

Evaluating the Genome and Resistome of Extensively Drug-Resistant *Klebsiella pneumoniae* using Native DNA and RNA Nanopore Sequencing

--Manuscript Draft--

Manuscript Number:	GIGA-D-19-00200	
Full Title:	Evaluating the Genome and Resistome of Extensively Drug-Resistant <i>Klebsiella pneumoniae</i> using Native DNA and RNA Nanopore Sequencing	
Article Type:	Research	
Funding Information:	Institute for Molecular Bioscience Centre for Superbug Solutions (610246)	Dr Lachlan Coin
Abstract:	<p>Background</p> <p><i>Klebsiella pneumoniae</i> frequently harbours multidrug resistance and current diagnostics struggle to rapidly identify appropriate antibiotics to treat these bacterial infections. The MinION device can sequence native DNA and RNA in real-time, providing an opportunity to compare the utility of DNA and RNA for prediction of antibiotic susceptibility. However, the effectiveness of bacterial direct RNA sequencing and base-calling has not previously been investigated. This study interrogated the genome and transcriptome of four extensively drug-resistant (XDR) <i>K. pneumoniae</i> clinical isolates, however, further antimicrobial susceptibility testing identified three isolates as pandrug-resistant (PDR).</p> <p>Results</p> <p>The majority of acquired resistance ($\geq 75\%$) resided on plasmids including several megaplasmids (≥ 100 kbp). DNA sequencing detected most resistance genes ($\geq 70\%$) within 2 hours of sequencing. Neural-network based base-calling of direct RNA achieved up to 86% identity rate, although only $\leq 23\%$ of reads could be aligned. Direct RNA sequencing (with approximately 6 times slower pore translocation) was able to identify (within 10 hours) $\geq 35\%$ of resistance genes, including those associated with resistance to aminoglycosides, β-lactams, trimethoprim and sulphonamide and also quinolones, rifampicin, fosfomycin and phenicol in some isolates. Polymyxin-resistant isolates showed a heightened transcription of <i>phoPQ</i> (≥ 2-fold) and the <i>pmrHFIJKLM</i> operon (≥ 8-fold). Expression levels estimated from direct RNA sequencing displayed strong correlation (Pearson: 0.86) compared to qRT-PCR across eleven resistance genes.</p> <p>Conclusion</p> <p>Overall, MinION sequencing rapidly detected the XDR/ PDR <i>K. pneumoniae</i> resistome and direct RNA sequencing revealed differential expression of these genes.</p>	
Corresponding Author:	Lachlan Coin AUSTRALIA	
Corresponding Author Secondary Information:		
Corresponding Author's Institution:		
Corresponding Author's Secondary Institution:		
First Author:	Miranda E. Pitt	
First Author Secondary Information:		
Order of Authors:	Miranda E. Pitt Son H. Nguyen	

	Tânia P.S. Duarte
	Haotian Teng
	Mark A.T. Blaskovich
	Matthew A. Cooper
	Lachlan Coin
Order of Authors Secondary Information:	
Additional Information:	
Question	Response
Are you submitting this manuscript to a special series or article collection?	No
<p>Experimental design and statistics</p> <p>Full details of the experimental design and statistical methods used should be given in the Methods section, as detailed in our Minimum Standards Reporting Checklist. Information essential to interpreting the data presented should be made available in the figure legends.</p> <p>Have you included all the information requested in your manuscript?</p>	Yes
<p>Resources</p> <p>A description of all resources used, including antibodies, cell lines, animals and software tools, with enough information to allow them to be uniquely identified, should be included in the Methods section. Authors are strongly encouraged to cite Research Resource Identifiers (RRIDs) for antibodies, model organisms and tools, where possible.</p> <p>Have you included the information requested as detailed in our Minimum Standards Reporting Checklist?</p>	Yes
<p>Availability of data and materials</p> <p>All datasets and code on which the conclusions of the paper rely must be</p>	Yes

either included in your submission or deposited in [publicly available repositories](#) (where available and ethically appropriate), referencing such data using a unique identifier in the references and in the “Availability of Data and Materials” section of your manuscript.

Have you have met the above requirement as detailed in our [Minimum Standards Reporting Checklist](#)?

1 **Evaluating the Genome and Resistome of Extensively Drug-Resistant *Klebsiella pneumoniae***
2 **using Native DNA and RNA Nanopore Sequencing**

3

4 Miranda E. Pitt¹, Son H. Nguyen¹, Tânia P.S. Duarte¹, Haotian Teng¹, Mark A.T. Blaskovich¹, Matthew A. Cooper¹,
5 Lachlan J.M. Coin¹

6 ¹ Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, 4072, Australia

7 Corresponding authors: Miranda Pitt (miranda.pitt@imb.uq.edu.au, ORCID: 0000-0002-8255-4036) and Lachlan
8 Coin (l.coin@imb.uq.edu.au, ORCID: 0000-0002-4300-455X)

9

10 **Abstract**

11 **Background:** *Klebsiella pneumoniae* frequently harbours multidrug resistance and current diagnostics struggle to
12 rapidly identify appropriate antibiotics to treat these bacterial infections. The MinION device can sequence native
13 DNA and RNA in real-time, providing an opportunity to compare the utility of DNA and RNA for prediction of
14 antibiotic susceptibility. However, the effectiveness of bacterial direct RNA sequencing and basecalling has not
15 previously been investigated. This study interrogated the genome and transcriptome of four extensively drug-resistant
16 (XDR) *K. pneumoniae* clinical isolates, however, further antimicrobial susceptibility testing identified three isolates
17 as pandrug-resistant (PDR).

18 **Results:** The majority of acquired resistance ($\geq 75\%$) resided on plasmids including several megaplasmids (≥ 100 kbp).
19 DNA sequencing detected most resistance genes ($\geq 70\%$) within 2 hours of sequencing. Neural-network based base-
20 calling of direct RNA achieved up to 86% identity rate, although only $\leq 23\%$ of reads could be aligned. Direct RNA
21 sequencing (with approximately 6 times slower pore translocation) was able to identify (within 10 hours) $\geq 35\%$ of
22 resistance genes, including those associated with resistance to aminoglycosides, β -lactams, trimethoprim and
23 sulphonamide and also quinolones, rifampicin, fosfomycin and phenicol in some isolates. Polymyxin-resistant isolates
24 showed a heightened transcription of *phoPQ* (≥ 2 -fold) and the *pmrHFIJKLM* operon (≥ 8 -fold). Expression levels
25 estimated from direct RNA sequencing displayed strong correlation (Pearson: 0.86) compared to qRT-PCR across
26 eleven resistance genes.

27 **Conclusion:** Overall, MinION sequencing rapidly detected the XDR/ PDR *K. pneumoniae* resistome and direct RNA
28 sequencing revealed differential expression of these genes.

29 **Introduction**

30 *Klebsiella pneumoniae* is one of the leading causes of nosocomial infections, with reports of mortality rates as high
31 as 50% [1-5]. This opportunistic pathogen commonly exhibits multidrug resistance which severely limits treatment
32 options [6]. A high abundance of resistance is commonly encoded on plasmids, accounting for the rapid global
33 dissemination of resistance [1,6]. Common therapeutic options for multidrug-resistant infections include
34 carbapenems, fosfomycin, tigecycline and polymyxins [7]. However, resistance is also rapidly developing against
35 these antibiotics resulting in the emergence of extensively drug-resistant (XDR) and subsequent pandrug-resistant
36 (PDR) strains [6-9].

37 One of the major contributors to the advent of antibiotic resistance is the inability for current detection methodologies
38 to readily and accurately assess bacterial infections in particular, the resistance profile [10]. Rapid sequencing has
39 been proposed as a way to determine antibiotic resistance, including approaches which utilise high accuracy short
40 reads, as well as those which exploit real-time single-molecule sequencing such as Oxford Nanopore Technologies
41 (ONT). The ONT MinION platform is a portable single-molecule sequencer which can sequence long fragments of
42 DNA and stream the sequence data for further data processing in real-time, detecting the presence of bacterial species
43 and acquired resistance genes [11-15]. Moreover, the long reads coupled with the ability to multiplex samples has
44 immensely aided with the assembly of bacterial genomes [16-18]. This capability allows for the rapid determination
45 of whether resistance is residing on the chromosome or plasmid/s. Of particular interest are high levels of resistance
46 encoded on plasmids, as these genes can rapidly be transferred throughout the bacterial population via horizontal gene
47 transfer. However, a limitation of DNA sequencing is accurately identifying whether the presence of an acquired
48 resistance gene or mutation is facilitating resistance.

49 ONT has recently released a direct RNA sequencing capability, which sequences native transcripts. Other sequencing
50 technologies rely on fragmentation, cDNA conversion and PCR steps that create experimental bias and hinder the
51 accuracy of determining gene expression [19, 20]. The ability for MinION sequencing to read long fragments enables
52 full length transcripts to be investigated. To date, only a few direct RNA sequencing publications exist which include
53 eukaryote transcriptomes, primarily yeast (*Saccharomyces cerevisiae* [19, 21]) and recently, *Homo sapiens* [22]. This
54 sequencing has additionally been implemented in viral transcriptomics [23-25]. Only one prior study by Smith AM *et*
55 *al.* has applied this sequencing to bacterial 16S ribosomal RNA (rRNA) to detect epigenetic modifications [26].
56 Bacterial transcription differs significantly from eukaryotes in that transcription and translation occur simultaneously.

57 As a result, bacterial mRNA transcripts lack poly(A) tails and alternative splicing [27]. The poly(A) tail is critical for
58 the library preparation for ONT sequencing thus, we have established a methodology for adding this component onto
59 transcripts.

60 In this study, we applied MinION sequencing to interrogate both the genome and the transcriptome (via direct RNA
61 sequencing) for XDR *K. pneumoniae* clinical isolates. Of interest was to compare the potential for RNA sequencing
62 to provide a better correlation to the resistance phenotype than DNA sequencing. These isolates have previously
63 undergone ‘traditional’ whole genome sequencing (Illumina) and antimicrobial susceptibility testing [28]. An
64 extended panel of antibiotics was tested in this study to identify PDR isolates. Three strains were selected from this
65 cohort which exhibited resistance to all 24 antibiotics or antibiotic combinations tested, a high abundance of antibiotic
66 resistance genes (≥ 26) and differing lineages (ST11 (16_GR_13), ST147 (1_GR_13) and ST258 (2_GR_12)).
67 Additionally, these isolates harbour polymyxin resistance which is facilitated by a disruption in or upstream of *mgrB*.
68 Variations in the *mgrB* gene result in increased expression of the *pmrCAB* and *pmrHFIJKLM* operon, enables the
69 addition of phosphoethanolamine and/ or 4-amino-4-deoxy-L-arabinose (Ara4N) to lipid A and subsequently
70 facilitates polymyxin resistance [29]. These pathways associated with polymyxin resistance were further explored
71 using direct RNA sequencing and compared against a polymyxin-susceptible XDR isolate (ST258; 20_GR_12). This
72 research aimed to assemble these genomes, discern the differential expression of resistance genes and ascertain the
73 time required for detection. Furthermore, we sought to compare DNA and RNA sequencing as modalities for the rapid
74 identification of acquired antibiotic resistance.

75

76 **Methods**

77 ***Bacterial strains and growth conditions***

78 XDR *K. pneumoniae* clinical strains were sourced through the Hygeia General Hospital, Athens, Greece [28].
79 Antimicrobial susceptibility assays (Supplementary Table S1), sequence typing and detection of acquired resistance
80 genes have previously been determined [28]. Strains were stored at -80°C in 20% (v/v) glycerol, the identical stock
81 was used as per the prior study and the extended panel of antimicrobial susceptibility testing conducted similarly [28].
82 When required for extractions, glycerol stocks were grown on lysogeny broth (LB) agar plates and 6 morphologically
83 similar colonies were selected for inoculation. The inoculum was grown in LB overnight at 37°C shaking at 220 rpm.
84 This overnight inoculum was used for both DNA and RNA extractions.

85 ***DNA extraction and high molecular weight DNA isolation***

86 DNA was extracted from 10 ml of overnight culture using the DNeasy Blood and Tissue Kit (Qiagen) according to
87 manufacturer's guidelines, with the addition of an enzymatic lysis buffer pre-treatment (60 mg/ml lysozyme). High
88 molecular weight (HMW) DNA from the prior extraction was selected using the MagAttract HMW DNA Kit (Qiagen)
89 as per manufacturer's instructions. Subtle changes included a further proteinase K treatment on the DNA extracts at
90 56°C for 10 min followed by supplementation of RNase A (1 mg) for 15 min at room temperature. Several attempts
91 at direct DNA extraction from bacterial cells were undertaken using the MagAttract HMW DNA kit, however, were
92 unsuccessful with these isolates. Due to several issues with potential carbohydrate contamination (260/230 ratio:
93 ≤ 0.3), 2_GR_12 was also purified with the Monarch[®] PCR & DNA Cleanup Kit (New England BioLabs) using the
94 protocol to isolate fragments >2000 bp. DNA and RNA contamination was quantitated using Qubit[®]2.0 (Thermo
95 Fisher Scientific) and purity determined with a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific). DNA
96 fragment sizes were determined using the Genomic DNA ScreenTape & Reagents (Agilent) and sizes from 200 to
97 >60000 bp were analyzed on a 4200 TapeStation System (Agilent) (Supplementary Figure S1).

98 ***RNA extraction, mRNA enrichment and poly(A) addition***

99 The overnight culture was sub-cultured in 10 ml of cation-adjusted Muller Hinton Broth (caMHB) to reflect conditions
100 used for minimum inhibitory concentration (MIC) assays. Cultures were grown to mid-log phase ($OD_{600} = 0.5-0.6$).
101 RNA was extracted via the PureLink[™] RNA Mini Kit (Thermo Fisher Scientific) as per manufacturer's protocols
102 which included using Homogenizer columns (Thermo Fisher Scientific). To remove DNA contamination, the TURBO
103 DNA-free[™] kit was implemented. A minor adjustment was an increased concentration of TURBO DNase (4 U)
104 incubated at 37°C for 30 min. The RNeasy Mini Kit (Qiagen) clean up protocol was additionally used to purify and
105 concentrate RNA samples. Ribosomal RNA was depleted via the MICROBExpress[™] Bacterial mRNA Enrichment
106 Kit (Thermo Fisher Scientific). Minor protocol changes included adding ≥ 2 μ g of DNA depleted RNA and the
107 enriched mRNA was precipitated for 3 h at -20°C. Poly(A) addition was performed using the Poly(A) Polymerase
108 Tailing Kit (Astral Scientific) as per the manufacturer's alternative protocol (4 U input of Poly(A) Polymerase). The
109 input RNA concentration was ≥ 800 ng and RNA samples were incubated at 37°C for 1 h. Poly(A) ligated RNA was
110 purified using Agencourt AmpureXP (Beckman Coulter Australia) beads (1:1 ratio). RNA and DNA contamination
111 was quantitated using the Qubit[®]2.0 (ThermoFisher Scientific) and purity determined with a NanoDrop 1000
112 Spectrophotometer (Thermo Fisher Scientific). RNA fragment size was checked using an Agilent RNA 6000 Pico kit

113 and run on a 2100 Bioanalyzer (Agilent Technologies) for the initial RNA extract, post ribosomal RNA depletion and
114 after poly(A) ligation (Supplementary Figure S2).

115 ***ONT library preparation and sequencing***

116 RNA libraries (≥ 600 ng poly(A) ligated RNA) were prepared using the Direct RNA Sequencing kit (SQK-RNA001).
117 The Rapid Barcoding Sequencing kit (SQK-RBK001) was used for HMW DNA samples (1_GR_13, 16_GR_13,
118 20_GR_12; 300 ng input each). Isolate 2_GR_12 (300 ng input) was prepared separately using the Rapid Sequencing
119 Kit (SQK-RAD003). Libraries were sequenced with MinION R9.4 flowcells and the raw data (fast5 files) were base-
120 called using Albacore 2.1.1 for DNA sequencing (Supplementary Figure S3). For benchmarking purposes, RNA reads
121 were additionally base-called with Albacore 2.2.7, Guppy 3.0.3 and the Chiron [30] RNA base-caller which was
122 trained in-house (<https://github.com/haotianteng/Chiron/releases/tag/v0.5>).

123 ***Real-time resistome detection emulation***

124 The real-time emulation was performed post sequencing and the time required to detect antibiotic resistance was
125 determined as previously described [14]. Briefly, this pipeline aligns Albacore base-called reads via BWA-MEM [31]
126 to an antibiotic resistance gene database. Antibiotic resistance genes were obtained from the ResFinder 3.0 database
127 [32]. This dataset comprises of 2131 genes which were clustered based on 90% identity to form 611 groups or gene
128 families. The detection of false positives is reduced using the multiple sequence alignment software kalign2 [33], a
129 probabilistic Finite State Machine [34] and once the alignment score reached a threshold, the resistance gene was
130 reported.

131 ***Assembly of genomes***

132 To assemble genomes with both Illumina and ONT reads, SPAdes v3.10.1 [35] was implemented. Hybrid assemblers
133 included npScarf [36] and Unicycler v0.3.1 [37]. Assemblers using only ONT reads included Canu v1.5 (excluding
134 reads < 500 bp) [38] and the combination of Minimap2 v2.1-r311 and Miniasm v0.2-r168-dirty; Racon (git commit
135 834442) were used in both cases to polish the assemblies [39, 40]. Consensus sequences were determined using Mauve
136 (snapshot_2015-02-13) to construct the final assembly [41]. The output from each assembly software is reported in
137 Supplementary Table S2. Genomes were annotated using the Rapid Annotation using Subsystem Technology (RAST)
138 which also provided a list of virulence genes [42]. The location of acquired antibiotic resistance genes were determined
139 using ResFinder 3.0 [32] and plasmids were identified via PlasmidFinder 1.3 [43]. To discern if plasmid sequences

140 have previously been reported, contigs underwent a BLASTn analysis against the National Center for Biotechnology
141 Information (NCBI) database (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

142 ***RNA alignment and expression profiling***

143 Base-called RNA reads were converted to DNA (uracil bases changed to thymine) and aligned using BWA-MEM
144 (parameters: -k 11 -W20 -r10 -A1 -B1 -O1 -E1 -L0 -Y) to the updated genome assemblies. Due to the lack of introns
145 and full length transcripts being obtained, BEDTools coverage [44] was used to ascertain the relative expression of
146 resistance genes. This was normalized to the number of counts obtained for the housekeeping gene, *rpsL* [45], to
147 compare against qRT-PCR results. Read alignments were further visualised using Integrative genomics viewer 2.3.59
148 [46].

149 ***Whole transcriptome differential gene expression***

150 To identify genes which were differentially expressed between a pair of samples (x and y), we used a beta-binomial
151 distribution to calculate the probability of observing less than or equal to x_g reads mapping to gene g in sample x,
152 conditional on the total number of reads mapping to all genes ($\sum_g(x_g)$), the number of reads in sample y mapping
153 to gene g (y_g) as well as the total number of reads mapping to all genes in sample y ($\sum_g(y_g)$). This was calculated
154 in R using the `pbetabinom.ab` function in the VGAM package, with $q = x_g$, $size = \sum_g'(x_g)$, $alpha = y_g + 1$; $beta$
155 $= \sum_g'(y_g) - y(g) + 1$. Genes for which this probability was less than a predefined threshold were deemed to be
156 significantly under expressed in sample x given sample y. A similar statistic was used to check for over-expression.

157 ***Quantitative real-time reverse transcriptase PCR (qRT-PCR)***

158 First strand synthesis to generate cDNA (1 μ g total DNase-depleted RNA) was performed using SuperScript III
159 (Thermo Fisher Scientific) which was also used for MinION direct RNA sequencing library preparations. Primers
160 used are displayed in Supplementary Table S3. Samples were prepared in triplicate via the SYBR Select Master Mix
161 (Thermo Fisher Scientific) and expression detected using a ViiA 7 Real-time PCR system (Thermo Fisher Scientific).
162 Cycling conditions include: Hold 50°C (2 min), 95°C (2 min) followed by 50 cycles of: 95°C (15 sec), 55°C (1 min).
163 A melt curve was included to determine the specificity of the amplification and a no template control to detect
164 contamination or primer dimers. Results were analysed with QuantStudio™ Real-Time PCR Software, triplicates
165 were averaged, normalised to the housekeeping gene *rpsL* and relative expression determined via the $2^{-\Delta\Delta CT}$ method
166 [47].

167

168 **Results**

169 *Antibiotic resistance and the location of acquired resistance in the genome*

170 This study assayed nine additional antibiotics or antibiotic combinations to further characterise the phenotypic
171 resistance of these isolates (Supplementary Table S1). Strains 1_GR_13, 2_GR_12 and 16_GR_13 were non-
172 susceptible to all antibiotics including the 24 antibiotics tested previously [28]. 20_GR_12 was only susceptible to
173 gentamicin and polymyxins.

174 Utilising the capacity for MinION sequencing to read long fragments of DNA, the location of antibiotic resistance
175 genes were clearly resolved (Table 1). All genomes were circular with the exception of 2_GR_12 where 3 plasmids
176 remained linear. This was partly due to difficulties extracting DNA and not retaining long fragments (Supplementary
177 Figure S1). Amongst the four isolates, the chromosome size ranged between 5.1-5.5 Mb which encoded resistance
178 genes *blaSHV-11*, *fosA* and *oqxAB*. The majority of resistance ($\geq 75\%$) mapped to plasmids.

179 At least one megaplasmid, defined as a plasmid larger than 100 kbp, was detected in all isolates (Table 1). These
180 commonly harboured the replicon IncA/C2 or InFIB and IncFIIK. The IncA/C2 plasmid was present in all samples
181 except 20_GR_12. This plasmid contained up to 16 resistance genes which conferred resistance towards
182 aminoglycosides, β -lactams, phenicols, rifampicin, sulphonamides, tetracyclines and trimethoprim, with the exception
183 of 16_GR_13. Isolate 16_GR_13 lacked trimethoprim resistance on its IncA/C2 plasmid. The plasmids containing
184 both replicons IncFIB and IncFIIK differed vastly between all four replicates. All contained IncFIB_{pKpn3} and IncFIIK,
185 however, 1_GR_13 differed with IncFII_{pKP91}. Additionally, a differing IncFIB replicon was detected on a separate
186 contig in 1_GR_13 (pKPHS1) and 2_GR_12 (pQil). The only instance where another dual replicon was identified was
187 in 1_GR_13 which harboured both IncR and IncN. This plasmid contained aminoglycoside, β -lactam, trimethoprim,
188 macrolide and sulphonamide resistance. 1_GR_13 also contained a 5.5 kb circular contig which was annotated as a
189 phage genome. Various regions of these megaplasmids were unique to these isolates compared to prior sequences
190 deposited on NCBI (Supplementary Table S5).

191 The ColRNAI plasmid was present in all except 1_GR_13 which encoded aminoglycoside and quinolone resistance
192 (*aac(6')-Ib*, *aac(6')-Ib-cr*) (Table 1). The ColRNAI plasmid in 2_GR_12 and 20_GR_12 was 13841 bp in size and
193 shared 75% similarity between the two isolates. This plasmid differed in 16_GR_13 which contained no resistance
194 genes and 35% the size. The same IncX3 plasmid (43380 bp) was apparent in isolates 2_GR_12 and 20_GR_12.
195 Unique to 16_GR_13 was the IncL/ M_{pOXA-48} plasmid containing *blaOXA-48* and the 50979 bp IncN plasmid in

196 20_GR_12 with resistance against 5 classes (aminoglycoside (*aph(3'')-Ib*, *aph(6)-Id*), β -lactam (*blaTEM-1A*),
197 sulphonamide (*sul2*), tetracycline (*tet(A)*), trimethoprim (*dfrA14*)) of antibiotics.

198 Multiple copies of acquired resistance genes were apparent across plasmids in several isolates. For 1_GR_13, up to
199 three copies were present of genes *aadA24*, *aph(3')-Ia*, *aph(6)-Id*, *dfrA1*, *dfrA14*, *strA* and *sul1* (Table 1). In 2_GR_12,
200 *sul1* and *blaTEM-1A* were duplicated and for 16_GR_13, only *sul1* was represented twice.

201 ***Real-time detection emulation of resistance genes via DNA sequencing***

202 The vast majority ($\geq 70\%$) of resistance genes were detected via DNA sequencing within the first 2 hours (Figure 1,
203 Supplementary Table S5). These genes confer resistance towards aminoglycosides, β -lactams, fosfomycin,
204 macrolides, phenicols, quinolones, rifampicin, sulphonamides, tetracyclines and trimethoprim. 20_GR_12 lacked
205 acquired resistance genes for macrolides, phenicols and rifampicin, however, all other classes were detected within 2
206 hours. All isolates, except 2_GR_12, were sequenced for 21 hours which was sufficient to obtain the complete genome
207 assembly. Only a few additional genes were detected after the first 10 hours across isolates (Supplementary Table S5).
208 For 2_GR_12, an extended run of 41 hours detected no further genes after 20 hours. Overall, the presence of these
209 resistance genes corresponded to a resistant phenotype towards aminoglycosides, β -lactams, fosfomycin, phenicols,
210 quinolones, sulphonamides (sulfamethoxazole), tetracyclines and trimethoprim (Supplementary Table S1). As
211 macrolides and rifampicin are not routinely used to treat *K. pneumoniae* infections, no breakpoints exist according to
212 CLSI and EUCAST guidelines, however, all isolates exhibit an MIC ≥ 128 $\mu\text{g/ml}$ towards erythromycin (macrolide)
213 and ≥ 64 $\mu\text{g/ml}$ for rifampicin.

214 Post 2 hours of sequencing, several genes not observed in the final assembly via ResFinder 3.0 were detected
215 (Supplementary Table S5). These were predominantly genes attributed to aminoglycoside, β -lactam, rifampicin and
216 phenicol resistance. Furthermore, resistance genes to additional differing classes were detected including fusidic acid
217 and vancomycin. This was evident in 2_GR_12 (*fusB*) and 16_GR_13 (*fusB*, *vanR*). However, these genes had less
218 than 30 reads and their phred-scale mapping quality (MAPQ) scores were less than 10 (misplaced probability greater
219 than 0.1). Furthermore, the majority of genes not observed in the final assembly nor observed in Illumina data
220 exhibited a MAPQ score of ≤ 10 which may indicate that a more stringent threshold is required to negate false positives.
221 However if this threshold increases, true positives would not be detected including *aadA1*, *aadA2* and *ARR-2* in
222 2_GR_12 and *blaOXA-48*, *blaCTX-M-15* and *ARR-2* in 16_GR_13.

223 Several genes found in the final assembly were not detected in the real-time emulation analysis (Supplementary Table
224 S5). This was mainly observed for aminoglycoside resistance encoding genes. For 1_GR_13, this included *aadA1*,
225 *ant(2'')-Ia*, *aph(6)-Id* and *aadA24*. Similarly, 2_GR_12 and 20_GR_12 lacked *aph(3'')-Ib* and *aph(6)-Id*. 2_GR_12
226 additionally had the absence of *ant(2'')-Ia*. Detection of *ant(2'')-Ia*, *aph(3'')-Ib*, *aph(6)-Id* was not present in
227 16_GR_13. 16_GR_13 further lacked *catB4* (phenicol) and *tet(A)* (tetracycline). Various phenicol resistance genes
228 were reported in the real-time emulation however, the incorrect gene was identified which may represent sequencing
229 errors accumulated over time and high similarity to other phenicol resistance genes. The tetracycline resistance gene,
230 *tet(A)*, was interestingly not reported in this emulation with 190 reads and the majority of reads exhibiting a high
231 mapping confidence (MAPQ = 60, equivalent to an error probability of 1×10^{-6}). This gene was only detected after 10
232 hours for 1_GR_13 and 2_GR_12 and this result may be influenced by the presence of only 1 copy of *tet(A)* encoded
233 on a low copy number megaplasmid (between 1 to 1.5, see Table 1).

234 ***Direct RNA sequencing resistance detection***

235 The time required to detect resistance was further interrogated using RNA sequencing. Rapid detection was possible
236 for several resistance genes via direct RNA sequencing (Figure 1). This was evident for genes conferring resistance
237 to aminoglycosides, β -lactams, sulphonamides and trimethoprim for all four isolates. Resistance towards these
238 antibiotics was commonly detected within 6 hours. In some instances, quinolone, rifampicin, fosfomycin and phenicol
239 resistance was detected. A similar result was obtained whether all reads or passed reads alone were analysed. The
240 most significant difference when analysing all reads was the detection of *fosA* in 1_GR_13 and *ARR-2* and *fosA* in
241 2_GR_12. Consistently absent from this analysis were genes attributed to macrolide (*mph(A)*) and tetracycline (*tet(A)*,
242 *tet(G)*) resistance, however, isolates exhibited high levels of resistance to tetracycline ($>64 \mu\text{g/ml}$) (Supplementary
243 Table S1). This may indicate that isolates require antibiotic exposure to enable transcription of these genes. Commonly
244 no new genes were detected after 12 hours of sequencing with the exception of *fosA* in 2_GR_12. Although *fosA* was
245 detected when including the failed reads, a low MAPQ score (≤ 10) was apparent. Similar to the DNA real-time
246 detection, several genes not found in the final assembly were identified (Supplementary Table S5). With the exception
247 of 20_GR_12, this included *aadB* and *strB* for all isolates. Additional genes detected included *ARR-7* in 1_GR_13,
248 *strA* in 2_GR_12 and for 16_GR_13, *blaCTX-M-64*, *blaOXA-436* and *strA*. Similar genes or gene families were
249 identified when comparing DNA and direct RNA sequencing. Overall, genes were detected more readily via DNA
250 rather than RNA sequencing, possibly due to a lack of RNA expression in the absence of the antibiotic to which

251 resistance is encoded. There were only a few instances where RNA sequencing detected resistance more quickly than
252 DNA sequencing: *aac(3')-IIa* in 16_GR_13 and *sul2* and *catA1* in 2_GR_12. Similar results were observed when
253 investigating data yield rather than time (Supplementary Figure S4).

254 *Levels of expression of resistance genes*

255 RNA sequencing accumulated over approximately 40 hours yielded between 0.9 and 1.7 million reads for these
256 isolates (Supplementary Figure S3). However, only a low proportion ($\leq 14.64\%$) of these reads passed base-calling
257 using Albacore 2.2.7 (Supplementary Table S6). Aligning passed reads alone to the final assembly, $\geq 98\%$ of reads
258 were mappable, however, $\leq 40\%$ of these had a MAPQ score ≥ 10 . When all reads (pass and fail) were aligned, the
259 majority were not mappable to the reference genome ($\geq 76.69\%$) and commonly exhibited a low MAPQ score (≤ 10).
260 In light of this, we decided to benchmark a number of different base-callers, including Albacore 2.2.7, Guppy 3.03 as
261 well as Chiron 0.5 which was trained in-house (Supplementary Table S6, Figure S5). Chiron base-called more reads
262 compared to Albacore 2.2.7 and Guppy 3.0.3, however, fewer reads aligned to the reference genome and had a slightly
263 lower identity rate. These results reflect the fact that base-calling algorithms have not yet been optimised for direct
264 RNA sequencing, and even less so for bacterial RNA sequencing. The poly(A) length was commonly found to be
265 approximately 400 to 700 bp across isolates (Supplementary Figure S6). Taking into consideration the Albacore 2.2.7
266 base-called reads, a proportion of these reads were found to map to rRNA including 1_GR_13 (18%), 2_GR_12 (37%),
267 16_GR_13 (24%) and 20_GR_12 (23%). Overall, at least 58% of genes (with at least 1 read mapping to the gene)
268 were identified to be expressed across isolates (1_GR_13 (68%), 2_GR_12 (58%), 16_GR_13 (75%) and 20_GR_12
269 (69%).

270 Amongst the four isolates, levels of expression for resistance genes on the chromosome (*blaSHV-11*, *fosA* and *oqxAB*)
271 were low (≤ 122 counts per million mapped reads (cpm)) (Figure 2). The remaining resistance genes were located on
272 plasmids. Resistance genes exhibiting high levels of expression (300 cpm) were apparent in 1_GR_13 (*blaTEM-1B*,
273 *blaVIM-27*, *sul1*, *aph(3')-Ia*), 2_GR_12 (*aac(6')-Ib*, *catA1*, *blaKPC-2*), 16_GR_13 (*aac(6')Ib-cr*, *aac(3)-IIa*, *blaCTX-*
274 *M-15*, *blaTEM-1B*, *blaOXA-48*) and 20_GR_12 (*blaKPC-2*, *aac(6')Ib*). Counts for *aac(6')-Ib* and *aac(6')-Ib-cr* in
275 2_GR_12 and 20_GR_12 were grouped. The gene *aac(6')-Ib-cr* is a shortened version of *aac(6')-Ib* and both were
276 identified in the same genome position, hence, only *aac(6')-Ib* is displayed in Figure 2. Relative expression did not
277 differ significantly when analysing passed reads alone or all reads. All highly expressed genes were detected within 6

278 hours as per the real-time detection emulation. As anticipated, low levels of expression were observed for fosfomycin
279 (*fosA*), tetracycline (*tet(A)*, *tet(B)*) and macrolide (*mph(A)*) resistance.

280 A subset of 11 resistance genes which represent resistance across various classes of antibiotics were investigated to
281 validate differential gene expression in these RNA extractions via qRT-PCR. These included resistance towards
282 aminoglycosides (*aac(6')Ib*, *strA*), β -lactams (*blaKPC-2*, *blaOXA-10*, *blaTEM-1*), phenicols (*cmlA1*), trimethoprim
283 (*dfrA14*), fosfomycin (*fosA*), quinolone (*oqxA*), sulphonamides (*sul2*) and tetracyclines (*tet(A)*). A similar trend was
284 observed between direct RNA sequencing and qRT-PCR results (Spearman's rank correlation coefficient: 0.83;
285 Pearson correlation: 0.86) (Figure 3). The highest expression of a resistance gene was observed for *blaKPC-2* although
286 only one copy was present in a lower copy number plasmid in 2_GR_12 and 20_GR_12 (Figure 2, Figure 3 and Table
287 1). Additionally, low levels of expression for *fosA* and *tet(A)* were apparent despite exhibiting resistance towards
288 fosfomycin and tetracycline (Figure 3, Supplementary Table S1). Direct RNA sequencing was unable to detect low
289 levels of expression whilst qRT-PCR could detect these genes (Figure 3).

290 Across the transcriptome, antibiotic resistance genes were identified to harbour high differential expression between
291 isolates (Figure 4). Virulence genes were comparable across these strains similar to all remaining or background genes.
292 The top differentially expressed genes were determined (Supplementary Figure S7) and several were associated with
293 polymyxin resistance pathways. Heightened expression was seen in polymyxin-resistant isolates 1_GR_13, 2_GR_12,
294 16_GR_13 in comparison to the single susceptible isolate in particular, genes associated with Ara4N synthesis. These
295 genes include 4-deoxy-4-formamido-L-arabinose-phosphoundecaprenol deformylase (*arnD*), UDP-4-amino-4-deoxy-
296 L-arabinose formyltransferase and UDP-4-amino-4-deoxy-L-arabinose-oxoglutarate aminotransferase.

297 ***Transcriptional biomarkers for polymyxin resistance***

298 Three of the isolates harboured resistance towards polymyxins via disruptions in *mgrB* which included 1_GR_13,
299 2_GR_12 and 16_GR_13. 1_GR_13. These isolates have an insertion sequence (IS) element, *ISKpn26*-like, at
300 nucleotide position 75 in the same orientation as *mgrB*. 2_GR_12 also contained an insertion at the same position,
301 however, in the opposite orientation and additional mutations in *phoP* (A95S) and *phoQ* (N253T). 16_GR_13
302 possessed an IS element, *ISIR*-like, 19 bp upstream of *mgrB*. Direct RNA sequencing revealed only low level
303 expression of *mgrB* in isolates (1_GR_13 (78.4 cpm), 2_GR_12 (16.3 cpm), 16_GR_13 (0 cpm), 20_GR_12 (2.3
304 cpm)). The expression levels of various genes associated with this pathway were verified via qRT-PCR which include
305 genes *phoP*, *phoQ*, *pmrA*, *pmrB*, *pmrC*, *pmrD*, *pmrE*, *pmrH* and *pmrK* (Figure 5). Direct RNA sequencing revealed a

306 slight increase in transcription of *phoPQ* (≥ 2 -fold) relative to the expression in 20_GR_12. A ≥ 13 -fold increase in
307 expression was observed for *pmrH* and ≥ 8 -fold elevation for *pmrK*. Similar trends for expression were also reported
308 using qRT-PCR (Figure 5B).

309

310 **Discussion**

311 XDR *K. pneumoniae* infections pose as a major threat to modern medicine. A rapid diagnostic would help to guide
312 appropriate treatment options [1, 6]. The MinION sequencing technology employed in this study has potential to detect
313 antibiotic resistance in a timely manner. Three of the four *K. pneumoniae* isolates examined in this study harboured
314 non-susceptibility to all antibiotics or antibiotic combinations assayed, and hence would be classified as PDR
315 according to published guidelines [48]. ONT sequencing was able to resolve both the assembly of plasmids harbouring
316 high levels of resistance (through DNA sequencing) and the expression from the resistome in the absence of antibiotic
317 treatment (via RNA sequencing).

318 The ability for ONT to sequence long fragments of DNA has significantly aided the assembly of bacterial genomes
319 and plasmids [16-18]. In this study, multiple megapasmids (≥ 100 kbp) were identified which were previously
320 unresolved via Illumina sequencing [28]. These harboured replicons IncA/C2 or a dual replicon, IncFIIK and IncFIB.
321 The IncA/C, IncF and IncN plasmids have been commonly associated with multidrug resistance [48]. Although several
322 plasmids in this study revealed similarity to previously reported isolates via NCBI, various sequences deviated. In
323 particular, the IncA/C2 plasmid exhibited multiple regions unique to these isolates. Several IncA/C2 megapasmids
324 have been previously described which harbour various resistance genes. However, the extent of resistance observed
325 in our study is extreme when compared to prior reports [50, 51]. Prior studies have shown the IncFIIK and IncFIB
326 replicons to localise on the same plasmid and also megapasmids with multidrug resistance [6]. The IncFIB_{pQil} plasmid
327 in this study contained various β -lactam resistance genes (*blaKPC-2*, *blaOXA-9*, *blaTEM-1A*) which has been
328 identified previously [52]. Similarly, *blaOXA-48* segregated with the IncL/M replicon [53,54], however, deviations in
329 this plasmid were identified.

330 The real-time analysis capability entailed in MinION sequencing has the potential to rapidly determine antibiotic
331 resistance profiles of pathogenic bacteria. Previously this device has been utilised to assemble bacterial genomes,
332 discern species and detect antibiotic resistance [12-15]. This study investigated the potential time required to discern
333 resistance via a real-time emulation as previously described [14]. The majority ($\geq 70\%$) of resistance genes were

334 detected via DNA sequencing within 2 hours. Several genes not identified in the final assembly were detected after 2
335 hours of sequencing. This may be attributed to the high similarity ($\geq 80\%$) amongst various genes, in particular, those
336 associated with aminoglycoside, β -lactam, rifampicin and phenicol resistance. Furthermore, the error rate associated
337 with ONT sequencing, and the accumulation of these errors over time, may result in the false annotation of these
338 genes. Nanopore DNA sequencing currently has an accuracy ranging from 80 to 90%, which limits its ability to detect
339 genomic variations [17]. Several resistance genes only differ by a few nucleotides which significantly impacts the
340 resistance phenotype and the antibiotics which can be utilised to treat these infections. Furthermore, direct RNA
341 sequencing has an average error rate of 12% [21]. Hence, it is essential for the technology to increase its accuracy in
342 order to correctly and rapidly diagnose antibiotic resistance.

343 Investigating the transcriptome of these isolates can potentially elucidate the correlation between genotype and the
344 subsequent resistant phenotype. One of the advantages of RNA sequencing is that it can identify conditions in which
345 a resistance gene is present but not expressed, potentially resulting in a susceptible phenotype. However, if expression
346 is only induced in the presence of an antibiotic, the absence of RNA transcripts may falsely suggest susceptibility.
347 Direct RNA sequencing revealed high levels of transcription from genes associated with aminoglycoside, β -lactam,
348 sulphonamide and trimethoprim resistance within 6 hours. The detection of quinolone, rifampicin, and phenicol
349 resistance correlated to the levels of transcription within samples. All isolates exhibited low levels of expression for
350 fosfomycin, macrolide and tetracycline resistance, despite exhibiting phenotypic resistance to fosfomycin and
351 tetracycline. Whether this transcription is due to prior exposure to these antibiotics in the clinic and the longevity of
352 this expression post exposure warrants further investigation. The changes in transcription levels in response to
353 antibiotic exposure also need to be assessed in future experiments. Furthermore, the time required to detect resistance
354 may be hindered by the slower translocation speed associated with direct RNA sequencing (70 bases/ second)
355 compared to DNA sequencing (450 bases/ second). Furthermore, insufficient rRNA depletion and low base-calling of
356 data could be impacting the detection of this low level expression.

357 Another variable to consider when evaluating differential expression is the operon or promoter which can further be
358 explored via cloning. In particular, the highest levels of expression were observed for *blaKPC-2* in 2_GR_12 and
359 20_GR_12. Alterations in the promoter region have previously been reported to influence high levels of expression
360 [55]. Furthermore, despite low levels of transcription for fosfomycin (*fosA*) and tetracycline (*tet(A)*, *tet(G)*),
361 phenotypically these isolates consistently retain resistance [28]. FosA, an enzyme involved in the degradation of

362 fosfomycin, is commonly encoded chromosomally in *K. pneumoniae* and a combination of expression and enzymatic
363 activity contributes to resistance [56]. Genes *tet(A)* and *tet(G)* encode efflux pumps which, in the absence of
364 tetracycline, are lowly expressed [57]. Detecting inducible resistance such as tetracycline resistance highlights one of
365 the advantages of investigating the transcriptome. Additionally, copy number of plasmids can further alter the levels
366 of expression detected for these resistance genes.

367 In this study we also investigated pathways attributed to polymyxin resistance. Three of these strains exhibited an IS
368 element upstream of within *mgrB*, the negative regulator of PhoPQ [29]. Elevated expression was apparent for *phoPQ*
369 and also the *pmrHFIJKLM* operon in our polymyxin-resistant isolates harbouring a disruption in *mgrB*. This has
370 previously been witnessed for other *K. pneumoniae* isolates harbouring *mgrB* disruptions and is a potential
371 transcriptional marker for polymyxin resistance [29, 44, 58, 59]. However, this study is limited to four isolates and
372 one mechanism associated with polymyxin resistance. Other pathways have previously been identified including the
373 role of other TCSs such as CrrAB [60]. The ability to use relative expression of key genes to detect polymyxin
374 resistance requires further validation, including an increased sample size of resistant and non-resistant isolates.
375 Furthermore, additional functional experiments such as complementation assays would be required in order to validate
376 the contribution of a certain mutation to the transcriptome and subsequent resistance.

377

378 **Conclusions**

379 This study has utilised MinION sequencing to assemble four XDR *K. pneumoniae* genomes and has revealed several
380 unique plasmids harbouring multidrug resistance. The vast majority of this resistance was detectable within 2 hours
381 of sequencing, though a number of resistance genes were identified that were not present in the final assembly.
382 Exploiting this analysis in real-time would allow for a rapid diagnostic, however, the presence of a resistance gene
383 does not necessarily indicate resistance is conferred and requires additional phenotypic characterisation. This research
384 also established a methodology and analysis for bacterial direct RNA sequencing. The differential expression of
385 resistance genes were successfully detected via this technology and can be exploited for bacterial transcriptomics.
386 Once base-calling algorithms have been optimised, this could allow for a whole transcriptome interrogation of full
387 length transcripts regulated by operons, where more than one gene is co-expressed in a transcript, and the evaluation
388 of the poorly characterised epitranscriptome. This research established a methodology and analysis for bacterial direct
389 RNA sequencing. The differential expression of resistance genes were successfully detected via this technology and

390 can be exploited for bacterial transcriptomics. Overall, this study has begun to unravel the association between
391 genotype, transcription and subsequent resistant phenotype in these XDR/ PDR *K. pneumoniae* clinical isolates,
392 establishing the groundwork for developing a diagnostic that can rapidly determine bacterial resistance profiles.

393

394 **Availability of supporting data**

395 The datasets supporting the results presented here are available in the National Center for Biotechnology Information
396 repository BioProject PRJNA307517 (www.ncbi.nlm.nih.gov/bioproject/PRJNA307517). ONT DNA sequencing
397 data has been deposited on the Sequence Read Archive (www.ncbi.nlm.nih.gov/sra/) under study SRP133040.
398 Accession numbers are as follows: 1_GR_13 (SRR6747887), 2_GR_12 (SRR6747886), 16_GR_13 (SRR6747885)
399 and 20_GR_12 (SRR6747884). ONT direct RNA sequencing data (pass and fail reads) have been deposited on the
400 Sequence Read Archive (www.ncbi.nlm.nih.gov/sra/) under study SRP133040. Accession numbers are as follows:
401 1_GR_13 (SRR7719054), 2_GR_12 (SRR7719055), 16_GR_13 (SRR7719052) and 20_GR_12 (SRR7719053).

402 **Abbreviations**

403 Ara4N: 4-amino-4-deoxy-L-arabinose; caMHB: cation-adjusted Muller Hinton Broth; CLSI: Clinical & Laboratory
404 Standards Institute; EUCAST: The European Committee on Antimicrobial Susceptibility Testing; HMW: High
405 molecular weight; LB: Lysogeny broth; MAPQ: Mapping quality; MIC: Minimum inhibitory concentration; NCBI:
406 National Center for Biotechnology Information; ONT: Oxford Nanopore Technologies; PDR: Pandrug-resistant;
407 RAST: Rapid Annotation using Subsystem Technology; rRNA: Ribosomal RNA; XDR: Extensively drug-resistant.

408 **Competing Interests**

409 The authors declare that there are no competing interests.

410 **Funding**

411 LJMC is an NHMRC career development Fellow APP1103384. MAC is an NHMRC Principal Research Fellow
412 (APP1059354) and currently holds a fractional Professorial Research Fellow appointment at the University of
413 Queensland with his remaining time as CEO of Inflazome Ltd. a company headquartered in Dublin, Ireland that is
414 developing drugs to address clinical unmet needs in inflammatory disease by targeting the inflammasome. MEP is an
415 Australian Postgraduate Award scholar. MATB is supported in part by a Wellcome Trust Strategic Award
416 104797/Z/14/Z. This work was supported by the Institute for Molecular Bioscience Centre for Superbug Solutions
417 (610246).

418 **Author Contributions**

419 MEP, LJMC, MATB and MAC conceived this study. MEP, SHN and HT performed the sequencing analysis.
420 Laboratory work was carried out by MEP and TPSD. MEP wrote the paper with input from all authors.

421

422 **Acknowledgements**

423 We would like to acknowledge Dr Ilias Karaiskos and Dr Helen Giamarellou for providing the bacterial strains in this
424 study. We also acknowledge Dr Evangelos Bellos for his guidance on the RNA sequencing analysis and Dr Devika
425 Ganesamoorthy for the initial advice on the direct RNA sequencing library preparation. We would like to acknowledge
426 Intawat Nookaew for providing yeast direct RNA sequence data.

427

428 **References**

- 429 1. Martin RM, Bachman MA. Colonization, Infection, and the Accessory Genome of *Klebsiella pneumoniae*.
430 *Front Cell Infect Microbiol*. 2018;8:4.
- 431 2. Magill SS, Edwards JR, Bamberg W, et al. Multistate point-prevalence survey of health care-associated
432 infections *N Engl J Med*. 2014;370:1198-208.
- 433 3. Kalanuria AA, Ziai W, Mirski, M. Ventilator-associated pneumonia in the ICU. *Crit Care*. 2014;18:208.
- 434 4. Talha KA, Hasan Z, Selina F, et al. Organisms associated with ventilator associated pneumonia in intensive
435 care unit. *Mymensingh Med J*. 2009;18:S93-7.
- 436 5. Podschun R, Ullmann U. *Klebsiella spp* as nosocomial pathogens: epidemiology, taxonomy, typing methods,
437 and pathogenicity factors. *Clin Microbiol Rev*. 1998;11:589-603.
- 438 6. Navon-Venezia S, Kondratyeva K, Carattoli A *Klebsiella pneumoniae*: a major worldwide source and shuttle
439 for antibiotic resistance. *FEMS Microbiol Rev*. 2017;41:252-75.
- 440 7. Karaiskos I, Giamarellou H. Multidrug-resistant and extensively drug-resistant Gram-negative pathogens:
441 current and emerging therapeutic approaches. *Expert Opin Pharmacother*. 2014;15:1351-70.
- 442 8. Chen L, Todd R, Kiehlbauch J, et al. Notes from the Field: Pan-Resistant New Delhi Metallo-Beta-
443 Lactamase-Producing *Klebsiella pneumoniae* - Washoe County, Nevada, 2016 *MMWR Morb Mortal Wkly*
444 *Rep*. 2017;66:33.

445 9 Zowawi HM, Forde BM, Alfaresi M, et al. Stepwise evolution of pandrug-resistance in *Klebsiella*
446 *pneumoniae*. *Sci Rep*. 2015;5:15082.

447 10 Sommer MOA, Munck C, Toft-Kehler RV, et al. Prediction of antibiotic resistance: time for a new preclinical
448 paradigm? *Nat Rev Microbiol*. 2017;15:689-96.

449 11 Gardy JL, Loman NJ. Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat*
450 *Rev Genet*. 2018;19:9-20.

451 12 Lemon JK, Khil PP, Frank KM, et al. Rapid Nanopore Sequencing of Plasmids and Resistance Gene
452 Detection in Clinical Isolates. *J Clin Microbiol*. 2017;55:3530-43.

453 13 Votintseva AA, Bradley P, Pankhurst L, et al. Same-Day Diagnostic and Surveillance Data for Tuberculosis
454 via Whole-Genome Sequencing of Direct Respiratory Samples. *J Clin Microbiol*. 2017;55:1285-98.

455 14 Cao MD, Ganesamoorthy D, Elliott AG, et al. Streaming algorithms for identification of pathogens and
456 antibiotic resistance potential from real-time MinION™ sequencing. *Gigascience*. 2016;5:32.

457 15 Quick J, Ashton P, Calus S, et al. Rapid draft sequencing and real-time nanopore sequencing in a hospital
458 outbreak of *Salmonella*. *Genome Biol*. 2015;16:114.

459 16 Wick RR, Judd LM, Gorrie CL, et al. Completing bacterial genome assemblies with multiplex MinION
460 sequencing. *Microb Genom*. 2017;3:e000132.

461 17 Li R, Xie M, Dong N, et al. Efficient generation of complete sequences of MDR-encoding plasmids by rapid
462 assembly of MinION barcoding sequencing data. *Gigascience*. 2018;7:1-9.

463 18 George S, Pankhurst L, Hubbard A, et al. Resolving plasmid structures in Enterobacteriaceae using the
464 MinION nanopore sequencer: assessment of MinION and MinION/Illumina hybrid data assembly
465 approaches. *Microb Genom*. 2017;3:e000118.

466 19 Garalde, DR, Snell, EA, Jachimowicz, D, et al. Highly parallel direct RNA sequencing on an array of
467 nanopores. *Nat Methods*. 2018;15:201-6.

468 20 Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet*. 2011;12:
469 87-98.

470 21 Jenjaroenpun P, Wongsurawat T, Pereira R, et al. Complete genomic and transcriptional landscape analysis
471 using third-generation sequencing: a case study of *Saccharomyces cerevisiae* CENPK113-7D. *Nucleic Acids*
472 *Res*. 2018;46:e38.

473 22 Workman RE, Tang A, Tang PS, et al. Nanopore native RNA sequencing of a human poly(A) transcriptome.
474 *bioRxiv*. 2018;459529.

475 23 Moldovan N, Tombacz D, Szucs A, et al. Third-generation Sequencing Reveals Extensive Polycistronism
476 and Transcriptional Overlapping in a *Baculovirus*. *Sci Rep*. 2018;8:8604.

477 24 Keller MW, Rambo-Martin BL, Wilson MM, et al. Direct RNA Sequencing of the Coding Complete
478 Influenza A Virus Genome. *Sci Rep*. 2018;8:14408.

479 25 Depledge DP, Puthankalam SK, Sadaoka T, et al. Native RNA sequencing on nanopore arrays redefines the
480 transcriptional complexity of a viral pathogen. *bioRxiv*. 2018;373522.

481 26 Smith AM, Jain M, Mulroney L, et al. Reading canonical and modified nucleotides in 16S ribosomal RNA
482 using nanopore direct RNA sequencing. *bioRxiv*. 2017;132274.

483 27 Sorek R, Cossart P. Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity.
484 *Nat Rev Genet*. 2010;11:9-16.

485 28 Pitt ME, Elliott AG, Cao, MD, et al. Multifactorial chromosomal variants regulate polymyxin resistance in
486 extensively drug-resistant *Klebsiella pneumoniae*. *Microb Genom*. 2018;mgen1090000158.

487 29 Olaitan AO, Morand S, Rolain JM. Mechanisms of polymyxin resistance: acquired and intrinsic resistance
488 in bacteria. *Front Microbiol*. 2014;5:643.

489 30 Teng H, Cao MD, Hall MB, et al. Chiron: translating nanopore raw signal directly into nucleotide sequence
490 using deep learning. *GigaScience*. 2018;7:10.1093/gigascience/giy037.

491 31 Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*.
492 2013;13033997.

493 32 Zankari E, Hasman H, Cosentino S, et al. Identification of acquired antimicrobial resistance genes. *J*
494 *Antimicrob Chemother*. 2012;67:2640-4.

495 33 Lassmann T, Frings O, Sonnhammer EL. Kalign2: high-performance multiple alignment of protein and
496 nucleotide sequences allowing external features. *Nucleic Acids Res*. 2009;37:858-65.

497 34 Allison L, Wallace CS, Yee CN. When is a string like a string? In: Artificial Intelligence and Mathematics.
498 *Ft Lauderdale FL*. 1990.

499 35 Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to
500 single-cell sequencing. *J Comput Biol*. 2012;19:455-77.

501 36 Cao MD, Nguyen SH, Ganesamoorthy D, et al. Scaffolding and completing genome assemblies in real-time
502 with nanopore sequencing. *Nat Commun.* 2017;8:14515.

503 37 Wick RR, Judd LM, Gorrie CL, et al. Unicycler: Resolving bacterial genome assemblies from short and long
504 sequencing reads. *PLoS Comput Biol.* 2017;13:e1005595.

505 38 Koren S, Walenz BP, Berlin K, et al. Canu: scalable and accurate long-read assembly via adaptive k-mer
506 weighting and repeat separation. *Genome Res.* 2017;27:722-36.

507 39 Li H. Minimap and miniiasm: fast mapping and *de novo* assembly for noisy long sequences. *Bioinformatics.*
508 2016;32:2103-10.

509 40 Vaser R, Sovic I, Nagarajan N, et al. Fast and accurate *de novo* genome assembly from long uncorrected
510 reads. *Genome Res.* 2017;27:737-46.

511 41 Darling AE, Tritt A, Eisen JA, et al. Mauve assembly metrics. *Bioinformatics.* 2011;27:2756-7.

512 42 Aziz RK, Bartels D, Best AA, et al. The RAST Server: rapid annotations using subsystems technology. *BMC*
513 *Genomics.* 2008;9:75.

514 43 Carattoli A, Zankari E, Garcia-Fernandez A, et al. *In silico* detection and typing of plasmids using
515 PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother.* 2014;58:3895-903.

516 44 Quinlan AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics.*
517 2014;47:11.12.1-34.

518 45 Cannatelli A, D'Andrea MM, Giani T, et al. *In vivo* emergence of colistin resistance in *Klebsiella pneumoniae*
519 producing KPC-type carbapenemases mediated by insertional inactivation of the PhoQ/PhoP *mgrB* regulator.
520 *Antimicrob Agents Chemother.* 2013;57:5521-6.

521 46 Robinson JT, Thorvaldsdottir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotechnol.* 2011;29;
522 24-6.

523 47 Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and
524 the 2⁻(-Delta Delta C(T)) Method. *Methods.* 2001;25:402-8.

525 48 Magiorakos AP, Srinivasan A, Carey RB, et al. Multidrug-resistant, extensively drug-resistant and pandrug-
526 resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance.
527 *Clin Microbiol Infect.* 2012;18,268–81.

528 49 Carattoli A. Resistance plasmid families in Enterobacteriaceae. *Antimicrob Agents Chemother.* 2009;53:
529 2227-38.

530 50 Desmet S, Nepal S, van Dijl JM, et al. Antibiotic Resistance Plasmids Cointegrated into a Megaplasmid
531 Harboring the *blaOXA-427* Carbapenemase Gene. *Antimicrob Agents Chemother.* 2018;62:e01448-17.

532 51 Papagiannitsis CC, Dolejska M, Izdebski R, et al. Characterisation of IncA/C2 plasmids carrying an In416-
533 like integron with the *blaVIM-19* gene from *Klebsiella pneumoniae* ST383 of Greek origin. *Int J Antimicrob*
534 *Agents.* 2016;47:158-62.

535 52 Chen L, Chavda KD, Melano RG, et al. Comparative genomic analysis of KPC-encoding pKpQIL-like
536 plasmids and their distribution in New Jersey and New York Hospitals. *Antimicrob Agents Chemother.*
537 2014;58:2871-7.

538 53 Poirel L, Bonnin RA, Nordmann P. Genetic features of the widespread plasmid coding for the carbapenemase
539 OXA-48. *Antimicrob Agents Chemother.* 2012;56:559-62.

540 54 Potron A, Poirel L, Nordmann P. Derepressed transfer properties leading to the efficient spread of the plasmid
541 encoding carbapenemase OXA-48. *Antimicrob Agents Chemother.* 2014;58:467-71.

542 55 Cheruvanky A, Stoesser N, Sheppard AE, et al. Enhanced *Klebsiella pneumoniae* Carbapenemase Expression
543 from a Novel Tn4401 Deletion. *Antimicrob Agents Chemother.* 2017;61:e00025-17.

544 56 Klontz EH, Tomich AD, Gunther S, et al. Structure and Dynamics of FosA-Mediated Fosfomycin Resistance
545 in *Klebsiella pneumoniae* and *Escherichia coli*. *Antimicrob Agents Chemother.* 2017;61:e01572-17.

546 57 Saenger W, Orth P, Kisker C, et al. The Tetracycline Repressor-A Paradigm for a Biological Switch. *Angew*
547 *Chem Int Ed Engl.* 2000;39:2042-52.

548 58 Cheng YH, Lin TL, Pan YJ, et al. Colistin resistance mechanisms in *Klebsiella pneumoniae* strains from
549 Taiwan. *Antimicrob Agents Chemother.* 2015;59:2909-13.

550 59 Haeili M, Javani A, Moradi J, et al. MgrB Alterations Mediate Colistin Resistance in *Klebsiella pneumoniae*
551 Isolates from Iran. *Front Microbiol.* 2017;8:2470.

552 60 Baron S, Hadjadj L, Rolain JM, et al. Molecular mechanisms of polymyxin resistance: knowns and
553 unknowns. *Int J Antimicrob Agents.* 2017;48:583-91.

554
555

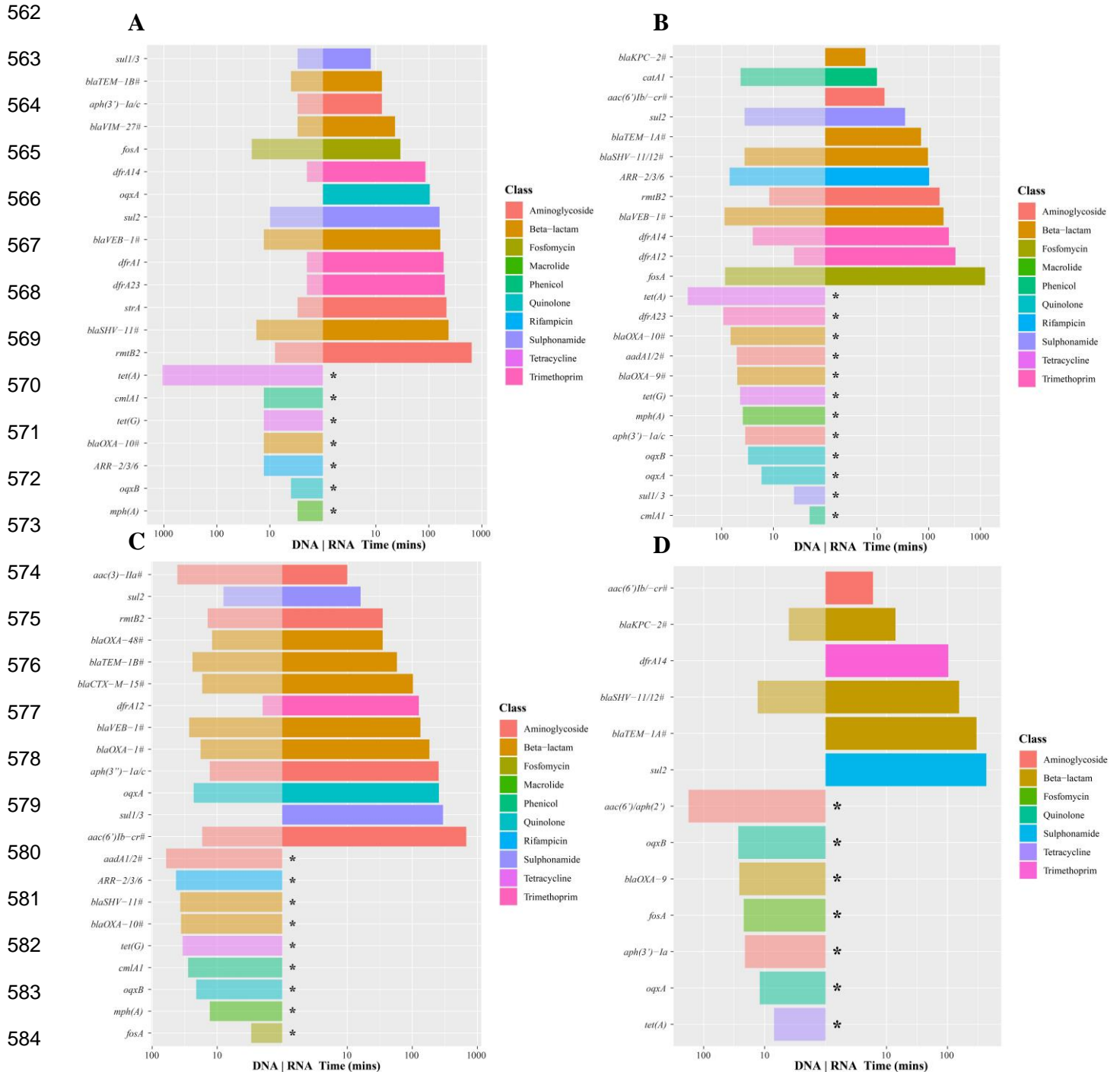
556 **Table and Figure Legends**557 **Table 1:** Final assembly of XDR *K pneumoniae* isolates and location of antibiotic resistance genes

Isolate	ST	Contig	Length (bp)	Coverage	Contig ID*	Resistance Genes**
1_GR_13	147	1	5181675	1	C	<i>blaSHV-11, fosA, oqxA, oqxB</i>
		2	192771	1.95	P: IncA/C2	<i>aadA1, ant(2'')-Ia, aph(6)-Id, ARR-2, blaOXA-10, blaTEM-1B, blaVEB-1, cmlA1, dfrA14, dfrA23, rmtB, strA, sul1, sul2, tet(A), tet(G)</i>
		3	168873	2	P: IncFIB _{pKpn3} , IncFII _{pKP91}	<i>aadA24, aph(3')-Ia, aph(6)-Id, dfrA1, dfrA14, strA</i>
		4	108879	1.53	P: IncFIB _{pKPHS1}	-
		5	55018	14.10	-	-
		6	53495	2.36	P: IncR, IncN	<i>aadA24, aph(3')-Ia, aph(6)-Id, blaVIM-27, dfrA1, mph(A), strA, sul1</i>
2_GR_12	258	1	5466424	1	C	<i>blaSHV-11, fosA, oqxA, oqxB</i>
		2	197872	1.3	P: IncFIB _{pKpn3} , IncFIIK	<i>aadA2, aph(3')-Ia, catA1, dfrA12, mph(A), sul1</i>
		3	175636	1.49	P: IncA/C2	<i>aadA1, ant(2'')-Ia, aph(3'')-Ib, aph(6)-Id, ARR-2, blaOXA-10, blaTEM-1A, blaVEB-1, cmlA1, dfrA14, dfrA23, rmtB, sul1, sul2, tet(A), tet(G)</i>
		4	95481	1.61	P: IncFIB _{pQil}	<i>blaKPC-2, blaOXA-9, blaTEM-1A</i>
		5	43380	1.91	P: IncX3	<i>blaSHV-12</i>
		6	13841	4	P: ColRNAI	<i>aac(6')-Ib, aac(6')Ib-cr</i>
16_GR_13	11	1	5426917	1	C	<i>blaSHV-11, fosA, oqxA, oqxB</i>
		2	187670	0.88	P: IncFIB _{pKpn3} ; IncFIIK	<i>aac(3)-IIa, aac(6')Ib-cr, aadA2, aph(3')-Ia, blaCTX-M-15, blaOXA-1, catB4, dfrA12, mph(A), sul1</i>
		3	155161	0.99	P: IncA/ C2	<i>aadA1, ant(2'')-Ia, aph(3'')-Ib, aph(6)-Id, ARR-2, blaOXA-10, blaTEM-1B, blaVEB-1, cmlA1, rmtB, sul1, sul2, tet(A), tet(G)</i>
		4	63589	1.49	P: IncL/ M _{pOXA-48}	<i>blaOXA-48</i>
		5	5234	188.49	-	-
		6	4940	97.77	P: ColRNAI	-
20_GR_12	258	1	5395894	1	C	<i>blaSHV-11, fosA, oqxA, oqxB</i>
		2	170467	1.77	P: IncFIB _{pKpn3} ; IncFIIK	<i>aph(3')-Ia, blaKPC-2, blaOXA-9, blaTEM-1A</i>
		3	50979	1.42	P: IncN	<i>aph(3'')-Ib, aph(6)-Id, blaTEM-1A, dfrA14, sul2, tet(A)</i>
		4	43380	1.78	P: IncX3	<i>blaSHV-12</i>
		5	13841	10.82	P: ColRNAI	<i>aac(6')-Ib, aac(6')Ib-cr</i>

558 *Contig ID represents chromosome (C) or plasmid (P): replicon determined via PlasmidFinder 1.3.

559 **Resistance genes identified using ResFinder 3.0 (≥90% sequence similarity, ≥60% minimum length) and displayed
560 in alphabetical order. **Bold** indicates a circular contig.

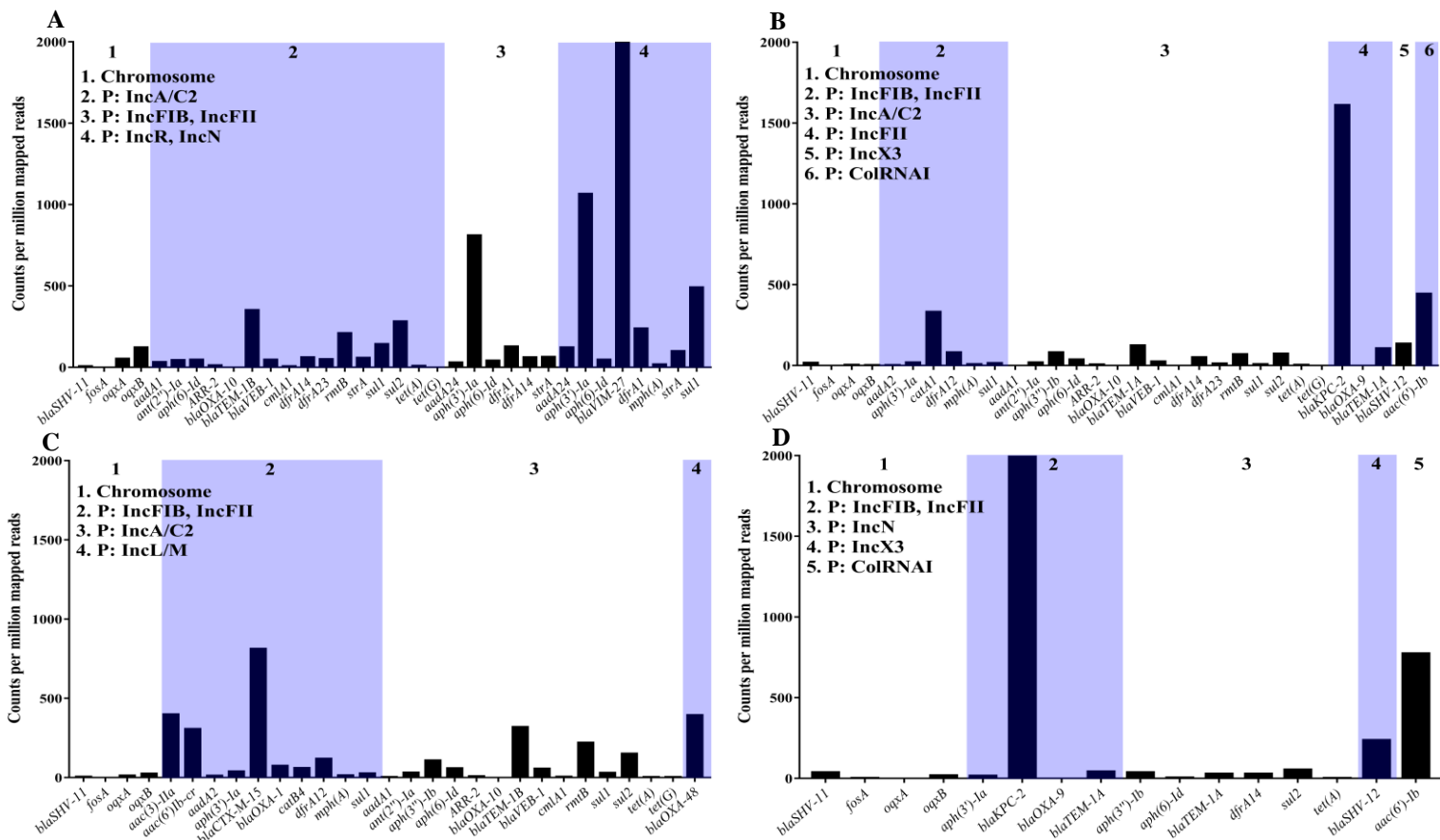
561



585 **Figure 1:** Time required to detect antibiotic resistance genes via the real-time emulation analysis using MinION DNA
 586 and direct RNA sequencing (**A**) 1_GR_13, (**B**) 2_GR_12, (**C**) 16_GR_13 and (**D**) 20_GR_12. Legend colours identify
 587 the class of antibiotic to which the gene confers resistance, / on y-axis indicates reads detected more than one resistance
 588 gene and # is a family of genes detected (>3). An asterisk (*) indicates the inability for direct RNA sequencing to
 589 detect this gene Albacore 2.2.7. base-called sequences were used and all reads (pass and fail) were included in this
 590 analysis.

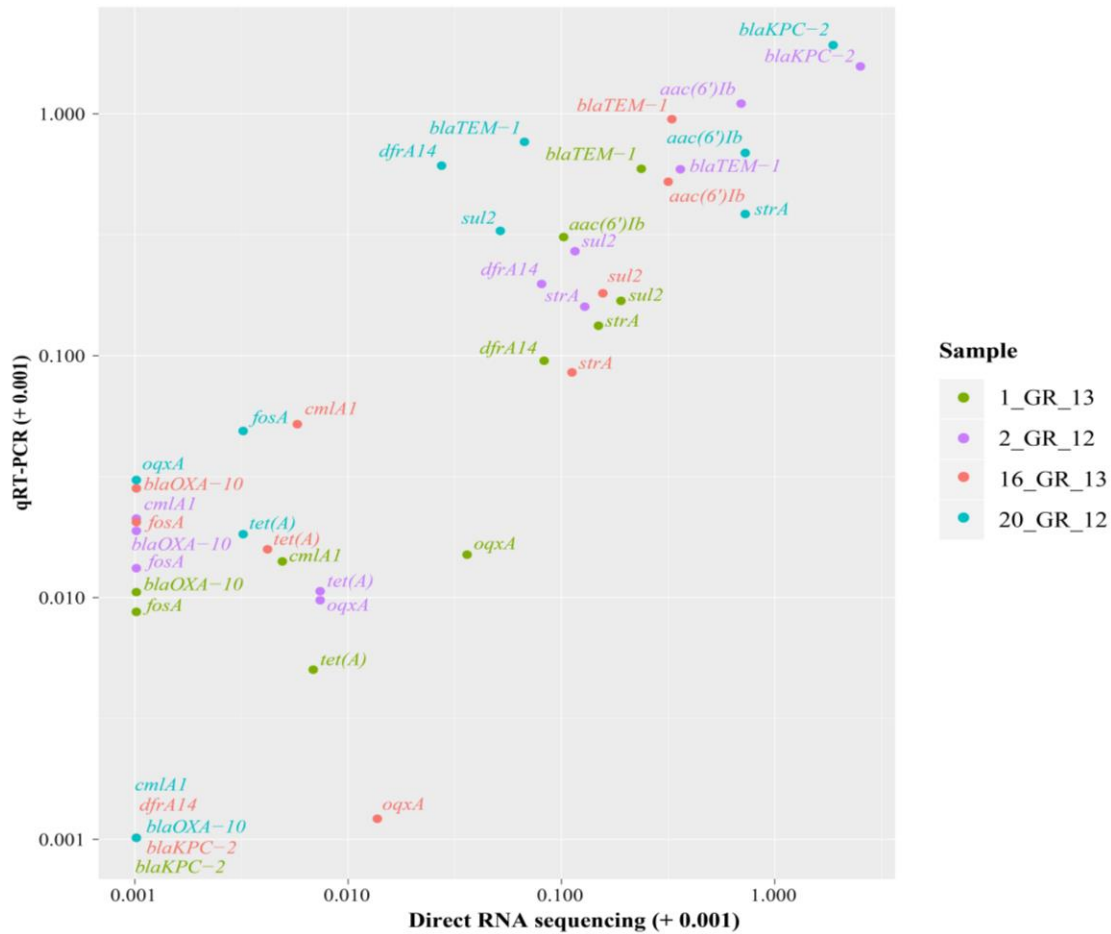
591

592
593
594



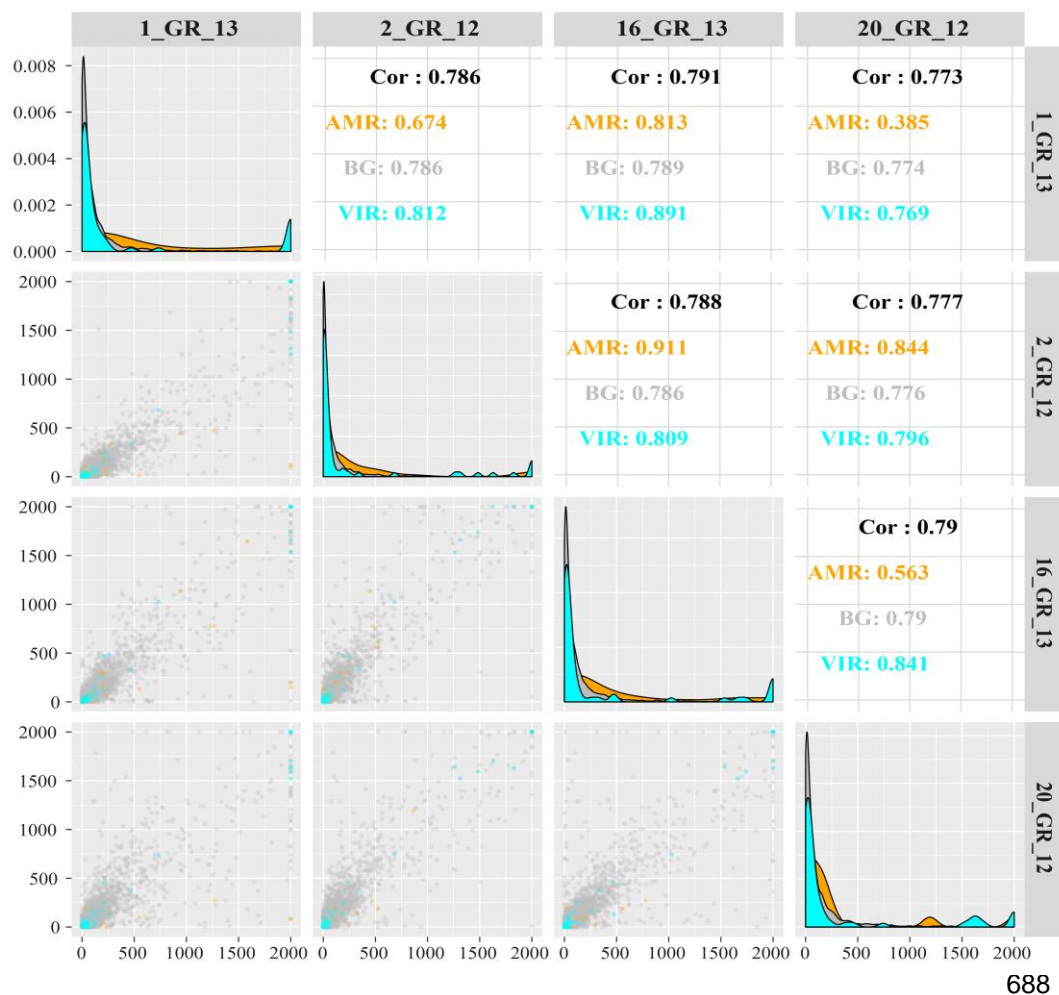
615 **Figure 2:** Relative expression of resistance genes normalised to housekeeping gene, *rpsL*, via direct RNA sequencing.
616 Strains investigated include: **(A)** 1_GR_13, **(B)** 2_GR_12, **(C)** 16_GR_13 and **(D)** 20_GR_12. X-axis depicts the
617 resistance genes and are grouped based on the location in the genome where P indicates a plasmid followed by replicon
618 identity. Albacore 2.2.7 base-called pass and fail reads were used for analysis.

619
620
621
622
623
624
625
626
627
628



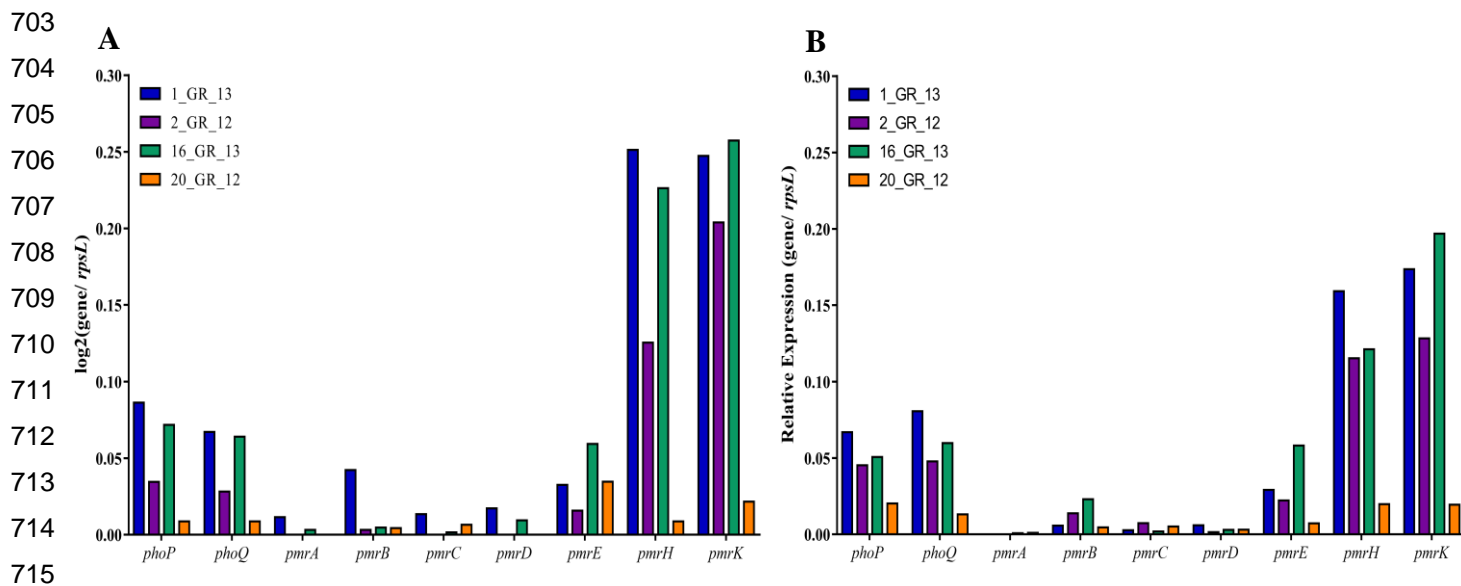
651 **Figure 3:** Correlation between resistance genes detected via direct RNA sequencing and validated using qRT-PCR.
 652 Relative expression was calculated via normalizing to the housekeeping gene, *rpsL* for both direct RNA sequencing
 653 ($\log_2(\text{gene}/rpsL)$) and qRT-PCR ($2^{-\Delta\Delta CT}$). Due to high similarity between certain genes, several primers recognise more
 654 than one gene. These include *aac(6')Ib*: *aac(6')Ib-cr*, *aadA24*; *strA*: *aph(3'')-Ib* and *blaTEM-1*: *blaTEM-1A*, *blaTEM-*
 655 *1B*.

656
 657
 658
 659
 660
 661
 662
 663
 664
 665

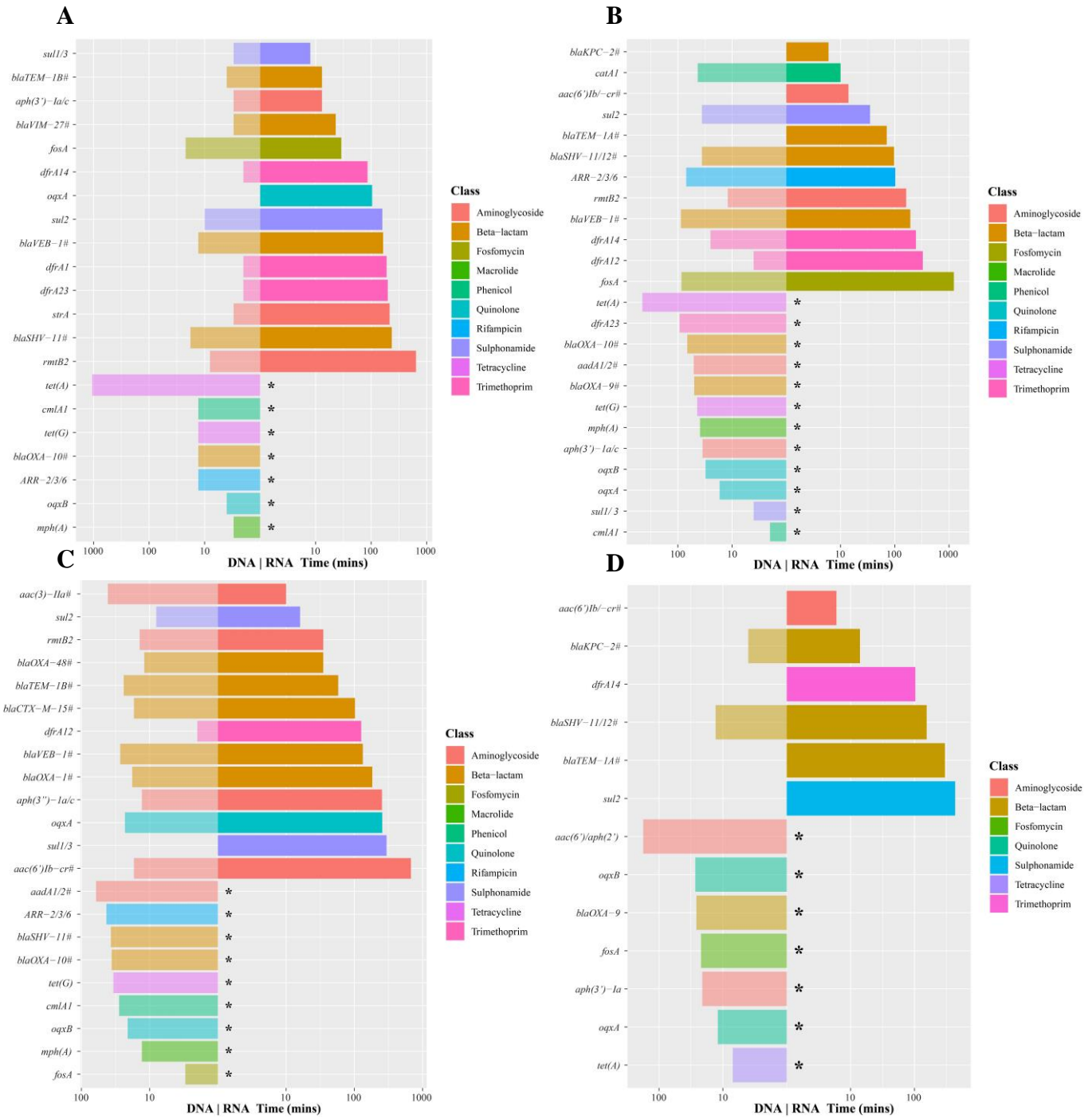


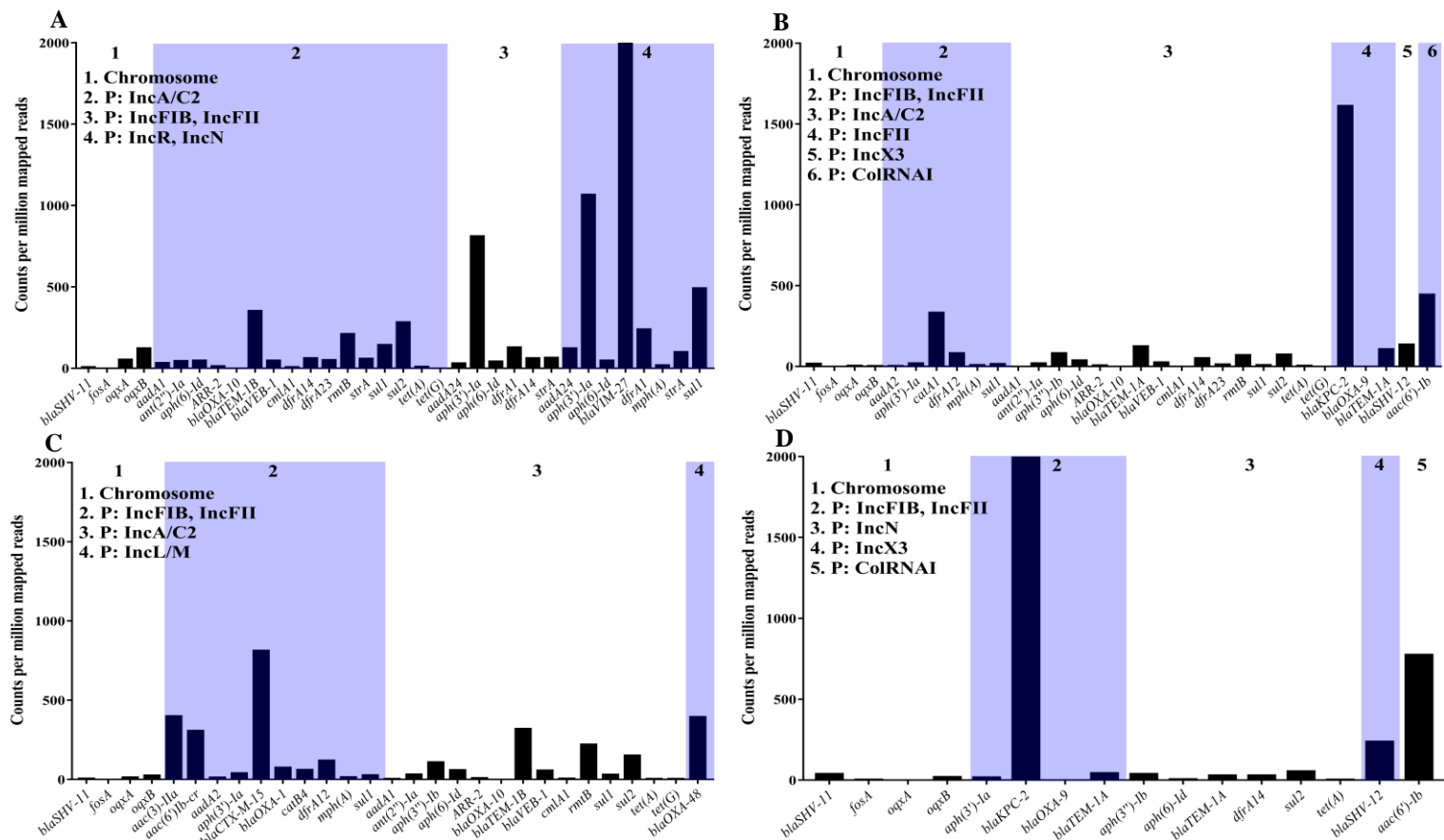
689 **Figure 4:** Correlation between the four XDR *K pneumoniae* isolates for gene expression via direct RNA sequencing.
 690 Top panels display spearman correlation coefficients. The diagonal panel shows the density of gene expression levels
 691 in counts per million mapped reads (cpm) for each sample. Colours indicate categorization of gene: antimicrobial
 692 resistance genes (AMR) as per ResFinder 3.0, virulence genes (VIR) determined via RAST and all other genes or
 693 background genes (BG) are displayed. Cpm was capped at 2000.

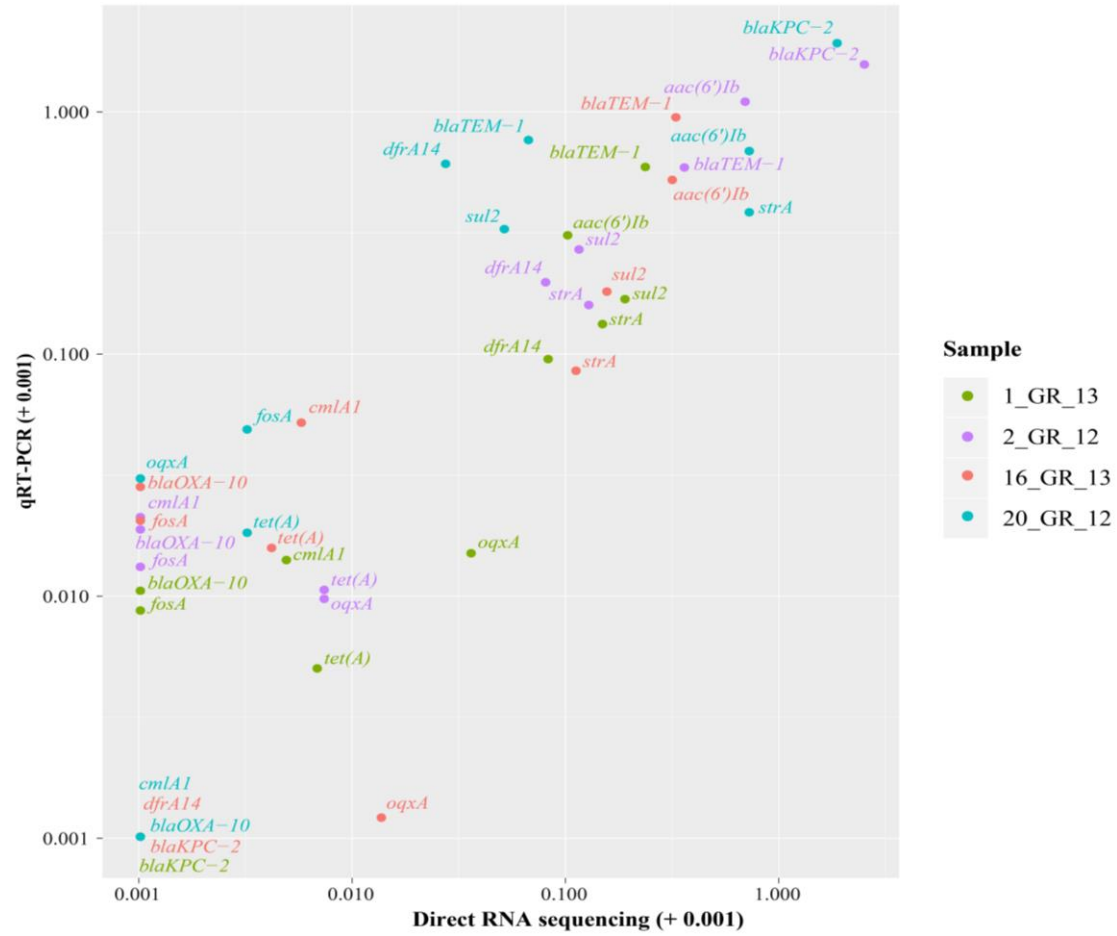
694
 695
 696
 697
 698
 699
 700
 701
 702

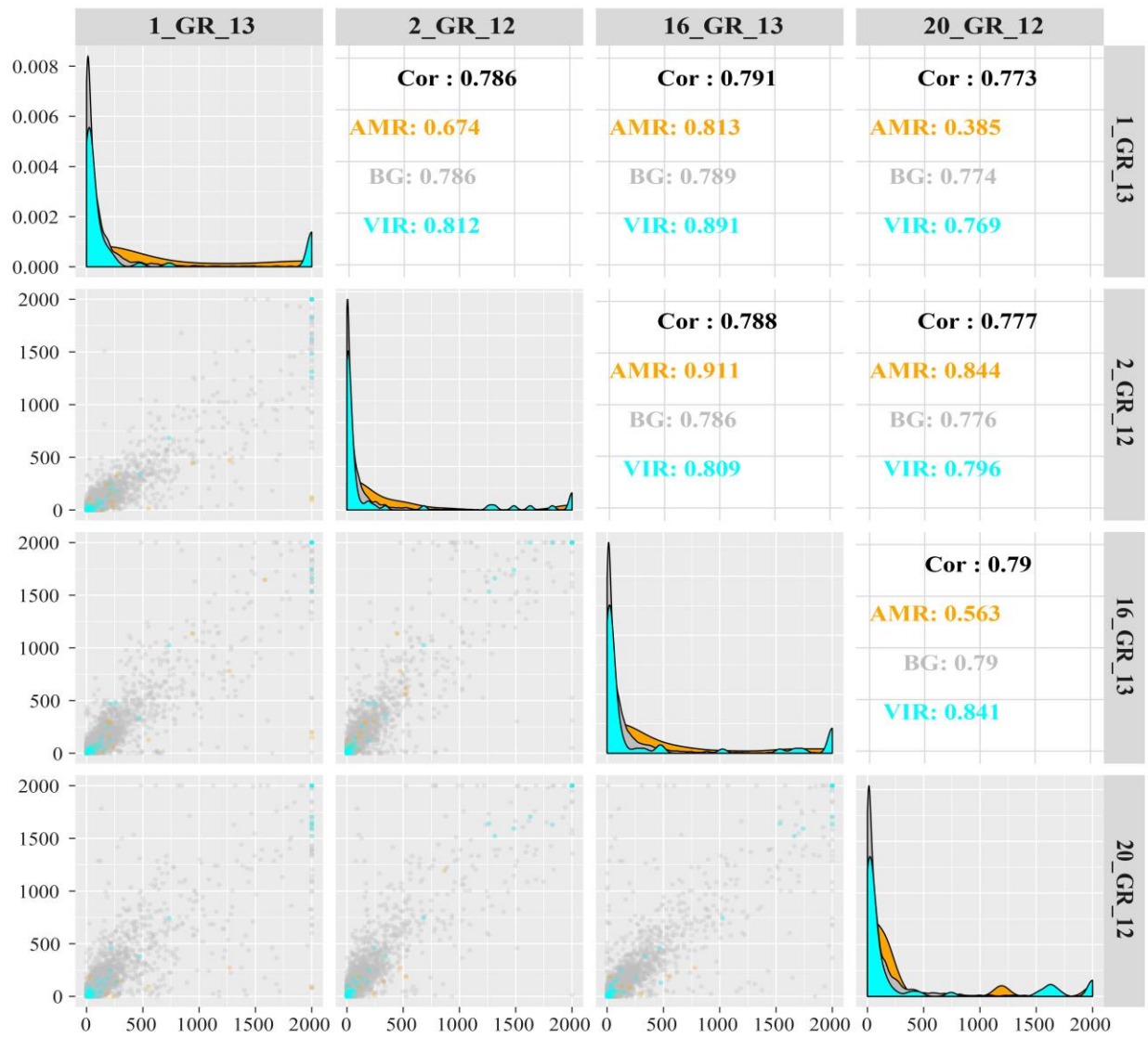


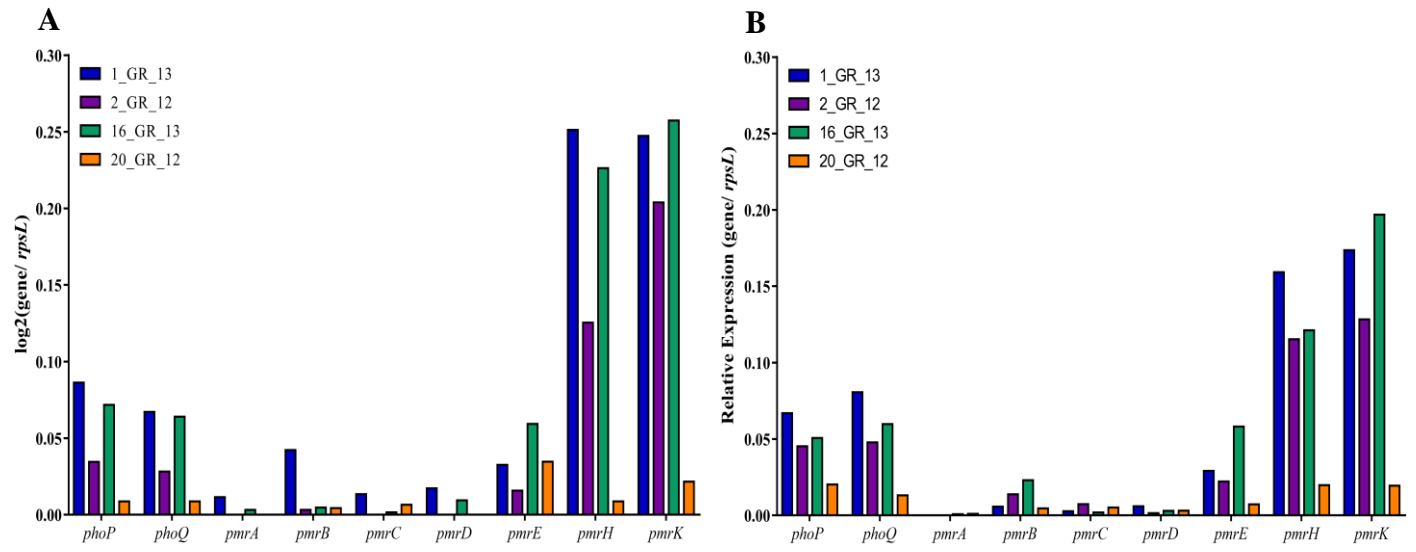
716 **Figure 5:** Expression of genes associated with the polymyxin resistance pathway. Comparison between (A) direct
 717 RNA sequencing and (B) qRT-PCR. Isolates harbouring resistance to polymyxins (MIC: >2 µg/mL) include
 718 1_GR_13, 2_GR_12 and 16_GR_13.













Click here to access/download
Supplementary Material
#SI_GS_Evaluating_XDRKP#.docx



3 June 2019

Dear Dr Scott Edmunds,

Re: Article submission to *GigaScience*

Please consider the enclosed manuscript entitled “**Evaluating the Genome and Resistome of Extensively Drug-Resistant *Klebsiella pneumoniae* using Native DNA and RNA Nanopore sequencing**” for publication in *GigaScience*.

Extensively drug-resistant bacteria, including *Klebsiella pneumoniae*, are an increasing cause of mortality globally. Rapid identification of the drug resistance profile of these bacteria is of critical importance in treating infections, so that the appropriate antibiotic can be given at the right time. Real-time nanopore sequencing of bacterial DNA has been investigated as a promising means to rapidly identify the drug resistance profile of these organisms. However, the potential for real-time direct RNA sequencing to inform identification of antibiotic resistance has yet to be explored.

In this manuscript, we develop the first ever protocol for sequencing native bacterial RNA using the Oxford Nanopore Technologies MinION sequencer, and use this approach to investigate the degree to which real-time RNA sequencing can be used to monitor the presence and expression levels of acquired antibiotic resistance genes. We show direct RNA sequencing can identify resistance to a number of antibiotics, including carbapenems and polymyxins. We also show that direct RNA sequencing provides an accurate estimate of gene expression levels, by validation using qRT-PCR.

We also use our data to evaluate the accuracy of Neural network base-calling on bacterial direct RNA, benchmarking a 3 proprietary base-callers as well as one which we trained in-house (Chiron), building on work previously published in *Gigascience*.

We confirm that no material herein has been published and that the manuscript is not under consideration for any other journal. All authors have approved the manuscript and agree with its submission.

Yours sincerely,



Associate Professor Lachlan Coin
Group Leader