular RNA research. *Cell* **177**, 231–242 (2019).
171. Lecrivain, A. L. & Beckmann, B. M. Bacterial RNA in extracellular vesicles: A new regulator of host-pathogen interactions? *Biochim. Biophys. Acta Gene Regul. Mech.* **1863**, 194519 (2020).
172. Baptista, M. A. P. & Dolken, L. RNA dynamics revealed by metabolic RNA labeling and biochemical nucleoside conversions. *Nat. Methods* **15**, 171–172 (2018).
173. Erhard, F. et al. scSLAM-seq reveals core features of transcription dynamics in single cells. *Nature* **571**, 419–423 (2019).
174. Crosetto, N., Bienko, M. & van Oudenaarden, A. Spatially resolved transcriptomics and beyond. *Nat. Rev. Genet.* **16**, 57–66 (2015).
175. Xia, C., Fan, J., Emanuel, G., Hao, J. & Zhuang, X. Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. *Proc. Natl Acad. Sci. USA* **116**, 19490–19499 (2019).
176. Ke, R. et al. In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* **10**, 857–860 (2013).
177. Wang, X. et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, eaat5691 (2018).
178. Liao, J., Lu, X., Shao, X., Zhu, L. & Fan, X. Uncovering an Organ's molecular architecture at single-cell resolution by spatially resolved transcriptomics. *Trends Biotechnol.* **39**, 43–58 (2021).
179. Wilbrey-Clark, A., Roberts, K. & Teichmann, S. A. Cell Atlas technologies and insights into tissue architecture. *Biochem. J.* **477**, 1427–1442 (2020).
180. Avital, G. et al. scDual-Seq: mapping the gene regulatory program of Salmonella infection by host and pathogen single-cell RNA-sequencing. *Genome Biol.* **18**, 200 (2017).
181. Sano, T., Smith, C. L. & Cantor, C. R. Immuno-PCR: very sensitive antigen detection by means of specific antibody-DNA conjugates. *Science* **258**, 120–122 (1992).
182. Ravikumar, V., Jers, C. & Mijakovic, I. Elucidating host-pathogen interactions based on post-translational modifications using proteomics approaches. *Front. Microbiol.* **6**, 1313 (2015).
183. Baddal, B. Next-generation technologies for studying host-pathogen interactions: a focus on dual transcriptomics, CRISPR/Cas9 screening and organs-on-chips. *Pathog. Dis.* **77**, ftz060 (2019).
184. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628 (2008).
185. Sultan, M. et al. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* **321**, 956–960 (2008).
186. Croucher, N. J. et al. A simple method for directional transcriptome sequencing using Illumina technology. *Nucleic Acids Res.* **37**, e148 (2009).
187. Guell, M. et al. Transcriptome complexity in a genome-reduced bacterium. *Science* **326**, 1268–1271 (2009).
188. Passalacqua, K. D. et al. Structure and complexity of a bacterial transcriptome. *J. Bacteriol.* **191**, 3203–3211 (2009).
189. Perkins, T. T. et al. A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus Salmonella typhi. *PLoS Genet.* **5**, e1000569 (2009).
190. Yoder-Himes, D. R. et al. Mapping the Burkholderia cenocepacia niche response via high-throughput sequencing. *Proc. Natl Acad. Sci. USA* **106**, 3976–3981 (2009).
191. Wurtzel, O. et al. A single-base resolution map of an archaeal transcriptome. *Genome Res.* **20**, 133–141 (2010).
192. Srikumar, S. et al. RNA-seq brings new insights to the intra-macrophage transcriptome of salmonella typhimurium. *PLoS Pathog.* **11**, e1005262 (2015).
193. Sittka, A. et al. Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet.* **4**, e1000163 (2008).
194. McNulty, N. P. et al. The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins. *Sci. Transl Med.* **3**, 106ra106 (2011).
195. Livny, J. et al. Comparative RNA-Seq based dissection of the regulatory networks and environmental stimuli underlying Vibrio parahaemolyticus gene expression during infection. *Nucleic Acids Res.* **42**, 12212–12223 (2014).

196. Galvez, E. J. C. et al. Distinct polysaccharide utilization determines interspecies competition between intestinal Prevotella spp. *Cell Host Microbe* **28**, 838–852.e6 (2020).
197. Ryan, D., Prezza, G. & Westermann, A. J. An RNA-centric view on gut Bacteroidetes. *Biol. Chem.* https://doi.org/10.1515/hsz-2020-0230 (2020).
198. Wilson, D. J. Insights from genomics into bacterial pathogen populations. *PLoS Pathog.* **8**, e1002874 (2012).
199. Sheppard, S. K., Guttman, D. S. & Fitzgerald, R. Population genomics of bacterial host adaptation. *Nat. Rev. Genet.* **19**, 549–565 (2018).
200. Moustafa, A. M., Lal, A. & Planet, P. J. Comparative genomics in infectious disease. *Curr. Opin. Microbiol.* **53**, 61–70 (2020).
201. Kellam, P. & Weiss, R. A. Infectogenomics: insights from the host genome into infectious diseases. *Cell* **124**, 695–697 (2006).
202. Khor, C. C. & Hibberd, M. L. Host-pathogen interactions revealed by human genome-wide surveys. *Trends Genet.* **28**, 233–243 (2012).
203. de Monerri, N. C. S. & Kim, K. Pathogens hijack the epigenome a new twist on host-pathogen interactions. *Am. J. Pathol.* **184**, 897–911 (2014).
204. Gomez-Diaz, E., Jorda, M., Peinado, M. A. & Rivero, A. Epigenetics of host-pathogen interactions: the road ahead and the road behind. *PLoS Pathog.* **8**, e1003007 (2012).
205. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. & Weissman, J. S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).
206. Fremin, B. J., Sberro, H. & Bhatt, A. S. MetaRibo-Seq measures translation in microbiomes. *Nat. Commun.* **11**, 3268 (2020).
A sensitive protocol for the performance of MetaRibo-seq to globally profile bacterial translation in complex consortia such as the human faecal microbiota.
207. Stern-Ginossar, N. Decoding viral infection by ribosome profiling. *J. Virol.* **89**, 6164–6166 (2015).
208. Holmes, M. J., Shah, P., Wek, R. C. & Sullivan, W. J. Jr. Simultaneous ribosome profiling of human host cells infected with toxoplasma gondii. *mSphere* **4**, e00292–19 (2019).
209. Leary, D. H., Hervey, W. J. T., Deschamps, J. R., Kusterbeck, A. W. & Vora, G. J. Which metaproteome? The impact of protein extraction bias on metaproteomic analyses. *Mol. Cell. Probes* **27**, 193–199 (2013).
210. Wilmes, P., Heintz-Buschart, A. & Bond, P. L. A decade of metaproteomics: where we stand and what the future holds. *Proteomics* **15**, 3409–3417 (2015).
211. Wang, Z., Han, Q. Q., Zhou, M. T., Chen, X. & Guo, L. Protein turnover analysis in Salmonella Typhimurium during infection by dynamic SILAC, Topograph, and quantitative proteomics. *J. Basic Microbiol.* **56**, 801–811 (2016).
212. Kaloyanova, D., Vogels, M., van Balkom, B. W. & Helms, J. B. Quantitative proteomic identification of host factors involved in the Salmonella typhimurium infection cycle. *Methods Mol. Biol.* **1225**, 29–45 (2015).
213. Selkrig, J. et al. Spatiotemporal proteomics uncovers cathepsin-dependent macrophage cell death during Salmonella infection. *Nat. Microbiol.* **5**, 1119–1133 (2020).
214. Auweter, S. D. et al. Quantitative mass spectrometry catalogues Salmonella pathogenicity island-2 effectors and identifies their cognate host binding partners. *J. Biol. Chem.* **286**, 24023–24035 (2011).
215. Jean Beltran, P. M., Federspiel, J. D., Sheng, X. & Cristea, I. M. Proteomics and integrative omic approaches for understanding host-pathogen interactions and infectious diseases. *Mol. Syst. Biol.* **13**, 922 (2017).
216. Byron, S. A., Van Keuren-Jensen, K. R., Engelthaler, D. M., Carpten, J. D. & Craig, D. W. Translating RNA sequencing into clinical diagnostics: opportunities and challenges. *Nat. Rev. Genet.* **17**, 257–271 (2016).
217. Seger, C. & Salzmann, L. After another decade: LC-MS/MS became routine in clinical diagnostics. *Clin. Biochem.* **82**, 2–11 (2020).
218. Hatzios, S. K. et al. Chemoproteomic profiling of host and pathogen enzymes active in cholera. *Nat. Chem. Biol.* **12**, 268–274 (2016).
219. Newsom, S. N. & McCall, L. I. Metabolomics: eavesdropping on silent conversations between hosts and their unwelcome guests. *PLoS Pathog.* **14**, e1006926 (2018).
220. Kentner, D. et al. Shigella reroutes host cell central metabolism to obtain high-flux nutrient supply for vigorous intracellular growth. *Proc. Natl Acad. Sci. USA* **111**, 9929–9934 (2014).
221. Beste, D. J. et al. 13C-flux spectral analysis of host-pathogen metabolism reveals a mixed diet for intracellular Mycobacterium tuberculosis. *Chem. Biol.* **20**, 1012–1021 (2013).
222. Garg, N. et al. Three-dimensional microbiome and metabolome cartography of a diseased human lung. *Cell Host Microbe* **22**, 705–716.e4 (2017).
223. Deatherage Kaiser, B. L. et al. A multi-omic view of host-pathogen-commensal interplay in salmonella-mediated intestinal infection. *PLoS ONE* **8**, e67155 (2013).
224. Lloyd-Price, J. et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* **569**, 655–662 (2019).
225. Bolyen, E. et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).
226. Shishkin, A. A. et al. Simultaneous generation of many RNA-seq libraries in a single reaction. *Nat. Methods* **12**, 323–325 (2015).
227. Wade, J. T. Where to begin? Mapping transcription start sites genome-wide in Escherichia coli. *J. Bacteriol.* **197**, 4–6 (2015).
228. Adiconis, X. et al. Comprehensive comparative analysis of 5'-end RNA-sequencing methods. *Nat. Methods* **15**, 505–511 (2018).
229. Tian, B. & Manley, J. L. Alternative cleavage and polyadenylation: the long and short of it. *Trends Biochem. Sci.* **38**, 312–320 (2013).
230. Ju, X., Li, D. & Liu, S. Full-length RNA profiling reveals pervasive bidirectional transcription terminators in bacteria. *Nat. Microbio.* **4**, 1907–1918 (2019).
231. Grünberger, F. et al. Exploring prokaryotic transcription, operon structures, rRNA maturation and modifications using Nanopore-based native RNA sequencing. Preprint at *bioRxiv* https://doi.org/10.1101/2019.12.18.880849 (2019).
232. Byrne, A., Cole, C., Volden, R. & Vollmers, C. Realizing the potential of full-length transcriptome sequencing. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **374**, 20190097 (2019).
233. Cain, A. K. et al. A decade of advances in transposon-insertion sequencing. *Nat. Rev. Genet.* **21**, 526–540 (2020).
234. Gawronski, J. D., Wong, S. M., Giannoukos, G., Ward, D. V. & Akerley, B. J. Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes required in the lung. *Proc. Natl Acad. Sci. USA* **106**, 16422–16427 (2009).
235. Gonyar, L. A. et al. In vivo gene essentiality and metabolism in Bordetella pertussis. *mSphere* **4**, e00694–18 (2019).
236. Sternon, J. F. et al. Transposon sequencing of brucella abortus uncovers essential genes for growth in vitro and inside macrophages. *Infect. Immun.* **86**, e00312–18 (2018).
237. Zhang, X. et al. RNA-seq and Tn-seq reveal fitness determinants of vancomycin-resistant Enterococcus faecium during growth in human serum. *BMC Genomics* **18**, 893 (2017).
238. Fu, Y., Waldor, M. K. & Mekalanos, J. J. Tn-Seq analysis of Vibrio cholerae intestinal colonization reveals a role for T6SS-mediated antibacterial activity in the host. *Cell Host Microbe* **14**, 652–663 (2013).
239. Chaudhuri, R. R. et al. Comprehensive assignment of roles for Salmonella typhimurium genes in intestinal colonization of food-producing animals. *PLoS Genet.* **9**, e1003456 (2013).
240. Vohra, P. et al. Retrospective application of transposon-directed insertion-site sequencing to investigate niche-specific virulence of Salmonella Typhimurium in cattle. *BMC Genomics* **20**, 20 (2019).
241. Shames, S. R. et al. Multiple Legionella pneumophila effector virulence phenotypes revealed through high-throughput analysis of targeted mutant libraries. *Proc. Natl Acad. Sci. USA* **114**, E10446–E10454 (2017).
242. Zhu, L. et al. Novel genes required for the fitness of streptococcus pyogenes in human saliva. *mSphere* **2**, e00460–17 (2017).
243. Subashchandrabose, S. et al. Acinetobacter baumannii genes required for bacterial survival during bloodstream infection. *mSphere* **1**, e00013–15 (2015).
244. Wang, H. et al. Hypermutation-induced in vivo oxidative stress resistance enhances Vibrio cholerae host adaptation. *PLoS Pathog.* **14**, e1007413 (2018).
245. Crabill, E., Schofield, W. B., Newton, H. J., Goodman, A. L. & Roy, C. R. Dot/Icm-translocated proteins important for biogenesis of the Coxiella burnetii-containing vacuole identified by screening of an effector mutant sublibrary. *Infect. Immun.* **86**, e00758–17 (2018).
246. Mann, B. et al. Control of virulence by small RNAs in Streptococcus pneumoniae. *PLoS Pathog.* **8**, e1002788 (2012).
247. Capel, E. et al. Comprehensive Identification of meningococcal genes and small noncoding RNAs required for host cell colonization. *mBio* **7**, e01173–16 (2016).
248. Carette, J. E. et al. Global gene disruption in human cells to assign genes to phenotypes by deep sequencing. *Nat. Biotechnol.* **29**, 542–546 (2011).
249. Friedrich, M. J. et al. Genome-wide transposon screening and quantitative insertion site sequencing for cancer gene discovery in mice. *Nat. Protoc.* **12**, 289–309 (2017).
250. Franceschini, A. et al. Specific inhibition of diverse pathogens in human cells by synthetic microRNA-like oligonucleotides inferred from RNAi screens. *Proc. Natl Acad. Sci. USA* **111**, 4548–4553 (2014).
251. Strich, J. R. & Chertow, D. S. CRISPR-cas biology and its application to infectious diseases. *J. Clin. Microbiol.* **57**, e01307-18 (2019).
252. Shalem, O. et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
253. Pacheco, A. R. et al. CRISPR screen reveals that EHEC's T3SS and shiga toxin rely on shared host factors for infection. *mBio* **9**, e01003–18 (2018).
254. Blondel, C. J. et al. CRISPR/Cas9 screens reveal requirements for host cell sulfation and fucosylation in bacterial type III secretion system-mediated cytotoxicity. *Cell Host Microbe* **20**, 226–237 (2016).
255. Chang, S. J., Jin, S. C., Jiao, X. & Galan, J. E. Unique features in the intracellular transport of typhoid toxin revealed by a genome-wide screen. *PLoS Pathog.* **15**, e1007704 (2019).
256. Tromp, A. T. et al. Human CD45 is an F-component-specific receptor for the staphylococcal toxin Panton-Valentine leukocidin. *Nat. Microbiol.* **3**, 708–717 (2018).
257. Tao, L. et al. Frizzled proteins are colonic epithelial receptors for C. difficile toxin B. *Nature* **538**, 350–355 (2016).
258. Vigouroux, A. & Bikard, D. CRISPR tools to control gene expression in bacteria. *Microbiol. Mol. Biol. Rev.* **84**, e00077–19 (2020).
259. Qu, J. et al. Modulating pathogenesis with mobile-CRISPRi. *J. Bacteriol.* **201**, e00304–19 (2019).
260. Liu, X. et al. Exploration of bacterial bottlenecks and Streptococcus pneumoniae pathogenesis by CRISPRi-Seq. *Cell Host Microbe* https://doi.org/10.1016/j.chom.2020.10.001 (2020).
261. Storz, G., Wolf, Y. I. & Ramamurthi, K. S. Small proteins can no longer be ignored. *Annu. Rev. Biochem.* **83**, 753-777 (2014).
262. Duval, M. & Cossart, P. Small bacterial and phagic proteins: an updated view on a rapidly moving field. *Curr. Opin. Microbiol.* **39**, 81–88 (2017).
263. Vogel, J. An RNA biology perspective on species-specific programmable RNA antibiotics. *Mol. Microbiol.* **113**, 550–559 (2020).
264. Bikard, D. et al. Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials. *Nat. Biotechnol.* **32**, 1146–1150 (2014).
265. Citorik, R. J., Mimee, M. & Lu, T. K. Sequence-specific antimicrobials using efficiently delivered RNA-guided nucleases. *Nat. Biotechnol.* **32**, 1141–1145 (2014).
266. Beisel, C. L., Gomaa, A. A. & Barrangou, R. A CRISPR design for next-generation antimicrobials. *Genome Biol.* **15**, 516 (2014).

Acknowledgements
The authors acknowledge financial support from the German Research Foundation (DFG): Individual Research Grant We6689/1-1 (to A.J.W.) and Leibnitz Award Vo875/18 (to J.V.).

Author contributions
The authors contributed equally to all aspects of the article.

Competing interests
The authors declare no competing interests.

Peer review information
Nature Reviews Genetics thanks M. Fällman, G. Myers, M. Waldor and J.-W. Veening for their contribution to the peer review of this work.

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