Rapid MinION metagenomic profiling of the preterm infant gut microbiota to aid in pathogen diagnostics

Richard M. Leggett[1][*][#], Cristina Alcon-Giner[2][#], Darren Heavens[1], Shabhonam Caim[2], Thomas C. Brook[3], Magdalena Kujawska[2], Lesley Hoyles[4], Paul Clarke[5], Lindsay J. Hall[2][*][†], Matthew D. Clark[1,6][*][†]

*Corresponding authors [#]Joint first author [†]Joint last author

[1] Earlham Institute, Norwich Research Park, Norwich, UK
[2] Quadram Institute Bioscience, Norwich Research Park, Norwich, UK
[3] University of Westminster, London, UK
[4] Imperial College London, London, UK
[5] Norfolk and Norwich University Hospital, Norwich, UK
[6] University of East Anglia, Norwich, UK

ABSTRACT

The Oxford Nanopore MinION sequencing platform offers direct analysis of DNA reads as they are generated, which combined with its low cost, low power and extremely compact size, makes the device attractive for in-field or clinical deployment, e.g. rapid diagnostics. We employed the MinION platform for shotgun metagenomic sequencing and analysis of mixed gut-associated microbial communities; firstly, we used a 20 species human microbiota mock community to show that Nanopore metagenomic data can be classified reliably and rapidly. Secondly, we profiled bacterial DNA isolated from faeces from preterm infants at increased risk of sepsis and necrotising enterocolitis to analyse their gut microbiota. Using longitudinal samples, and comparing Illumina to MinION, we captured the diversity of the immature gut microbiota and observed how its complexity changes over time in response to interventions, i.e. probiotic, antibiotics and episodes of suspected sepsis. Finally, we performed a 'real-time' run from sample to analysis using a faecal sample of a critically ill infant. Real-time analysis was facilitated by our new NanoOK RT software package. We determined that we can reliably identify potentially pathogenic taxa (i.e. *Klebsiella pneumoniae*) along with corresponding AMR gene profiles in as little as one hour, post sequencing start. Furthermore, data obtained revealed insights into how antibiotic treatment decisions may be rapidly modified in response to specific AMR profiles, which was validated using pathogen isolation, whole genome sequencing and antibiotic susceptibility testing. Our results demonstrate that MinION sequencers offer the ability to progress from clinical samples to a potential tailored patient antimicrobial treatment in just a few hours.

INTRODUCTION

Next generation sequencing (NGS) has revolutionised profiling of environmental and clinical microbial communities. In particular, the culture-independent, sensitive, data-rich nature of metagenomic sequencing, combined with powerful bioinformatics tools, have allowed researchers to start to differentiate patient groups from healthy individuals based on their microbial profiles[1-5]. Microbiota linked health effects include the ability to support immune system development, facilitate dietary metabolism, modulate the metabolome and provide antimicrobial protection[6-8]. Disturbances in the microbiota by, for example, external influences such as antibiotics, have been associated with an increased risk of diseases including ulcerative colitis[9], obesity, and autoimmune conditions[10, 11], and within an infectious disease context increased risk of pathogen overgrowth[12].

NGS platforms are often large capital investments with a considerable physical footprint, and generate huge quantities of data using a pipeline which combines many samples into a single sequencing run, which can take days or weeks to run and analyse. These are not ideal attributes for clinical settings. In contrast, the MinION platform released by Oxford Nanopore Technologies in 2014 is a pocket-sized sequencing device powered from a laptop's USB port capable of producing gigabase yields of long reads in real time (see recent review of MinION technology and it's applications[13]). This low cost, $1000 platform may allow scientists working within a clinical setting a readily portable and easy-to-use NGS platform that shows great potential for rapid pathogen screening, diagnosis, and treatment strategy design. Crucially, the device operates in 'real-time' allowing sequenced reads to be analysed immediately after they are generated. However, due to the 'real-time' nature of the platform and its differing error profile, it is critical to develop new bioinformatics pipelines to take full advantage of the data, particularly in the clinical arena.

The huge rise in antimicrobial resistance (AMR)[14-16] highlights the need for rapid methodologies to identify at-risk individuals, diagnose infectious agents and evaluate treatments. The MinION may represent such a technology for clinical and healthcare settings. One such at-risk patient cohort is premature infants, defined as born before the start of the 37th week of pregnancy, and accounting for 1 in 10 live births globally, with this number increasing[17]. These infants are born with underdeveloped gut physiology and immunity, which is associated with an increased risk of life threatening infections[18]. Furthermore, preterm infants are often born via Caesarean-section and receive multiple antibiotic courses during their hospitalisation that deplete their gut microbiota, including the beneficial genus *Bifidobacterium*, that may correlate with overgrowth of pathogens linked to diseases such as necrotising enterocolitis (NEC)[19, 20]. Notably, this disease may be difficult to diagnose in its early stages and is often associated with fulminant deterioration. Pathogens currently linked to NEC include *Clostridium perfringens* and *Klebsiella pneumoniae*[21]. From a diagnostic perspective it is therefore important to be able to confidently detect (i) microbes at the species level to provide accurate diagnosis, (ii) species abundance within the microbiota (as these bacteria can also reside within the wider community, but not cause disease when at low levels) and (iii) AMR gene repertoires.

Despite technical challenges, MinIONs have been successfully used in medical research on low complexity samples. To date this includes surveillance of the Ebola outbreak in West Africa (using RT-PCR amplicons)[22], characterisation of bacterial isolates[23] and DNA spiked into patient samples at known levels[24] or from samples heavily (~90%) infected with a single species[25, 26]. Diagnostics in metagenomics samples is still challenging due to lower MinION sequence yields and accuracy, but is essential as most clinical samples are complex. Here we demonstrate MinION based metagenomics, first in a controlled (i.e. mock community) setting and then with clinical samples including in 'real time'. These studies allowed us to determine longitudinal microbiota profiles, gut-associated pathogens linked with sepsis or NEC and their AMR profiles. We benchmarked these data against conventional Illumina sequencing, whole-genome sequencing (WGS) on pathogen isolates and phenotypic antibiotic susceptibility testing.

RESULTS

Accurate classification of a microbial mock community using MinION sequencing

We benchmarked the accuracy of MinION technology by profiling a bacterial mock community of a staggered abundance mixture (HM-277D, BEI Resources). Sequencing was carried out on R7.3 flowcells using a "2D" library protocol that involves sequencing both strands of DNA. One flowcell produced 148,441 total reads, with 71,675 reads passing default quality filter and a mean size of 3,047 bp (read size is driven largely by the size of input DNA molecules) and a longest read size of 40,561 bp (**Table 1**). Reads were analysed with NanoOK[27] and produced alignments to the 20 reference sequences with 82-89% identity (consistent with previous published analysis[28]) including long error-free sequences up to 223 bases (the full NanoOK report is available at https://github.com/richardmleggett/bambi). Coverage of each genome ranged from almost 0x (8 reads of *Actinomyces odontolyticus*) to 13x (7,695 reads of *Streptococcus mutans*), which is consistent with their mock concentrations (**Supplementary table 5**). We benchmarked to 'conventional' (Illumina based) metagenomic sequencing using 1 million reads (equivalent to a MiSeq nano flowcell). To simulate an unknown community, reads were BLAST searched against the NCBI nt database and then classified taxonomically using MEGAN6[29]. This showed broadly similar abundance levels across both platforms, but in some cases, MEGAN's Lowest Common Ancestor algorithm was able to assign a greater proportion of Nanopore reads to species level rather than genus or family, while in other cases, a greater proportion of Illumina reads were able to be assigned to species (**Fig. 1a**). This apparent contradiction is likely because in some cases the longer length of Nanopore reads will provide better specificity, however these longer reads may also contain errors which may lessen specificity, depending on sequence distance between two species within the same genus. Comparing MinION and Illumina species abundance resulted in a Pearson correlation coefficient of 0.91 (**Fig. 1b**), which is comparable to previously published work assessing mock community composition on Illumina technology[30].

Monitoring microbial disturbances in the preterm gut microbiota using MinION

We next sought to determine the relevance of MinION technology in a more clinical context; profiling at-risk preterm infants. We studied faecal samples from one preterm patient (P10) from days 13, 28 and 64 after birth (samples P10N, P10R and P10V respectively, **Fig. 2a**), and compared MinION platform (version R7.3) to Illumina shotgun and Illumina 16S rRNA gene amplicon sequencing. We initially ran one flowcell for P10N, but it was necessary to run two flowcells for P10R and P10V in order to generate sufficient yield on this earlier MinION chemistry. This study generated between 145,342 and 234,453 reads per sample across 5 flowcells (**Table 1**). We confirmed that the MinION sequencing depth was sufficient to capture the complete species diversity of the samples, by computing rarefaction curves (**Supplementary Fig. 1**). For all sequencing technologies and samples, the vast majority of species diversity was captured by ~20,000 reads. This analysis indicated that the yield and accuracy of R7.3 MinION flowcells was sufficient for analysis of low bacterial diversity preterm samples[31].

When we compared MEGAN's taxonomic assignments obtained using MinION vs. Illumina data we again observed that the results were comparable for the majority of the bacterial genera present in preterm infant P10 (e.g. *Klebsiella*, *Enterobacter*, *Veillonella*, *Staphylococcus* and *Bifidobacterium*) (**Fig. 2b**). Interestingly, for sample P10N (collected when the infant was receiving probiotic supplementation) both platforms indicated presence of one probiotic strain, *Bifidobacterium bifidum*. MinION and Illumina also detected the presence of *Enterobacter cloacae* (in sample P10R), a notorious pathogen causing late onset

sepsis (LOS) in preterm infants[32], which correlated with the clinical diagnosis of suspected sepsis (**Fig. 2a,** however no clinical microbiology testing was carried out to confirm this). Corresponding 16S rRNA gene data (**Supplementary Fig. 2**) provided similar profiles to our shotgun results (albeit with different abundances), but unsurprisingly the short 16S rRNA gene reads failed to differentiate some bacteria taxa even at genus level e.g. members of the family *Enterobacteriaceae*, which comprises commensal gut bacteria as well as opportunistic pathogens and whose full-length 16S rRNA genes are often indistinguishable from one another. Overall, these data highlight the potential for MinION technology and shotgun sequencing to be used in a clinical setting to confirm the impact of microbiota intervention studies (i.e. probiotics) and elucidate potential causative pathogens which may lead to clinical infection diagnoses.

Administration of antibiotics can lead to disruption of early colonisation by gut microbes, and may also contribute to the reservoir of AMR genes; the 'resistome'[33]. Here, we also determined the AMR profile in preterm infant P10 using the Comprehensive Antibiotic Resistance Database (CARD) and compared MinION sequencing data to HiSeq Illumina data. **Supplementary Fig. 3** represents a summary of the AMR genes detected using both sequencing technologies. Overall, if we classify the AMR genes according to mode of action, the detection efficiency of MinION and Illumina was comparable, with only four genes with unique resistance mechanisms (*mphC*, *fusB, sat-4 and vanRG*) were detected exclusively by Illumina. Focusing on gene abundance, four main groups: efflux pumps, β-lactamases, aminoglycosides and fluoroquinolones were particularly prevalent (**Supplementary Fig. 4**). Importantly, MinION technology was able to detect AMR genes specific for certain species such as *ileS* encoding for mupirocin[34] resistance in *Bifidobacterium* (sample P10N), or *fosA2*[35] encoding for fosfomycin resistance, which is specific for *E. cloacae* (sample P10R, **Supplementary Table 1**). These data illustrate the ability of MinION technology to detect the pool of AMR genes present in the samples tested, and also determine species-specific AMR genes.

Use of MinION for rapid characterisation of gut-associated pathogenic bacteria in a preterm infant with necrotising enterocolitis
Having successfully detected a specific beneficial bacterium (i.e. *B. bifidum* after probiotic supplementation) and a potential opportunistic pathogen (i.e. *E. cloacae*) using MinION technology (**Fig. 2b**), we next performed a 'real time' run to evaluate how rapidly the MinION could do this in a clinical setting. Current rapid clinical microbiology tests, including determining antibiotic susceptibility, take between 36 and 48 h. We performed a 'real-time' run (from sample preparation to data analysis) using the most up-to-date MinION R9.4 flowcells on a faecal sample from a preterm infant (P8) clinically diagnosed with suspected NEC (Bell stage 1A) (**Supplementary Fig. 5**). Notably, this infant was exposed, before sample collection, to 43 days of non-concurrent antibiotic treatments (i.e. benzylpenicillin, gentamicin, meropenem, tazocin, vancomycin, flucloxacillin, metronidazole and amoxicillin).

We executed the 'real-time' run and timed all stages used for this pipeline including sample preparation (90 min), DNA quality control (45 min), 2D MinION library preparation and loading onto the MinION flowcell (2 h), and sequencing-and-data analysis (40 min for first non-specific AMR hit) (**Fig. 3**). Thus, the overall time from initiating DNA extraction to obtaining first clinically relevant data was less than 5 h. In this run, we also evaluated two different basecalling approaches for the nanopore sequencing. Initially we ran ONT's Metrichor cloud-

based service, but we encountered significant time delays on the experimental day. Having generated 787,601 reads in just over 11 h, but basecalled only a small proportion of these, we then abandoned the attempt and reloaded the same flowcell using MinKNOW's local basecalling option, which was able to keep pace with read generation. Using this approach, we generated 423,422 reads in approximately 23.5 h, and it is these that we considered in the following analysis. Shortly after completing the experiment, ONT announced the discontinuation of the Metrichor basecalling service, explaining the lag we had observed.

The sequenced reads which successfully passed the quality filter (mean quality 9 or greater giving 230,494 reads, **Table 1**) were analysed using the NanoOK RT pipeline which groups 500 read batches for analysis against the NCBI nt and CARD databases (see Online Methods for further details). The first 500 reads immediately indicated a dominance (230 reads) of *K. pneumoniae* (a potential causative organism that has been associated with NEC pathogenesis in preterm infants[36]), as well as *Proteus mirabilis* (42 reads). By 1 h after sequencing start (5 h 15 min total), the pipeline had analysed 18,000 pass reads and *K. pneumoniae* accounted for around 75% of reads. To further verify we had sequenced enough of the bacterial diversity existing in the sample at this time point, we compared taxonomic profiles from analysis completed at 6 h (91,000 reads, 10 h 15 min total time) and at 26 h (229,500 reads, 30 h 15 min total time, sequencing finished and all analysis complete). This comparison verified that there were no significant qualitative differences between the three taxonomic profiles (**Fig. 4a**).

As highlighted previously, it is clinically important to detect AMR genes in metagenomic samples from preterm infants to guide appropriate antibiotic prescription. In our 'real-time' run we determined how rapidly we could map AMR genes from the CARD database with respect to sequencing depth over time. **Fig. 4b** shows the classes of AMR genes detected throughout the run, including quinolone resistance, β-lactamases and efflux pumps (similar to observations in preterm infant P10) were detected in as little as 1 h after sequencing start. We were able to detect *K. pneumoniae*-specific SHV variants[37] as early as 1 h 8 min (at 21,000 reads, 5 h 23 min total time), whilst other lower abundance AMR genes in the sample, such as those conferring trimethoprim, streptothricin and colistin resistance, were not detected until 3-9 h post sequencing (7-13 h total).

As MinION reads are typically longer than Illumina reads we reasoned we could extract additional information by examining flanking sequences either side of each AMR hit and searching the NCBI nt database for hits that were independent (defined as ≳ 50 bp) from the AMR sequence. This 'walkout' study indicated that the majority of AMR genes within the whole metagenomic sample mapped to *K. pneumoniae* (**Supplementary Fig. 6, Supplementary Table 2**), including multidrug exporters such as *acrB* or *oqxAB* associated to *K. pneumoniae* and conferring resistance to tetracycline, chloramphenicol, and fluoroquinolones. We also correlated specific AMR gene cassettes including *aadA*, which can confer aminoglycoside resistance to *P. mirabilis*. These data indicated that relevant AMR genes detected in a metagenomic sample and further mapped to known pathogenic species may facilitate tailored antibiotic treatment strategies for critically ill patients.

Isolation of *Klebsiella pneumoniae* colonies from preterm infant P8 and phenotypic characterisation of antibiotic resistance

To validate the genotypic data obtained from our 'real time' MinION nanopore run, we isolated

nine *K. pneumoniae* colonies from patient P8, and performed alignment analysis on their 16S rRNA gene sequences which showed similarity levels ranging from 99.8% to 100% (**Supplementary Table 3**). To further confirm our metagenomics findings, we performed WGS on one *K. pneumoniae* isolate (assembled using Prokka, 69 contigs, ~5.7MB genome, 57% GC and 23.5x coverage). Notably, we observed five AMR genes (*FosA5* [encoding Fosfomycin resistance], *acraA* [efflux pump], *oqxA*, *oqxB* [quinolone resistance], and several different SHV allelic variants [encoding extended-spectrum β-lactamases, ESBLs]), which matched to the same genes as analysed from MinION data, suggesting metagenomic sequencing and our 'walk-out' analysis was sufficient to detect relevant AMR genes (**Supplementary Fig. 8**).

In order to demonstrate antibiotic resistance phenotypes correlating to presence of AMR genes we tested the susceptibility of the same *K. pneumoniae* isolate with the seven most commonly used antibiotics in NICUs (**Supplementary Fig. 9**). Interestingly, the *K. pneumoniae* isolate was found to have a higher minimum inhibitory concentration (MIC) breakpoint value for those antibiotics that were prescribed to the preterm infant P8 (i.e. benzylpenicillin, amoxicillin, metronidazole, gentamicin and vancomycin, **Supplementary Fig. 5**). In contrast, the MIC breakpoint values obtained for antibiotics that were not administered to the infant (i.e. cefotaxime) were lower and similar to those put forward by the European Committee on Antimicrobial Susceptibility Testing (EUCAST[38]). These data correlate with the AMR genes determined from the MinION and WGS data; SHV variants encode resistance to β-lactam antibiotics (i.e. benzylpenicillin and amoxicillin). Interestingly, this *K. pneumoniae* isolate was also found to be vancomycin resistant by MIC testing, but we did not detect any specific vancomycin resistance genes encoded in its genome. This resistance may therefore be explained by the presence of various multidrug exporter genes such as *acrA* and *oqxAB* (**Supplementary Fig. 9**).

DISCUSSION

There are pressing clinical needs for faster microbial diagnostics and tailored, rather than widespread broad spectrum, antibiotic use. Gold standard microbial laboratory screening is limited by the growth rates of possible pathogens on selective media and the combinatorial tests that can be conducted. While microbiological tests take days, newer sequencing platforms, e.g. Illumina, could manage this in a faster manner and provide rich data that can also be used to track hospital epidemics[39], but requires specialised laboratories and large capital investment. The new MinION sequencer seems an ideal tool because it is smaller, faster and cheaper than Illumina machines. Using a combination of improved Nanopore sequencing chemistries, and our own open source Nanopore analysis packages, NanoOK RT and NanoOK Reporter, we showed that the MinION platform is able to successfully profile known metagenomes (i.e. mock community), and even clinical samples. Importantly, MinION sequencing data using the new R9.4 flowcells were comparable in discriminatory power to the conventional Illumina sequencing platform, and provided clinically relevant information within just 5 h from sample receipt. We were able to demonstrate three clinically relevant, and actionable, pieces of data: 1) abundance of microbial species present in the sample, 2) overall antibiotic resistance gene profile and 3) species-specific antibiotic profiles.

Our first findings, using a known metagenomic sample (mock community), indicated that the

MinION could be suitable for detection of microbes from a mixed clinical sample and so we next tested longitudinal samples from a preterm infant residing in NICU. We found MinION and Illumina analyses were comparable in discriminating species and indicated one of the probiotic strains (i.e. *B. bifidum*) during the supplementation period was present (**Fig. 2b**). Furthermore, these data provided information on probiotic therapy indications after antibiotic administration; absence of these beneficial species may correlate with a 'disturbed' microbial profile[40]. Specifically, *E. cloacae* completely dominated the microbiota (sample P10R) and is a known sepsis pathogen (able to disseminate to the blood) in immunocompromised, including preterm, patients[41]. Therefore, profiling opportunistic gut-associated pathogens may provide key clinical data to prevent subsequent systemic infections.

One of the critical tests carried out by clinical microbiology labs is the detection of pathogens and their antibiotic sensitivities. This is particularly pertinent due to the increasing prevalence of antibiotic-resistant bacteria and lack of novel antibiotic development. We believe MinION sequencing may provide a rapid tool for crafting bespoke antibiotic treatment strategies in time-critical patients. When we investigated AMR genes detected by MinION and Illumina, both sequencing platforms generated reads mapping to genes with similar antibiotic resistance mechanisms (**Supplementary Fig. 3**), and only 4 genes (*mphC*, *fusB*, *sat-4 and vanRG*) out of 146 genes were detected exclusively by Illumina with unique resistance mechanisms. This result may be correlated to the lower MinION read count, and so could be mitigated by improvements in MinION technology for subsequent studies. Notably, we observed presence of AMR genes that corresponded to prescribed antibiotics; β-lactamase and aminoglycoside genes conferring resistance to benzylpenicillin and gentamicin, while fluoroquinolone resistance genes did not correlate to any prescribed antibiotics, which may relate to AMR gene transfer of strains from other sources (**Supplementary Fig. 4**). Thus, AMR profiling may guide clinical treatment decisions at an earlier stage of patient care.

Whilst our comparisons of preterm gut-associated microbial profiles highlighted the significant scope for MinION in a relevant clinical context, we wanted to demonstrate that the entire pipeline of sample preparation, library construction, sequencing and analysis could be carried out rapidly and in 'real-time'. Importantly, we used the most recent flowcells (R9.4) for this study, which have an improved error rate (~8% for 1D, and 4% for 2D reads) and yield. Specifically, for this study we added significant functionality to NanoOK for the 'real-time' analysis of species abundance and antibiotic resistance genes. The sample chosen for this run-through was from an extremely ill preterm infant P8 (born after 26 weeks' gestation with a birthweight of only 508 g), who had received multiple courses of antibiotics since birth (46 days of antibiotic treatment out of 63 days of life at sample collection), and presented with clinical NEC observations at the time of sample collection. Whilst MinION allows generation of long reads, DNA extraction to maximise this type of data is more time-consuming, thus in this study we utilised a rapid DNA extraction protocol (including a bead-beating step). Furthermore, previous studies, including our own (Alcon-Giner et al, unpublished), indicate that incomplete DNA extraction significantly biases metagenomic profiles obtained[42], which may in turn limit pathogen detection and AMR analysis. Therefore, we acknowledge a limitation here which would be linked to the relatively short (N50 of 1,052 bp) reads produced during the MinION run and we expect improved DNA extraction methods would yield more powerful datasets.

We also wanted to provide as robust sequencing data as possible, thus we used the 2D library preparation for the run, which allows both strands to be sequenced and a concensus

generated with greater accuracy. Our initial attempt was unsuccessful, with no reads passing filter, potentially down to the quality of flowcells. Additionally, our second attempt was slowed down by the Metrichor cloud base calling service (subsequently decommissioned), causing us to re-load the same flowcell to carry out local base calling and obtain impressively high yields (423,422 reads). This second attempt at a complete MinION run, revealed a *K. pneumoniae*-dominated profile, after just 1 h of sequencing (~18,000 reads), enabling us to confidently 'call' this potential pathogen. This analysis was further strengthened as more sequencing, e.g. time-points 6 h and 26 h (after sequencing start), gave almost identical microbial profiles (**Fig. 4a**), as did Illumina sequencing (**Supplementary Fig. 7**). However, it should be noted that the single species domination in this sample facilitates early detection at lower read depth and low level pathogen abundance would require deeper sequencing. Nevertheless, these data highlight how rapid diagnosis of pathogen overgrowth is possible using R9.4 Nanopore flowcells. Our data indicate the gut was colonised with *K. pneumoniae* in an infant with suspected NEC, which suggests that this bacterium may be causative in this illness. Importantly, *K. pneumoniae* has been linked to preterm NEC (and is supported by corresponding clinical observations of Bells diagnostic staging criteria, **Supplementary Fig. 5**), with overgrowth in the intestine linked to pathological inflammatory cascades, facilitated by a 'leaky' epithelial barrier[43].

Whilst detection of individual pathogens is important, a critical additional requisite is identification of AMR profiles so that tailored antibiotic treatment can be used. Real-time analysis of MinION data highlighted the presence of a significant metagenomic 'resistome', including presence of colistin resistance, a last resort antibiotic, by the detection of gene *arnA*[44]. We noted the greater the read depth the greater the number of AMR genes detected, although importantly we were able to detect a significant number of AMR genes as rapidly as 1 h after sequencing start, including β-lactamases, quinolone, aminoglycoside and tetracycline resistance genes (**Supplementary Table 4**). *Klebsiella* is of particular concern in an AMR context due to the increasing emergence of multidrug-resistant isolates that cause severe infection, which represent a real threat to patient outcomes[45]. Therefore, to improve the potential for guiding antibiotic treatment decisions we performed the 'walked-out' study from the AMR gene sequences to determine which bacterial species were carrying which AMR genes. Our data indicated that these genes primarily mapped back to *K. pneumoniae* (**Supplementary Fig. 6**), and secondly *P. mirabilis*. This level of data may significantly enhance clinical diagnosis and antibiotic regimens to be determined in an intensive care environment. Notably, although infant P8 did not receive quinolone-type antibiotics, we detected quinolone resistance genes in these data (potentially acquired via horizontal gene transfer), which may indicate that prescription of, for example, ciprofloxacin may be ineffective in this patient.

Based on these data, a clinical decision (as determined by neonatologist Dr Paul Clarke, in a blinded fashion) at 1 h after sequencing start (5 h 15 min total) would be continuation of gentamicin and flucloxacillin or cefotaxime treatment (NICE guidance[46] states antibiotics should be given within 1 h of presentation of signs of sepsis). However, using 6 h AMR hit data (10 h 15 min total), the presence of *acrD* (a transporter belonging to the resistance-nodulation-division family, known to participate in the efflux of aminoglycosides[47]) might suggest gentamicin resistance and thus the infant may be moved onto meropenem (a carbapenem) that is gaining popularity as a single agent for neonatal infections[48]. Interestingly,

we first detected *acrD* at 2 h 49 min after sequencing started (50,500 reads analysed, 7 h 4 min since sample receipt) therefore an earlier change in prescribed antibiotic, i.e. to meropenem, could have been possible. At the final 26 h time-point (30 h 15 min total) the guidance would be to continue with meropenem, which is also a last resort antibiotic against *Enterobacteriaceae*, including *K. pneumoniae*[49]. In NEC cases, where the patient has deteriorated and sepsis is suspected, the presence of *K. pneumoniae* in the blood stream would be required for a complete diagnosis. However, few patients with NEC symptoms are also normally positive for this sepsis diagnostic, with sample processing taking approx. 36-48 h. Thus, MinION faecal profiling is potentially very important clinically because of the speed of the test, and may allow clinical decisions from first suspicions of disease, to rapidly identify potential gut-associated causative organisms and corresponding AMR profiles.

We sought to further benchmark our MinION pipeline by correlating our metagenomic observations with isolated *K. pneumoniae* strains that had undergone WGS and MIC testing. Notably, whole genome analysis indicated a close agreement of resistance genes determined by MinION analysis, specifically when comparing results from the ´walk-out´ analysis (**Supplementary Fig. 8**). When subjecting strains to MIC testing (gold standard for profiling AMR), we observed phenotypic resistance to all main groups of antibiotics that had been prescribed to infant P8, including gentamicin (which would have continued to be used using 1h post sequencing AMR hits; **Supplementary Fig. 10**). There was good association between AMR gene sequence detection and MIC testing, i.e. SHV and β-lactam antibiotics, which highlights that MinION could be extremely useful for rapid AMR profiling. However, other resistances identified via MIC testing, such as vancomycin and metronidazole, did not correlate to presence of specific resistance genes in either the MinION, Illumina or WGS sequence data. These differences are not only restricted to MinION, and could relate a need of updating AMR databases with phenotype testing. As such we expect that rapid NGS, via MinION, would inform early and more appropriate antibiotic choices for patient care, halting the rapid deterioration observed in critically ill patients. Subsequently, phenotypic testing via standard clinical microbiology labs would further refine clinical management of patients. Notably, MIC testing also suggested susceptibility to meropenem, which is the same antibiotic that we discussed may be prescribed (2 h 49 min post sequencing start, 7 h 4 min from sample receipt), based on the lack of resistance determinants in the metagenomic MinION analysis. Importantly, it is clear that preterm infants harbour a significant reservoir of AMR genes within the wider microbiome as well as in potentially pathogenic bacteria such as *K. pneumoniae*, which makes bespoke antibiotic treatment decisions non-trivial. Thus, further highlighting the urgent requirement for comprehensive AMR databases, new antibiotic development and alternative treatments such as novel antimicrobials or microbiota therapies.

We obtained MinION bioinformatics results in under 5 hours. In comparison, standard Illumina MiSeq sequencing (paired 250 bp reads) alone normally takes 39 hours, excluding sample preparation and analysis (https://www.illumina.com/systems/sequencing-platforms/miseq/specifications.html), and PacBio (i.e. long-read sequencing) with the same sample prep and quality as our MinION experiment, using the rapid (4 hour) library method (http://www.pacb.com/wp-content/uploads/2015/09/Guide-Pacific-Biosciences-Template-Preparation-and-Sequencing.pdf), 30 minute diffusion and the shortest sequencing run (30 minute) would take over 7 hours even without base calling or bioinformatics, highlighting the significantly quicker detection potential of MinION in a clinical setting.

A number of products announced by ONT could make Nanopore sequencing in clinical settings even more attractive. The SmidgION, expected end of 2018, is an even more compact device than the MinION that can be connected to a mobile phone. At the other extreme, the GridION X5 and PromethION are respectively capable of running between 5 and 48 flowcells simultaneously, facilitating better economies of scale. Sample preparation is already relatively simple, particularly for 1D libraries, but this is likely to be further simplified by VolTrax, a compact automated library preparation system from ONT that can manipulate fluids around an array of pixels. ONT have demonstrated 1D library preparation directly from an *Escherichia coli* cell culture entirely with VolTrax (https://nanoporetech.com/publications/voltrax-rapid-programmable-portable-disposable-device-sample-and-library-preparation). Balancing the excitement about new products is that the technology is still relatively immature and is frequently updated, issues that must be addressed for a true diagnostic tool. Even within the time-frame of this study we have observed significant advances in optimisation of MinION technology, and we hope this could lead to a stable platform for large-scale clinical testing.

Whilst we show the results of single patient diagnostics alone, if this platform were to be broadly adopted, even within a single ward or hospital, additional epidemiological analysis should be possible. Transmission chain analysis of hospital-acquired pathogens has been shown using Illumina data[39], and Nanopore sequence-based epidemiology was used to monitor the West Africa Ebola virus outbreak[50].

CONCLUSION

In conclusion, we have demonstrated that MinION technology is an easy-to-use and rapid sequencing platform that has the ability to detect gut-associated pathogens that have been associated with potentially life-threatening preterm-associated infections in 'real time'; identification of specific pathogenic taxa (i.e. *K. pneumoniae*) and corresponding AMR profiles. Data obtained may allow clinicians to rapidly tailor antibiotic treatment strategies (i.e. change from gentamicin to meropenem) in a rapid (~7 h decision from sample receipt) and timely manner. The utility of this approach was confirmed when compared to Illumina metagenomic sequencing and isolation and characterisation of *K. pneumoniae* strain including WGS and phenotypic (i.e. MIC) testing. This suggests that MinION may be used in a clinical setting, potentially improving health care strategies and antibiotic stewardship for at-risk preterm infants in the future.

METHODS
Methods and associated references are available in the online version of the paper.

Accession codes
The Illumina and MinION read data supporting the conclusions of this article are available in the European Nucleotide Archive (http://www.ebi.ac.uk/ena) under study accession PRJEB22207.

ACKNOWLEDGEMENTS

AUTHORS CONTRIBUTIONS

LJH, MDC and RML designed the research; CAG, DH, MK, RML, SC, and TCB performed research; RML contributed new software tools; CAG, LH, LJH, MDC, PC, RML and SC analysed data; and CAG, LJH, MDC and RML wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors have not received direct financial contributions from ONT, but RML and MDC have received a small number of free flowcells as part of the MAP and MARC programs. RML is in receipt of travel and accommodation expenses to speak at an ONT conference and is on a PhD student advisory team with a member of ONT staff.

TABLES

Table 1: Nanopore flowcell versions and yields. Yield and length metrics for Nanopore runs. Flowcell for sample P8 (2D) was used for two experiments: (a) Initial Metrichor basecalled run which was abandoned due to network lag and (b) Local basecalled run which was used for results presented in the analysis. Sample P8 (1D) utilised the rapid transposase-based library preparation protocol but was ultimately unsuccessful in our experiment. Only pass reads were used for analysis - these are defined as having a mean Q score of 9 or greater.

Sample	Flowcell	Sequencing kit	Total number of reads	Number of pass reads	Mean length of pass reads (bp)	Pass read N50	Longest pass read
Mock community	R7.3	SQK-MAP006	148,441	71,675	3,047	5,497	40,561
P10N	R7.3	SQK-MAP006	145,342	82,734	2,926	3,910	17,979
P10R (1)	R7.3	SQK-MAP006	103,705	35,560	2,958	3,967	18,069
P10R (2)	R7.3	SQK-MAP006	45,486	12,118	3,832	4,832	23,423
P10V (1)	R7.3	SQK-MAP006	165,026	53,437	925	1,087	11,300
P10V (2)	R7.3	SQK-MAP006	69,427	21,780	2,027	2,318	15,206
P8 (2D)	R9.4	SQK-LSK208	(a) 787,601 (b) 423,422	(a) 407,818 (b) 230,494	(a) 606 (b) 622	(a) 1,768 (b) 1,052	(a) 18,769 (b) 18,095
P8 (1D)	R9.4	SQK-RAD002	69,442	44	1,480	3,089	7,848

ONLINE METHODS

Mock Community

DNA

We used genomic DNA from a microbial mock community used in the Human Microbiome Project (HM-277D, BEI Resources, Manassas, VA). This mock community contains a mixture of 20 bacterial strains containing staggered RNA operon counts. Details of the strains present in the community are indicated in **Supplementary Table 5**.

Illumina sequencing of mock community

Illumina compatible amplification-free paired end libraries were constructed with inserts spanning from 600 bp to >1000 bp. A total of 600 ng of DNA was sheared in a 60 µl volume on a Covaris S2 (Covaris, Massachusetts, USA) for 1 cycle of 40 s with a duty cycle of 5%, cycles per burst of 200 and intensity of 3. Fragmented DNA was then end-repaired using the NEB End Repair Module (NEB, Hitchin, UK), size selected with a 0.58x Hi Prep bead clean-up (GC Biotech, Alphen aan den Rijn, The Netherlands) and followed by A tailing using the NEB A tailing module (NEB) and ligation of adapters using the NEB Blunt/TA Ligase Master Mix (NEB). Three 1x bead clean-ups were then undertaken to remove all traces of adapter dimers. Library quality control was performed by running an Agilent BioAnalyser High Sensitivity chip and quantified using the Kappa qPCR Illumina quantification kit. Based on the qPCR quantification libraries were loaded at 9 pM on an Illumina MiSeq and sequenced with 300 bp paired reads.

MinION sequencing of mock community

MinION 2D libraries were constructed targeting inserts >8 kbp. A total of 1 µg of DNA was fragmented in a 46 µl volume in a Covaris G-tube (Covaris, Massachusetts, USA) at 6,000 rpm in an Eppendorf centrifuge 5417. Sheared DNA was then subjected to a repair step using the NEB FFPE repair mix (NEB, Hitchin, UK) and purified with a 1x Hi Prep bead clean-up (GC Biotech, Alphen aan den Rijn, The Netherlands). A DNA control was added to the repaired DNA and then end-repaired and A-tailed using the NEBNext Ultra II End Repair and A-Tailing Module (NEB), purified with a 1x Hi Prep bead clean-up and then the AMX and HPA MinION adapters ligated using the NEB Blunt/TA Ligase Master Mix (NEB). An HP tether was then added and incubated for 10 min at room temperature followed by a further 10 min room temperature incubation with an equal volume of pre-washed MyOne C1 beads (Thermo Fisher, Cambridge, UK). The library bound beads were washed twice with bead binding buffer (ONT) before the final library eluted via a 10 min incubation at 37 ºC in the presence of the MinION Elution Buffer. The final library was then mixed with running buffer, fuel mix and nuclease free water and loaded onto an R7.3 flowcell per the manufacturer's instructions and sequencing data collected for 48 h.

Mock community data analysis

MinION reads were basecalled using the Metrichor service and downloaded as FAST5 files. NanoOK v0.54[27] was used to extract FASTA files, to align (via the LAST aligner[51]) against a reference database of the 20 genomes and to generate an analysis report (**Supplementary**

Note 1). A QC of the Illumina data was carried out with FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) to ensure read quality was within expected bounds. This demonstrated a mean QC of 30 up to base 250. We subsampled a random set of 1,000,000 reads (subsample.pl script, https://github.com/richardmleggett/scripts) to represent the yield of a MiSeq nano flowcell and ran Trimmomatic[52] to remove remaining adaptor content and apply a sliding window quality filter (size 4, mean quality greater than or equal to 15). Illumina and MinION reads were then BLASTed against the NCBI nt database and results were imported into MEGAN6[29] for taxonomic analysis. The Pearson coefficient was computed in Excel using species abundance counts exported from MEGAN.

Clinical samples

Ethical approval and preterm sample collection
The Ethics Committee of the Faculty of Medical and Health Sciences in the University of East Anglia (Norwich, UK) approved subject recruitment for this study. Protocol for faeces collection was laid out by the Norwich Research Park (NRP) Biorepository (Norwich, UK) and was in accordance with the terms of the Human Tissue Act 2004 (HTA), and approved with license number 11208 by the Human Tissue Authority. Infants admitted to the Neonatal Intensive Care Unit (NICU) of the Norfolk and Norwich University Hospital (NNUH, Norwich, UK) were recruited by doctors or nurses with informed and written consent obtained from parents. Oral probiotic supplementation provided to the infants in this study contained *Bifidobacterium bifidum* and *Lactobacillus acidophilus* (i.e. Infloran®, Desma Healthcare, Switzerland) strains with a daily dose of 2×10^9 of each species. Collection of faecal samples was carried out by researchers and stored at -80 °C prior to DNA extraction.

DNA extraction of faeces samples from preterm infants
Bacterial DNA was extracted using the FastDNA Spin Kit for Soil (MP) following the manufacturer's instructions but extending the bead-beating step to 1 min, and eluting the DNA with 55 °C DES. The starting faecal material used for the DNA extraction was between 100 and 150 mg. The purity and concentration of the DNA was assessed using a NanoDrop 2000c Spectrophotometer and Qubit® 2.0 fluorometer. Samples with DNA concentrations higher than 25 ng/µl were considered acceptable.

MinION shotgun library preparation
MinION 2D libraries were constructed as outlined for the mock (see above) except that for the R9.4 flowcells the final library was mixed with running buffer containing fuel mix, library loading beads and nuclease free water and loaded onto the flowcell per the manufacturer's instructions. MinION 1D libraries were prepared by incubating 200 ng of DNA for with 2.5 µl FRM 1 min at 30 °C then 1 min at 75 °C followed by the addition of 1 µl RAD and 0.2 µl NEB Blunt/TA Ligase Master Mix (NEB) and a room temperature incubation for 5 min. The final library was then mixed with running buffer containing fuel mix, library loading beads and nuclease free water and loaded onto the flowcell per the manufacturer's instructions.

Libraries from infant P10 (P10N, P10V and P10R) were prepared using the SQK-MAP006 Genomic Sequencing Kit and sequenced on R7.3 flowcells for 48 h. ONT's MinKNOW software was used to collect signal data and the Metrichor cloud-based service was used for

basecalling.

Illumina HiSeq 2500 shotgun library preparation

Libraries for samples (P10N, P10R and P10V) were prepared using TruSeq Nano DNA Library Prep Kit according to the manufacturer's instructions and sequenced using the HiSeq Illumina 2500 machine with 150 bp paired end reads. The library for P8 was prepared as for the Amplification Free library for the mock (see above) and run at 9 pM on an Illumina MiSeq with a 2x 250 bp read metric.

16S rRNA gene library preparation and bioinformatics analysis

16S rRNA gene libraries were prepared using bacterial DNA normalised to 5 ng ml-1 for sample P10N and P10V. V4 region of the 16S rRNA gene was amplified using primers, 5' AAT GAT ACG GCG ACC ACC GAG ATC TAC A and, 5' CAA GCA GAA GAC GGC ATA CGA GAT AAC T. PCR conditions used for this amplification were: 1 cycle of 94°C 3 min and 25 cycles of 94°C for 45 s, 55°C for 15 s and 72°C for 30 s using a 96 well Thermal Cycler PCR machine. Sequencing of the 16S RNA gene libraries was performed on the Illumina MiSeq platform with 250 bp paired end reads.

Bioinformatic analysis started by processing raw reads through quality control using FASTX-Toolkit[53] keeping a minimum quality threshold of 33 for at least 50% of the bases. Reads that passed the threshold were aligned against SILVA database (version: SILVA_119_SSURef_tax_silva)[54] using BLASTN[55] (ncbi-blast-2.2.25+; Max e-value 10e-3) separately for both pairs. After performing the BLASTN alignment, all output files were imported and annotated using the paired-end protocol of MEGAN[29].

Time series study for infant P10

Illumina and MinION sequencing data for samples P10N, P10V and P10R from infant P10 were studied. For the Illumina samples, we removed PCR duplicates (remove_pcr_duplicates.pl script from https://github.com/richardmleggett/scripts), ran Trimmomatic[52] to remove adaptors and applied a sliding window quality filter (size 4, mean quality greater than or equal to 15) and then randomly sub-sampled 1 million reads (subsample.pl script from same location). These reads were used as the input to a blastn search of NCBI's nt database. For the Nanopore sequencing, we took only the reads classified as 'pass' reads (defined as 2D reads with a mean Q value >9) and performed no further pre-processing before running blastn. Using MEGAN6[29], we removed reads matching Homo sapiens (accounting for < 0.1% per sample) and performed taxonomic analysis.

Real-time diagnostic study for preterm P8 using MinION and NanoOK RT

One sample from infant P8 was sequenced. Both 1D and 2D Nanopore libraries were prepared using the SQK-RAD002 Rapid Sequencing Kit 1D and SQK-LSK208 Ligation Sequencing Kit 2D, respectively, and each library was sequenced on a R9.4 flowcell. MinKNOW software was used to collect signal data. As described in the main text, an attempt was made to use the Metrichor base calling service for the 2D library but failed, instead local basecalling through MinKNOW was used, though at the time only basecalled the first, or template strand, so could not generate more accurate 2D consensus reads. MinKNOW was also used for local basecalling of the 1D library, which cannot generate 2D reads.

To enable real-time analysis of MinION data new functionality was added to NanoOK[29]. The

new software, NanoOK RT, monitored a specified directory for FAST5 files as they are created. For each new file, a FASTA file was extracted automatically. For efficiency FASTA files were grouped into batches of 500, and each batch was BLAST searched against the NCBI nt database (downloaded February 2017) and the CARD database (v1.1.1, downloaded October 2016) of antibiotic resistance genes[56]. NanoOK RT can also write out command files for MEGAN[29], which allows more detailed analysis of community composition, either as the run proceeds, or on completion. NanoOK RT is available as an extension to NanoOK, selectable as a run-time option, from https://github.com/TGAC/NanoOK.

Another bioinformatic tool, NanoOK Reporter, was also developed for this project, and provided a graphical user interface to monitor the run, view summaries of species, and antibiotic resistance genes identified. The tool allowed the user to browse through data in real time as batches were processed, or after all of the results were in using their timestamps to indicate when a result was first obtained. This summary data can also be exported as plain text files and these were subsequently used for later analysis (below). NanoOK Reporter is available from https://github.com/richardmleggett/NanoOKReporter. Documentation for NanoOK Reporter as well as a tutorial utilising the data from this publication is available at https://documentation.tgac.ac.uk/display/NANOOK/NanoOK+Reporter.

Generation of resistance heat map in Figure 4b
We opened the CARD results using NanoOK Reporter and used the option to save summary data as a plain text file. This saved a text file for the analysis at each time point (here batches of timestamped 500 reads) summarising the counts of resistance genes identified up to that point. We took the latest time point file (chunk 459, available at https://github.com/richardmleggett/bambi) and extracted a list of the ARO (Antibiotic Resistance Ontology) numbers from the ID column. Each unique ARO was manually assigned to one of 13 groups (as displayed in the figure). We wrote a script (gather_heatmap_data.pl, same GitHub repository) to take the summary files, together with this mapping and to generate a file (BAMBI_P8_2D_Local_070317_hits.txt) summarising hits per group at each time point. An R script (plot_card_heatmap.R, same GitHub repository) takes this file and renders the heat map.

Walking out from resistance genes to identify the encompassing bacteria
We wrote a shell script to go through all the CARD BLAST hits and wrote each CARD hit and the corresponding nt hits for the same read (walk_out_preprocess.sh, available at https://github.com/richardmleggett/bambi). A second script (walk_out.pl) took the output from the first script and parsed it, read-by-read. If there was a hit in nt that began at least 50 bases before the start of the CARD hit, or at least 50 bases after the end of the CARD hit, then this species was taken as the encompassing species. The script also recorded count of the number of times each species was seen.

Statistical analysis on number of reads
Read counts at different stages of the bioinformatics analysis are provided in **Table 1**. For comparative analysis, sequences were normalised using values from the sample with the lowest number of reads.

Isolation and biochemical characterisation of P8 *Klebsiella pneumoniae* strains
An aliquot (100 mg) of faecal sample was homogenised in 1 mL TBT buffer (100 mM Tris/HCl,

pH 8.0; 100 mM NaCl; 10 mM $MgCl_2 \cdot 6H_2O$) by pipetting and plate mixing at 1500 rpm for 1 h. Homogenates were serially diluted to 10^{-4} in TBT buffer. Aliquots of 50 μl were spread on MacConkey (Oxoid) agar plates in triplicate and incubated aerobically at 37 °C overnight. Colonies were selectively screened for lactose-positive (i.e. pink) colonies. One colony of each morphology type was re-streaked on MacConkey agar three times to purify. Biochemical characterisation was performed using API 20E tests (Biomerieux) according to manufacturer's instructions.

16S rRNA phylogenetic analysis of P8 *Klebsiella pneumoniae* isolates

Sequences of the 16S rRNA gene from nine *Klebsiella pneumoniae* isolates were prepared to perform the phylogenetic analysis. We extracted DNA using FastDNA Spin Kit for Soil (MP) following the manufacturer's instructions and then amplified the 16S rRNA gene employing Veriti™ 96 well Thermal Cycler (Applied Biosystems, USA), master mix from Kapa2G Robust PCR reagents (Kapa Biosystems, USA) and primers: fD1 (5'- AGA GTT TGA TCC TGG CTC AG - 3'), fD2 (5'- AGA GTT TGA TCA TGG CTC AG - 3') and rP1 (5' - ACG GTT ACC TTG TTA CGA CTT - 3')[57]. PCR amplification conditions were: 1 cycle of 94 °C for 5 min, followed by 35 cycles of 94 °C for 1 min, 43 °C for 1 min, and 72 °C for 2 min followed by a final strand extension at 72 °C for 7 min. Amplicons were sequenced using automated Sanger sequencing service (Eurofins Genomics, Luxembourg).

Partial 16S ribosomal RNA sequences (~900 positions) of nine isolates of *Klebsiella pneumoniae*, obtained using automated Sanger sequencing service (Eurofins Genomics, Luxembourg), were compared for similarity. Multiple sequence alignments were performed with SINA (SILVA Incremental Aligner) (Pruesse et al., 2012) (https://www.arb-silva.de/aligner/), and manually curated for quality. The nucleotides were coloured using Boxshade version 3.21 (http://www.ch.embnet.org/software/BOX_form.html). The similarity/identity matrix between the sequences was calculated using MatGAT, version 2.01 (Matrix Global Alignment Tool) with BLOSUM 50 alignment matrix[58].

Determination of Minimal Inhibitory Concentration (MIC) for P8 *Klebsiella pneumoniae* isolate

Calculation of the antibiotic Minimal Inhibitory Concentration (MIC) of growth of a *Klebsiella pneumoniae* isolate from baby P8 (sample P8E) was performed using the broth microdilution method[59]. Serial two fold dilutions of the most common antibiotics used at NICU (benzylpenicillin, gentamicin, vancomycin, metronidazole, meropenem and cefotaxime) were added to sterile nutrient broth. Bacterial inoculum of the isolate was prepared using 10 μl from a fresh overnight culture. Microplates were incubated for 24 h at 37 °C under aerobic conditions. Cell density was monitored using a plate reader (BMG Labtech, UK) at 595 nm. MICs were determined as the lowest concentration of antibiotic inhibiting any bacterial growth.

DNA extraction from P8 *Klebsiella pneumoniae* isolate for WGS analysis

An overnight (10 ml) culture of the isolate was centrifuged at 4000 rpm for 10 min, re-suspended in 30 ml of Phosphate Buffered Saline (PBS) (Sigma-Aldrich, UK) and centrifuged again. The pellet was then re-suspended in 2 ml of 25% sucrose (Fisher Scientific, USA) in TE (10 mM Tris (Fisher Scientific, USA) and 1 mM EDTA at pH 8.0 (VWR Chemicals, USA); 50 μl of Roche Lysozyme (Roche Molecular Systems, UK) at 100 mg/ml in 0.25 M Tris pH 8.0

was added. The mixture was incubated at 37 °C for 1 h, and 100 µl of Proteinase K at 20 mg/ml (Roche Molecular Systems, UK), 30 µl of RNase A at 10 mg/ml (Roche Molecular Systems, UK), 400 µl of 0.5 M EDTA pH 8.0 (VWR Chemicals, USA), and 250 µl of freshly prepared 10% Sarkosyl NL30 (Sigma-Aldrich, UK) were added. The mixture was then incubated on ice for 2 h and subsequently transferred to a water bath at 50 °C overnight. Next, E Buffer (10 mM Tris pH8.0 (Fisher Scientific, UK)) was added to the sample to a final volume of 5 ml, mixed with 5 ml Phenol:Chloroform:Isoamyl Alcohol (PCIA) (25:24:1) (Sigma-Aldrich, UK) in a Qiagen MaXtract High Density tube (Qiagen, DE) and centrifuged for 15 min at 4000 rpm. The aqueous phase was transferred into a new Qiagen MaXtract High Density tube, made up with E Buffer to the volume of 5 ml if necessary, mixed with 5 ml of PCIA, and centrifuged for 10 min at 4000 rpm. This procedure was repeated, with 5 min centrifugation time. Next, the aqueous phase was transferred into a Qiagen MaXtract High Density tube, made up to 5 ml with E Buffer if necessary, mixed with 5 ml of Chloroform:Isoamyl Alcohol (CIA) (24:1) (Sigma-Aldrich, UK), and centrifuged for 5 min at 4000 rpm. The CIA step was repeated once more, after which the final aqueous phase was transferred into a sterile Corning ™ 50ml centrifuge tube, and 2.5 volumes of ethanol (Ethanol absolute AnalaR NORMAPUR®, VWR Chemicals, USA) were added. The sample was incubated for 15 min at -20 °C, then centrifuged for 10 min at 4000 rpm and 4 °C. Finally, the DNA pellet was washed with 10 ml of 70% ethanol and centrifuged at 4000 rpm for 10 min twice, dried overnight, and re-suspended in 300 µl of E Buffer.

Whole genome sequencing library preparation and sequencing of P8 *Klebsiella pneumoniae* isolate

DNA samples containing 500 ng genomic DNA were analysed. DNA was sheared into fragments of 400-600 bp using a Covaris plate with glass wells and AFA fibres. SPRI clean-up was used to remove smaller sized fragments and concentrate the sheared DNA samples. Whole genome library construction performed by a liquid handling robot comprised end repair, A-tailing and adapter ligation reactions. Adapter ligated samples were subsequently amplified using the following PCR conditions: 5 min 95°C, 10 cycles of (30 sec 98°C, 30 sec 65°C, 1 min 72°C) and 10min at 72°C. LabChip GX was then used to size and assess the quality of the libraries and determine pooling volumes for each library using Beckman Coulter Biomek NXp (span-8). Libraries were prepared using Agilent Technologies Sure Select Custom Library Prep kit. Final pools were finally loaded on the HiSeq 2500 sequencers.

Assembly and bioinformatics analysis of P8 *Klebsiella pneumoniae* isolate's WGS

Presence or absence testing of AMR genes was performed on one *Klebsiella pneumoniae* isolate assembled using Velvet. Contigs were aligned using BLAST (v2.2.30) against the CARD database under double filtering criteria of expected e-value $1e^{-10}$ and 90% identity. Customised in-house scripts were used to generate a data matrix that was then used to construct a heat map using the R heatmap2 package.
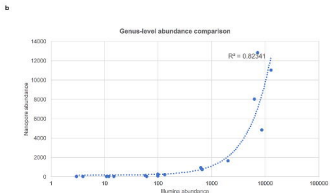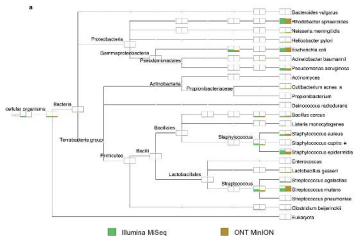
REFERENCES

1.   Ley, R.E. et al. Obesity alters gut microbial ecology. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 11070-11075 (2005).
2.   Claesson, M.J. et al. Gut microbiota composition correlates with diet and health in the

elderly. *Nature* **492**, 178 (2012).

3. Qin, J. et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**, 55-60 (2012).

4. Karlsson, F.H. et al. Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature* **498**, 99 (2013).

5. Qin, N. et al. Alterations of the human gut microbiome in liver cirrhosis. *Nature* **513**, 59 (2014).

6. Blander, J.M., Longman, R.S., Iliev, I.D., Sonnenberg, G.F. & Artis, D. Regulation of inflammation by microbiota interactions with the host. *Nature immunology* **18**, 851 (2017).

7. Lewis, B.B. & Pamer, E.G. Microbiota Therapy for Clostridium difficile and Antibiotic Resistant Enteric Infections. *Annual Review of Microbiology* **71** (2017).

8. Pedersen, H.K. et al. Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature* **535**, 376-381 (2016).

9. Lane, E.R., Zisman, T.L. & Suskind, D.L. The microbiota in inflammatory bowel disease: current and therapeutic insights. *Journal of inflammation research* **10**, 63 (2017).

10. Lippert, K. et al. Gut microbiota dysbiosis associated with glucose metabolism disorders and the metabolic syndrome in older adults. *Beneficial microbes*, 1-12 (2017).

11. Wekerle, H. Brain Autoimmunity and Intestinal Microbiota: 100 Trillion Game Changers. *Trends in immunology* (2017).

12. Khoruts, A. & Sadowsky, M.J. Understanding the mechanisms of faecal microbiota transplantation. *Nature Reviews Gastroenterology & Hepatology* **13**, 508-516 (2016).

13. Leggett, R.M. & Clark, M.D. (PeerJ Preprints, 2017).

14. Organization, W.H. Antimicrobial resistance: global report on surveillance. (World Health Organization, 2014).

15. Hoffman, S.J. et al. An international legal framework to address antimicrobial resistance. *Bulletin of the World Health Organization* **93**, 66-66 (2015).

16. Globally, T.D.-R.I. Final Report and Recommendations. *Review on Antimicrobial Resistance Report* (2016).

17. birth, W.P. WHO. (2016).

18. Melville, J.M. & Moss, T.J. The immune consequences of preterm birth. *Frontiers in neuroscience* **7** (2013).

19. Raymond, S.L., Rincon, J.C., Wynn, J.L., Moldawer, L.L. & Larson, S.D. impact of early-Life exposures to infections, Antibiotics, and vaccines on Perinatal and Long-term Health and Disease. *Frontiers in Immunology* **8** (2017).

20. Underwood, M.A. & Sohn, K. The Microbiota of the Extremely Preterm Infant. *Clinics in Perinatology* (2017).

21. Pammi, M. et al. Intestinal dysbiosis in preterm infants preceding necrotizing enterocolitis: a systematic review and meta-analysis. *Microbiome* **5**, 31 (2017).

22. Quick, J. et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530**, 228 (2016).

23. Ashton, P.M. et al. MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. *Nature biotechnology* **33**, 296-300 (2015).

24. Votintseva, A.A. et al. Same-day diagnostic and surveillance data for tuberculosis via whole-genome sequencing of direct respiratory samples. *Journal of clinical microbiology* **55**, 1285-1298 (2017).

25. Schmidt, K. et al. Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. *Journal of Antimicrobial Chemotherapy* **72**, 104-114 (2016).

26. Greninger, A.L. et al. Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. *Genome medicine* **7**, 99 (2015).

27. Leggett, R.M., Heavens, D., Caccamo, M., Clark, M.D. & Davey, R.P. NanoOK: multi-reference

alignment analysis of nanopore sequencing data, quality and error profiles. *Bioinformatics* **32**, 142-144 (2015).

28.   Ip, C.L. et al. MinION Analysis and Reference Consortium: Phase 1 data release and analysis. *F1000Research* **4** (2015).

29.   Huson, D.H. et al. MEGAN community edition-interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS computational biology* **12**, e1004957 (2016).

30.   Jovel, J. et al. Characterization of the gut microbiome using 16S or shotgun metagenomics. *Frontiers in microbiology* **7** (2016).

31.   DiBartolomeo, M.E. & Claud, E.C. The developing microbiome of the preterm infant. *Clinical therapeutics* **38**, 733-739 (2016).

32.   Chen, H.-N., Lee, M.-L., Yu, W.-K., Lin, Y.-W. & Tsao, L.-Y. Late-onset Enterobacter cloacae sepsis in very-low-birth-weight neonates: experience in a medical center. *Pediatrics & Neonatology* **50**, 3-7 (2009).

33.   Hu, Y., Gao, G.F. & Zhu, B. The antibiotic resistome: gene flow in environments, animals and human beings. *Frontiers of Medicine*, 1-8 (2017).

34.   Serafini, F. et al. Insights into physiological and genetic mupirocin susceptibility in bifidobacteria. *Applied and environmental microbiology* **77**, 3141-3146 (2011).

35.   Xu, H., Miao, V., Kwong, W., Xia, R. & Davies, J. Identification of a novel fosfomycin resistance gene (fosA2) in Enterobacter cloacae from the Salmon River, Canada. *Letters in applied microbiology* **52**, 427-429 (2011).

36.   Sim, K. et al. Dysbiosis anticipating necrotizing enterocolitis in very premature infants. *Clinical Infectious Diseases* **60**, 389-397 (2014).

37.   Liakopoulos, A., Mevius, D. & Ceccarelli, D. A review of SHV extended-spectrum β-lactamases: neglected yet ubiquitous. *Frontiers in microbiology* **7** (2016).

38.   EUROCAST, E.C.o.A.S.T. (2017).

39.   Harris, S.R. et al. Evolution of MRSA during hospital transmission and intercontinental spread. *Science* **327**, 469-474 (2010).

40.   Thomas, J.P., Raine, T., Reddy, S. & Belteki, G. Probiotics for the prevention of necrotizing enterocolitis in very-low-birth-weight infants: A meta-analysis and systematic review. *Acta Paediatrica* (2017).

41.   Anand, R.J., Leaphart, C.L., Mollen, K.P. & Hackam, D.J. The role of the intestinal barrier in the pathogenesis of necrotizing enterocolitis. *Shock* **27**, 124-133 (2007).

42.   Gerasimidis, K. et al. The effect of DNA extraction methodology on gut microbiota research applications. *BMC research notes* **9**, 365 (2016).

43.   Hodzic, Z., Bolock, A.M. & Good, M. The Role of Mucosal immunity in the Pathogenesis of Necrotizing enterocolitis. *Frontiers in pediatrics* **5** (2017).

44.   Poirel, L., Jayol, A. & Nordmann, P. Polymyxins: antibacterial activity, susceptibility testing, and resistance mechanisms encoded by plasmids or chromosomes. *Clinical microbiology reviews* **30**, 557-596 (2017).

45.   Wyres, K.L. & Holt, K.E. Klebsiella pneumoniae population genomics and antimicrobial-resistant clones. *Trends in microbiology* **24**, 944-956 (2016).

46.   Osvald, E.C. & Prentice, P. NICE clinical guideline: antibiotics for the prevention and treatment of early-onset neonatal infection. *Archives of Disease in Childhood-Education and Practice* **99**, 98-100 (2014).

47.   Rosenberg, E.Y., Ma, D. & Nikaido, H. AcrD of Escherichia coli is an aminoglycoside efflux pump. *Journal of bacteriology* **182**, 1754-1756 (2000).

48.   Final Report Summary - NEOMERO (European multicenter network to evaluate pharmacokinetics, s.a.e.o.M.i.n.s.a.m. (2016).

49.   El-Gamal, M.I. et al. Recent updates of carbapenem antibiotics. *European journal of medicinal chemistry* (2017).

50.   Hoenen, T. et al. Nanopore sequencing as a rapidly deployable Ebola outbreak tool.
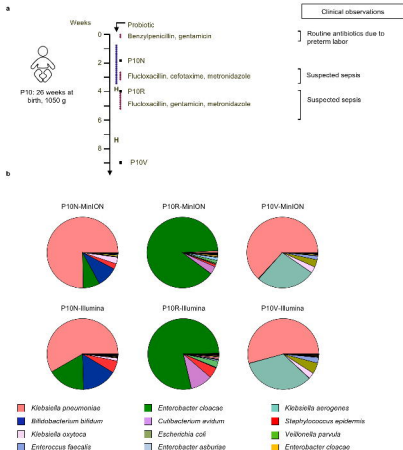
Emerging infectious diseases **22**, 331 (2016).

51. Kiełbasa, S.M., Wan, R., Sato, K., Horton, P. & Frith, M.C. Adaptive seeds tame genomic sequence comparison. *Genome research* **21**, 487-493 (2011).

52. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).

53. Gordon, A. & Hannon, G. Fastx-toolkit. *FASTQ/A short-reads preprocessing tools (unpublished)* http://hannonlab*. cshl. edu/fastx_toolkit* **5** (2010).

54. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research* **41**, D590-D596 (2012).

55. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *Journal of molecular biology* **215**, 403-410 (1990).

56. McArthur, A.G. et al. The comprehensive antibiotic resistance database. *Antimicrobial agents and chemotherapy* **57**, 3348-3357 (2013).

57. Weisburg, W.G., Barns, S.M., Pelletier, D.A. & Lane, D.J. 16S ribosomal DNA amplification for phylogenetic study. *Journal of bacteriology* **173**, 697-703 (1991).

58. Campanella, J.J., Ludwig-Mueller, J., Bakllamaja, V., Sharma, V. & Cartier, A. ILR1 and sILR1 IAA amidohydrolase homologs differ in expression pattern and substrate specificity. *Plant Growth Regulation* **41**, 215-223 (2003).

59. Reller, L.B., Weinstein, M., Jorgensen, J.H. & Ferraro, M.J. Antimicrobial susceptibility testing: a review of general principles and contemporary practices. *Clinical infectious diseases* **49**, 1749-1755 (2009).

a

- Bacteroides vulgatus
- Rhodobacter sphaeroides
- Neisseria meningitidis
- Helicobacter pylori
- Escherichia coli
- Acinetobacter baumannii
- Pseudomonas aeruginosa
- Actinomyces
- Cutibacterium acnes *
- Propionibacterium
- Deinococcus radiodurans
- Bacillus cereus
- Listeria monocytogenes
- Staphylococcus aureus
- Staphylococcus cuptis *
- Staphylococcus epidermidis
- Enterococcus
- Lactobacillus gasseri
- Streptococcus agalactiae
- Streptococcus pneumoniae
- Streptococcus mutans
- Clostridium beijerinckii
- Eukaryota

Proteobacteria, Gammaproteobacteria, Flavobacteriales, Actinobacteria, Propionibacteriaceae, Bacilli, Bacillales, Staphylococcus, Firmicutes, Lactobacillales, Streptococcus — cellular organisms, Bacteria, Terrabacteria group

Illumina MiSeq | ONT MiniON

b

Genus-level abundance comparison

$R^2 = 0.82341$

Figure 1: Sequencing of microbial mock community (HM-277D, BEI Resources) using Illumina MiSeq and Oxford Nanopore Technologies MiniON sequencing.

(a) MEGAN taxonomy tree representing species assigned from the mock community control sequenced by Illumina MiSeq (green) and ONT MiniON (brown). The height of the bars represents the number of reads assigned for each species class. (*) represents species assigned by Megan but not specified as members of the community.

(b) Genus level abundance correlation plot, with log-based x-axis. Points represent abundance in Nanopore vs. Illumina. Calculated Pearson coefficient 0.91.

a

Weeks

Clinical observations

P10: 26 weeks at birth, 1050 g

Probiotic

Benzylpenicillin, gentamicin ⎤ Routine antibiotics due to ⎦ preterm labor

0 — P10N

Flucloxacillin, cefotaxime, metronidazole ⎤ Suspected sepsis ⎦

2 — P10R

Flucloxacillin, gentamicin, metronidazole ⎤ Suspected sepsis ⎦

4 —

6 —

H

8 — P10V

b

P10N-MinION

P10R-MinION

P10V-MinION

P10N-Illumina

P10R-Illumina

P10V-Illumina

🟥 *Klebsiella pneumoniae*	🟩 *Enterobacter cloacae*
🟦 *Bifidobacterium bifidum*	🟪 *Cutibacterium avidum*
🟧 *Klebsiella oxytoca*	🟨 *Escherichia coli*
⬜ *Enterococcus faecalis*	🟦 *Enterobacter asburiae*

🟩 *Klebsiella aerogenes*
🟥 *Staphylococcus epidermidis*
🟩 *Veillonella parvula*
🟨 *Enterobacter cloacae*

Figure 2: Longitudinal study on preterm P10 using MinION and Illumina sequencing.
(a) Timeline diagram of preterm P10 indicating times of faecal sample collection (P10N, P10R and P10V), duration of antibiotic and probiotic treatment, and relevant clinical observations. The timeline diagram is divided into weeks and dots represent days within the scale. Blue dots represent days of probiotic treatment, red dots antibiotic treatment, black squares the points for sample collection and letter H transfer of the preterm to another hospital. More clinical information of preterm P10 can be found in **Supplementary Table 6.**
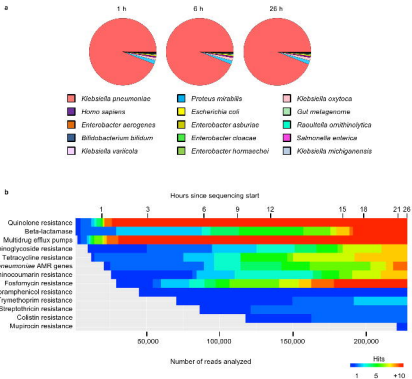(b) Sequencing data from ONT MinION and Illumina HiSeq 2500 from preterm P10. Pie charts represents taxonomic profiles at different time points from preterm P10N, P10R and P10V, as assigned by MEGAN. The top row corresponds to results obtained using MinION sequencing and bottom row displays results using Illumina HiSeq. The figure legend shows all species with abundance ≥ 0.01% in at least one sample.
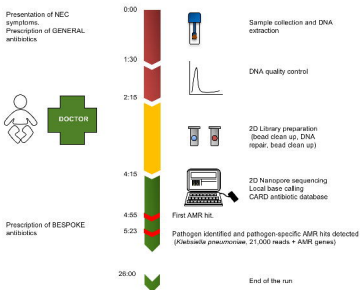
Figure 3: Timeframe diagram for 'real time' run performed for rapid diagnostic of a preterm infant (P8) suffering from NEC.
Step 1 (red, 2h 25min): Sample collection, DNA extraction and quality control. Step 2 (yellow, 2h): 2D library preparation using bead clean up and DNA repair. Step 3 (green): data analysis using local base calling and NanoOK RT. Pathogen detection (*Klebsiella pneumoniae*) and specific AMR genes specific were first detected at 5 hours and 23 minutes (21,000 reads analysed). Left side of the panel indicates clinical symptoms and general guidelines for antibiotic prescription. More clinical information of preterm P8 can be found in **Supplementary Table 7**.

a

1 h	6 h	26 h

- Klebsiella pneumoniae
- Homo sapiens
- Enterobacter aerogenes
- Bifidobacterium bifidum
- Klebsiella variicola
- Proteus mirabilis
- Escherichia coli
- Enterobacter asburiae
- Enterobacter cloacae
- Enterobacter hormaechei
- Klebsiella oxytoca
- Gut metagenome
- Raoultella ornithinolytica
- Salmonella enterica
- Klebsiella michiganensis

b

Figure 4: Rapid diagnostic of preterm P8 clinically diagnosed with suspected NEC.
(a) Taxonomic profiles obtained using MinION technology at 1h, 6h and 26h since sequencing started. Figure legend comprises the 15 most abundant taxonomic taxa obtained. (b) Evolution along the sequencing time of the number of CARD database hits detected among the most common groups of antibiotic resistance genes found in preterm P8. Top and lower panel indicate the hours since sequencing started and the number of reads analyzed, respectively within this timeframe.