www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277-7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2023; 12(7): 122-132

www.thepharmajournal.com
Received: 21-04-2023
Accepted: 23-05-2023

Syam S
Research Scholar, Division of Microbiology, ICAR-Indian Agricultural Research Institute (IARI), New Delhi, India

Bhavya K
Research Scholar, Division of Microbiology, ICAR-Indian Agricultural Research Institute (IARI), New Delhi, India

Hirudhaya Ravi
Research Scholar, Division of Microbiology, ICAR-Indian Agricultural Research Institute (IARI), New Delhi, India

Teenu Paul
Research Scholar, Division of Microbiology, ICAR-Indian Agricultural Research Institute (IARI), New Delhi, India

JD Khumukcham
Research Scholar, Division of Genetics and Plant Breeding, ICAR-Indian Agricultural Research Institute (IARI), New Delhi, India

Sowmya P
Research Scholar, Division of Microbiology, ICAR-Indian Agricultural Research Institute (IARI), New Delhi, India

Corresponding Author:
Syam S
Research Scholar, Division of Microbiology, ICAR-Indian Agricultural Research Institute (IARI), New Delhi, India

Exploring microbial phenotypic multifariousness with single-Cell RNA sequencing

Syam S, Bhavya K, Hirudhaya Ravi, Teenu Paul, JD Khumukcham and Sowmya P

Abstract
Even with bacterial cells grown under various circumstances, bacterial gene expression is very diverse; bacteria differentiate into subpopulations that may play various roles in the community's survival. Researchers found a distinctive microbial gene regulation mechanism in the quantitative single-cell RNA sequencing using bioinformatics pipelines. Advances in single-cell RNA sequencing (scRNA-seq) have allowed for comprehensive analysis of the bacterial phenotypic community analysis. Prokaryotes' use of scRNA-seq has been limited by their extremely low mRNA quantity, absence of mRNA polyadenylation, and thick cell walls. In the present scenario, as a complementary technique, microSPLiT and PETRIseq use combinatorial indexing-based scRNA-seq procedures to analyse hundreds of bacteria at the same time. Researchers developed MicroSPLiT, a scalable single-cell RNA sequencing technological advancement designed exclusively for bacteria. MicroSPLiT detected a wide range of gene expression levels in over 25,000 single B. subtilis cells, including unusual and unexpected cell states not before observed at the population level. MicroSPLiT is a new approach for analysing single-cell gene expression in complex natural and synthetic microbial communities that has high scalability and resolution. PETRI-seq is a low-cost, high-throughput prokaryotic scRNA-sequencing pipeline. PETRI-seq barcodes transcripts from tens of thousands of cells in a single experiment via in situ combinatorial indexing. PETRI-seq efficiently and accurately collects single-cell transcriptomes of Gram-negative and Gram-positive bacteria. We anticipate that MicroSPLiT and PETRI-seq along with other technologies will be useful in identifying single-cell states and their dynamics in complex microbial diversity.

Keywords: Single-cell RNA sequencing, microSPLiT, PETRIseq, bacterial heterogeneity

1. Introduction
Microbes have developed complex strategies to respond to their environment and escape the immune system by individualizing their behaviour. While single-cell RNA sequencing has become instrumental for studying mammalian cells, its use with fungi, protozoa and bacteria is still in its infancy. RNA sequencing (RNA-seq) has become an essential tool in molecular biology. It is used in studies of all aspects of microbiology, from fundamental prokaryote physiology and RNA-based regulation processes to advance antibiotic susceptibility diagnostics. More recently, single-cell RNA-seq (scRNA-seq) has flourished as a popular tool to map eukaryotic cellular states by providing an inventory of all transcripts at single-nucleotide resolution in a single cell. scRNA-seq is a powerful technique that allows researchers to study the transcriptomes of individual cells. In recent years, this technology has been applied to the study of microbial communities, enabling researchers to investigate the phenotypic diversity within these communities at the single-cell level.
Single-cell RNA sequencing (scRNA-seq) is an agnostic approach to examine heterogeneity and has been successfully applied to eukaryotic cells. Unfortunately, current extensively utilized methods of eukaryotic scRNA-seq present difficulties when applied to bacteria. Specifically, bacteria have a cell wall, which makes eukaryotic lysis methods incompatible, bacterial mRNA has a shorter half-life and lower copy numbers, and isolating an individual bacterial species from a mixed community is difficult.
Bulk RNA sequencing technologies have provided invaluable insights into host and bacterial gene expression and associated regulatory networks. Nevertheless, the majority of these approaches report average expression across cell populations, hiding the true underlying expression patterns that are often heterogeneous in nature. Due to technical advances, single-cell transcriptomic in bacteria has recently become reality, allowing exploration of these heterogeneous populations, which are often the result of environmental changes and stressors.

Prokaryotic cells are much smaller than eukaryotic cells, leading to less input material per cell. Single bacteria contain RNA in the femtogram range and the average mRNA copy number is low, at only 0.4 copies/cell. Further challenges include efficient cell lysis, which is hampered by the bacterial cell wall, and capture of non-polyadenylated bacterial transcripts.

Evolution

Microbial diversity has recently been explored in a great variety of environments, including soil, sea, air, and the human body, including from a medical perspective, the gastrointestinal tract. These microbial communities have been characterized in terms of community structure, composition, metabolic function, and ecological roles. Investigations of environmental microbial diversity have employed the 16S rRNA (16S) gene marker, which offers phylogenetic taxonomic classification without requiring isolation and cultivation.

Metagenomics Approach

Metagenomics is a culture-independent genomic analysis of entire microbial communities inhabiting a particular niche. Metagenomics investigations aimed at finding "who's there and what are they doing" (Board on Life Sciences, 2007) are providing new insight into the genetic variability and metabolic capabilities of unknown or uncultured microorganisms. the metagenomic approach from 16S high-throughput methods, as the principal purpose of metagenomic is to explore the entire gene content of metagenomes for metabolic pathways and to understand microbial community interactions (Tringe *et al*., 2005), rather than targeting a single gene such as 16S.

Pyrosequencing Technology

Pyrosequencing is a DNA sequencing method based on the sequencing-by-synthesis principle, which was first described in 1985. This method relies on efficient detection of the sequential incorporation of natural nucleotides during DNA synthesis.

During the reaction, the Klenow fragment of DNA polymerase I releases inorganic pyrophosphate molecules (PPi) upon the addition of one nucleotide to a primer hybridized to a single-stranded DNA template. The second reaction, catalysed by ATP sulfurase, produces ATP, using the released PPi as a substrate. The ATP molecules are then converted to a luminometric signal by the luciferase enzyme. Therefore, the signal light is detected only if a base pair is formed with the DNA template, and the signal strength is proportional to the number of nucleotides incorporated in a single nucleotide flow. Finally, the unincorporated nucleotides and excess ATP are degraded between base additions by a nucleotide-degrading enzyme such as apyrase (Ronaghi *et al*., 1998); at this point, another dNTP is added and a new cycle begins.

Single-cell isolation following time-lapse imaging (SIFT) is an integrated platform combining long-term bacterial culture over consecutive generations under uniform and stable growth conditions, time-lapse imaging, optical trapping, and collection for downstream genomics analysis

Watterson *et al*. introduced a droplet-based microfluidic device that anaerobically isolates and cultivates microbial cells for several days in millions of picoliter chambers. Droplets containing bacteria can be individually sorted using an image-based analysis and enriched for rare slow growing taxa. These end-to-end systems (isolation, culture, imaging and sorting) pave the way for the analysis of rare phenotypes or rare populations at the single-cell level.

'Targeted' single-cell profiling technologies

A large number of techniques have leveraged advances in microscopy, cytometry, molecular biology and, most recently, next-generation sequencing to profile single cells. Many of these approaches have been developed and optimized to be used in studies that aim to immune cell heterogeneity, but they can differ by orders of magnitude in terms of the number of cells that can be analysed per experiment (the breadth of cellular profiling) and the number of genes per cell that can be detected (The depth of cellular profiling). "Targeted' technologies can assess a pre-selected set of molecular dimensions (Pre-selected genes for mRNA expression studies and protein-level detection) across hundreds to millions of cells using known molecular baits — such as fluorescently labelled oligonucleotide probes, fluorescent or metal-conjugated antibodies, or PCR primers — to profile genes or proteins with single-cell resolution.

Dating to 2009, pioneering work established single-cell RNA-seq in eukaryotes. While this field rapidly evolved, the development of scRNA-seq in bacteria was slow to progress due to several challenges. The development of scRNA-seq approaches has allowed for unbiased single-cell transcriptome profiling enabling the discovery of new cellular states, the profiling of genetic heterogeneity ranging from single nucleotide polymorphisms to diverse bacterial sequences, and the study of the transcriptomes of non-model organisms.

Nevertheless, thanks to technical advances, bacterial single-cell transcriptomics has recently become a reality. Three general types of approaches are currently available.

1. Bacterial multiple annealing and deoxycytidine (dC) tailing-based quantitative scRNA seq (MATQ-seq) is a workflow originally developed for eukaryotes that relies on cell isolation by fluorescence-activated cell sorting (FACS) and random priming of cellular transcripts.
2. A second type, also previously established for eukaryotes and termed split-pool ligation transcriptomics sequencing (SPLiT-seq), is based on combinatorial barcoding. It was adapted for bacterial scRNA-seq in two independent studies introducing the so-called prokaryotic expression profiling by tagging RNA in situ and sequencing (PETRI-seq) and microbial SPLiT-seq (microSPLiT) protocols. In comparison to MATQ-seq, bacterial split-pool barcoding workflows enable the analysis of thousands of cells instead of a few hundred, offsetting the lower transcript capture rate and higher rate of cell loss in these protocols.
3. The third type is a microscopy- and probe-based approach that does not employ RNA-seq. It is called parallel sequential fluorescence in situ hybridization (par-seqFISH) and allows spatial transcriptomics on the level of single bacteria.

The emergence of single-cell RNA sequencing (scRNA-seq) techniques make it now possible to observe transcriptome-wide heterogeneity agnostically, without any preconceived bias introduced in studies when choosing a limited number of genes studied simultaneously. Spatial transcriptomic tools including seqFISH, MERFISH, and others (Lubeck *et al*., 2014; Wang *et al*., 2018) [11, 12] combine single-cell transcriptomics with spatial mapping of each cell.

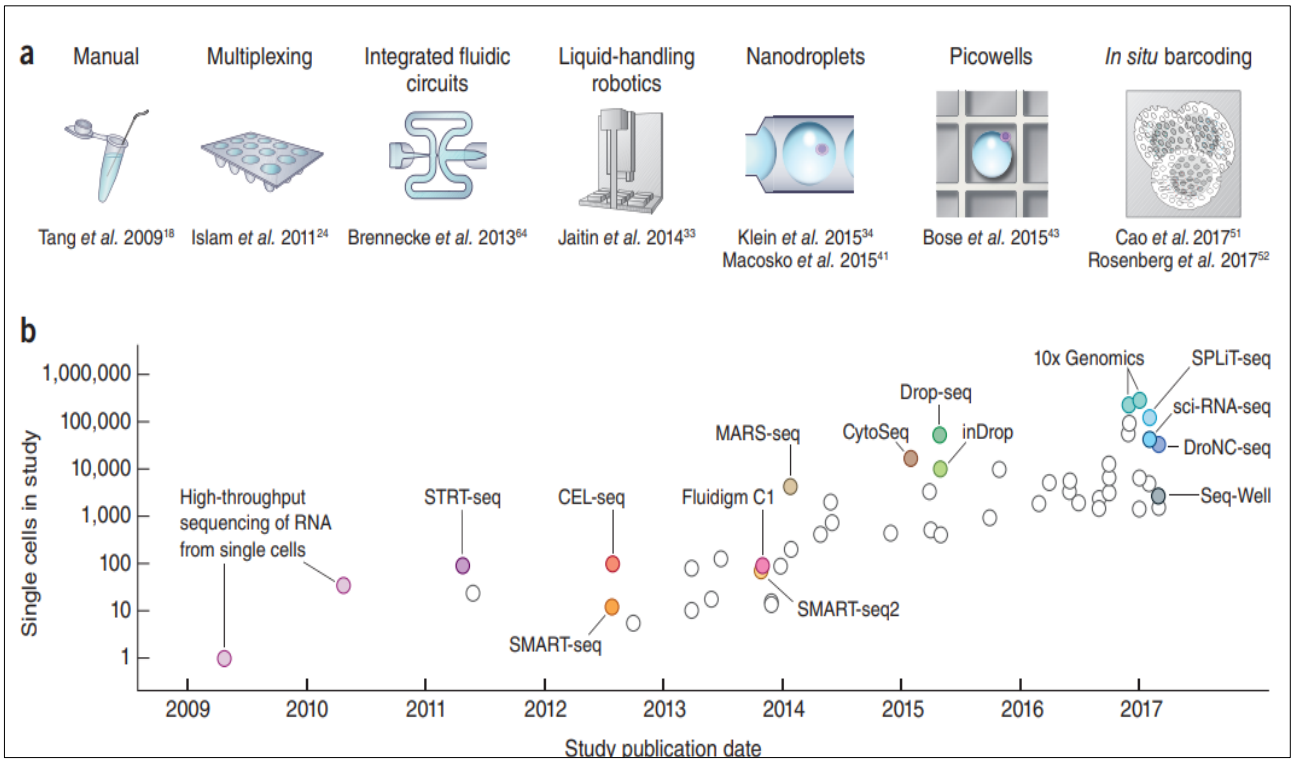

Fig 1: Scaling of scRNA-seq experiments.

(a) Key technologies that have allowed jumps in experimental scale. A jump to ~100 cells was enabled by sample multiplexing, and then a jump to ~1,000 cells was achieved by large-scale studies using integrated fluidic circuits, followed by a jump to several thousands of cells with liquid-handling robotics. Further orders-of-magnitude increases bringing the number of cells assayed into the tens of thousands were enabled by random capture technologies using nanodroplets and picowell technologies. Recent studies have used in situ barcoding to inexpensively reach the next order of magnitude of hundreds of thousands of cells.
(b) Cell numbers reported in representative publications by publication date. Key technologies are indicated.
In the single-cell tagged reverse transcription sequencing (STRT-seq) method, full-length cDNA is amplified by template switching, but only the 5′-end fragment is captured and sequenced. Fragments from the full length cDNA are sequenced in the SMART-seq method. The CEL-seq (Cell Expression by Linear amplification and sequencing) protocol leveraged this for linear mRNA amplification from single cells.
The RT adaptor also contains a T7 promoter sequence, which allows the final double-stranded DNA to be transcribed into multiple copies of antisense RNA32. Once enough amplified antisense RNA has been produced, it can be fragmented and again reverse-transcribed to cDNA. This method was also used for MARS-seq (massively parallel scRNA-seq), and later for InDrop (droplet sequencing) and Seq-Well (a well-array-based variation on MARS-seq).
The first commercial tool for passive cell capture for scRNAseq was the microfluidic C1 system released by Fluidigm. Feature of the Fluidigm technology is the design of microfluidics devices (or chips) that allow the sequential delivery of very small and precise volumes into tiny reaction chambers. Cells are loaded onto the chip and are passively captured in (up to) 96 isolated chambers in about half an hour. “Multiplexing” refers to the practice of processing and analyzing multiple samples at once. In sequencing assays this is achieved through the addition of molecular barcodes to cDNA fragments. the first single-cell RNA-seq protocol for multiplexing cells from a single 96-well plate by using a unique template-switching oligo (TSO) in each well via their STRT-seq method (a later version of STRT-seq removed early multiplexing in favor of passive cell capture with the Fluidigm C1 platform).
The first, combinatorial approach was first described with the targeted CytoSeq method. The authors used initial pools of 96 barcodes into which beads were split-pooled three times to generate a set of 963 = 884,736 barcoded beads.
The second approach was taken in the Drop-seq and Seq-Well protocols, where they used 12 rounds of split-pool single base DNA synthesis on the beads to generate 4^{12} = 16.7 million barcodes. A remarkable strategy that allows a user to avoid both of these pitfalls is combinatorial in situ barcoding. This technique was initially devised for single-cell assay for transposase-accessible chromatin using sequencing (scATAC seq) and later was adapted for whole-genome sequencing and single-cell Hi-C. Recently, these concepts were adapted to scRNA-seq in the single-cell combinatorial-indexing RNAseq and split-pool ligation-based transcriptome sequencing methods. Here single cells are never individually isolated and lysed; instead, cells are fixed and the mRNA is manipulated in situ inside each cell.

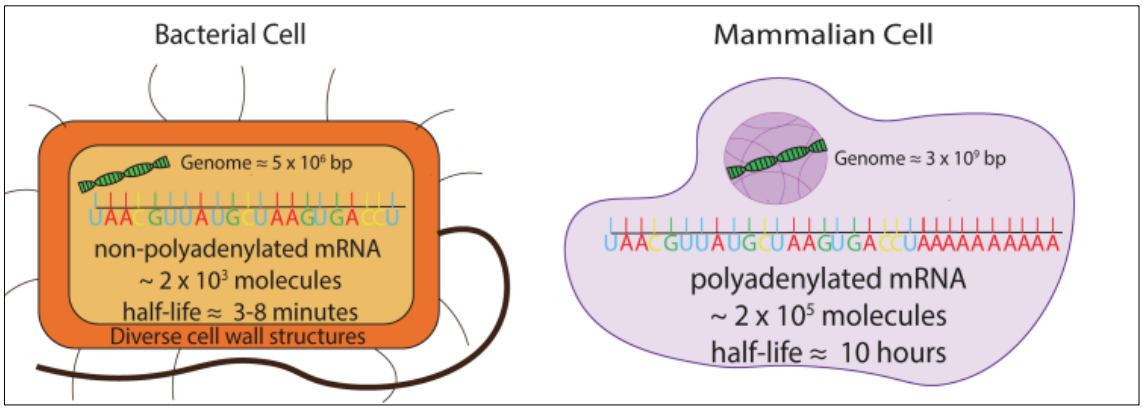

Fig 2: Comparison study of bacterial and mammalian cell

Drawbacks

Challenges in studying the RNA content of a single microbe (Imdahl, F., & Saliba, A. E.2020) [4].

Performing single-cell transcriptomics of microbes is constrained by major technical hurdles. First, one needs to access microbial RNA. Yet, there is no universal method that streamlines single-microbe lysis in a fast (few seconds) and high-throughput manner.

Bacterial mRNA is non-polyadenylated, making it hard to separate from ribosomal RNA that accounts for more than 90% of all RNA. Systematic investigation of the mean copy number of mRNA molecules per bacteria and yeast per gene has reported barely more than a few copies per cell. These methods also suffer from a high dropout rate (i.e. the percentage of mRNA molecules that fail to be captured), which ranges from 26% to 74%.

A general limitation of current scRNA-seq workflows is efficient cell lysis, which might require species-specific customization. Consequently, analysis of mixed bacterial communities is a challenge. This is especially true if their cell wall compositions vary, as this necessitates different enzymatic disruption and poses the risk of introducing bias based on varying lysis efficiency. (Imdahl, F *et al*., 2020) [5]

Technical barriers exist regarding bacterial scRNA-seq. Because

- Bacterial cells contain only femtograms of RNA12—that is, >100-fold less than typical eukaryotic cells—a sensitive complementary DNA synthesis and amplification protocol is required.
- Bacterial messenger RNAs are intrinsically labile (with half-lives of several minutes, as compared to hours for eukaryotes), meaning that perforation of the thick bacterial envelope, cell lysis and subsequent RNA stabilization must all occur rapidly.
- the absence of a poly(A) tail in functional bacterial transcripts precludes the straightforward reverse transcription (RT) strategy that is commonly used for eukaryotes to selectively enrich mRNAs and concomitantly deplete ribosomal RNAs.
- Lastly, whereas most current eukaryotic scRNA-seq protocols have a lower detection limit of between five and ten copies per transcript per cell, bacterial scRNA-seq must deal with a much lower average mRNA copy number (0.4copies per cell).

Investigating Microbial Heterogeneity

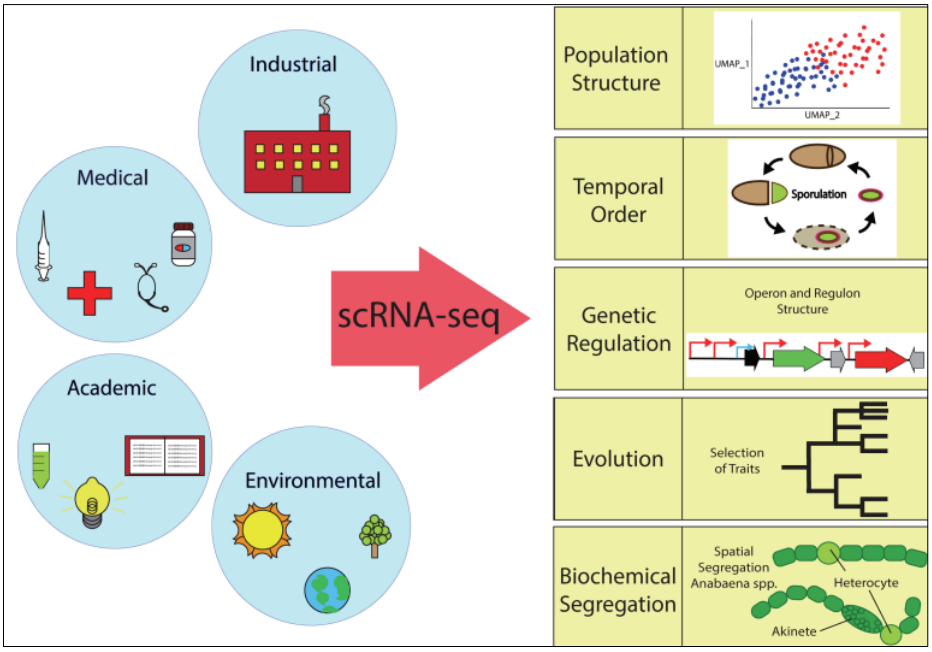

Fig 3: Single-Cell RNA-seq Allows High-Throughput Characterization of Population Structure

In bacterial systems, the first reports were able to provide this information and the structure and presence of known populations, such as sporulation and competence, have been reported (Kuchina *et al*., 2021; McNulty *et al*., 2021) [6, 10]. Similarly a rare subpopulation was discovered in Bacillus subtilis expressing PBSX prophage genes (Kuchina *et al*., 2021) [6], a rare prophage-induced subpopulation of S. aureus (Blattman *et al*., 2020) [2], and a new cell type expressing arginine synthesis genes was discovered in Bacillus subtilis (McNulty *et al*., 2021) [10]. Single Cell RNA-Seq Ties Genotype and Phenotype Together in Lab-Based Evolution Studies and Provides Insight Into the Underlying Forces That Govern Cellular Heterogeneity

Experimental evolution studies have uncovered many rules of natural selection in many organisms (Kawecki *et al*., 2012). Experimental evolution studies using microbes have made many contributions to better understand evolutionary processes because of their fast generation time and small genomes, coupling quick adaptation with straightforward genetic analysis.

Methodology

Single-bacteria RNA-seq

The RNA content of a bacterium is tenfold to hundred-fold lower than that of single mammal cell and a fungus, reaching levels of 1–100 femtograms. As mRNA, transcripts are not polyadenylated, yet another challenge emerges in discriminating mRNA from ribosomal RNA.

Multiple Annealing and Tailing-based Quantitative single cell RNA sequencing (MATQ-seq) is a scRNA-seq protocol that initiates reverse transcription using random primers and uses multiple annealing rounds without intermediate melting steps to enhance the capture of rare transcripts. Imdahl *et al*. built on this protocol to achieve single-bacteria RNA-seq using *Salmonella enterica serovar Typhimurium* and *Pseudomonas aeruginosa*. A protocol remains lab-intensive and can handle only a few hundred bacteria. (Imdahl, F *et al*., 2020) [5].

Starts with the isolation of single cells from culture, followed by one-tube cell lysis and cDNA synthesis and amplification. To obtain sufficient cDNA from single bacteria, we used the poly (A)-independent multiple annealing and dC-tailing-based quantitative scRNA-seq (MATQ-seq) protocol, in which RT primers are hybridized to internal transcript regions at low temperature, thus enabling the theoretical detection of all transcripts including those of low abundance.

Within the MATQ-seq protocol, library preparation and quality control consist of a series of different labour-intensive pipetting steps. We implemented a user-friendly and highly flexible automation process by establishing protocols for all pipetting steps on the I.DOT dispensing robot, with the exception of clean up and quality control steps (Fig. 1). This decreased turnaround times and the amount of consumables needed. Concurrently, automation increased sample throughput. For cDNA analysis and quality control, we integrated the high-throughput Qubit Flex fluorometer for a precise and fast procedure. To facilitate sample processing, we applied a miniaturization step for the final PCR volume by skipping the splitting step of the PCR implemented in the original MATQ-seq protocol. This allowed processing of up to 96 single cells per MATQ-seq passage, compared to a maximum of 24 cells in the previous protocol, and decreased the overall processing time from about 10 to 8 h. More importantly, the hands-on time was reduced from about 6 to 3 h. Finally, we updated our data processing and analysis pipeline to improve data quality by implementing a better trimming approach, alignment, normalization, and identification of outliers. Overall, all these steps discussed above led to higher cell throughput, improved accuracy (Fig.2 &3).

MATQ-seq allows the transcriptome of a single bacterium to be captured. Reverse transcription (RT) with multiple cycles and ramping temperatures is followed by polyC tailing of the first strand of cDNA. PCR is launched after the second strand synthesis, followed by library preparation.

Detailed workflow of the MATQ-seq protocol separated into two main steps: cell isolation and cDNA synthesis (left) and library preparation including DASH for rRNA depletion (right).

The use of SuperScript IV (SS IV) for reverse transcription, reaction optimization, and integration of DASH into the library preparation.

All pipetting steps were automated using the I.DOT dispensing robot, with the exception of all clean up and quality control steps.

Due to the high transcript capture rate of MATQ-seq, this method is particularly well suited for experimental settings in which the starting material is limited, such as the analysis of small subpopulations of bacterial cells in host niches or intracellular bacteria. In these settings, the application of dual RNA-seq allows the study of host-pathogen interactions through the simultaneous analysis of the transcriptomes of both the bacteria and the host.

Single-cell dual RNA-seq (scDual-Seq) has been attempted, but so far, either bacterial gene detection has been inefficient or the experiments were performed under a high multiplicity of infection, which does not reflect physiological conditions. Since MATQ-seq was initially developed for scRNA-seq in eukaryotes, we see high potential in establishing scDual-Seq with MATQ-seq to capture both eukaryotic and prokaryotic transcript.

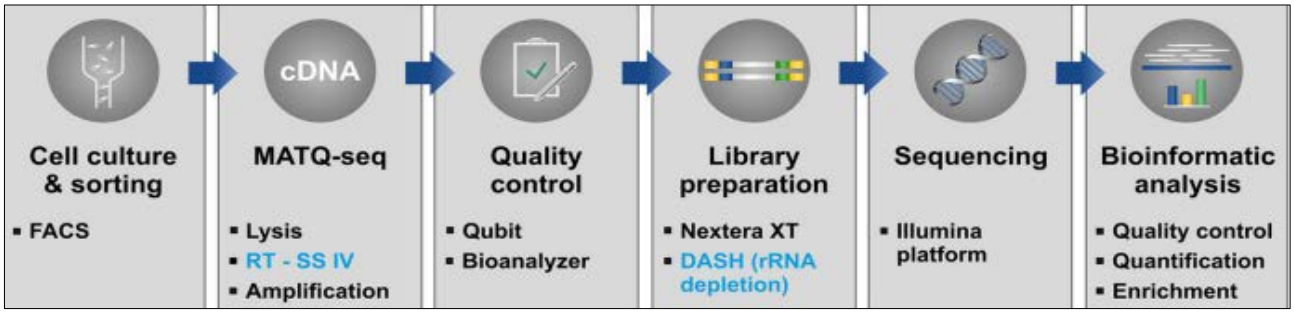

Fig 4: Overview of bacterial scRNA-seq pipeline including major steps from cell culture to bioinformatics analysis.

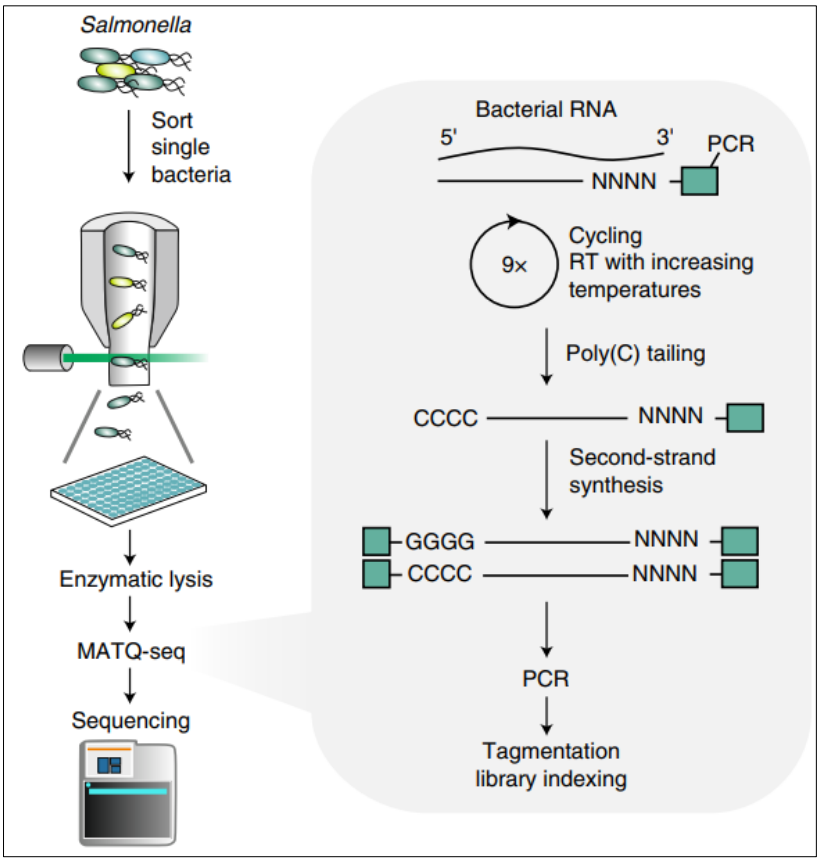

Fig 5: MATQ-seq Steps from cell culture to bioinformatic analysis.

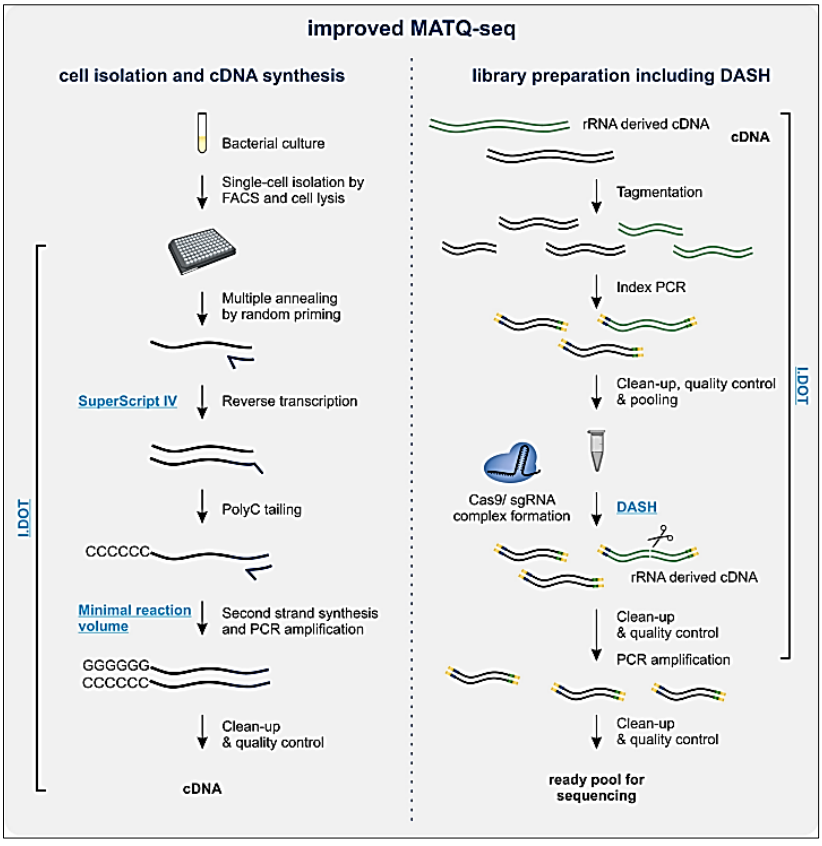

Fig 6: Steps from cell culture to bioinformatic analysis. Changes from the previous MATQ-seq protocol are highlighted in blue.

As a complementary approach, microbial split-pool ligation transcriptomics (microSPLiT) and prokaryotic expression profiling by tagging RNA in situ and sequencing (PETRIseq) build on the combinatorial indexing-based scRNA-seq protocols to analyze thousands of bacteria simultaneously. Here, cDNA with its associated barcodes is generated in situ.

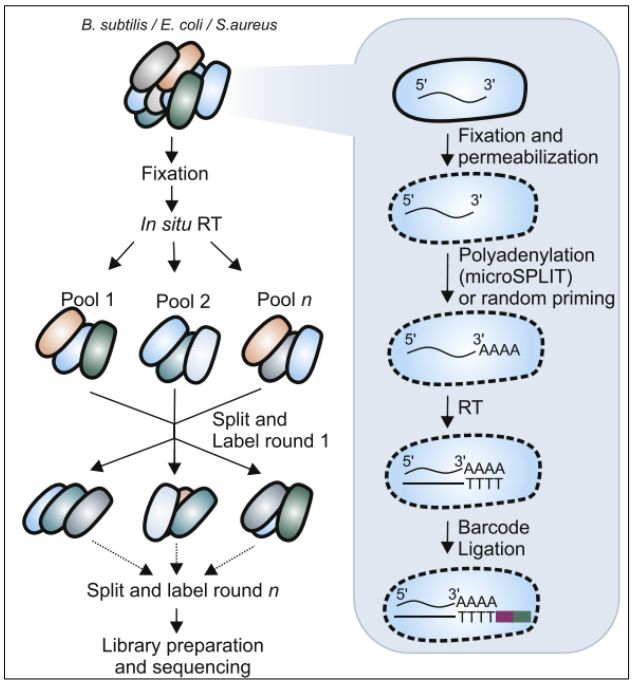

Fig 7: Microbial split-pool ligation transcriptomics (microSPLiT) and prokaryotic expression profiling by tagging RNA in situ and sequencing (PETRIseq) steps based on combinatorial indexing.

MicroSPLiT and PETRI-seq use the principle of combinatorial indexing (left; split and pool) to achieve single-bacteria RNA-seq of a microbial community. Each bacterium (right insert) is fixed and permeabilized, and transcripts are poly adenylated before RT or RT is initiated using random primers. cDNA is barcoded successively at each round, which gives each cell a unique identity.

MicroSPLiT includes an mRNA-specific polyadenylation reaction so that rRNA molecules are not carried along with the reaction. Together, these protocols demonstrate that single bacteria RNA-seq is possible across Gram positive and Gram negative species (*Escherichia coli*, *Bacillus subtilis*, *Salmonella Typhimurium*, *Pseudomonas aeruginosa*, *Staphylococcus aureus*) and can capture a few hundred transcripts per bacteria reporting for a specific growth condition (e.g. heat, salt concentration, optical density).

Combinatorial indexing-based protocols can cope better with efficient cell lysis limitation than MATQ-seq, because these protocols can process many more cells at once. A higher number of input cells therefore compensates for inefficient lysis, although the danger of introducing bias remains. In contrast, MATQ-seq is more limited in throughput because of the cell-sorting step and therefore inefficient lysis will lead to a high rate of failure. Nevertheless, MATQ-seq can in principle be applied to mixed cell populations if lysis conditions can be optimized.

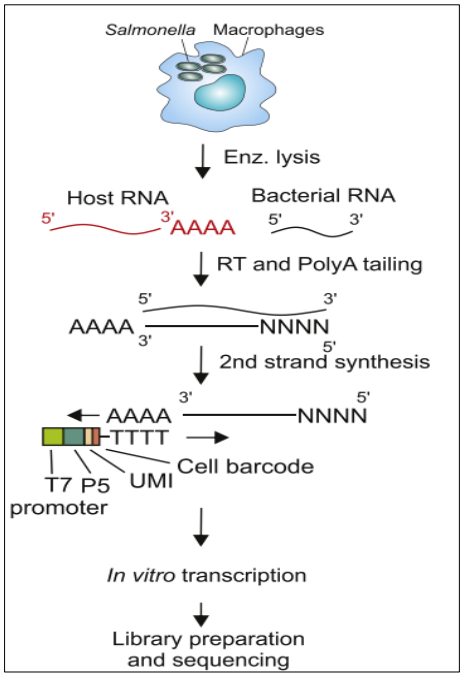

Fig 8: Cell Expression by Linear amplification and Sequencing (CEL-Seq2) Steps

CEL-seq2 aims to capture the transcriptome of the host and invading intracellular bacteria simultaneously. After the second strand synthesis and the integration of a T7 promoter, *in vitro* transcription amplifies the initial input, before library preparation and sequencing. (Yanai, I., & Hashimshony, T., 2019) [9].

CEL-Seq2—Single-Cell RNA Sequencing by Multiplexed Linear Amplification. CEL-Seq uses a polyT primer that introduces a cell-specific barcode next to the transcripts' polyA and selects the 30 ends of the transcripts for sequencing. This reduces sequencing costs, as less reads are needed compared to methods that cover the entire length of the transcript. With CEL-Seq2, we implemented the unique molecular identifiers (UMI) approach, which further reduced amplification biases. CEL-Seq works well on manually collected single cells and cells sorted by FACS but has also been combined with automation and fluidics to increase throughput.

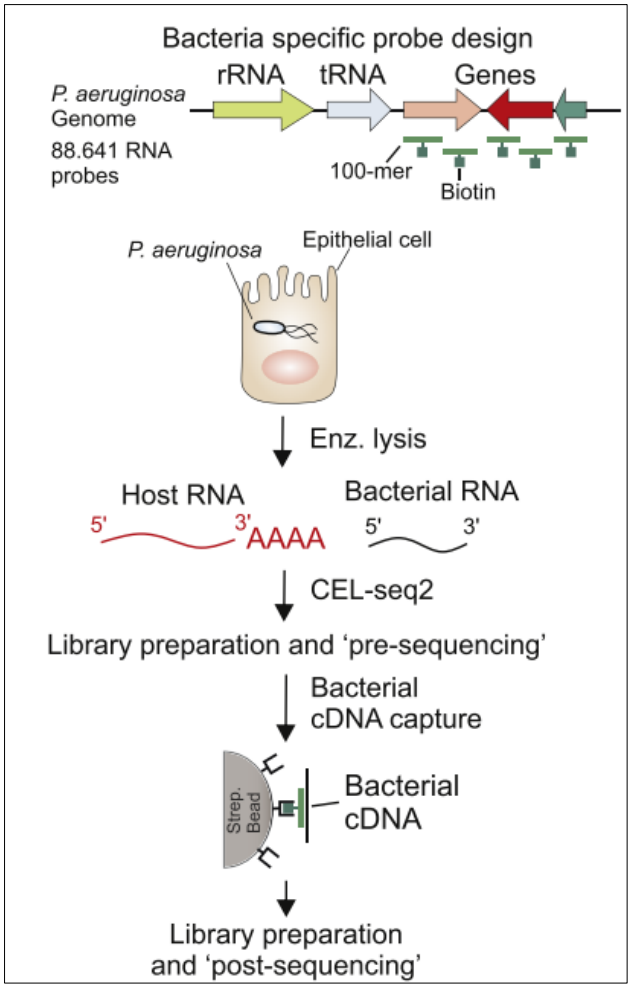

Fig 9: Steps of Pathogen Hybrid Capture (PatH-Cap)

PatHCap allows enhancing the detection of bacterial transcripts by capturing bacterial cDNA with a set of pre-designed biotinylated RNA-probes as illustrated here for P. aeruginosa. cDNA is generated using CEL-seq2 (panel (c)) and is sequenced two times before and after bacterial cDNA capture

Droplet-based method is CEL-seq2, CEL-seq2 supports unique molecular identifier (UMI), and to lower amplification biases in CEL-seq2, mRNA amplification is often completed through *in vitro* transcription rather than PCR.For the profiling of full-length RNA, the MARS-seq2.0 method, applied in a plate-based setup, has always been used. As bacterial mRNA represents as little as 0.05% of the total RNA of both organisms [61], the challenge is immense. Yanai *et al*. adapted a mammalian scRNA-seq approach called cell expression by linear amplification and sequencing (CEL-Seq2) to capture the transcriptome of macrophages infected with intracellular Salmonella cells (Figure 5). To reach the single bacterium level, Betin *et al*. engineered probe sets to capture the pathogen specifically, a method they call Pathogen Hybrid Capture (PatH-Cap) (Figure 6)

scRNAseq protocols are highly prone to dropouts, while single-molecule fluorescence in situ hybridization (smFISH) remains the gold standard for quantifying transcripts. smFISH to track the expression of genes at the single-cell level in the major bacterial models E. coli and B. subtilis.

Bioinformatics and its pipelines (Hwang, B *et al*., 2019) [3]

Mapping genotypes to phenotypes is one of the long-standing challenges in biology and medicine, and a powerful strategy for tackling this problem is performing transcriptome analysis. scRNA-seq is increasingly being utilized to delineate cell lineage relationships in early development, myoblast differentiation, and lymphocyte fate determination.

Common steps required for the generation of scRNA seq libraries include cell lysis, reverse transcription into first-strand cDNA, second-strand synthesis, and cDNA amplification. In general, cells are lysed in a hypotonic buffer,

and poly (A)+ selection is performed using poly (dT) primers to capture messenger RNAs (mRNAs). It has been well established that due to Poisson sampling, only 10–20% of transcripts will be reverse transcribed at this stage. This low mRNA capture efficiency is an important challenge that remains in existing scRNA-seq protocols and necessitates a highly efficient cell lysing strategy.

Pre-processing the data

Once reads are obtained from well-designed scRNA-seq experiments, quality control (QC) is performed. Low-quality bases (usually at the 3′ end) and adapter sequences can be removed at this pre-processing step. Existing QC tools available, FastQC (Babraham Institute, http://www.bioinformatics.babraham.ac.uk/projects/ fastqc/) is a popular tool for inspecting quality distributions across entire reads

Read alignment is the next step of scRNA-seq analysis

The tools available for this procedure, including the Burrows-Wheeler Aligner (BWA) and STAR are the same as those used in the bulk RNA seq analysis pipeline. When UMIs are implemented, these sequences should be trimmed prior to alignment.

The RNA-seQC program provides post-alignment summary stats, such as uniquely mapped reads, reads mapped to annotated exonic regions, and coverage patterns associated with specific library preparation protocols.

After alignment, reads are allocated to exonic, intronic, or intergenic features using transcript annotation in General Transcript Format. Only reads that map to exonic loci with high mapping quality are considered for generation of the gene expression matrix (N (cells)×m (genes)). A distinctive feature of scRNA-seq data is the presence of zero-inflated counts due to reasons such as dropout or transient gene expression

Raw expression counts are normalized using scaling factor estimates by standardizing across cells, assuming that most genes are not differentially expressed. The most commonly used approaches include RPKM50, FPKM, and transcripts per kilobase million (TPM).

RPKM, for example, is calculated as (exonic read×10^9)/(exon length×total mapped read). The only difference between RPKM and FPKM is that FPKM considers the read count in one of the aligned mates if paired-end sequencing is performed. TPM is a modification of RPKM in which the sum of all TPMs in each sample is consistent across samples (exonic read×mean read length×10^6 /exon length×total transcript).

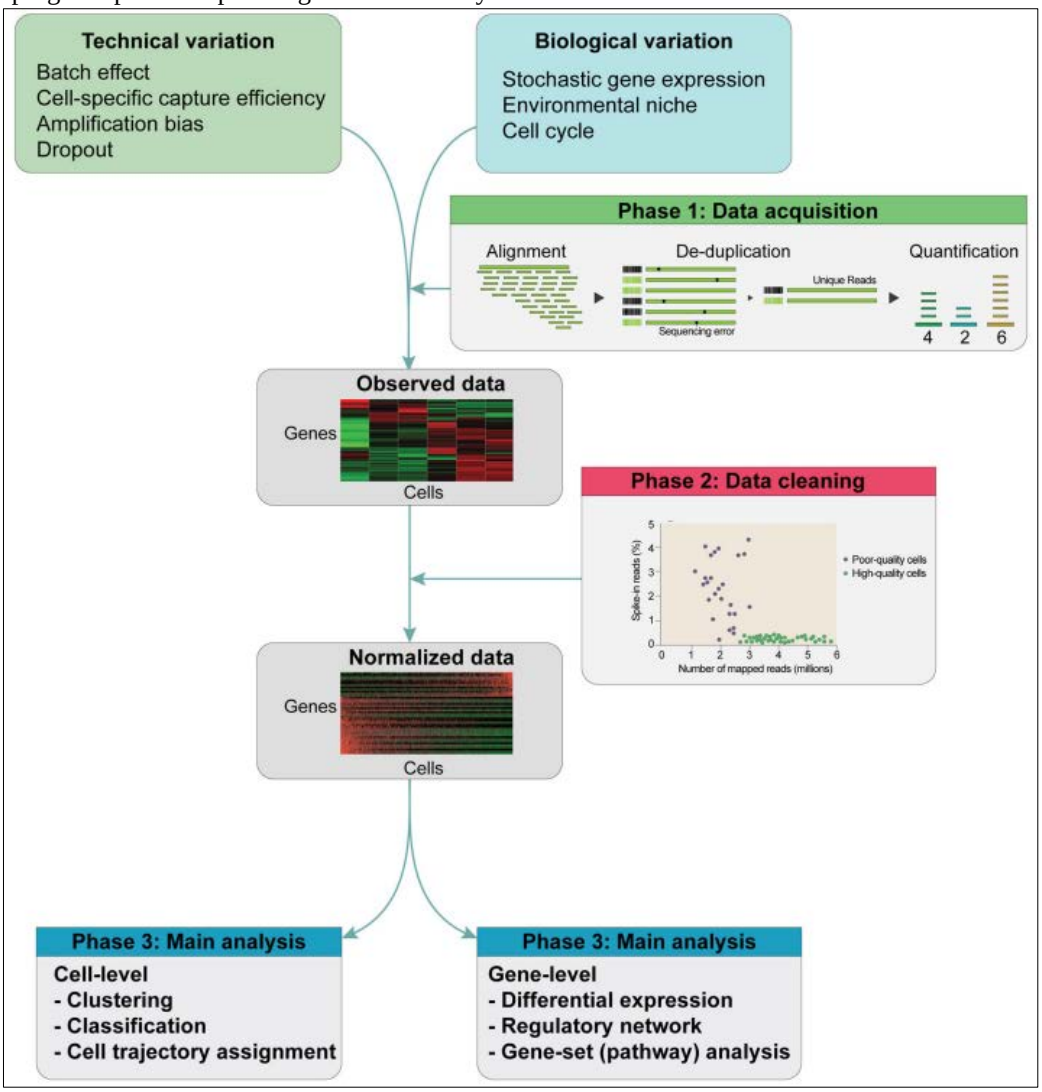

Fig 10: A schematic overview of scRNA-seq analysis pipelines.

scRNA-seq data are inherently noisy with confounding factors, such as technical and biological variables. After sequencing, alignment and de-duplication are performed to quantify an initial gene expression profile matrix. Next, normalization is performed with raw expression data using various statistical methods. Additional QC can be performed when using spike-ins by inspecting the mapping ratio to discard low-quality cells. Finally, the normalized matrix is then subjected to main analysis through clustering of cells to identify subtypes. Cell trajectories can be inferred based on these data and by detecting differentially expressed genes between clusters.

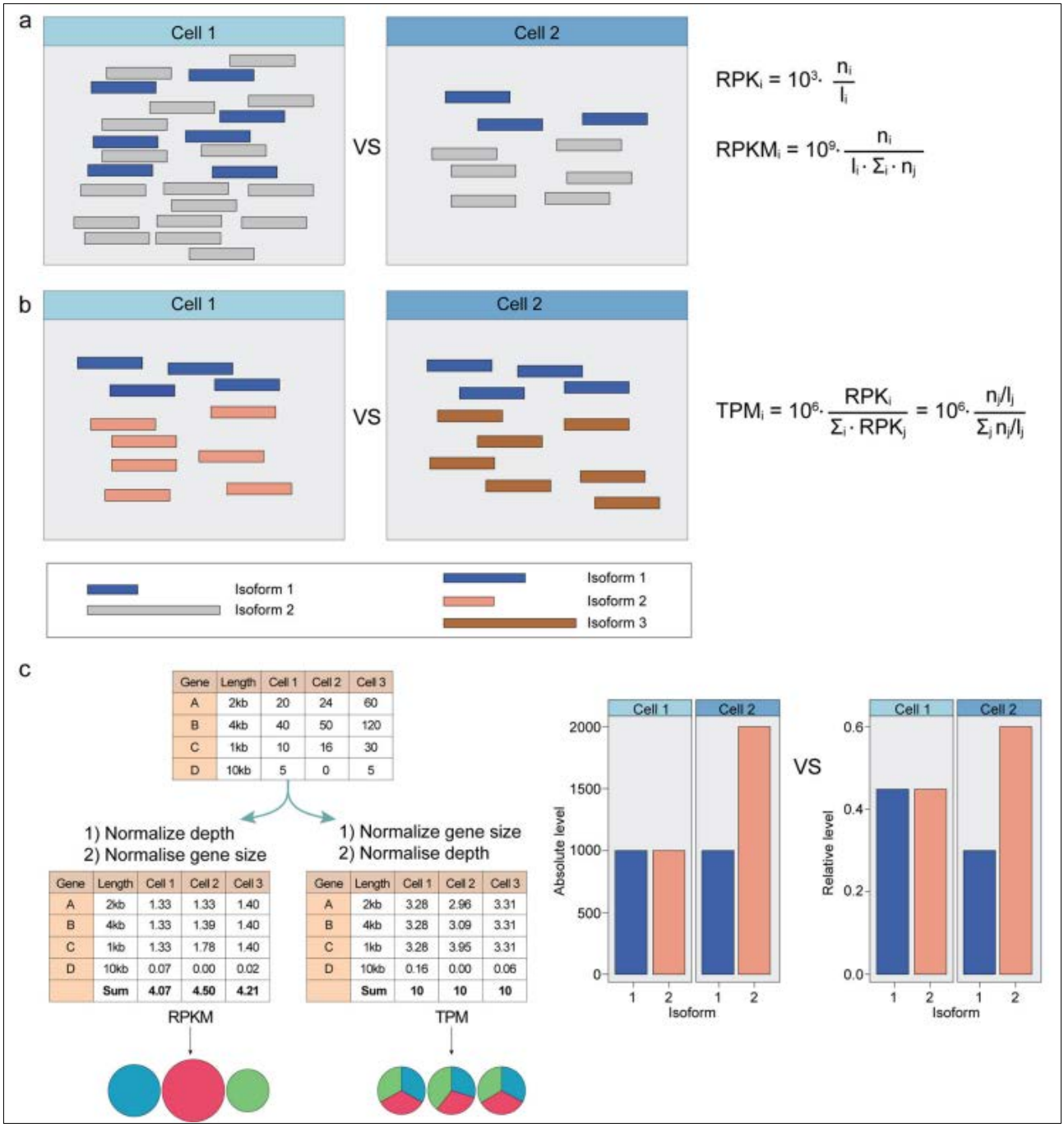

Fig 11: Methods for the quantification of expression in scRNA-seq

Applications

1. Characterizing distinct cell subsets.

2. Characterizing the heterogeneity of a population.

3. Cell Type identification

By projecting cells into 2D space, we can easily visualize samples with increased interpretability (Fig.). Additional non-linear dimensionality reduction methods, such as t-distributed stochastic neighbor embedding (tSNE), multidimensional scaling, locally linear embedding (LLE), and Isomap , can also be utilized.

4. Inferring regulatory networks

The elucidation of gene regulatory networks (GRNs) can enhance our understanding of complex cellular process in living cells, and these networks generally reveal regulatory interactions between genes and proteins. GRN determination is not the final outcome of a biological study.

5. Cell hierarchy reconstruction

Individual cells are continually undergoing dynamic processes and responding to various environmental stimuli. Some of these responses are fast, whereas others can be much slower and can occur over the course of many years (e.g., pathogenesis). This dynamic process is particularly reflected in a cell's molecular profile, including RNA and protein content. In single-cell systems, however, cells are unsynchronized, which enables the capture of different instantaneous time points along an entire trajectory. We can then apply algorithms to reconstruct dynamic cellular trajectories with respect to differentiation or cell cycle progression.

6. Lineage tracing is a long-standing fundamental question in biology. As a proof-ofconcept, researchers at Caltech have recently developed a method using the sequential readout of mRNA levels in a single cell to reconstruct lineage phylogeny over many generations. Another interesting potential application of scRNA-seq includes identifying genes involved in stem cell regulatory networks.

Conclusion and Outlook

Overall, these studies demonstrate the power of scRNA-seq for exploring the phenotypic diversity of microbial communities. By analyzing the transcriptomes of individual cells, researchers can gain insights into the functional roles of different bacterial subpopulations within these communities and the mechanisms that underlie their interactions.

Microbial single-cell RNA-seq is flourishing and has enabled key studies over the past two years using model organisms and major pathogens. Single-bacteria RNA-seq (for both Gram negative and Gram-positive species) is still emerging. Although microbial single-cell RNA-seq technologies are becoming increasingly robust, benchmark studies will be essential to establish common ground for the community. We need to move towards studies of natural microbiomes in which the number of mRNA per bacterial cell might not exceed a few hundred. Efficient lysis of a single bacterial cell and rRNA depletion using CRISPR based protocols will require to be developed along with microfluidics-based devices to streamline sample processing. Microbiomes might have defined spatial organizations with associated functions in health and diseases that will require the combination of single-cell transcriptomics and imaging to decipher the patterns of microbe organisations. Finally, defining the role of microbial subsets in space and time will be key to designing targeted therapeutic approaches in a species-specific manner.

References

1. Brennan MA, Rosenthal AZ. Single-cell RNA sequencing elucidates the structure and organization of microbial communities. Frontiers in Microbiology. 2021;12:713128.
2. Blattman SB, Jiang W, Oikonomou P, Tavazoie S. Prokaryotic single-cell RNA sequencing by in situ combinatorial indexing. Nature microbiology. 2020;5(10):1192-1201.
3. Hwang B, Lee JH, Bang D. Single-cell RNA sequencing technologies and bioinformatics pipelines. Experimental & molecular medicine. 2018;50(8):1-14.
4. Imdahl F, Saliba AE. Advances and challenges in single-cell RNA-seq of microbial communities. Current Opinion in Microbiology. 2020;57:102-110.
5. Imdahl F, Vafadarnejad E, Homberger C, Saliba AE, Vogel J. Single-cell RNA-sequencing reports growth-condition-specific global transcriptomes of individual bacteria. Nature Microbiology. 2020;5(10):1202-1206.
6. Kuchina A, Brettner LM, Paleologu L, Roco CM, Rosenberg AB, Carignano A. Microbial single-cell RNA sequencing by split-pool barcoding. Science. 2021;371(6531):eaba5257.
7. Papalexi E, Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity. Nature Reviews Immunology. 2018;18(1):35-45.
8. Svensson V, Vento-Tormo R, Teichmann SA. Exponential scaling of single-cell RNA-seq in the past decade. Nature protocols. 2018;13(4):599-604.
9. Yanai I, Hashimshony T. CEL-Seq2—Single-cell RNA sequencing by multiplexed linear amplification. Single Cell Methods: Sequencing and Proteomics, 2019, 45-56.
10. McNulty JK, Meltzer AL, Neff LA, Karney BR. How both partners' individual differences, stress, and behavior predict change in relationship satisfaction: Extending the VSA model. Proceedings of the National Academy of Sciences. 2021 Jul 6;118(27):e2101402118.
11. Lubeck E, Coskun AF, Zhiyentayev T, Ahmad M, Cai L. Single-cell in situ RNA profiling by sequential hybridization. Nature methods. 2014 Apr;11(4):360-1.
12. Wang G, Wei Y, Qiao S, Lin P, Chen Y. Generalized inverses: theory and computations. Singapore: Springer; 2018 May 12.