One pot RNA:DNA assembly for ribosomal RNA detection of pathogenic bacteria with single-molecule sensitivity

Filip Boskovic[1,#], Sarah Sandler[1,#], Simon Brauburger[1], Yuan Shui[1], Bhavik Kumar[1], Joana Pereira Dias[2,3], Plamena Naydenova[2,3], Jinbo Zhu[1], Stephen Baker[2,3], Ulrich F. Keyser[1,*]

[1]Cavendish Laboratory, University of Cambridge, Cambridge, UK

[2]University of Cambridge School of Clinical Medicine, Cambridge Biomedical Campus, Hills Road, Cambridge, UK

[3]Department of Medicine, University of Cambridge School of Clinical Medicine, Cambridge Biomedical Campus, Hills Road, Cambridge, UK

[#]Co-first authors

*E-mail: ufk20@cam.ac.uk

SUMMARY

Effective treatment of infectious diseases hinges on timely and accurate diagnosis. Current methods face challenges in distinguishing between pathogens that cause a similar symptom and where appropriate treatment may be hindered by antimicrobial resistance (AMR). Here, we introduce a PCR-free system that identifies and quantifies ribosomal RNA transcripts (rRNA) in pathogenic bacteria instead of DNA-encoding rRNA genes. Our method leverages the potential of >1 million rRNAs found in replicating bacterial cells in comparison to the limited number of rRNA DNA copies per cell. We combined nanopores with RNA nanotechnology to identify rRNA from bacteria and other cells. We developed a simultaneous protocol for rRNA isolation, assembly and enrichment with assembled RNA identifiers (IDs), which prevents heat degradation and eliminates background RNA. Our method detects multiple bacterial species, including AMR variants of *Salmonella enterica* and *Acinetobacter baumannii*, and coinfection. This approach offers unmatched specificity, with the ability to identify a single bacterium without amplification. The integration of CRISPR-dCas9 binding to RNA IDs leads to serovar-specific identification and opens new avenues in identifying pathogens up to their variants using abundant rRNAs.

MAIN TEXT

The identification of pathogens plays a critical role in the successful treatment of infections and the control of antimicrobial resistance (AMR)[1]. Pathogen detection allows doctors to select the correct treatment, start it early, and prevent antimicrobial overuse, which may facilitate AMR[1–4]. However, many pathogens induce comparable symptoms, complicating the differentiation of causative agent[5,6]. As AMR increases the development of precise, sensitive, and highly specific diagnostic tools become increasingly urgent[7]. The ability to detect and discriminate among pathogens with high specificity not only has implications for individual patient care but also has broader ramifications for public health and the monitoring of infectious disease outbreaks[3].

Bacterial identification mainly relies on amplicon sequencing of 16S ribosomal RNA (rRNA) genes[1,6,8–10]. However, there are typically approximately ten 16S rRNA genes per one bacterial cell[11]; therefore, they are difficult to detect[9]. PCR has a theoretical limit of detecting one molecule; however, the preparation procedure inevitably leads to loss of single bacterium sensitivity just by mere dilution. Conversely, 16S rRNA may produce 10^4-10^7 copies per cells[12], as it is one of the key structures for protein synthesis. The principal issue with using rRNA is the required amplification and reverse transcription, which can introduce errors. Enzyme-introduced errors can lead to artificial rRNA diversity (up to 70%), misassigning species as well as introducing quantification errors[8,9,13]. Additionally, pathogen diagnostics utilizing PCR have a practical challenge due to sequence variations, which can result in false negatives[14]. This problem necessitates regular updates of PCR primers to accurately target the gene of interest, ensuring the reliability and accuracy of the diagnostic process[14]. In addition, PCR of DNA does not provide information if pathogens are dead or alive[15].

Recently, there has been progress regarding viral variant detection using nanotechnology-based approaches[16–18] employing the resistive-pulse detection with nanopores[2,19–22], as well as CRISPR-Cas9-based assays[23–25]. Although these methods have proven valuable, they suffer from limited multiplexing capabilities and specificity[25]. We are still in need of a method that reliably identifies multiple pathogens simultaneously. Additionally, methods for identifying bacteria not only on the level of species but on the level of clinically relevant serological variants (serotypes).

Here, we introduce a PCR-free system to directly identify and quantify pathogens, employing the use of full-length ribosomal RNA (rRNA) rather than ribosomal DNA. Leveraging the abundance of rRNA in the cell[12], our approach achieves serotype-specificity and sensitivity, even paving a way to the theoretical detection of a single bacterium without preamplification. By utilizing rRNA assembly to RNA identifiers (IDs) with dCas9 binding to a variant-specific manner, we have developed a nanopore method that detects multiple species, distinguishes AMR variants and bacterial serotypes, or coinfections in parallel. We assembled rRNA IDs by binding short complementary oligonucleotides to rRNA while some of oligonucleotides bear structural elements to create a unique identifier for that molecule. We demonstrate simultaneous rRNA isolation,

rRNA ID assembly, and the enrichment from the background RNA with <10-minute heating protocol. RNA IDs offers serotype specificity, amplification-free sensitivity, and rapid read out, hence a promising alternative to traditional techniques, paving the way for more effective control of bacterial infections and AMR. The potential impact of this technology extends beyond immediate healthcare applications, providing a new avenue for research, epidemiological studies, and global health initiatives.

RESULTS

Pathogen serotype identification using full-length rRNAs to make RNA identifiers

We utilized the whole sequence of abundant 16S rRNA, reshaping it with short complementary DNA strands to create an RNA:DNA nanostructure that we named RNA identifier (RNA ID) (**Fig.1a**). Some of these DNA strands possessed overhanging sequences with biotin, which binds to streptavidin and allows for sequence-specific barcoding (green and red)[26].

We hypothesized that RNA ID assembly with the DNA oligonucleotides should be possible as the bacterial cells are lysed. Therefore, we tested if high temperature required for rapid rRNA isolation from bacteria would degrade rRNA while in the process of rRNA assembly. We heated *E.coli* cells to 75, 85, 90 and 100 °C for 5 minutes in the presence of DNA oligonucleotides (oligos) (**Fig.1**; sequences are listed in **Table S1**). 16S RNA from *E. coli* was selected as target in this example case. The design uses both the `1` and `3` labels, where the denotes the size of structures, so called structural[26]. The result is a unique RNA:DNA hybrid duplex, named here as RNA ID `1131`. `1131` also indicates the shape of the specific signal in nanopore measurements for this RNA ID (**Fig.1b**). The unique `1131` current signal allows us to detect the 16S RNA even in the presence of cell biomacromolecules and debris like other RNA strands, DNA, and proteins. In addition, we predicted that RNA that is not hybridized with DNA would be degraded during the heating step.

We validated assembly of RNA IDs using both gel electrophoresis and nanopore measurements by heating them for 5 minutes at different temperatures using *E. coli* DH5α total RNA (**Fig.1c,d**). The assembly state of RNA before and after the heat treatment is shown by gel electrophoresis shown in **Fig.1c**. Before treatment, total RNA (lane 3) exhibited three distinct bands, corresponding to 23S, 16S, and 5S rRNAs. After applying our assembly method on 16S rRNA to create the ID `1311` (lanes 4-7) the other two forms of rRNA were not visible, indicating that the 23S and 5S were degraded during the heating step. For all four heat treatment temperatures (70, 85, 90, and 100 °C) successful RNA ID assembly is indicated by the remaining band in lanes 4, 5, 6, and 7, respectively. The results specify that our process enriched the sample predominantly with the targeted 16S rRNA IDs, highlighted by the red dashed box in **Fig.1c**.

We next evaluated all temperature conditions for 5 minutes to assess the success of ID assembly at the single-molecule level (**Fig.1d**; additional events are presented in **Fig.S1**). In all four

conditions we could successfully detect the `1131` barcode during the 10-minute nanopore measurement. The successful single-molecule detection using our nanopores microscope corresponded with the gel electrophoresis results on total RNA extracts. Nanopore detection requires 1000 × less material than gel electrophoresis[2].

Finally, we heated the *E. coli* bacterial samples mixed with DNA oligonucleotides at 100 °C to assemble 16S rRNA ID `1131` (**Fig.1e**). After the heating protocol we observed the presence of 16S rRNA ID `1311`. Lane 2 on the agarose gel in **Fig.1e** shows the sample after the heat assembly treatment corresponding to a 1.6 kbp long RNA ID. Finally, in **Fig.1f** we demonstrated single molecule events after heating *Salmonella* Typhimurium culture with oligonucleotides. Again, we targeted its 16S rRNA and produced rRNA ID `3113`. The RNA IDs were then identified within a 10-min nanopore measurement (**Fig.1f** and additional events detected are presented in **Fig.S2)**. The heating protocol accelerates the bacterial identification process by omitting the need for RNA isolation kits with multistep treatment, and in essence performing RNA isolation, RNA ID assembly and enrichment in parallel.

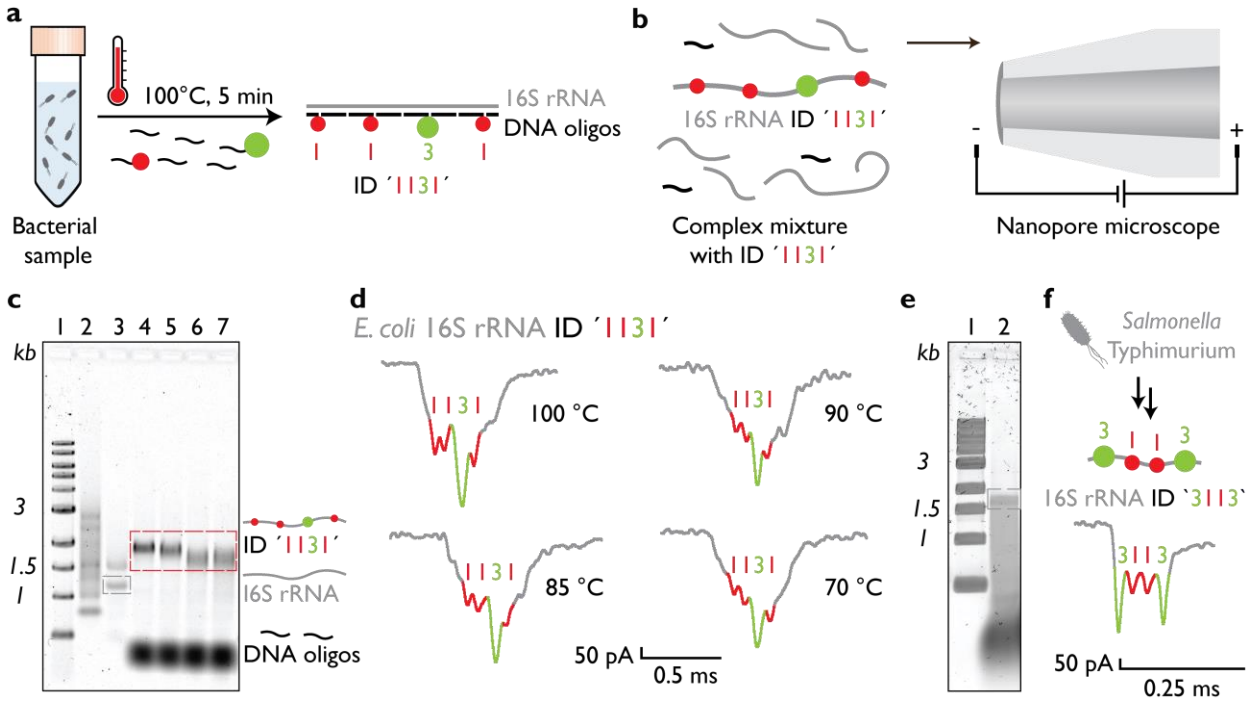

Fig.1 Simultaneous isolation and assembly of ribosomal RNA ID facilitate pathogen diagnostics with nanopores. (**a**) An *E. coli* bacterial culture was heated for 5 minutes at 100 °C, with DNA oligos. This process yields both rRNA isolation and rRNA ID assembly. 16S rRNA transcript undergoes linearization by complementary DNA oligos to yield an RNA:DNA hybrid duplex so called RNA ID. By placing a docking overhang in the specific positions, we can create structural colors e.g. `1` and `3`. In this case 16S rRNA ID `1311` is assembled uniquely for *E. coli*. (**b**) A complex mixture of proteins and potential DNA and RNA remains can be directly analyzed with nanopore microscopy. (**c**) 1 % (w/v) agarose gel electrophoresis displays total RNA before and after heat treatment, as seen in lane 3 and lanes 4-7, respectively. In lane 3, *E. coli* total RNA is

loaded hence three bands are visible, representing 23S rRNA at the top, followed by 16S rRNA, and then 5S rRNA. Upon assembly of the 16S rRNA ID `1311` the other two unpaired rRNAs, 23S and 5S, degraded, thus enriching the sample solely for 16S rRNA IDs. We tested temperatures of 70, 85, 90, and 100 °C for RNA ID assembly, corresponding to lanes 4, 5, 6, and 7, respectively. Lane 1 – 1 kbp ladder; Lane 2 – ssRNA ladder. (**d**) Nanopore ID ionic current traces for all four temperature conditions demonstrate successful assembly of IDs in each case when measured for 10 minutes. (**e**) Directly from *E. coli* culture we demonstrated isolation of 16S rRNA, assembly of rRNA ID, and ID enrichment. Agarose gel electrophoresis shows in lane 1 – 1 kbp ladder (NEB), and in lane 2 – is *E. coli* sample prepared with the heating protocol. (**f**) 16S rRNA ID ´3113` from *Salmonella* Typhimurium culture were successfully assembled and verified by nanopore measurements.

Discriminating between antimicrobial susceptible and resistant variants of A. baumannii

Following the successful development of our heating protocol, we decided to detect *A. baumannii*, a highly antimicrobial resistant bacterium that poses a significant threat in hospital-acquired infections[27,28]. Part of its pathogenesis can be attributed to a unique sugar coat associated with the gene *gtr100*[28]. To prescribe the appropriate treatment, it is helpful to determine whether the gtr100 antigen is present. Ideally, a diagnostic assay should not only identify the species through 16S rRNA but also detect the presence of gtr100 messenger RNA (mRNA). We targeted both the 16S rRNA and the gtr100 mRNA specific to a KL49 organism, assembling them into RNA IDs `331` and `3113`, respectively (**Fig.2a**; sequences are listed in **Table S2** and **S3**, respectively). Using nanopores to analyze total RNA (**Fig.2b**) of two organisms, *A. baumannii* (control) and *A. baumannii* KL-49 (with *gtr100*). In our nanopore measurements, we identified the 16S rRNA ID in both samples (**Fig.1c**; additional events are shown in **Fig.S3**). However, the gtr100 mRNA ID was only detected in the KL49 sample (**Fig.2c**). The mRNA was predicted to be present in a lower concentration than the 16S RNA. Indeed, we found that we had to measure 30,000 events to identify two events for gtr100 mRNA. We have shown that we only need a handful of RNA IDs signals to determine the presence of the specific target sequence[29].

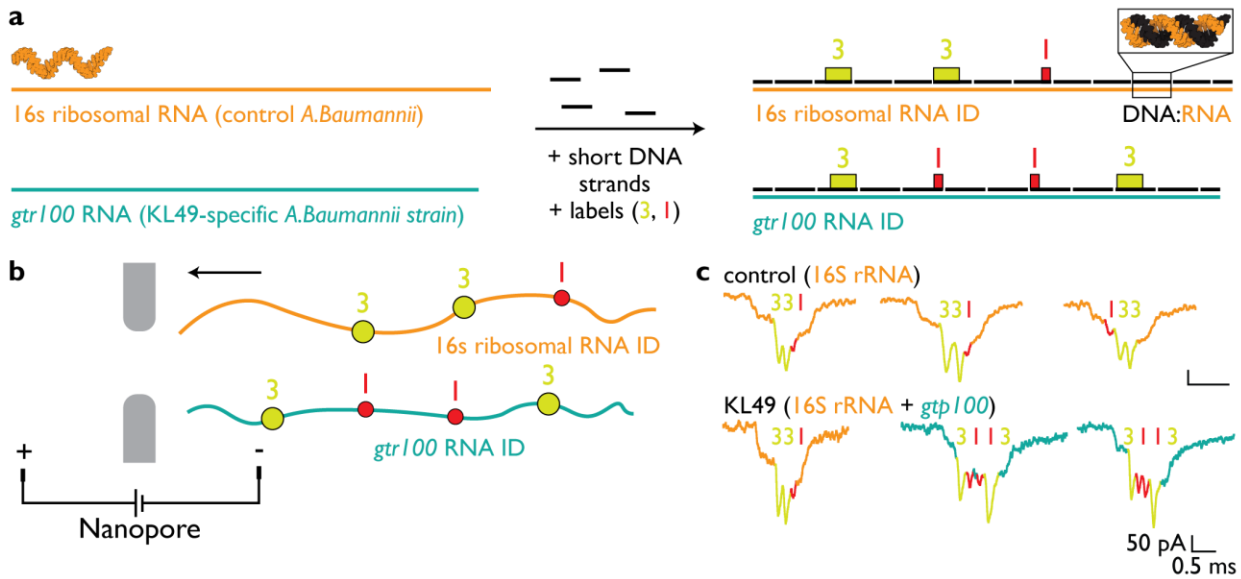

Fig.2 rRNA identifiers distinguish *Acinetobacter baumannii* antimicrobial-resistant strain KL59 from antimicrobial-susceptible KL58 strain. (**a**) We targeted two RNA molecules, 16s rRNA that is specific to *A. baumannii* and messenger RNA from *gtr100* gene linked to KL49 strain. 16s rRNA ID and gtr100 RNA ID were assembled to show `331` and `3113`, respectively. (**b**) RNA IDs from two different strains were identified with nanopores. (**c**) Nanopore events indicate presence of the targeted RNA molecules 16S rRNA and *gtp100* mRNA. For control strain we observed only 16S rRNA `331` ID, for KL49 strain we observed in addition *gtp100* gene transcript `3113` ID.

Identifiers enable multiplexed species determination and coinfection tracking

RNA IDs are also suitable for species identification and co-infection diagnostics. Initially, we verified that rRNA can be used for RNA IDs by using total RNA samples from four diverse sources: humans, mice, *E. coli*, and *S.* Typhimurium (**Fig.S4**). These samples were selectively targeted for eukaryotic 18S or prokaryotic 16S rRNA transcripts, serving as unique molecular identifiers for each species (**Fig.S4a**). We designed four IDs `1311`, `1111`, `3232`, and `3113` for 16S rRNA *E. coli*, human 18S rRNA, mice 18S rRNA, and 16S rRNA *S.* Typhimurium, respectively (**Fig.S4b**; sequences are listed in **Table S1, 4, 5,** and **6**). Notably, the RNA ID designs and corresponding nanopore readouts (additional events are depicted in **Fig.S1, S5, S6,** and S7) demonstrate the capability of the method for parallel identification, highlighting its utility for multiplex diagnostics.

The experiment was further extended to study co-infection by targeting *E. coli* and its associated bacteriophages, 7,228 nt long single-stranded DNA M13 and 3,569 nt long single-stranded RNA MS2, simultaneously (**Fig.S8**). RNA IDs targeting *E. coli* 16s rRNA (ID `1131`; sequences are listed in **Table S1**), MS2 RNA (ID `111` sequences are listed in **Table S7**), and M13 DNA (ID `111111` sequences are listed in **Table S8 and 9**) were assembled using the nucleic acid mixture containing total RNA from *E. coli*, MS2 RNA and M13 DNA. The nanopore traces for these targets

(**Fig.S8**), provide evidence for successful parallel identification of *E. coli* and ssDNA for phage M13 and ssRNA for phage MS2. Our results substantiate the potential of our approach to serve as a robust and versatile platform for both species and co-infection diagnostics.

To validate our specificity of RNA ID assembly, we mixed five hundred oligonucleotides targeting human 18S rRNA, mice 18S rRNA, *E. coli* 16S rRNA, *Salmonella* spp. 16S and 23S RNA, with total RNA from *A. baumannii* (**Fig.S9a**). We performed both agarose gel electrophoresis and nanopore measurements to check for the presence of any events that indicate non-specific RNA IDs assembly including `331` for *A. baumannii* (**Fig.S9b** and **S9c**, respectively). By measuring for more than 10 hours with 2,047 RNA/DNA-like nanopore events, that were filtered for RNA or dsDNA molecules, we did not observe any RNA ID nanopore events. Consequently, the designed set of >500 oligonucleotides yields not a single positive detection. These results demonstrate the specificity of RNA IDs for the target (pathogen) of interest. We attribute this to the combined effect of RNA degradation during the heating/assembly step and the high specificity of the nanopore signals associated with RNA IDs.

Salmonella enterica serovar identification by fragmented 23S rRNA

After the detection of species with rRNA we now expand the use of RNA IDs to analyse rRNA fragments. It has been shown that that some *Salmonella* serovars can bear two intervening sequences (IVS) in their 23S rRNA genes[30]. A variant that does not have IVS produces a full-length 23S rRNA. In contrast, a variant with IVS produces naturally fragmented 23S rRNA because of RNase III excision of IVSs[30] (**Fig.3**). *Salmonella enterica* serotypes can have two intervening sequences (IVS) in their 23S rRNA. The presence of one or both of IVSs in pre-processed 23S rRNA leads to fragmentation by excision at the IVS sequence (**Fig.3a**). To identify all possibilities of 23S rRNA fragmentation we made ID `32213`so that the full length 23S, IVS2, or IVS1 have codes `32213`, `322` and `13`, and `3` and `2213`, respectively (**Fig.3b**). The length of all probable 2.9 kbp 23S rRNA transcripts with these codes are shown in **Fig.3b**.

We tested our identification strategy using total RNA from three *Salmonella* serotypes: Typhimurium, Newport, and Hadar. *Salmonella* serotypes have 7 genes for 23S rRNA, meaning that they produce different 23S rRNA. Typhimurium has 6 genes that have IVS2, and 1 gene IVS1, in this sample using the same mixture of oligonucleotides we expect four 23S rRNA transcripts `322`, `13`, `3`, and `2213` but not full length 23S rRNA. Conversely, Newport has no genes with IVS meaning we should mainly observe `32213` full length 23S rRNA ID (**Fig.3c**). After assembly 23S rRNA IDs all these three samples were run with nanopores. Example nanopore events for Typhimurium, Newport, and Hadar are represented in **Fig.3d,e,** and **f**, respectively. We can observe the presence of specific transcripts labelled in colors regarding the gene type they originated from. IDs can translocate starting from either end hence the ID readout can be e.g. `32213` or `31223` for the full length 23S rRNA ID.

The nanopore results in Fig.3d-f show that we observe `32213` RNA IDs for Newport (**Fig.3d**), a mixture of `32213`, `3`, `322`, `13`, `2213`, for Hadar (**Fig.3e**) and only `322`, `13`, `3122` and `3` for Typhimurium (**Fig.3f**). Our results show that fragmented 23S rRNA can be used to identify potentially antimicrobial resistant serotypes of *Salmonella enterica*. The identification of bacterial serotypes by fragmented 23S has great potential for detection; however, fragmentation after or during sample extraction may present challenges to exclude false positive or negative detections. Consequently, we sought to identify resistant organisms and serotypes by exploiting our recent development of nanopore sensing with single-nucleotide specificity[31].

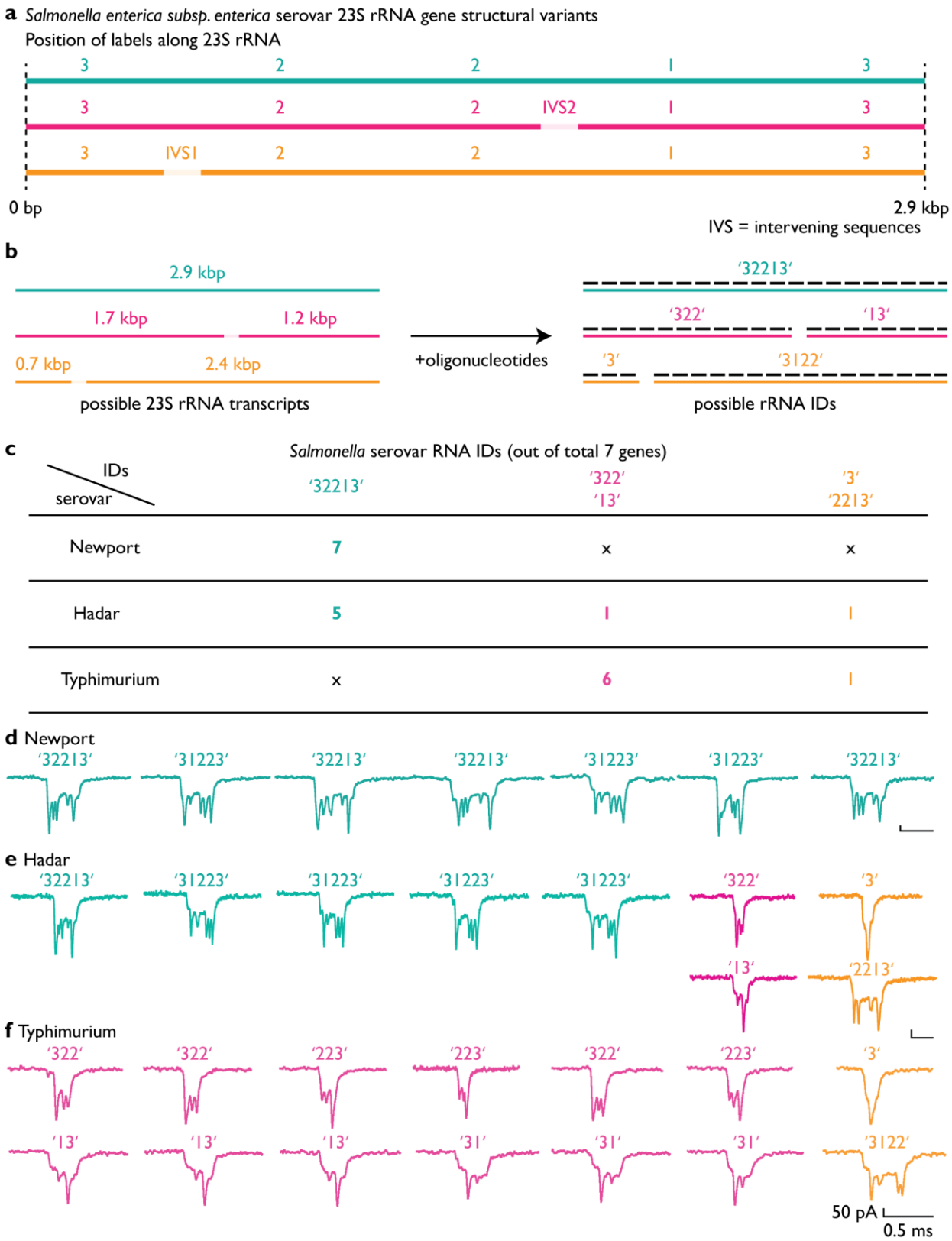
a Salmonella enterica subsp. enterica serovar 23S rRNA gene structural variants
Position of labels along 23S rRNA
3 2 2 1 3
3 2 2 IVS2 1 3
3 IVS1 2 2 1 3
0 bp
2.9 kbp
IVS = intervening sequences
b
2.9 kbp
1.7 kbp
1.2 kbp
0.7 kbp
2.4 kbp
possible 23S rRNA transcripts
+oligonucleotides
'32213'
'322'
'13'
'3'
'3122'
possible rRNA IDs
c
Salmonella serovar RNA IDs (out of total 7 genes)
IDs
serovar
'32213'
'322' '13'
'3' '2213'
Newport 7 x x
Hadar 5 1 1
Typhimurium x 6 1
d Newport
'32213' '31223' '32213' '32213' '31223' '31223' '32213'
e Hadar
'32213' '31223' '31223' '31223' '31223' '322' '3'
'13' '2213'
f Typhimurium
'322' '322' '223' '223' '322' '223' '3'
'13' '13' '13' '31' '31' '31' '3122'
50 pA
0.5 ms

Fig.3 Fragmented 23S rRNA transcripts from different *Salmonella enterica subsp. enterica* serotypes enable serotype discrimination using RNA IDs and nanopore analysis. (**a**) Three 23S rRNA gene variants subsequently produce multiple distinct 23S rRNA transcripts. The first variant (turquoise) is a full length 2.9 kbp 23S rRNA with ID `32213` (oligos listed in the **Supplementary Table S15**). Two variants of Typhimurium have intervening sequences (IVS1 and IVS2) that are cleaved from 23S rRNA yielding two fragments each. The second one (purple) yields `322` and `13` rRNA fragments while the third variant (orange) yields `3` and `2213`. (**b**) RNA ID assembly with the same oligos will create distinct IDs for all three variants. (**c**) *Salmonella* serotypes have 7 genes for 23S rRNA for each of analyzed serotypes the number of expected 23S rRNA variants is presented as N of 7. *Salmonella* Newport has only the full length 23S rRNA ID `32213` (7/7). *Salmonella* Hadar has all three 23S rRNA variants with predominately the full-length variant (5/7). *Salmonella* Typhimurium has in all its genes IVSs, hence we can discriminate them by not having the full length 23S rRNA ID `32213` (0/7). Example nanopore events for respective 23S rRNA IDs (from (b)) for serotype Newport, Hadar, and Typhimurium, are shown in (**d**), (**e**), and (**f**), respectively. IDs translocate through nanopores bidirectionally hence IDs can be read from either end.

Catalytically-inactive Cas9 enabled serotype discrimination

Now we demonstrate the ability to identify single nucleotide rRNA variants and discriminate between two *Salmonella enterica* serotypes by employing catalytically inactive Cas9 (dCas9) (**Fig.4**). CRISPR-dCas9 is a ribonucleic protein-complex (RNP) that can be designed to specifically target a 20 bp region of DNA or RNA through use of a guide RNA (gRNA). Depending on the design and target, the complexes often can bind to target sequences with single-nucleotide specificity. Herein, we exploit the programmability of the CRISPR gRNA to discriminate single serotype rRNA binding. We selected *Salmonella* Typhi as an example target (**Fig.4a)**. *S.* Typhi serotype can be distinguished from Newport using an RNP consisting of dCas9 and a gRNA targeting position 1,258 in the 16S Typhi rRNA (**Fig.4b)**. 16S Typhi was assigned a barcode `22C`, where `C` corresponds to dCas9, and each single-digit number corresponds to the number of streptavidin molecules bound to the RNA ID as explained above. In contrast, the Newport strand does not have the target sequence for the dCas9 probe, there is a single-base pair change in the sequence (dA to dG). Hence, the Newport variant only has a `220` barcode. Hence, the presence or absence of dCas9 facilitates the discrimination of these two serotypes with single base resolution. Example events measured with our nanopore microscope can be seen in **Fig.4c**.

Following the successful discrimination of Typhi from Newport, we tested the ability to distinguish other serotypes. The best discrimination method is achieved when the single-nucleotide variations were located close to a protospacer-adjacent motif (PAM). The latter PAM motif is required for dCas9 binding. Via sequence alignment we found several sites in different serotypes (**Fig.S10 and Fig.4b).** These dCas9 RNPs were then tested for their single-nucleotide specificity using a previously described method.[31] The test measurements are shown in **Fig.S11**

(sequences are listed in **Tables S8, S10,** and **S11)**. Based on the specificity and the sequence alignments for the full-length rRNA sequences, the most specific guide RNA with the single-nucleotide pair change that was present in the highest quantity amongst rRNA genes was selected. We again aimed to distinguish the Typhi serotype from the other serotypes, namely Enteritidis and Newport.

For the experiments, the selected dCas9 RNP was mixed with 16S rRNA IDs (sequences for rRNA for assembly of `022` are listed in **Table S12**) from the different *Salmonella* serotypes and measured with nanopores. The number of events containing three spikes (barcode plus `C`) are shown in the bar chart in **Fig.4d**. We plotted the presence of `22C` events described with three downward spikes in nanopore events for different variants. The other variants do show this difference, yielding `220` and missing dCas9 binding. *S*. Typhi serotype showed a higher presence of `22C` compared to the control values. Both the assembly and measurement were repeated for *S*. Typhi, and the same trend was observed (**Fig.S12)**.

Additionally, we show that in the mix of different serotypes we could observe dCas9 binding in the presence of the right serotype of *S*.Typhi. Two samples were tested, one without the addition of serotype *S*. Typhi (N=32 linear nanopore events) and with serotype *S*. Typhi (N=30 linear nanopore events) as seen in **Fig.S13a**. Again, events were quantified based on whether they had two peaks `220` (no dCas9 bound) or three peaks `22C` (dCas9 bound). Example events can be seen in **Fig.S13b**. As expected, we can distinguish between *S*. Typhi and the other serotypes (Newport and Enteritidis).

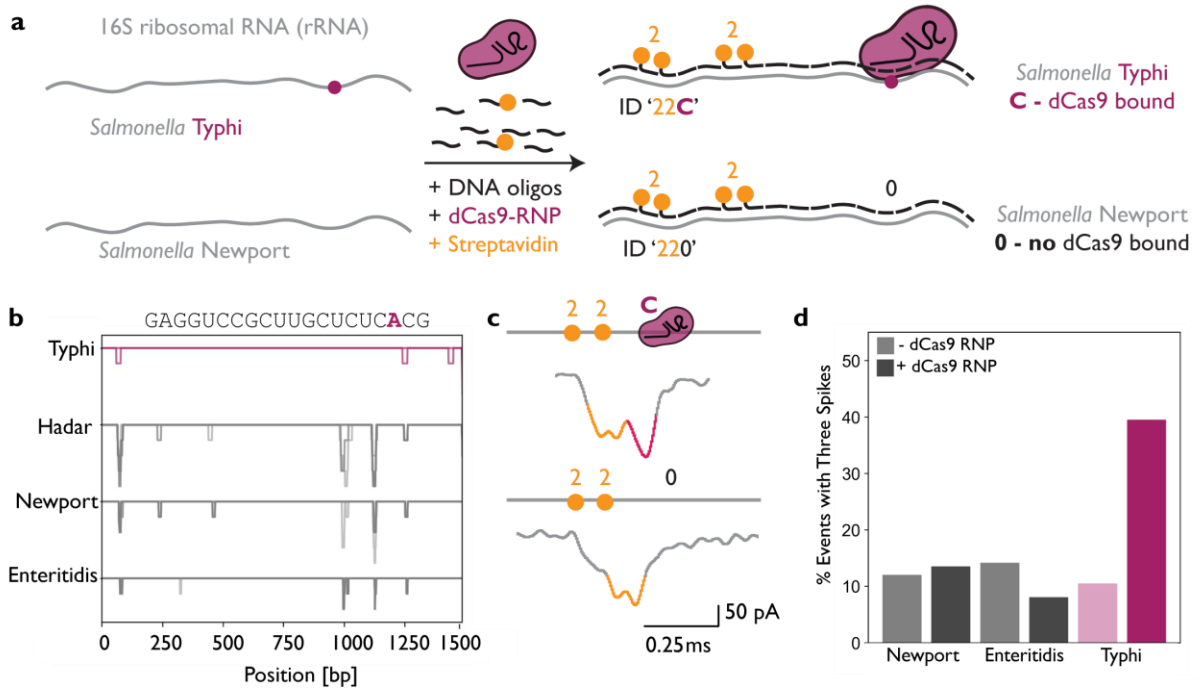

Fig.4 Single *Salmonella* serotype variant discrimination with inactive Cas9 (dCas9). (**a**) Two distinct 16S rRNAs from *Salmonella* serotypes Newport and Typhi feature a unique sequence

difference (in purple). These rRNA strands serve as scaffolds for binding complementary DNA oligonucleotides, forming double-stranded RNA:DNA identifiers (IDs) with specific structural elements representing bits `2` (in orange). Each bit `2` has two streptavidins bound to biotinylated ends of oligonucleotides. The binding of dCas9 introduces an additional `C` bit, which is exclusive to *Salmonella* Typhi. As a result, two unique IDs are generated: `22C` for Typhi and `220` for Newport. (**b**) Single nucleotide variants were determined via sequence alignment of 16S rRNA for serotypes Typhi, Hadar, Newport and Enteritidis. (**c**) `22C` ID specific to Typhi and `220` ID present in all serotypes. An increase in nanopore events with three downward spikes, representing `22C` ID, was observed with serotype Typhi (N=32 in sample without Typhi, N= 30 in sample with Typhi). (**d**) Percentage of events with three peaks before and after dCas9 is added. The percentage is similar for when no dCas9 is added for Newport (without dCas9 N=40, with dCas9 N=54) and Enteritidis (without dCas9 N=14, with dCas9 N=37). For Typhi, the percentage of events with three peaks changes from ~10% without dCas9 (N=45) to ~35% with dCas9 (N=33). The various bars are repeats amongst different nanopores. The points in the final bar are Typhi samples tested in the same nanopore but prepared in different sample preparations representing independent repeats.

Discussion

Here, we have introduced an amplification-free system that opens new avenues in pathogen diagnostics by employing ribosomal RNA (rRNA) for the direct identification and quantification of pathogenic bacteria. This novel approach exploits the high abundance of rRNA in bacterial cells, which is estimated to be 10^4 to 10^7 times more prevalent than ribosomal RNA genes in DNA. Utilizing the capabilities of nanopores and RNA-protein nanotechnology, our method has demonstrated the successful identification of rRNAs from a range of pathogenic bacteria, including serotype different *Salmonella enterica* serotypes and antimicrobial resistant *Acinetobacter baumannii*, as well as the detection of coinfections within a single measurement.

Bacterial culture has long been the cornerstone of bacterial diagnostics and testing antimicrobial susceptibility. Culture-based approaches are limited by their time-consuming nature, the complexity of culturing conditions needed for various pathogens, and the challenge in defining clinically relevant resistance breakpoints[32]. Still such testing has unique capability to distinguish between alive and dead cells[33]. This differentiation is crucial, as it enables the avoidance of false positives in pathogen detection, a common issue with methods that do not account for cell viability[33,34]. The use of RNA identifiers, which target the rRNA transcript, leverages this aspect by focusing on RNA that is both easily degradable and indicative of replicating pathogenic bacterium [33], thus avoiding false positives.

Following bacterial culture, PCR techniques represent a current cornerstone of molecular diagnostics[4,14,32,34], offering the ability to detect even minimal amounts of target DNA or RNA in a sample. Despite high sensitivity, PCR faces challenges, including a susceptibility to artificial variability and the tendency to produce false negatives in the presence of mismatches in target sequences[32]. These mismatches are particularly problematic in pathogenic organisms, which

evolve rapidly[32]. The theoretical limit of PCR to detect target molecules highlights its sensitivity but also underscores limitations, such as potentially missing the detection of a few copies of rRNA gene from one bacterium[13]. Nevertheless, considering the abundance of rRNA transcripts per cell, rRNA ID can, in theory, achieve detection of less than one bacterium`s sensitivity, ensuring that the genetic signature of bacteria reaches the detection threshold.

Nanopore sensing technologies mark the latest advancement in diagnostic techniques[35–39], offering unique benefits in detecting a wide array of biomolecules, including DNA[40–42], RNA[2,19,26,43,44], proteins[45–48], and metabolites[49–51]. Despite their innovative approach to molecular detection, nanopore sensors have encountered challenges related to specificity[2,39,52]. The similarity in physical properties among molecules can lead to non-specific background signals, a significant issue when dealing with complex biological samples laden with diverse biomacromolecules. To counter this, DNA nanostructures enhance the specificity of RNA detection[2,19]. These encoded structures are designed to produce distinct current readouts upon binding with the short target RNA and employing single molecule nanopore readout.

Our approach leverages entire rRNA sequences, transformed through RNA nanotechnology to unique species-specific identifiers, to achieve unparalleled specificity. By targeting the full extent of the rRNA sequence, we ensure robustness against mismatches that are common in rapidly evolving pathogens, thereby overcoming a significant challenge in molecular diagnostics. To attain single-nucleotide specificity, essential for identifying variants of pathogens, we utilized the CRISPR-dCas system single-nucleotide variant specificity[23–25,53]. RNA IDs allow for the simultaneous identification of multiple pathogens enabling panels for pathogen diagnostics. The integration of these techniques (RNA nanotechnology species-specific targeting capability and CRISPR-dCas serotype specificity) enables the development of identifiers that maintain strict specificity, a critical factor for reliable diagnostics. This melding of RNA-protein nanotechnology heralds a new era in diagnostics, promising significantly enhanced specificity and multiplexing capabilities for pathogen detection.

Online Methods

Salmonella enterica and Acinetobacter baumannii cultures and total RNA isolation

Bacteria were cultured overnight in LB and 500 µL of overnight culture was treated with two volumes of RNAprotect reagent (Qiagen). Cultures were digested with 15 mg/mL lysozyme (Sigma) for 20 min and RNA extraction was performed using the RNeasy Mini Kit (Qiagen) following the manufacturer's guidelines. Total RNA was quantified, and quality assessed using Nanodrop (ThermoFisher). Total RNAs from other samples were obtained from ThermoFisher.

RNA identifier assembly

RNA ID was assembled using 5-10 µL of total RNA (ng/µL), 2.4 µL of oligopool (Integrated DNA Technologies; concentration of each oligo is 1 µM), 4 µL of 1 M LiCl, 4 µL of 10 × TE (100 mM Tris-HCl pH 8.0, 10 mM EDTA), 1 µL of 25 µM 3`-biotinylated imaging strand (IDT; 5`-ACCACTAATGAGTGATATCC-TEG-biotin-3`), add nuclease-free water was added to 40 µL[26]. The mixture was subjected to the following heating protocol using a ProFlex PCR System (Applied Biosystems) firstly it was heated to 70 °C for 30 s, then it was slowly cooled to 20 °C over 45 min, it was stored at 4 °C or filtered immediately. The 40 µL mixture was mixed twice with 460 µL of a washing buffer (10 mM Tris-HCl pH 8.0, 0.5 mM $MgCl_2$) and centrifuged in 0.5 mL Amicon ultra centrifugal filter 100 kDa (Sigma-Millipore, catalogue number UFC510008) at 9,400 × g for 10 min. The filtered RNA IDs were retrieved by inverting the filter to a new 2 mL tube and centrifuged at 1,000 × g for 2 min and stored at 4 °C prior the use. All buffers were filtered using Millipore 0.22 µm syringe filters.

Short heating protocol for RNA ID assembly

RNA ID was assembled using the identical procedure except the heating step as mentioned for RNA ID. The heating protocol was changed to 70, 80, 90, or 100 °C for 5 min, placed on ice for 5 min before ultrafiltration with Amicon 0.5 mL filter.

DNA carrier assembly

The DNA constructs contained different "barcodes" plus a customizable overhang sequence for which is a DNA form of the target RNA sequence. The DNA construct was synthesized from pairing a linearized 7.2 kbp single-stranded (ss) M13mp18 DNA (Guild Biosciences) and with 190 complimentary oligonucleotides via Watson-Crick base pairing to produce full dsDNA over the period of 1-hour in a thermocycler. All oligonucleotides were synthesized by IDT and dissolved in IDTE (10 mM Tris–HCl, 0.1 mM EDTA, pH 8.0). The sequences are listed in Table S1[52]. The sample is then filtered using a 100 kDa Amicon filter and measured with a ThermoFisher Scientific Nanodrop™ 2000 Spectrophotometer for concentration information. All samples were stored in a storage buffer of 10 mM Tris HCl 0.5 mM $MgCl_2$ after ultrafiltration. The design contains five groups of equally spaced simple dumbbell hairpin motifs within the 190 oligos to create the spikes which act as a barcode on the DNA nanostructure. Each group consists of 11 DNA dumbbells to create a single spike[52]. The sequences modified to contain DNA dumbbells are listed in Table S2. The overhang used for the target was created by replacing oligos No. 142 and 143 from Table S1 with 50 bp target sequences as shown in Table S3.

CRISPR-dCas9 and Streptavidin binding assay

Catalytically deactivated Cas9 D10A/H10A (dCas9) from Streptococcus pyogenes binds with a tracrRNA and a sequence specific RNA (crRNA), both synthesized by Integrated DNA Technologies (IDT) and dissolved in IDTE (10 mM Tris–HCl, 0.1 mM EDTA, pH 8.0). The target sequences for the crRNA for the probes were designed using online software (http://chopchop.cbu.uib.no/)[54]. To assemble the dCas9 RNPs, first the tracrRNA (200 nM) was incubated at 90 degrees for 2 minutes in a 1X low salt buffer (25 mM HEPES-NaOH (pH 8.0), 150 mM NaCl, 1 mM $MgCl_2$) and then quickly cooled by placing the tube on ice. After 3 minutes, the dCas9 (100 nM) was added and allowed to sit at 25 degrees Celsius (room temperature) with tracrRNA for at least 5 minutes. After 5 minutes, the crRNA (250 nM) was added and allowed to sit for at least 10 minutes at 25 degrees Celsius. Finally, the assembled dCas9 probes were then incubated with the DNA nanostructures or RNA-IDs for at least 20 minutes at 25 degrees Celsius, with the dCas9 probes added in excess of typically 15 dCas9 probes per DNA binding site. For the RNA IDs, after dCas9 binding, streptavidin was added in 10x excess per binding site directly before measurement and allowed to sit at 25 degrees Celsius for 10 minutes. The samples are then diluted to 0.1 nM DNA nanostructure or RNA ID into a 2 M LiCl, 1 × TE buffer solution immediately before the beginning of the measurement in the nanopore system.

Agarose gel electrophoresis

Our RNA ID constructs were run on an 0.8-1 % (w/v) agarose gel (agarose for molecular biology, Sigma, catalogue number A9539), 1 × TBE (0.089 M Tris-Borate, 0.002 M EDTA, pH 8.3; Sigma, catalogue number SRE0062) in an ice bath, for 1.5 h at 70 V. The gel was poststained in 3 × GelRed solution (Biotium, catalogue number 41001) for 10 min on a horizonal shaker. The stained gel was imaged with a GelDocIt™(UVP).

Gel images were analysed using ImageJ (Fiji)[55] by inverting grayscale and homogenous background subtraction with 100 pixels rolling ball.

Nanopore measurements

Conical nanopores with diameters of 10-15 nm are made of commercially available quartz capillaries (0.2 mm inner diameter/0.5 mm outer diameter Sutter Instruments, CA). They are created using a laser-assisted pipette puller (P-2000, Sutter Instrument, CA). Sixteen nanopores are then placed in a custom-templated polydimethylsiloxane (PDMS) chip containing a communal cis reservoir and individual trans reservoirs. In order to generate the current, silver/silver chloride (Ag/AgCl) electrodes are connected to the cis and trans reservoirs in the PDMS chip. As only one nanopore can be measured at a time due to the electronics, the trans reservoir contains the electrode with a 500 mV bias voltage, and the central cis reservoir is connected to a ground.

Using an Axopatch 200B patch-clamp amplifier (Molecular Devices, CA), current signals are collected. The setup is operated in whole-cell mode with the internal filter set to 100 kHz. An 8-pole analogue low-pass Bessel filter (900CT, Frequency Devices, IL) with a cutoff frequency of 50 kHz is used to reduce noise. The applied voltage is controlled through an I/O analog-to-digital converter (DAQ-cards, PCIe-6251, National Instruments, TX). A LabView program records the current signal at a bandwidth of 1 MHz.

Nanopore data analysis

For data using DNA carrier and dCas9 the following procedure was followed. First, a translocation finder python script (https://gitlab.com/keyserlab/nanopyre) is used that identifies the events from the raw traces using user-defined thresholds (minimum 0.1 ms duration, minimum 0.1 nA current drop) and stores them in an hdf5 file. Next, the hdf5 file is loaded into the GUI categorizer python script, found here: https://gitlab.com/keyserlab/nanopycker. Events are then sorted based on the number of peaks or the barcodes and presence or absence of proteins bound.

The specificity as seen in Fig.4b. was calculated using the following formula:

$$\%\, dCas9\ Events_{\ 11111} = \frac{N_{11111\ dCas9}}{N_{11111\ dCas9} + N_{11001\ dCas9}} * 100$$

$$\%\, dCas9\ Events_{\ 11001} = \frac{N_{11001\ dCas9}}{N_{11111\ dCas9} + N_{11001\ dCas9}} * 100$$

The binding of dCas9 to RNA-DNA salmonella samples:

$$\%\, Events_{\ 3\ peaks} = \frac{N_{3\ peaks}}{N_{3\ peaks} + N_{2\ peaks}} * 100$$

Where 3 peaks correspond to `22C` ID and 2 peaks correspond to `22` ID. Events containing 1 peak or no peaks were discarded.

Acknowledgements

F.B. acknowledges research funding from the George and Lilian Schiff Foundation Studentship, the Winton Programme for the Physics of Sustainability PhD Scholarship and St John`s College Benefactors` Scholarship. S.E.S. acknowledges funding from Oxford Nanopore Technologies, Engineering and Physical Sciences Research Council (EPSRC) and Cambridge Trust. U.F.K. acknowledge funding from a European Research Council (ERC) consolidator grant (DesignerPores no. 647144) and an ERC-2019-PoC grant (PoreDetect no. 899538).

Author contributions

Competing interests

F.B. and U.F.K. are inventors of two patents related to RNA analysis with nanopores (UK patent application no. 2113935.7, in process; UK Patent application nos. 2112088.6 and PCT/GB2022/052171, in process) submitted by Cambridge Enterprise on the behalf of the University of Cambridge. S.E.S. is partially funded by Oxford Nanopore Technologies for her PhD. U.F.K. is a co-founder of Cambridge Nucleomics. All other authors have no competing interests.

REFERENCES

1. Burnham, C.-A. D., Leeds, J., Nordmann, P., O'Grady, J. & Patel, J. Diagnosing antimicrobial resistance. *Nat Rev Microbiol* **15**, 697–703 (2017).

2. Bošković, F. *et al.* Simultaneous identification of viruses and viral variants with programmable DNA nanobait. *Nat Nanotechnol* **18**, 290–298 (2023).

3. Caliendo, A. M. *et al.* Better tests, better care: improved diagnostics for infectious diseases. *Clin Infect Dis* **57**, S139–S170 (2013).

4. Deliver on diagnostics to save lives. *Nat Microbiol* **3**, 847–847 (2018).

5. Daniel, P., Bewick, T., Welham, S., Mckeever, T. M. & Lim, W. S. Adults miscoded and misdiagnosed as having pneumonia: results from the British Thoracic Society pneumonia audit. *Thorax* **72**, 376–379 (2017).

6. Murray, P. R. ; Baron, E. J. ; Jorgensen, J. H. ; Landry, M. L. ; Pfaller, M. A. *Manual of Clinical Microbiology*. (ASM Press, Washington, 2006).

7. Dougan, G. & Baker, S. Salmonella enterica Serovar Typhi and the Pathogenesis of Typhoid Fever. *Annu Rev Microbiol* **68**, 317–336 (2014).

8. Hoshino, T. & Inagaki, F. Application of Stochastic Labeling with Random-Sequence Barcodes for Simultaneous Quantification and Sequencing of Environmental 16S rRNA Genes. *PLoS One* **12**, e0169431 (2017).

9. Jin, J. *et al.* High-throughput identification and quantification of single bacterial cells in the microbiota. *Nat Commun* **13**, 863 (2022).

10. Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences* **87**, 4576–4579 (1990).

11. Rainey, F. A., Ward-Rainey, N. L., Janssen, P. H., Hippe, H. & Stackebrandt, E. Clostridium paradoxum DSM 7308T contains multiple 16S rRNA genes with heterogeneous intervening sequences. *Microbiology (N Y)* **142**, 2087–2095 (1996).

12. Ortiz, J. O., Förster, F., Kürner, J., Linaroudis, A. A. & Baumeister, W. Mapping 70S ribosomes in intact cells by cryoelectron tomography and pattern recognition. *J Struct Biol* **156**, 334–341 (2006).

13. Yang, S. & Rothman, R. E. PCR-based diagnostics for infectious diseases: uses, limitations, and future applications in acute-care settings. *Lancet Infect Dis* **4**, 337–348 (2004).

14. Järvinen, A.-K. *et al.* Rapid identification of bacterial pathogens using a PCR- and microarray-based assay. *BMC Microbiol* **9**, 161 (2009).

15. Lee, K.-M., Runyon, M., Herrman, T. J., Phillips, R. & Hsieh, J. Review of Salmonella detection and identification methods: Aspects of rapid emergency response and food safety. *Food Control* **47**, 264–276 (2015).

16. Bengtson, M. *et al.* CRISPR-dCas9 based DNA detection scheme for diagnostics in resource-limited settings. *Nanoscale* **14**, 1885–1895 (2022).

17. Bharadwaj, M., Bengtson, M., Golverdingen, M., Waling, L. & Dekker, C. Diagnosing point-of-care diagnostics for neglected tropical diseases. *PLoS Negl Trop Dis* **15**, e0009405 (2021).

18. Kaittanis, C., Santra, S. & Perez, J. M. Emerging nanotechnology-based strategies for the identification of microbial pathogenesis. *Adv Drug Deliv Rev* **62**, 408–423 (2010).

19. Zhu, J. *et al.* Multiplexed Nanopore-Based Nucleic Acid Sensing and Bacterial Identification Using DNA Dumbbell Nanoswitches. *J Am Chem Soc* **145**, 12115–12123 (2023).

20. Taniguchi, M. *et al.* Combining machine learning and nanopore construction creates an artificial intelligence nanopore for coronavirus detection. *Nat Commun* **12**, 3726 (2021).

21. Sethi, K. *et al.* Direct Detection of Conserved Viral Sequences and Other Nucleic Acid Motifs with Solid-State Nanopores. *ACS Nano* **15**, 8474–8483 (2021).

22. Stein, U. *et al.* Quantification of mRNA expression using single-molecule nanopore sensing. *ACS Nano* **14**, 13964–13974 (2020).

23. Kellner, M. J., Koob, J. G., Gootenberg, J. S., Abudayyeh, O. O. & Zhang, F. SHERLOCK: nucleic acid detection with CRISPR nucleases. *Nat Protoc* **14**, 2986–3012 (2019).

24. Quan, J. *et al.* FLASH: a next-generation CRISPR diagnostic for multiplexed detection of antimicrobial resistance sequences. *Nucleic Acids Res* **47**, e83–e83 (2019).

25. Kaminski, M. M., Abudayyeh, O. O., Gootenberg, J. S., Zhang, F. & Collins, J. J. CRISPR-based diagnostics. *Nat Biomed Eng* **5**, 643–656 (2021).

26. Bošković, F. & Keyser, U. F. Nanopore microscope identifies RNA isoforms with structural colours. *Nat Chem* **14**, 1258–1264 (2022).

27. Singh, J. K., Adams, F. G. & Brown, M. H. Diversity and Function of Capsular Polysaccharide in Acinetobacter baumannii. *Front Microbiol* **9**, (2019).

28. Wyres, K. L., Cahill, S. M., Holt, K. E., Hall, R. M. & Kenyon, J. J. Identification of Acinetobacter baumannii loci for capsular polysaccharide (KL) and lipooligosaccharide outer core (OCL) synthesis in genome assemblies using curated reference databases compatible with Kaptive. *Microb Genom* **6**, (2020).

29. Zhu, J., Ermann, N., Chen, K. & Keyser, U. F. Image Encoding Using Multi-Level DNA Barcodes with Nanopore Readout. *Small* **17**, 2100711 (2021).

30. Pabbaraju, K., Miller, W. L. & Sanderson, K. E. Distribution of Intervening Sequences in the Genes for 23S rRNA and rRNA Fragmentation among Strains of

the Salmonella Reference Collection B (SARB) and SARC Sets. *J Bacteriol* **182**, 1923–1929 (2000).

31. Sandler, S. E. *et al.* Sensing the DNA-mismatch tolerance of catalytically inactive Cas9 via barcoded DNA nanostructures in solid-state nanopores. *Nat Biomed Eng* (2023) doi:10.1038/s41551-023-01078-2.

32. Antibiotic susceptibility diagnostics for the future. *Nat Microbiol* **4**, 1603–1603 (2019).

33. Cangelosi, G. A. & Meschke, J. S. Dead or Alive: Molecular Assessment of Microbial Viability. *Appl Environ Microbiol* **80**, 5884–5891 (2014).

34. Kralik, P. & Ricchi, M. A Basic Guide to Real Time PCR in Microbial Diagnostics: Definitions, Parameters, and Everything. *Front Microbiol* **8**, (2017).

35. Lastra, L. S., Sharma, V., Farajpour, N., Nguyen, M. & Freedman, K. J. Nanodiagnostics: A review of the medical capabilities of nanopores. *Nanomedicine* **37**, 102425 (2021).

36. Xue, L. *et al.* Solid-state nanopore sensors. *Nature Reviews Materials* vol. 5 931–951 Preprint at https://doi.org/10.1038/s41578-020-0229-6 (2020).

37. Deamer, D., Akeson, M. & Branton, D. Three decades of nanopore sequencing. *Nat Biotechnol* **34**, 518–524 (2016).

38. James, P. *et al.* LamPORE: rapid, accurate and highly scalable molecular screening for SARS-CoV-2 infection, based on nanopore sequencing. *medRxiv* 2020.08.07.20161737 (2020) doi:10.1101/2020.08.07.20161737.

39. Ying, Y.-L. *et al.* Nanopore-based technologies beyond DNA sequencing. *Nat Nanotechnol* **17**, 1136–1146 (2022).

40. Dekker, C. Solid-state nanopores. *Nat Nanotechnol* **2**, 209–215 (2007).

41. Maglia, G., Heron, A. J., Stoddart, D., Japrung, D. & Bayley, H. Analysis of single nucleic acid molecules with protein nanopores. *Methods Enzymol* **475**, 591 (2010).

42. Zhang, S. *et al.* Detection of small-sized DNA fragments in a glassy nanopore by utilization of CRISPR-Cas12a as a converter system. *Analyst* **147**, 905–914 (2022).

43. Wanunu, M. *et al.* Rapid electronic detection of probe-specific microRNAs using thin nanopore sensors. *Nat Nanotechnol* **5**, 807–814 (2010).

44. Tripathi, P. *et al.* Discrimination of RNA Fiber Structures Using Solid-State Nanopores. *Nanoscale* **14**, 6866–6875 (2022).

45. Bell, N. A. W. & Keyser, U. F. Specific Protein Detection Using Designed DNA Carriers and Nanopores. *J Am Chem Soc* **137**, 2035–2041 (2015).

46. Steinbock, L. J. *et al.* Probing the size of proteins with glass nanopores. *Nanoscale* **6**, 14380–14387 (2014).

47. Plesa, C., Ruitenberg, J. W., Witteveen, M. J. & Dekker, C. Detection of Individual Proteins Bound along DNA Using Solid-State Nanopores. *Nano Lett* **15**, 3153–3158 (2015).

48. Lucas, F. L. R., Versloot, R. C. A., Yakovlieva, L., Walvoort, M. T. C. & Maglia, G. Protein identification by nanopore peptide profiling. *Nat Commun* **12**, 5795 (2021).

49. Kong, J., Zhu, J., Chen, K. & Keyser, U. F. Specific Biosensing Using DNA Aptamers and Nanopores. *Adv Funct Mater* **29**, 1807555 (2019).

50. Beamish, E., Tabard-Cossa, V. & Godin, M. Programmable DNA Nanoswitch Sensing with Solid-State Nanopores. *ACS Sens* **4**, 2458–2464 (2019).

51. He, L. *et al.* Digital immunoassay for biomarker concentration quantification using solid-state nanopores. *Nat Commun* **12**, 5348 (2021).

52. Bell, N. A. W. & Keyser, U. F. Digitally encoded DNA nanostructures for multiplexed, single-molecule protein sensing with nanopores. *Nat Nanotechnol* **11**, 645–651 (2016).

53. Abudayyeh, O. O. *et al.* RNA targeting with CRISPR–Cas13. *Nature 2017 550:7675* **550**, 280–284 (2017).

54. Labun, K. *et al.* CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res* **47**, W171–W174 (2019).

55. Schindelin, J. *et al.* Fiji: An open-source platform for biological-image analysis. *Nat Methods* **9**, 676–682 (2012).