Clinical validation of a novel metagenomic nanopore sequencing method for detecting viral respiratory pathogens

Nicholas D. Sanderson*[#1,2], Kate Dingle[#1,2], Katie M.V. Hopkins[1,2], Ali Vaughan[1,2], Matthew Colpus[2], Melody Parker[1,2], Elisabeth Vilde Dietz[2], Jessica Gentry[3], Anita Justice[3], Sarah Oakley[3], Lucinda Barrett[3], T. Phuong Quan[1,2,4], Nicole Stoesser[1,2,3,4], David W. Eyre[1,2,4], Philip Bejon[1,2,3], Ann Sarah Walker†[1,2,4], Bernadette C. Young†[1,2,3]

[#] contribution considered equal

† contribution considered equal

Affiliations

[1] NIHR Oxford Biomedical Research Centre, University of Oxford, Oxford, UK

[2] Modernising Medical Microbiology Unit, Nuffield Department of Medicine, University of Oxford, Oxford, UK

[3] Oxford University Hospitals NHS Foundation Trust, Oxford, UK

[4] NIHR Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance at University of Oxford, Oxford, UK

Contributions

Conception or design; N.D.S, K.D, D.E, A.S.W, B.C.Y. Acquisition, analysis or interpretation of data; N.D.S, K.D, K.M.V.H, M.C, M.P, V.D, T.P.Q, N.S., D.E, P.B., A.S.W, B.C.Y. Drafted the work or substantively revised it: N.D.S, A.S.W, B.C.Y.

Corresponding author

22 *Correspondence to: nicholas.sanderson@ndm.ox.ac.uk

23 https://orcid.org/0000-0001-6370-1961

24 Abstract

25 Clinical metagenomics (CMg) offers high-throughput respiratory pathogen detection with a wider
26 range than targeted, probe-dependent diagnostics. We optimised a long-read sequencing workflow to
27 detect RNA viruses in nasopharyngeal swabs. We derived quality control (QC) criteria and diagnostic
28 thresholds against a gold standard comprising 23 pathogen targets detected by routine multiplex PCR.
29 We validated this workflow using 344 prospectively collected respiratory samples submitted for
30 routine testing. Using pre-defined QC and positivity criteria, the workflow's sensitivity versus PCR
31 was 51% (95%CI: 45%-57%) (133/260 positive targets detected) (ranging from 19%-85% across
32 pathogens with >20 gold-standard detections), and specificity 99.8% (95%CI: 99.6%-99.9%)
33 (3836/3845 negative targets not detected). Sensitivity improved to 58% using post-hoc optimised
34 thresholds, 61% only considering RNA pathogens, 70% excluding rhinovirus/enterovirus and 83%
35 excluding samples with qPCR Ct values ≥35. Multiplexed nanopore sequencing costs were
36 £112GBP/sample, however read crossover from multiplex sequencing contributed importantly to
37 false-positives. Although CMg has substantial diagnostic potential, this validation study demonstrates
38 the technical limitations of current metagenomic sequencing methods applied to viral detection in
39 upper respiratory tract samples with high host and low pathogen biomass: its higher reliability at
40 higher viral loads demonstrates the importance of identifying the most appropriate use cases to
41 maximise its utility.

42 Introduction

43 Respiratory tract infections remain a leading cause of illness with substantial impact on health care
44 systems[1]. Diagnostic testing in respiratory infections continues to be a major challenge for medical
45 microbiology[2–5]. A wide range of pathogens can cause respiratory tract infection, and while much of
46 the burden is caused by different viral pathogens, co-infections are common[6]. Still, most routine
47 diagnostic testing remains targeted[7]. Targets may include single pathogens (eg urinary Legionella

antigen), or be configured as 'syndromic' panels which test for a pre-defined array of pathogens[8–10]. While offering high sensitivity, and convenient workflows with minimal hands-on time, these assays are constrained in scope to target species or serogroups[11,12]. Further, molecular tests may be vulnerable to naturally occurring variations in the assay target (such as mutations altering primer binding regions or antigen conformation)[13,14], and may lack ability to distinguish between important species (eg. rhinovirus vs enterovirus) or subtypes. Ideal pathogen workflows would be robust to sequence changes, including the target failures which can compromise PCR assays (such as the S-gene drop out seen with evolving lineages of SARS-CoV-2)[15], and would not be limited to a list of target species.

Clinical metagenomics (CMg) has emerged as a promising modality to deliver rapid, untargeted and simultaneous pathogen detection directly from clinical samples. Early adoption of this approach has focussed particularly on critical care settings and the analysis of lower respiratory tract specimens such as bronchoalveolar lavage (BAL)[16,17]. In critically ill patients, a broad differential diagnosis, preponderance of bacterial infections and extensive antibiotic exposure may give CMg an advantage against conventional diagnostics in terms of speed and yield[18]. Studies reporting CMg use in critical care have focussed on bacterial detection and antimicrobial resistance profiling[19,20], but its flexibility means it could be extended to include viral and fungal infections[21].

Routinely applying a respiratory CMg workflow in broader patient groups has technical challenges. Specimens containing high host but limited pathogen biomass (such as nose/throat swabs) can make accurate detection of signal from noise particularly difficult, particularly as some CMg workflows count only a single pathogen read as detection[22]. Human nucleic acid dominates many samples, though host depletion strategies can improve relative pathogen abundance[23]. The human sequence dominance has been overcome by short-read sequencing to very high depths, but at a relatively high per sample cost: a well validated workflow incurred sequencing costs of $260USD per test (estimated £198GBP on 2025 average exchange rate)[24]. Pathogen enrichment strategies include targeted pathogen baits, targeted amplification of preserved motifs (such as bacterial detection with 16s rRNA), or untargeted amplification (sequence independent, single primer amplification (SISPA), or phi29 polymerase)[25,26]. These strategies can improve sensitivity but potentially introduce detection bias, especially where

amplification is targeted, and increase the cost of processing. The introduction of contaminating material from reagents or the environment, and barcode/index misclassification, can all produce false positive results[27]. Distinguishing true detection of low pathogen levels from artefact relies on the use of well-constructed controls as well as optimised laboratory practices and bioinformatic sequence data processing[28]. Other approaches include multidisciplinary review of results for interpretation, but this is costly and may limit an assay's reproducibility and throughput[17]. Hence, there are multiple technical challenges for metagenomic tests. Importantly, whist ultimately the goal is agnostic detection of any pathogen, validation of sensitivity and specificity against known pathogen targets is necessary (but not sufficient) before using any metagenomic assay untargeted.

In order to scale CMg in routine care, it will not be cost-effective to depend on high read numbers, and the extent to which transparent and robust bioinformatic criteria can be set and consistently applied for the detection of known, common pathogens is currently unclear. We therefore present an end-to-end clinical validation of a metagenomic workflow for enhanced detection of common RNA respiratory viruses from naso-pharyngeal samples. The workflow uses a combination of low-cost virus enrichment and non-target depletion approaches, embedding positive and negative controls, prior to a modernised SISPA step. Samples are sequenced on the Oxford Nanopore Technologies (ONT) platform. As the necessary first step, a specific set of 23 known pathogens and pathogen subtypes comprising the diagnostic targets of the most commonly used commercial multiplex PCR platform were identified by mapping reads to a curated reference genome database. We used a derivation sample set to develop bioinformatic thresholds for pathogen identification, and then evaluated their performance in an independent validation set, using commercial multiplex PCR assays as the gold standard and within laboratory and bioinformatic quality control frameworks, including delineation of analytic sensitivity and specificity.

Results

Respiratory samples for pathogen detection

100 The study used nose and/or throat samples submitted for routine clinical microbiology PCR testing,
101 which was done on one of three platforms: two four-target assays (Alinity Resp-4-plex assay (Abbott),
102 or Xpert Xpress CoV-2/Flu/RSV plus (Cepheid), or the 23-target assay BIOFIRE FilmArray
103 Respiratory 2.1 panel (Biomerieux) (**Table S1**). The four-target assays (covering influenza A/B,
104 respiratory syncytial virus (RSV), and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-
105 2)) were performed routinely, while 23-target PCR testing was undertaken for samples from
106 immunocompromised patients or those admitted to critical care. Surplus material from samples with a
107 positive result on any platform were prospectively collected, along with a smaller number of
108 temporally matched samples testing negative by FilmArray (**Table 1(a)).**

109 In total 382 respiratory samples were retrieved, divided temporally into separate sample sets: a) a
110 derivation (n=38 clinical samples, tested repeatedly to produce n=104 individual sequences); and b) a
111 validation set (n=344 clinical samples, each tested once to produce n=344 individual sequences).
112 Because each sample had a positive or negative result from PCR gold standard testing for multiple
113 pathogens which may or may not be detected by metagenomic sequencing, each sample contributed
114 multiple observations to the analysis (i.e. each pathogen tested for using gold standard PCR
115 contributed one observation with the outcome being presence or absence of that pathogen on
116 sequencing results). The derivation set was used to define bioinformatic criteria for pathogen detection
117 by sequencing and thresholds for quality control (QC) for both positive and negative controls, criteria
118 which were then fixed and tested using the validation set, with all validation assays and sequencing
119 conducted by laboratory staff blinded to validation study results throughout.

120 Sample preparation and sequencing followed a single optimised laboratory method in the validation
121 set (**Figure 1, Methods**) (derivation sample preparation and sequencing used a similar near final
122 method, see **Methods**). Validation samples were sequenced in batches with batch positive and
123 negative controls, and internal amplification controls in each sample (**Table 2**, **Methods**). Samples
124 were positive for a range of pathogens by PCR, dominated by influenza A, RSV, human
125 rhinovirus/enterovirus (HRE) and SARS-CoV-2, other RNA viruses and a small number of DNA
126 viruses and bacteria (**Figure 2**). These were randomly mixed across sequencing runs for both

127 derivation and validation sets. Samples came from patients with a wide range of ages (**Table 1(b)**) and
128 a range of inpatient and emergency locations.

129 Derivation Set: Defining thresholds for quality control

130 The derivation set included 104 sequencing results from 38 samples, all with at least one pathogen
131 detected on multiplex PCR testing (**Table 1**, **Figure 2**). Although the CMg workflow was focussed on
132 RNA virus detection, we evaluated its performance with respect to all pathogens included in the
133 FilmArray panel. Total data obtained per run and reads per sample (**Figure S1**) were used to set
134 heuristic thresholds for sequencing run quality checks (**Table 2**).

135 Derivation Set: Improved base-calling accuracy and read thresholds reduce
136 false-positives from barcode crossover

137 To test the ability of the CMg workflow to reproduce results from commercial platforms at the highest
138 possible resolution[29], pathogen detection was based on mapping reads to a curated database of known
139 respiratory pathogen genome sequences using minimap2 (**Methods**). In the derivation set the median
140 (IQR) length of all reads was 292bp (278-308bp) (**Table 1**). Using standard high accuracy base-calling
141 (HAC) in the derivation set initially, we identified evidence of barcode crossover leading to
142 sequencing reads being mis-attributed to another sample within a multiplexed pool, alongside broader
143 issues with demultiplexing (**Figure S2**). For example, a median (IQR) 0.1% (0.08%-0.11%) of reads
144 were assigned to barcodes not used in the library preparation, and 21% (17%-32%) barcode sequences
145 identified remained unclassified. Among the derivation set, 65/104 (62%) sequences had reads
146 mapping to the pathogen identified by PCR in that sample; but 63/104 (61%) sequences also had one
147 or more reads mapping to a pathogen for which the sample tested negative by PCR. Use of the super
148 accuracy base-calling (SUP) model improved but did not eliminate this problem, reducing the median
149 reads assigned to unused barcodes to 0.01% (0.01%-0.013%), and resulting in 64/104 (62%)
150 sequences with reads mapping to the pathogen identified by routine PCR testing and 34/104 (33%)
151 with one or more reads mapping to a pathogen for which the sample was PCR-negative before any
152 diagnostic thresholds were implemented. All these 34 initial false-positives in the derivation set

occurred on sequencing runs which included another PCR-positive sample with a high number of reads for the pathogen detected, and were based on single read in 12/104 (12%) sequences. We therefore implemented a threshold of "≥2 reads " for target pathogen detection (**Table 2**), resulting in 56/104 (54%) derivation set sequences with ≥2 reads mapped to the pathogen identified by routine PCR testing and 18/104 (17%) with ≥2 reads mapping to a pathogen for which the sample was PCR-negative.

This "≥2 reads" quality threshold was also then applied to sequencing batch positive and negative controls, and internal sample controls (**Table 2**). MS2 was added to each sample as a control of nucleic acid extraction and subsequent cDNA synthesis and SISPA. Three viruses were added to each sample following nucleic acid extraction as internal controls of cDNA synthesis and SISPA. Each batch of 16 samples was accompanied by a positive amplification control consisting of the same 3 spiked-in viruses a negative (no template) control. 2/6 (33%) derivation sequencing runs, containing 24/104 (23%) sequences, failed sequencing batch negative control thresholds(i.e. had ≥2 reads mapping to MS2 or one the three viral internal amplification controls). In those runs that passed sequencing batch controls, 19/80 (24%) sequences failed internal controls, of which 8/19 (42%) failed to detect two internal control viruses and 11/19 (58%) failed to detect all three internal control viruses (**Figure 3C**). In total 61/104 (59%) derivation set sequences successfully passed the negative and positive batch control and internal controls and were used to subsequently establish bioinformatic thresholds for pathogen detection.

Derivation Set: Defining bioinformatic thresholds for pathogen detection

Within the 61 sequences in the derivation set passing QC, 67 pathogen targets were detected by routine PCR (i.e. gold standard positives), and 1154 pathogen targets were not detected by PCR (i.e. gold standard negatives), given the variable numbers of pathogen targets tested per sample, depending on the multiplex PCR assay used (**Table 1**). We evaluated 28 different bioinformatic metrics against true-positives and true-negatives to optimise detection (**Tables S2**, **S3**), individually and in combination. Many individual metrics and combinations of metrics had very similar performance

179 (**Tables S2, S3, S4**), and were highly collinear (**Figure S3**). Maximising Youden's index and taking
180 into account different sequencing aspects (**Table S2**), we chose a final combination of three separate
181 metrics (details in **Methods**), encompassing coverage breadth, depth and the proportion of mapped
182 reads, to address known sources of error in metagenomics[30,31]. Specifically, a pathogen was defined as
183 detected in the sequence read data if a) the percentage of the reference genome with coverage of at
184 least 1 read was >0.25%, or b) the mean read depth (truncated at a maximum of 10 at all positions)
185 across the entire genome (denoted "area under the genome") was >0.003, or c) the number of reads in
186 the sample mapping to the target reference genome as a percentage of reads on the run mapping to any
187 pathogen reference genome >0.006% (**Table 2**).

188 Combining these three "OR" criteria with an additional "AND" criterion for ≥2 target pathogen reads
189 led to 28/1929 false-positives including all sequenced samples in the derivation set (including those
190 not passing batch/internal QC control), and 4/1154 false-positives after restricting to sequenced
191 samples that passed QC checks (run, batch and internal). False-positives had proportionately lower
192 "area under the genome" for the percentage of reads in the sample mapping to the target reference
193 genome as a percentage of reads on the run mapping to any pathogen reference genome (**Figure 4A**).
194 We therefore added a third "AND" "ratio criterion" for positivity, specifically that the ratio between
195 these criteria should be >0.1 (**Table 2,** red line on **Figure 4A**).

196 Hence the final pathogen detection criteria determined on the derivation set were: percentage of
197 reference genome with at least 1 read >0.25% OR "area under the genome" >0.003 OR percentage of
198 reads mapping to reference genome>0.006%) AND >2 target pathogen reads AND "ratio
199 criterion">0.1 (**Table 2**).

200 43/67 (64%) PCR-positives in the derivation set were detected by these three criteria (**Figure 3E**): of
201 the 24 PCR-positives classified as negative, 18 (27% of the total) had no reads mapped to the target
202 pathogen, and 6 (9%) had only a single read mapped to the target pathogen (failed the ≥2 reads
203 threshold). 1151/1154 (99.7%) of the pathogens tested for, but not detected by multiplex PCR, were
204 not detected by sequencing. The three false-positive target detections in 2 of 38 original samples were

205 SARS-CoV2 (two reads) and influenza A, further subtyped as H1-2009 (five reads; two target
206 detections).

207 Validation Set: Diagnostic performance

208 To validate the performance of the metagenomic workflow with the derived thresholds, we used 344
209 independent respiratory samples with known multiplex PCR results (199 tested for 23 targets with
210 FilmArray, 136 tested for four targets with Xpert, and 9 tested for four targets with Resp-4-Plex,
211 **Table 1**). These were extracted in 22 batches and sequenced on 11 runs (**Figure 5**). 2/22 (9%) batches
212 (containing 31/344 (9%) samples) failed batch QC, with one amplification control sample failing all
213 three internal control viruses (<2 mapped reads), and one batch amplification control sample failing
214 due to detection of reads from a target pathogen (**Figure 5**). 39 samples failed internal quality checks
215 of extraction or amplification, 38/39 (97%) failing to detect MS2 extraction and amplification control;
216 in addition, 2/39 (8%) did not detect two internal control viruses and 9/39 (23%) failed to detect any
217 internal control viruses (**Figure 3D**). Overall, 273/344 (79%) respiratory samples in the validation set
218 passed negative and positive control QC criteria.

219 In these 273 samples passing QC, there were 260 pathogen target detections by multiplex PCR (i.e. the
220 gold standard), with 212 (78%) samples positive for a single pathogen, 21 (8%) positive for two
221 pathogen targets (42 in total), 2 (1%) positive for 3 pathogen targets (6 in total). Within these
222 detections, six (2%) samples were positive for influenza A with additional detection of the H1 subtype
223 target (counted as 2 detections each, see Methods) (**Table 1**). 38/273 (14%) samples passing QC were
224 PCR-negative for all pathogens.

225 133/260 (51%, 95%CI: 45-57%) PCR-positive results (i.e. gold standard positives) were also detected
226 by sequencing (**Table 3**, **Figures 3D, 4B**). Of samples with multiple pathogens found by PCR, 1/21
227 (4.8%) had two pathogens detected by sequencing, while 0/2 (0%) had three pathogen detections by
228 sequencing. Within the 127 false negative results, 89 (34% of the 260 total results) had no reads
229 mapped to the target pathogen, 17 (7% of the total) had only a single mapped read and therefore fell
230 below the threshold designed to minimise barcode crossover, 20 (8%) had ≥2 reads and met at least

231 one of the three main criteria but failed the ratio criteria (as defined in **Table 2**), and 1 (0.4%) had ≥2
232 mapped reads and achieved the ratio criteria but did not meet any of the three main criteria (**Figures**
233 **3D, 4B**).

234 3901/3910 (99.8%) targets negative by PCR (as tested in the original assays, see **Methods**) were also
235 not detected from sequencing (**Table 3**); 38 samples were PCR negative for all pathogens, of which 36
236 (95%) were also negative by sequencing. There were nine false-positive pathogen detections by
237 sequencing, all potentially attributable to barcode crossover (**Figure 4B**): these were: a) four
238 detections of SARS-CoV2 with two reads each, all on a single run that contained a sample with a very
239 early cycle detection of SARS-CoV2 on PCR (Resp-4-Plex Cn 13.54; 51,158 reads); b) four detections
240 of coronaviruses (3 HCoV-HKU1, 1 HCoV-OC43) from one run with 2-38 reads (Run 11) that
241 included a sample with 33,347 reads mapping to HCoV-HKU1 (PCR done by FilmArray, hence
242 lacking a Cn/Ct result); and a metapneumovirus from a run with 79 reads that also included two other
243 positive metapneumovirus samples with 48 and 810 reads (PCR also done by FilmArray).

244 The ratio criterion performed considerably less well in the validation set compared with the derivation
245 set, resulting in 20 false negatives solely from applying this criterion (**Figures 3D, 4B**). We compared
246 the main and ratio criteria (**Table 2**) with an alternative criterion which had relatively similar
247 performance to these criteria in the derivation set (**Tables S2, S4**). This criterion confirmed detection
248 if the number of reads in the sample mapping to the target reference genome as a percentage of reads
249 on the run mapping to the specific target genome (rather than all reference genomes) was >0.037%.
250 Replacing the main and ratio criteria with this alternative criterion (but retaining the ≥2 read criterion)
251 increased the number of samples passing batch controls from 260 to 274, increased sensitivity to 58%
252 (95% CI 52-64%; 159/274), and improved specificity to 99.9% (95% CI 99.8-100%; 4015/4017) with
253 only 2 false positive detections (**Table 3, Figure S4**), the metapneumovirus and coronavirus OC43
254 detections.

255 Performance by pathogen

256 While overall sensitivity using the criteria from the derivation set was 51%, assay performance varied
257 substantially by pathogen. Sequencing detected fourteen different RNA viruses in samples passing
258 QC, but no bacteria or DNA viruses. One *Chlamydia pneumoniae* was detected (3 reads) in a sample
259 whose positive controls failed (**Figure 6A**). Of the pathogen targets with more than 20 PCR-positive
260 results, sensitivities from highest to lowest were: RSV 85% (95% CI: 75-95%); SARS-CoV-2 65%
261 (95% CI: 53-78%); influenza A 58% (95% CI: 40-75%), and rhinovirus/enterovirus 19% (95% CI: 8-
262 29%) (**Figure 6B**) using the main criteria, with small improvements using the alternative criteria for
263 all except SARS-CoV-2.

264 The sample workflow was designed to maximise RNA pathogen detection. Excluding DNA pathogen
265 targets (adenovirus and bacterial species) from the evaluation increased sensitivity modestly to 61%
266 (159/260) (95% CI 55-67%) (**Table 3**). The clinical significance of some low viral level results for
267 enterovirus and rhinovirus is unclear[32] and in our study qPCR results frequently indicated low viral
268 loads (**Figure S5B**). Removing these two targets increased overall sensitivity using the alternative
269 criteria to 70% (145/207) (95% CI 63%-76%) (**Table 3**).

270 CMg provides greater taxonomic resolution than multiplex PCR

271 The sequencing workflow classified pathogens at higher resolution than multiplex PCR methods. Ten
272 rhinovirus/enterovirus FilmArray positive samples were further identified as Rhinovirus B3 (n=6),
273 Rhinovirus C (n=1), Rhinovirus A89 (n=1), Enterovirus A (n=2). All 47 samples with RSV detected
274 on PCR were further identified as either RSV-A (n=25) or RSV-B (n=22).

275 Additional pathogen targets detected

276 We detected two additional pathogens using CMg sequencing of targets not assayed in the original
277 PCR test (i.e. samples originally tested for only 4 pathogens), in 2/111 (2%) samples. These two also
278 passed the alternative criteria and were not on runs with other strong positives, representing plausible
279 additional detection. Human rhinovirus/enterovirus (HRE) was detected in one RSV-positive sample,
280 with genome mapping revealing it to be human rhinovirus B3. Coronavirus OC43 was identified in
281 one sample with SARS-CoV-2 detected by PCR. In addition, 13 samples with influenza A detected on

282 PCR were able to be subtyped with sequencing: 12 as influenza A/H1-2009 and one as influenza
283 A/H3.

284 Performance by viral load (Cn,Ct) and other factors

285 PCR Cn/Ct values were reported from the original Resp-4-Plex/Xpert platforms for influenza, RSV
286 and SARS-CoV-2: these showed high correlation with Ct values ascertained separately by qPCR as
287 part of this study (**Figure S5A**), although study qPCR Ct values were 5-10 cycles higher than original
288 Cn/Ct values, potentially due to loss of material during processing and/or to the use of different assays.
289 Other pathogens show similar study qPCR Ct values across targets (Coronavirus HKU,
290 rhinovirus/enterovirus) where detected (**Figure S5B**). Study qPCR detection rates did not vary
291 substantially by pathogen (**Figure S5C**).

292 The number of mapped reads was strongly associated with Ct value, with samples with higher Ct
293 values generating fewer reads as expected, whether Ct was from the original Resp-4-Plex/Xpert assay
294 (not available for Filmarray) (**Figure 7A**) or from study qPCR (**Figure 7B**). A valid Ct value from the
295 study qPCR assay indicated nucleic acid was extracted during the lab protocol; sensitivity using the
296 alternative criteria for samples with any Ct value from the study qPCR assay increased to 75% (95%
297 CI, 68-81%) (**Table 3**). Restricting analysis further to those with a study qPCR Ct <35 (i.e. higher
298 viral loads) increased sensitivity using the alternative criteria to 83% (95% CI, 76-88%) (**Table 3**).

299 Sequencing assay cost

300 The sequencing and consumables costs were calculated based on multiplexing 32 clinical samples on
301 each sequencing run at £102.41/sample (**Table S6**). Sequencing platform costs were estimated
302 conservatively (assuming no other use of sequencing platform and amortising purchase cost over five
303 years) and added a further £9.61/sample. The analysis pipeline can be performed using GridION
304 onboard computing capacity, so no additional server or compute costs are required. In addition to
305 £112.02 of sequencing and infrastructure costs, there was 70 minutes of labour required per sample.

306 # Discussion

307 Here, we have developed and assessed the performance of an end-to-end metagenomic workflow
308 which detects a range of RNA respiratory viruses. By applying stringent quality thresholds, we
309 achieved high specificity for both target pathogens (99.8%) and at an individual sample level (>96%).
310 Despite integrating pathogen amplification targeting the most common RNA viruses, sensitivity in a
311 large validation set was modest compared to multiplex PCR, ranging from 51% overall, and as high as
312 85% for RSV and as low as 19% for rhino/enterovirus using pre-defined detection criteria. This could
313 be improved modestly to 58% by using post-hoc exploratory alternative bioinformatic criteria, and
314 then further to 61% and 70% by again post-hoc excluding DNA pathogen targets and
315 enterovirus/rhinovirus respectively. By limiting our analysis to RNA viruses and samples with
316 pathogen detections on study qPCR with Ct <35 and using the alternative bioinformatic criteria,
317 sensitivity improved to 83%

318 We deliberately designed this ONT-based workflow to be achieved with a small footprint platform and
319 modest costs. The final reagent costs were £82GBP per sample, and reagent plus sequencing and
320 analysis platform costs are estimated at £112 per sample. This is achieved by multiplexing 32 clinical
321 samples across a single sequencing flow cell. CMg delivered multiple pathogen detections from a
322 single assay, and provided additional sub-typing and lineage information compared with PCR. While
323 cost compares favourably with extended multiplex PCR testing, the hands-on and labour time of the
324 assay was substantial, estimated at 70 minutes a sample, compared with a few minutes. The
325 sequencing cost of this assay was 45% less than the sequencing costs of an Illumina based respiratory
326 virus sequencing assay[24].

327 The ultimate goal of the assay was agnostic pathogen detection: however, a necessary first step was
328 the ability to reliably detect known targets. For this study, we therefore developed an initial simplified
329 mapping-based analysis pipeline for a panel of common and important respiratory pathogens tested in
330 routine microbiological laboratories, arguing mapping is the most sensitive approach to demonstrate
331 performance against known targets[29]. We validated this workflow on a prospective collection of

332 samples including 23 different PCR targets, representing a substantially broader validation than the 4-
333 8 targets validated in recent RNA virus CMg workflows[23,24]. Had the assay shown better performance
334 against known targets, the sequencing output generated could be further interrogated for novel and
335 emerging viruses with adaptations to the analysis pipeline reference database, or by sharing sequenced
336 reads with public health repositories[33].

337 The varying sensitivities obtained highlight the major challenges of direct from sample sequencing to
338 detect infections with high host and low pathogen biomass, as from nose and throat swabs rather than
339 the BALs most commonly used in ICU studies. Despite human cell depletion and RNA amplification,
340 34% of samples had no reads detected for the pathogen detected by routine PCR testing, and 20% of
341 gold-standard detections were not found by study qPCR following RNA extraction. At the same time,
342 imperfect barcode de-multiplexing of nanopore sequencing reads meant stringent thresholds were
343 needed to prevent false-positive detections. Reads were relatively short (~300bp), potentially due to
344 degradation of RNA genomes or short SISPA products. One recent CMg workflow required 3 non-
345 overlapping pathogen reads from a minimum of 5 million reads with short-read sequencing[24], while
346 another reported a single positive read on ONT sequencing as the threshold for virus detection[17]. For
347 the latter, sequencing was limited to 3 clinical samples per run, reducing the potential for cross-over,
348 but this increases per sample costs up to 2.7-fold, compared with testing 32 samples on a run.

349 One interesting question is why the assay performed relatively more poorly for rhinovirus/enterovirus
350 than other RNA viruses. We found that differences between study sequences and the references used
351 for mapping (**Table S5**) were greater for rhinovirus/enterovirus than other pathogens (**Figure S6**),
352 suggesting that alternative choice of reference strains might improve performance, and reflecting the
353 diversity of this pathogen group. The small genome size (~8kB vs ~13-30kb for other pathogen
354 targets) may also play a role in the low rate of detection (**Table S5**).

355 Our findings contrast with a recent UK study, which reported sensitivity for viral detection of 89%
356 (95%CI: 71-98) using respiratory CMg[21]. The differential performance is likely multifactorial. The
357 study by Alcolea-Medina et al was performed in patients in critical care, and >85% of viral detections
358 were from lower respiratory tract samples rather than the upper respiratory tract samples in this study.

359 For common respiratory viruses, lower respiratory tract samples show a higher viral load[34–36] than
360 nasopharyngeal samples. Moreover the low mass of host material in BAL samples allowed
361 investigators to omit depletion steps[21]. Finally, the study multiplexed ≤5 samples per sequencing run,
362 and 66% of included samples had no respiratory viruses detected by PCR testing. The lower capacity
363 for barcode or sample cross-over facilitated relaxed thresholds for sequencing detection, with a single
364 read only required for positivity[17].

365 For this specific analysis, we used a competitive mapping-based approach rather than a metagenomic
366 classifier because this allowed us to directly evaluate the performance of the workflow against a well-
367 defined gold standard of known pathogen targets. The mapping approach enabled fast curation and
368 masking of references and provided coverage metrics. However, the laboratory method is potentially
369 pathogen agnostic and therefore capable of detecting a wide range of known and unknown RNA
370 pathogens. A metagenomic classification tool, such as Kraken2[37] or Sylph[38], might be more
371 appropriate for the analysis if the method were to be used for pathogen surveillance: the performance
372 of an agnostic workflow would require further evaluation.

373 Our workflow was optimised for RNA pathogen detection, including two DNAse steps in the
374 laboratory methods: while reads mapping to DNA viruses and bacteria were detected they fell below
375 thresholds for detection in samples passing QC. This lack of sensitivity is likely due to the reliance on
376 detecting these pathogens via RNA transcripts. Nucleic acid extraction from clinical samples followed
377 by SISPA, as employed here, has a higher hands-on time and requires more specialised laboratory
378 practitioners than commercial kit-based multiplex-PCRs. These technical aspects and the modest
379 performance of the workflow mean that it will not replace currently available methods as it stands.
380 However, a metagenomic workflow using established laboratory methods and limited specialised
381 equipment could be a valuable asset to pandemic preparedness. RNA viruses are regarded as likely
382 candidates for emerging pathogens[39,40], and experience with MERS-CoV, SARS-CoV-2, highly
383 pathogenic avian influenza viruses and mpox virus have all highlighted the critical importance of
384 flexible diagnostic testing for pathogen detection and surveillance[41].

385 CMg undoubtedly has substantial diagnostic potential. However, with a laboratory workflow
386 optimised for RNA extraction, and bioinformatics optimised to stringent recognition of sample and
387 barcode cross-over, this study demonstrates the technical limitations of current nanopore metagenomic
388 sequencing applied to viral detection in upper respiratory tract samples with high host and low
389 pathogen biomass. Its higher reliability at higher viral loads demonstrates the importance of
390 identifying the most appropriate use cases to maximise its utility.

391

392 # Methods

393 ## Sample collection

394 Nasopharyngeal samples collected with flocked swabs were placed in viral transport medium (VTM)
395 (UTM (Copan, Italy) or Sigma-Virocult (MWE, UK)) for routine clinical testing in the Microbiology
396 Laboratory, Oxford University Hospitals NHS Foundation Trust (Oxford, UK). Samples surplus to
397 clinical testing were retrieved. Inclusion criteria were: (i) nose and throat swabs, (ii) <7 days since
398 patient sampling, iii) pathogen detected at cycle threshold (Ct) or cycle number (Cn) value ≤30 for
399 Resp-4-Plex (Abbott, USA) and Xpert assays (Cepheid, USA) or sampled tested by BIOFIRE
400 FilmArray Respiratory 2.1 Panel (Biomerieux, France); the latter assay does not report a Ct value.

401 For the derivation set, samples were selected to cover the broadest range of pathogen detections by
402 PCR between March and July 2024. For the validation set, all positive samples meeting inclusion
403 criteria between October 2024 and March 2025 were retrieved, until a threshold of n=60 positives per
404 virus was reached (met for RSV, SARS-CoV-2 and human rhino/enterovirus). Approximately one
405 FilmArray negative sample was collected for every eight positives (by any method).

406 Clinical laboratory results were confirmed by retrieving data from the Laboratory Information
407 Management System. This included date collected, test type, pathogen result, and Cn/Ct value if
408 available.

409 **Host Nucleic Acid Depletion and Virus Enrichment**

410 Prior to RNA extraction, samples were pre-treated to reduce host nucleic acid and enrich for virus
411 particles. Swab-containing VTM was vortexed (30 seconds) and 1mL removed. Samples were
412 clarified by low speed centrifugation (3,200g, 30 min), pelleting intact human cells and debris. The
413 supernatant was transferred to a fresh lo-bind Eppendorf tube and free nucleic acid was digested using
414 M-SAN HQ (1000U) (ArcticZymes Technologies, Tromsø, Norway) with final $MgCl_2$ and Tris HCl
415 pH 8.0 concentrations adjusted to 10mM and incubation at 37°C for one hour. 500µl Intact Virus
416 Precipitation Reagent (Thermo Fisher Scientific, Waltham, MA, USA) was added, mixed by pipetting
417 and incubated overnight (4°C). Precipitants were collected by high speed centrifugation (13,200g, 30
418 min) and the supernatant was discarded. Phosphate buffered saline (PBS, Sigma Aldrich, USA) (20µl)
419 and MS2 extraction control (10µl) (Thermo Fisher Scientific) were added and the pellet resuspended
420 gently by pipetting, prior to RNA extraction.

421 **RNA Extraction**

422 Samples (30µl) were extracted using the Direct-zol RNA Microprep kit (Zymo Research) according to
423 the manufacturer instructions, including the optional on-column DNase treatment. RNA was eluted in
424 DEPC treated water (10µl) (Ambion, Thermo Fisher Scientific).

425 **cDNA Synthesis and SISPA**

426 First strand cDNA was synthesized using the Maxima H Minus 1st Strand cDNA synthesis kit with ds
427 DNase (Thermo Fisher Scientific), in a 20µl reaction volume, according to the manufacturer's
428 protocol, with the following modifications. Extracted RNA (8µl) was treated with DNAse, as
429 recommended, then spiked with three reverse transcription/amplification controls (1µl each,
430 containing approximately 10^4 genome copies each of Zika virus (ATCC-VR-1838DQ), murine
431 respirovirus (ATCC-VR-907DQ) and orthoreovirus (ATCC-VR-824DQ) (ATCC Manassas, Virginia,
432 USA). cDNA synthesis was primed randomly using the 3' terminal N_9 segment of custom
433 oligonucleotide primers (5'-GAT-GAT-AGT-AGG-GCT-TCG-TCA-CNN-NNN-NNN N-3';
434 Integrated DNA Technologies, Leuven, Belgium), final concentration 5µM. This incorporated the
435 SISPA primer sequence (5'-GAT-GAT-AGT-AGG-GCT-TCG TCA-3') into the first strand cDNA.

436 Following the recommended denaturation at 65°C for 5 min, 5x RT buffer and Maxima H minus
437 Enzyme Mix were added and primer annealing, reverse transcription and inactivation took place under
438 recommended conditions, 25°C for 10 min, 50°C for 30 min and 85°C for 5 min. Following reverse
439 transcription, cDNA fragments >~100 nucleotides were purified using SPRI beads (AMPure XP
440 beads, Beckman Coulter, Indianapolis, Indiana, USA) with two 70% ethanol washes and elution in
441 16µl DEPC-treated water (Ambion, Thermo Fisher Scientific). A 3µl cDNA aliquot was reserved for
442 separate study qPCR, and 11µl taken forward into second strand cDNA synthesis.

443 Second strand cDNA was synthesized using Sequenase™ V2.0 DNA Polymerase (Thermo Fisher
444 Scientific) and the same custom oligonucleotide primer, ((5'-GAT-GAT-AGT-AGG-GCT-TCG-TCA-
445 CNN-NNN-NNN N-3') randomly priming second strand synthesis and incorporating the SISPA
446 oligonucleotide primer. Reaction conditions were as described previously[42] comprising 11µl ss cDNA,
447 4µl 5x sequenase buffer, and 1µl oligonucleotide primer (40 µM). Samples were denatured at 97˚C for
448 2 minutes and cooled to 4˚C, then 1µl DTT (100mM), 1µl dNTPs (10mM) and 1.8 µl DEPC-treated
449 water were added, followed by incubation at 37˚C for 8 minutes. A second identical incubation was
450 performed following the addition of 0.55µl sequenase dilution buffer and 0.2 µl Sequenase. Following
451 SPRI bead purification, dsDNA >100 bp was eluted in 22µl nuclease free water, 20µl of which was
452 amplified by SISPA PCR.

453 Each reaction (50µl) comprised 20µl ds cDNA, 5µl SISPA 1 oligonucleotide (5'-GAT-GAT-AGT-
454 AGG-GCT-TCG-TCA-3') (20 µM) and 25µl NEBNext® Ultra™ II Q5® Master Mix (New England
455 Biolabs, Ipswich, Massachusetts, USA). A no template negative control replacing cDNA with 20µl
456 DEPC-treated water was included at this stage. Amplification conditions were denaturation at 98°C for
457 30 seconds (1 cycle), followed by 30 cycles of denaturation at 98°C for 10 seconds, annealing 64°C
458 for 30 seconds, and extension 72°C for 1 min 30sec (30 cycles), and a final extension of 72°C for 2
459 minutes. A final SPRI beads purification was followed by elution in 15µl DEPC-treated water. The
460 amplification products were quantified by Qubit fluorometer using the dsDNA 1X BR Assay (Thermo
461 Fisher Scientific).

462 **Library preparation and sequencing**

463 Libraries were prepared according to the Rapid Sequencing V14-Amplicon Sequencing protocol from
464 ONT. Samples were sequenced using Rapid Barcoding Kit V14 (SQK-RBK114.96; ONT, Oxford,
465 UK) on R10.4.1 Flow Cells (ONT). A maximum of 36 samples were run on a single flow cell (32
466 clinical samples plus 4 positive/negative controls), and were run for 72 hours (with SUP base calling
467 enabled for the validation set).

468 qPCR of extracted cDNA

469 During cDNA synthesis, 3µl of single-stranded cDNA was reserved for quantification. Samples were
470 tested by qPCR assay targeting respiratory viruses detected on gold standard testing. Gold standard
471 RSV positive underwent separate qPCR for RSV-A and RSV-B. Gold standard HRE positives
472 underwent separate qPCR for enterovirus and rhinovirus. Due to limited cDNA availability, only 3
473 qPCR reactions were possible per sample. Samples with gold standard detection of >1 pathogen target
474 were prioritised for quantification with enterovirus and rhinovirus if HRE detected, followed by other
475 RNA viruses. Reaction conditions were as recommended by the manufacturer; 0.5µl TaqMan Microbe
476 Detection Assay (Thermo Fisher Scientific), 5µl TaqMan Fast Advanced Master Mix (Thermo Fisher
477 Scientific), 3.5µl H_2O and 1µl cDNA. Thermocycling was performed on the Quant Studio 5 PCR
478 machine with the following conditions; initial denaturation at 95°C for 20 seconds followed by 40
479 cycles of denaturation at 95°C for 1 second and annealing/extension at 60°C for 20 seconds. Standard
480 curves for quantification were generated by serially diluting True Mark Respiratory Panel 2.0
481 Amplification Control (Thermo Fisher Scientific) for a dilution range between 10^5 to 10^1 copies per
482 reaction. Standard curves were included on every qPCR reaction plate in triplicate. Quantification and
483 Ct values were determined using the Design and Analysis Software (Thermo Fisher Scientific) with
484 the "Quantification-TaqMan, Standard Curve" settings enabled.

485

486 Bioinformatic analysis

487 Raw ONT sequencing reads were basecalled on a GridION Mk1 with using Dorado (ONT) with the
488 current basecalling models at the time of sequencing: dna_r10.4.1_e8.2_400bps_hac@v4.3.0

489 (derivation set) or dna_r10.4.1_e8.2_400bps_sup@v4.3.0 (validation set). The derivation set ONT
490 sequences were subsequently basecalled again with the then current standalone model
491 dna_r10.4.1_e8.2_400bps_sup@v5.0.0 model and hardware (Nvidia RTX2080 GPU and Intel(R)
492 Xeon(R) Gold 5217 CPU @ 3.00GHz). It was not possible to re-call the validation set with this model
493 as the ONT output data was changed in July 2025 and the relevant pod5 files were not retained by
494 default.

495 Human host reads were filtered from the output fastq files before downstream analysis using kraken2
496 and the k2_pluspf_20240605 database (https://benlangmead.github.io/aws-indexes/k2). Further read
497 removal was performed using Deacon[43] with default settings before uploading to the ENA (project
498 number PRJEB102181).

499 The analysis workflow is provided on https://github.com/oxfordmmm/AD_analysis (v1.0). Read and
500 base numbers for each run and sample were calculated with SeqKit[44] v2.9.0. SISPA primers were
501 identified with BLASTn[45] v2.16.0+ (using parameters: -task blastn-short, word_size 7, gapopen 2,
502 max_target_seqs 10000000) and trimmed using the trim_primer.py script. Reads with more than 10
503 SISPA primers were removed before mapping. Reads were mapped using minimap2[46] v2.22-r1101
504 (using parameters: -ax map-ont -N 1000) to a curated composite reference fasta file including
505 representative genomes from a range of respiratory viral and bacterial pathogens (**Table S5**). Due to
506 non-specific mapping to the larger genomes (**Figure S6**), reads mapping to the bacterial genomes with
507 a mapQ score <40 were excluded. This was not applied to the reads mapping to the viral genomes.

508 We considered 28 potential bioinformatic metrics created from mapping the raw sequencing reads to
509 the composite references in samples passing QC in the derivation set (**Tables S2, S3**). Sequencing
510 coverage metrics used for species classification were calculated using samtools[47] v1.21 and the
511 coverageStats.py script. Metrics considered covered the following aspects

512 - Coverage: at 1x, 3x, 5x or 10x depth, in absolute number of bases and as a percentage of
513 reference length (ie normalised)

514 • Read-depth coverage across the whole genome ("area under the genome") and divided by the
515 reference length: at overall, and truncating read depth to 5x or 10x to reduce the impact of
516 mis-mapped spikes of coverage

517 • Mean read-depth coverage across the genome with read depth at least one, excluding from the
518 denominator regions with no coverage, overall and truncating to 5 or 10 read depth.

519 • Bases mapped to the reference, absolute and as a percentage of reference length (ie
520 normalised)

521 • Reads mapped to reference, excluding multiple supplementary alignments on the same
522 reference but including secondary alignments on different references, overall, and restricting
523 to reads with alignment length ≥200, ≥300, ≥400, or ≥500.

524 • Reads mapped to a target in a sample as a percentage of the total number of reads in the run,
525 the total number of reads in the run mapping to any reference or the total number of reads in
526 the run mapping to references from the target species (the last will be 100% where there only
527 one reference from a species and only one positive for that species on the run and no false-
528 positives on the run).

529 • Reads mapped to a target in a sample as a percentage of the total number of reads in the
530 sample mapping to references from the target species (100% where there only one reference
531 from a species and only one positive for that species on the sample).

532 We excluded the most collinear based on Spearman correlations, leaving eight metrics for detailed
533 consideration (**Table S2**, **Figure S3**). Area under the receiver operating curves (AUROC) for these
534 eight ranged from 0.854-0.864 (**Table S2**). We chose individual thresholds for each of these eight
535 metrics based on Youden's index (**Table S2**) and then calculated the sensitivity and specificity of all
536 255 combinations of detection based on sequences meeting any one to eight of these metrics (**Table**
537 **S4**). The maximum sensitivity of 0.731 was achieved by more than 30 combinations of three metrics
538 (**Table S4**). We therefore selected a combination which covered three different aspects of sequencing

539 performance, namely coverage breadth, depth of the mapped reads and the mapped reads as a
540 percentage of reads mapping to any reference in the run.

541 ## Statistical analysis

542 Detection thresholds for each metric were determined using Youden's index performed on all QC
543 passing samples in the derivation set. The gold standard was the multiplex PCR results (FilmArray,
544 Resp-4-Plex, Xpert, **Table 1**). Each sample contributed N observations, one for each of the potential
545 pathogen targets that could have been detected by the PCR test used (FilmArray 23, Resp-4-
546 Plex/Xpert 4). For influenza A, FilmArray reports positive/negative overall, and then can report none
547 or one of three subtypes if positive. These subtypes were only included as targets in analysis if
548 detected by FilmArray. For sequencing detection, the subtype (H1, H1-2009, H3) was determined by
549 the highest number of mapped sequencing reads, and Influenza A was counted as positive if any
550 subtype was positive. ROC analyses for each metric were performed using Python 3.10.4 and
551 Sklearn[48] v1.1.1. 95% confidence intervals were calculated using binomial proportion with Clopper-
552 Pearson exact interval using statsmodels[49] v0.15.0.

553

554 # Data Availability

555 Sequence data are available at the ENA (project number PRJEB102181).

556

557 # Code Availability

558 https://github.com/oxfordmmm/AD_analysis/tree/v1.0

559 # References

560 1. Infectious diseases impacting England, 2025 report.

561 2. Stojanovic, Z. *et al.* Advances in diagnostic tools for respiratory tract infections: from
562 tuberculosis to COVID-19 - changing paradigms? *ERJ Open Res.* **8**, 00113–02022 (2022).
563 3. Charlton, C. L. *et al.* Practical Guidance for Clinical Microbiology Laboratories: Viruses
564 Causing Acute Respiratory Tract Infections. *Clin. Microbiol. Rev.* **32**, e00042-18 (2019).
565 4. Chalmers, J. D. The Modern Diagnostic Approach to Community-Acquired Pneumonia in
566 Adults. *Semin. Respir. Crit. Care Med.* **37**, 876–885 (2016).
567 5. Wolf, J. & Daley, A. J. Microbiological aspects of bacterial lower respiratory tract illness
568 in children: typical pathogens. *Paediatr. Respir. Rev.* **8**, 204–210, quiz 210–211 (2007).
569 6. Musher, D. M. & Thorner, A. R. Community-acquired pneumonia. *N. Engl. J. Med.* **371**,
570 1619–1628 (2014).
571 7. UK Health Security Agency. UK Standards for Microbiology Investigations.
572 https://www.rcpath.org/profession/publications/standards-for-microbiology-
573 investigations.html (2025).
574 8. Rytter, H., Jamet, A., Coureuil, M., Charbit, A. & Ramond, E. Which Current and Novel
575 Diagnostic Avenues for Bacterial Respiratory Diseases? *Front. Microbiol.* **11**, 616971
576 (2020).
577 9. Yanagihara, K. The role of molecular diagnosis in acute respiratory tract infection.
578 *Respir. Investig.* **57**, 511 (2019).
579 10. Evans, S. E. *et al.* Nucleic Acid-based Testing for Noninfluenza Viral Pathogens in
580 Adults with Suspected Community-acquired Pneumonia. An Official American Thoracic
581 Society Clinical Practice Guideline. *Am. J. Respir. Crit. Care Med.* **203**, 1070–1087
582 (2021).
583 11. Bouzid, D. *et al.* Added value of rapid respiratory syndromic testing at point of care
584 versus central laboratory testing: a controlled clinical trial. *J. Antimicrob. Chemother.* **76**,
585 iii20–iii27 (2021).

586 12. Markussen, D. L. *et al.* Diagnostic Stewardship in Community-Acquired Pneumonia With
587 Syndromic Molecular Testing: A Randomized Clinical Trial. *JAMA Netw. Open* **7**,
588 e240830 (2024).
589 13. Del Vecchio, C. *et al.* Impact of antigen test target failure and testing strategies on the
590 transmission of SARS-CoV-2 variants. *Nat. Commun.* **13**, 5870 (2022).
591 14. Blais, J. *et al.* Risk of Misdiagnosis Due to Allele Dropout and False-Positive PCR
592 Artifacts in Molecular Diagnostics: Analysis of 30,769 Genotypes. *J. Mol. Diagn. JMD*
593 **17**, 505–514 (2015).
594 15. Castañeda-Mogollón, D. *et al.* A metagenomics workflow for SARS-CoV-2
595 identification, co-pathogen detection, and overall diversity. *J. Clin. Virol. Off. Publ. Pan*
596 *Am. Soc. Clin. Virol.* **145**, 105025 (2021).
597 16. Heitz, M. *et al.* Metagenomics for the microbiological diagnosis of hospital-acquired
598 pneumonia and ventilator-associated pneumonia (HAP/VAP) in intensive care unit (ICU):
599 a proof-of-concept study. *Respir. Res.* **24**, 285 (2023).
600 17. Charalampous, T. *et al.* Routine Metagenomics Service for ICU Patients with Respiratory
601 Infection. *Am. J. Respir. Crit. Care Med.* **209**, 164–174 (2024).
602 18. Hsu, W.-Y. *et al.* Performance of a hybrid capture-based target enrichment next-
603 generation sequencing for the identification of respiratory pathogens and resistance-
604 associated genes in patients with severe pneumonia. *Microbiol. Spectr.* **13**, e0213024
605 (2025).
606 19. Gan, M. *et al.* Antimicrobial resistance prediction by clinical metagenomics in pediatric
607 severe pneumonia patients. *Ann. Clin. Microbiol. Antimicrob.* **23**, 33 (2024).
608 20. Chen, H. *et al.* Assessment and clinical utility of metagenomic next-generation
609 sequencing for suspected lower respiratory tract infections. *Eur. J. Med. Res.* **29**, 213
610 (2024).

611 21. Alcolea-Medina, A. *et al.* Rapid pan-microbial metagenomics for pathogen detection and
612 personalised therapy in the intensive care unit: a single-centre prospective observational
613 study. *Lancet Microbe* **6**, (2025).
614 22. Alcolea-Medina, A. *et al.* Unified metagenomic method for rapid detection of
615 microorganisms in clinical samples. *Commun. Med.* **4**, 135 (2024).
616 23. Gauthier, N. P. G. *et al.* Validation of an Automated, End-to-End Metagenomic
617 Sequencing Assay for Agnostic Detection of Respiratory Viruses. *J. Infect. Dis.* **230**,
618 e1245–e1253 (2024).
619 24. Tan, J. K. *et al.* Laboratory validation of a clinical metagenomic next-generation
620 sequencing assay for respiratory virus detection and discovery. *Nat. Commun.* **15**, 9016
621 (2024).
622 25. Yin, Y. *et al.* Enhancing lower respiratory tract infection diagnosis: implementation and
623 clinical assessment of multiplex PCR-based and hybrid capture-based targeted next-
624 generation sequencing. *EBioMedicine* **107**, 105307 (2024).
625 26. Guo, Y. *et al.* A dual-process of targeted and unbiased Nanopore sequencing enables
626 accurate and rapid diagnosis of lower respiratory infections. *EBioMedicine* **98**, 104858
627 (2023).
628 27. Diao, Z. *et al.* Assessing the Quality of Metagenomic Next-Generation Sequencing for
629 Pathogen Detection in Lower Respiratory Infections. *Clin. Chem.* **69**, 1038–1049 (2023).
630 28. Torres Montaguth, O. E., Buddle, S., Morfopoulou, S. & Breuer, J. Clinical
631 metagenomics for diagnosis and surveillance of viral pathogens. *Nat. Rev. Microbiol.* 1–
632 15 (2025) doi:10.1038/s41579-025-01223-5.
633 29. Marić, J., Križanović, K., Riondet, S., Nagarajan, N. & Šikić, M. Comparative analysis of
634 metagenomic classifiers for long-read sequencing datasets. *BMC Bioinformatics* **25**, 15
635 (2024).

636 30. Cane, J. *et al.* Nanopore sequencing of influenza A and B in Oxfordshire and the United
637 Kingdom, 2022–23. *J. Infect.* **88**, 106164 (2024).
638 31. Street, T. L. *et al.* Nanopore adaptive sampling for bacterial identification from
639 periprosthetic joint replacement tissue. *Microb. Genomics* **11**, (2025).
640 32. Wright, P. F. *et al.* Comparison of Results of Detection of Rhinovirus by PCR and Viral
641 Culture in Human Nasal Wash Specimens from Subjects with and without Clinical
642 Symptoms of Respiratory Illness. *J. Clin. Microbiol.* **45**, 2126–2129 (2007).
643 33. Venkatesan, P. UK launch metagenomic pathogen surveillance programme. *Lancet*
644 *Microbe* **6**, (2025).
645 34. Bouzid, D. *et al.* Agreement of respiratory viruses' detection between nasopharyngeal
646 swab and bronchoalveolar lavage in adults admitted for pneumonia: a retrospective study.
647 *Clin. Microbiol. Infect.* **29**, 942.e1-942.e6 (2023).
648 35. Lin, C. *et al.* Comparison of throat swabs and sputum specimens for viral nucleic acid
649 detection in 52 cases of novel coronavirus (SARS-Cov-2)-infected pneumonia (COVID-
650 19). *Clin. Chem. Lab. Med.* **58**, 1089–1094 (2020).
651 36. Mei, F. *et al.* Diagnostic Role of Bronchoalveolar Lavage in Patients with Suspected
652 SARS-CoV-2 Pneumonia and Negative Upper Respiratory Tract Swab: A Systematic
653 Review and Meta-Analysis. *J. Clin. Med.* **11**, 4656 (2022).
654 37. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2.
655 *Genome Biol.* **20**, 257 (2019).
656 38. Shaw, J. & Yu, Y. W. Rapid species-level metagenome profiling and containment
657 estimation with sylph. *Nat. Biotechnol.* **43**, 1348–1359 (2025).
658 39. Ameni, G. *et al.* A Narrative Review on the Pandemic Zoonotic RNA Virus Infections
659 Occurred During the Last 25 Years. *J. Epidemiol. Glob. Health* **14**, 1397–1412 (2024).

660 40. Carrasco-Hernandez, R., Jácome, R., López Vidal, Y. & Ponce de León, S. Are RNA
661 Viruses Candidate Agents for the Next Global Pandemic? A Review. *ILAR J.* **58**, 343–
662 358 (2017).
663 41. Garcia-Blanco, M. A., Ooi, E. E. & Sessions, O. M. RNA Viruses, Pandemics and
664 Anticipatory Preparedness. *Viruses* **14**, 2176 (2022).
665 42. K, L. *et al.* Metagenomic Nanopore Sequencing of Influenza Virus Direct from Clinical
666 Respiratory Samples. *J. Clin. Microbiol.* **58**, (2019).
667 43. Constantinides, B., Lees, J. & Crook, D. W. Deacon: fast sequence filtering and
668 contaminant depletion. 2025.06.09.658732 Preprint at
669 https://doi.org/10.1101/2025.06.09.658732 (2025).
670 44. Shen, W., Sipos, B. & Zhao, L. SeqKit2: A Swiss army knife for sequence and alignment
671 processing. *iMeta* **3**, e191 (2024).
672 45. BLAST+: architecture and applications | BMC Bioinformatics | Full Text.
673 https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-10-421.
674 46. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–
675 3100 (2018).
676 47. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008
677 (2021).
678 48. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**,
679 2825–2830 (2011).
680 49. Seabold, S. & Perktold, J. statsmodels: Econometric and statistical modeling with python.
681 in *9th Python in Science Conference* (2010).
682

683 **Acknowledgements (optional)**

This project was funded by the NIHR Biomedical Research Centre, Oxford, and supported by the NIHR Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance at Oxford University in partnership with the UK Health Security Agency (UKHSA) (NIHR207397). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health and Social Care or UKHSA. ASW is an NIHR Senior Investigator. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript (AAM) version arising.

Ethics declarations

The use of surplus material from respiratory samples processed by the Oxford University Hospitals NHS Foundation Trust Clinical Microbiology laboratory, and use of de-identified patient information from these samples, without patient-level consent, was approved by the Queens Square NHS Research Ethics Committee (REC), London, reference 17/LO/1420, as part of work on the development of rapid diagnostics.

Competing interests

The authors have no competing interests to declare.

702 **Table 1 Characteristics of patient samples (nose and throat swabs) and sequences in derivation**
703 **and validation sets**

704 **(a)**

	Derivation set samples (N=38)	**Derivation set sequences† (N=104)**	**Derivation set sequences QC pass†† (N=61)**	**Validation set samples/ sequences (N=344)**	**Validation set samples/ sequences QC pass†† (N=273)**
Original test by (n (%))					
Xpert Xpress CoV-2/Flu/RSV plus*	2 (5%)	4 (4%)	4 (7%)	9 (3%) (2-°,7+)	9 (3%) (2-°,7+)
Alinity Resp-4-Plex*	6‡ (16%)	18‡ (17%)	8‡ (13%)	136 (40%) (0-,136+)	102 (37%) (0-,102+)
BIOFIRE FilmArray Respiratory 2.1**	31‡ (82%)	85‡ (82%)	51‡ (84%)	199 (58%) (45-,154+)	162 (59%) (36-,126+)
Number of pathogen targets detected‡‡, n (%)					
0	0	0	0	47 (14%)	38 (14%)
1	34 (89%)	94 (90%)	55 (90%)	270 (78%)	212 (78%)
2	4 (11%)	10 (10%)	6 (10%)	25 (7%)	21 (8%)
3	0	0	0	2 (1%)	2 (1%)
Total pathogen targets detected	42	114	67	326	260
Thousand reads generated	-	92 (69-127)	89 (68-108)	115 (81-151)	115 (82-146)
Median read length	-	292 (278-308)	290 (275-300)	270 (246-292)	276 (256-295)
N50	-	428 (375-469)	448 (382-473)	386 (336-429)	393 (348-432)
Read PHRED quality	-	10.9 (10.0-12.3)	10.5 (9.9-11.2)	11.4 (11.0-11.7)	11.4 (11.1-11.7)
Any target pathogen reads identified, n (%) - median (IQR) reads if >0	-	77 (74%) 35 (5-166)	49 (80%) 18 (5-98)	230 (67%) 6 (2-151)	191 (70%) 8 (2-188)
Any gold standard (target) pathogen reads identified, n (%) – median (IQR) reads if >0	-	64 (62%) 70 (6-2817)	45 (74%) 46 (6-135)	198 (58%) 44 (4-514)	162 (59%) 58 (4-603)

705 * tests for influenza A/B, RSV and SARS-CoV-2 (**Table S1**).
706 ** tests for adenovirus, coronaviruses (229E, HKU1, NL63, OC43, SARS-CoV-2, MERS-CoV),
707 human metapneumovirus, human rhino/enterovirus, influenza viruses (including subtypes influenza A,
708 A/H1, A/H3, A/H1-2009), influenza B virus, parainfluenza viruses (1-4), respiratory syncytial virus,
709 *Bordatella parapertussis*, *Bordetella pertussis*, *Chlamydia pneumoniae* and *Mycoplasma pneumoniae*
710 (**Table S1**).
711 † Original derivation samples were split and sequenced multiple times. Some sequenced samples
712 included pooled samples tested with multiple platforms, see ‡. For the validation set each sample was
713 sequenced once only.
714 †† Samples that passed run, batch and sample internal controls, see **Table 2**.
715 ‡ One derivation set sample was pooled and tested using both Resp-4-plex and FilmArray; this was
716 sequenced 3 times.
717 ‡‡ Includes influenza subtypes as separate pathogen targets.
718 ° Included in error
719
720

721 **(b)**

	Validation set samples/ sequences (N=344)	**Validation set samples/ sequences QC pass†† (N=273)**
Patient age (years)	48 (3-70) n=342*	38 (2-68) n=271*
<1	46 (13%)	37 (14%)
1-12	70 (20%)	65 (24%)
13-49	56 (16%)	45 (17%)
50-70	86 (25%)	66 (24%)
71-84	58 (17%)	39 (14%)
85+	26 (8%)	19 (7%)
Male	176 (51%) n=342*	141 (52%) n=271*
Location where sample taken		
Haematology or oncology	103 (30%)	83 (30%)
Other inpatient adult ward	77 (22%)	53 (19%)
Other inpatient paediatric ward	38 (11%)	30 (11%)
Ambulatory or emergency department	92 (27%)	76 (28%)
Critical care	19 (6%)	18 (7%)
Community	13 (4%)	11 (4%)
Trial	2 (1%)	2 (1%)
Days between collection and extraction date	5 (3-6)	5 (3-6)
Copy number (Cn) (Alinity)	22.5 (19.7-25.3)	22.9 (20.3-25.5)
Cycle threshold (Ct) (Xpert)	19.5 (18.8-22.1)	19.5 (18.8-22.1)
Total pathogens detected by routine testing	326	260
Study qPCR performed**	306 (94%)	244 (94%)
Study qPCR run, but Ct undetermined (>40) (% of routine testing targets with qPCR)	68 (22%)	50 (20%)
Study qPCR run, Ct determined (<40) (% of routine testing targets with qPCR) - Median (IQR) Ct	238 (78%) 30.2 (IQR: 27.2-33.7)	194 (80%) 30.0 (IQR: 27.2-33.9)

722 * missing values from anonymised trial samples. Percentages of non-missing values.
723 ** DNA target or multiple pathogens identified on routine testing and insufficient sample to run qPCR
724 for all (see **Methods**)
725
726 Note: showing n (%) or median (IQR). QC=quality control. See **Figure S1** for run characteristics.

Table 2 Quality check, controls and pathogen detection thresholds for CMg workflow

Type	Threshold for failure (any)	Applied to	Rationale
Control (original prespecified before validation study)			
Run sequencing	<400 Mb data	Sequencing run (two batches of 16 samples plus controls)	Minimum level of sequencing data acquisition given variability across runs (**Figure S1**)
Sample sequencing	<25,000 reads	Each sample (unique barcode)	Minimum level of sequencing data acquisition given variability across samples. Minimum read number ensures sample QC is not biased by few very long reads (**Figure S1**)
Batch PCR negative control	≥2 reads mapped to any of the four positive spikes* or any target pathogen not failing (i.e. passing) main or ratio positivity criteria below	Negative control on each batch	Negative (no template) control included with all samples undergoing SISPA PCR in a batch. Detecting target amplification in this control indicates cross contamination.
Batch positive amplification control	<2 reads mapped to 2 or 3 of the 3 amplification controls	Positive control on each batch	Control with only 3 RNA virus spikes included in double stranded (ds) cDNA synthesis and SISPA amplification process for a batch of samples. Failure to detect the virus spike indicates these processes have failed across the batch.
Batch positive amplification control	MS2 or any target pathogen not failing (i.e. passing) main or ratio positivity criteria below	Positive control on each batch	Control with only 3 RNA virus spikes included in ds cDNA synthesis and SISPA amplification process for a batch of samples. Detecting MS2 or a target pathogen indicates cross-contamination.
Internal extraction, cDNA synthesis and SISPA control	<2 reads mapped to MS2**	Each sample	MS2 is added prior to RNA extraction. Failure to detect in a sample indicates that RNA extraction and/or subsequent amplification has failed within the sample
Internal cDNA synthesis and SISPA control	<2 reads mapped to 2 or 3 of the 3 amplification controls	Each sample	Three RNA virus spikes added prior to ds cDNA synthesis and downstream amplification. Failure to detect indicates these processes have failed within the sample.
Positivity to a target pathogen (original pre-specified before validation study)			
Barcode crossover	<2 reads mapped to a target pathogen	Each sample	Reduce impact of barcode crossover, **Figure S2**
Main criteria	Area under the genome truncated at 10 <0.003 AND coverage at depth 1 <0.25% AND sample reads as a percentage of all references <0.006%	Each sample	Combination of three metrics performance with high sensitivity and AUROC in derivation set, reducing potential impact of kitome and sample contamination observed in previous multiplexed sequencing studies.

Type	Threshold for failure (any)	Applied to	Rationale
Ratio criteria	Ratio sample reads as a percentage of all references / area under the genome truncated at 10<0.1	Each sample	Excluded 23 false-positives in the full derivation set (**Figure 3**)
Positivity to a target pathogen (alternative exploratory after validation study)			
Barcode crossover	<2 reads mapped to a target pathogen	Each sample	Reduce impact of barcode crossover, **Figure S2**
Main criteria	Sample reads as a percentage of reads mapping to all target references <0.037%	Each sample	Single metric optimising sensitivity, specificity and AUROC and sensitivity in derivation set

* MS2 and internal viral amplification controls (Zika virus, murine respirovirus and orthoreovirus). Each sequencing run included 32 samples on a single flow cell. Samples were extracted and amplified in batches of 16 samples plus batch controls, and two batches included in a run.

** Quality check included for the validation set only due to inconsistent performance of the MS2 stock in the derivation experiments.

Note: All run, batch individual sample quality check criteria had to pass in order to have pathogen reported as detected by sequencing; target pathogen declared present if all positivity criteria met. See **Figure 4** for a summary of the combinations of criteria passed and failed.

Table 3 Sensitivity and specificity of different criteria for positivity in validation set

Analysis	Sensitivity (per target)	Pathogen targets not detected on qPCR*	Sensitivity in true-positive targets detected by qPCR	Sensitivity in true-positive targets with study qPCR Ct < 35	Pathogen targets with 0 mapped reads	Sensitivity in true-positive targets with >0 mapped reads	Sensitivity (per sample)	Specificity (per target)	False-positive detections	Specificity (per sample, excluding true positive targets)	Specificity (per negative sample)
Derivation criteria (**Table 2**)	51% [45-57%] (133/260)	66	66% [59-73%] (129/194)	73% [65-80%] (116/159)	89	78% [71-84%] (133/171)	55% [48-61%] (129/235)	99.8%[99.6%-99.9%] (3901/3910)	9	96.7%[93.8%-98.5%] (264/273)	94.7%[82.3%-99.4%] (36/38)
Replace main and ratio criteria with detection if the number of reads mapping to the target reference as a percentage of the number of reads mapping to the same reference in the run >0.037% (alternative criteria)	58% [52-64%] (159/274)	68	75% [68-81%] (154/206)	83% [76-88%] (140/169)	89	86% [80-91%] (159/185)	61% [55-67%] (152/249)	99.9%[99.8%-100.0%] (4015/4017)	2	99.3%[97.5%-99.9%] (287/289)	100%[91.2%-100%] (40/40)
Alternative criteria, only RNA pathogens	61% [55-67%] (159/260)	54	75% [68-81%] (154/206)	83% [76-88%] (140/169)	75	86% [80-91%] (159/185)	61% [55-67%] (152/249)	99.9%[99.8%-100.0%] (3204/3206)	2	99.3%[97.5%-99.9%] (287/289)	100%[91.2%-100%] (40/40)
Alternative criteria, only RNA pathogens other than rhinovirus/enterovirus	70% [63-76%] (144/207)	47	87% [81-92%] (139/160)	95% [90-98%] (125/131)	46	89% [84-94%] (144/161)	56% [50-62%] (140/249)	99.9%[99.8%-100.0%] (3092/3094)	2	99.3%[97.5%-99.9%] (287/289)	100%[91.2%-100%] (40/40)

* reflecting loss of genomic material during processing or study qPCR not conducted due to insufficient material remaining, see **Table 1**.

Figure 1 Clinical metagenomic workflow overview

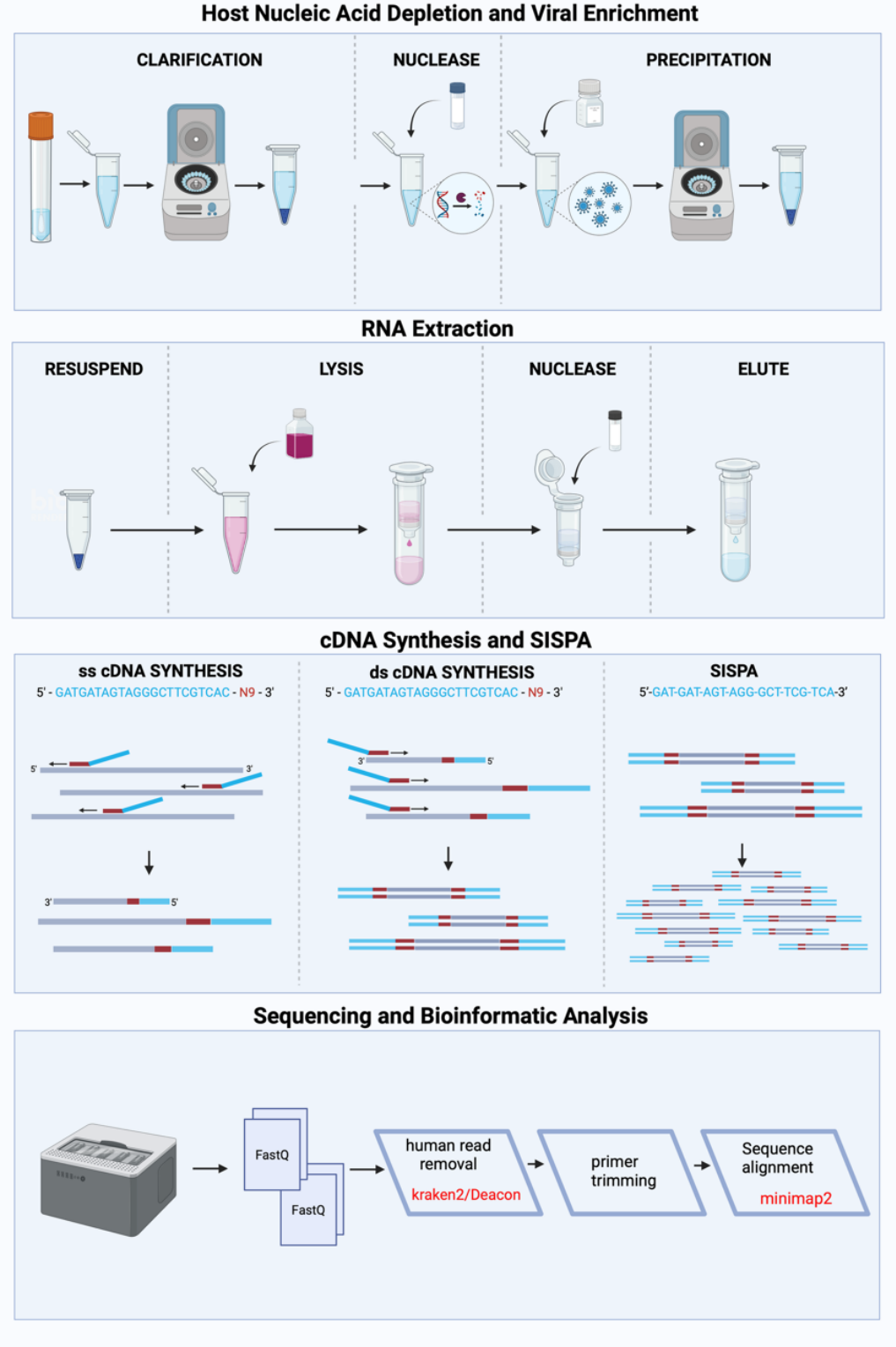


Figure 2 Pathogens identified (A, B) and not identified (C,D) on routine diagnostic testing in (A, C) derivation and (B, D) validation sets. Each sample could be positive for multiple pathogens, see **Table 1**.

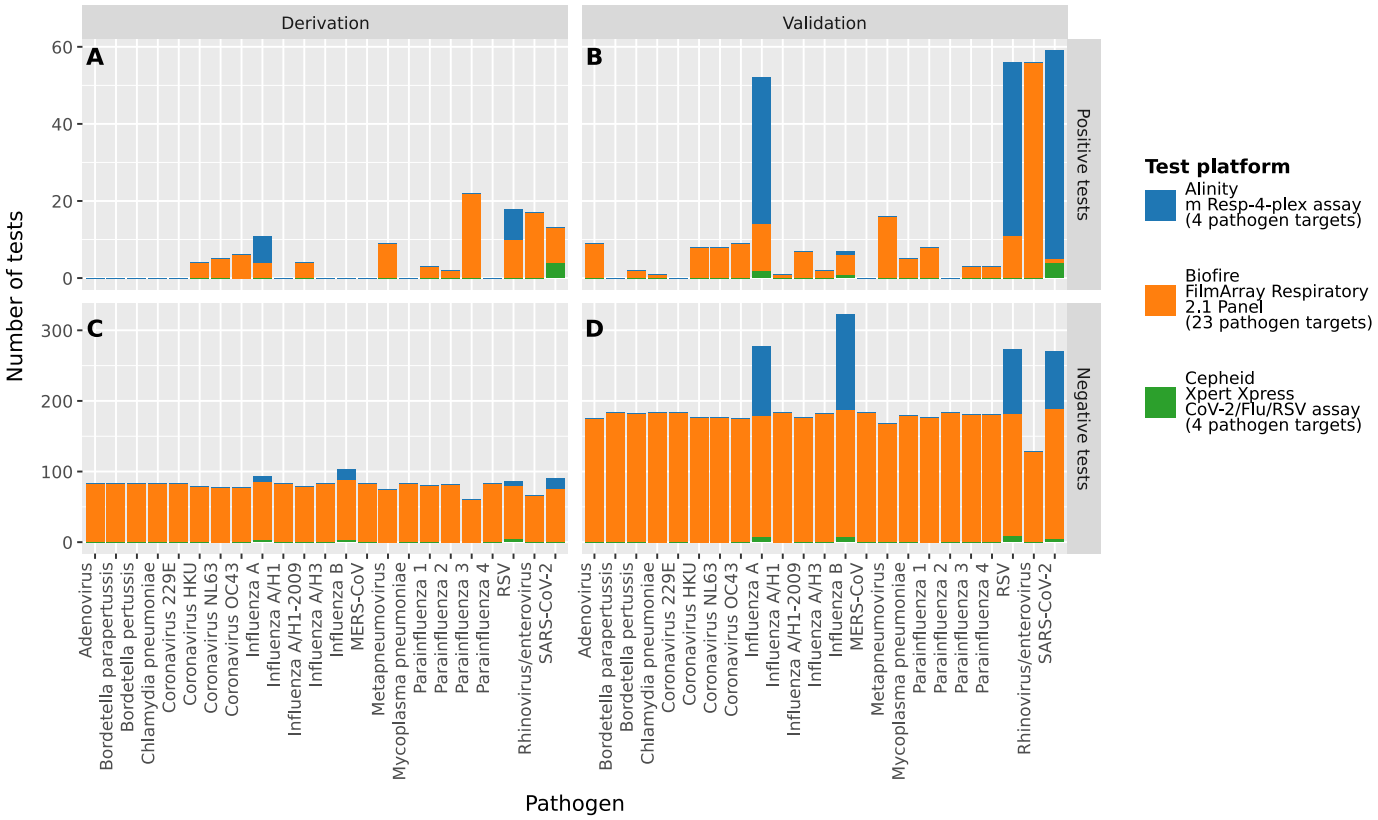


Figure 3 Batch control (A, B) and positive control (C, D) passed per sample and target detection (E, F) criteria passed for (A, C, E) derivation and (B, D, F) validation sets. Red indicates passed and black failed. RNA from murine respirovirus, orthoreovirus and Zika virus were spiked as cDNA synthesis and SISPA controls, MS2 phage was spiked as an extraction, cDNA synthesis and SISPA control. Within the derivation set MS2 was not required to pass QC due to poor performance of the laboratory MS2 stock; in both validation and derivation only detection of 2/3 virus spikes was required to pass QC.

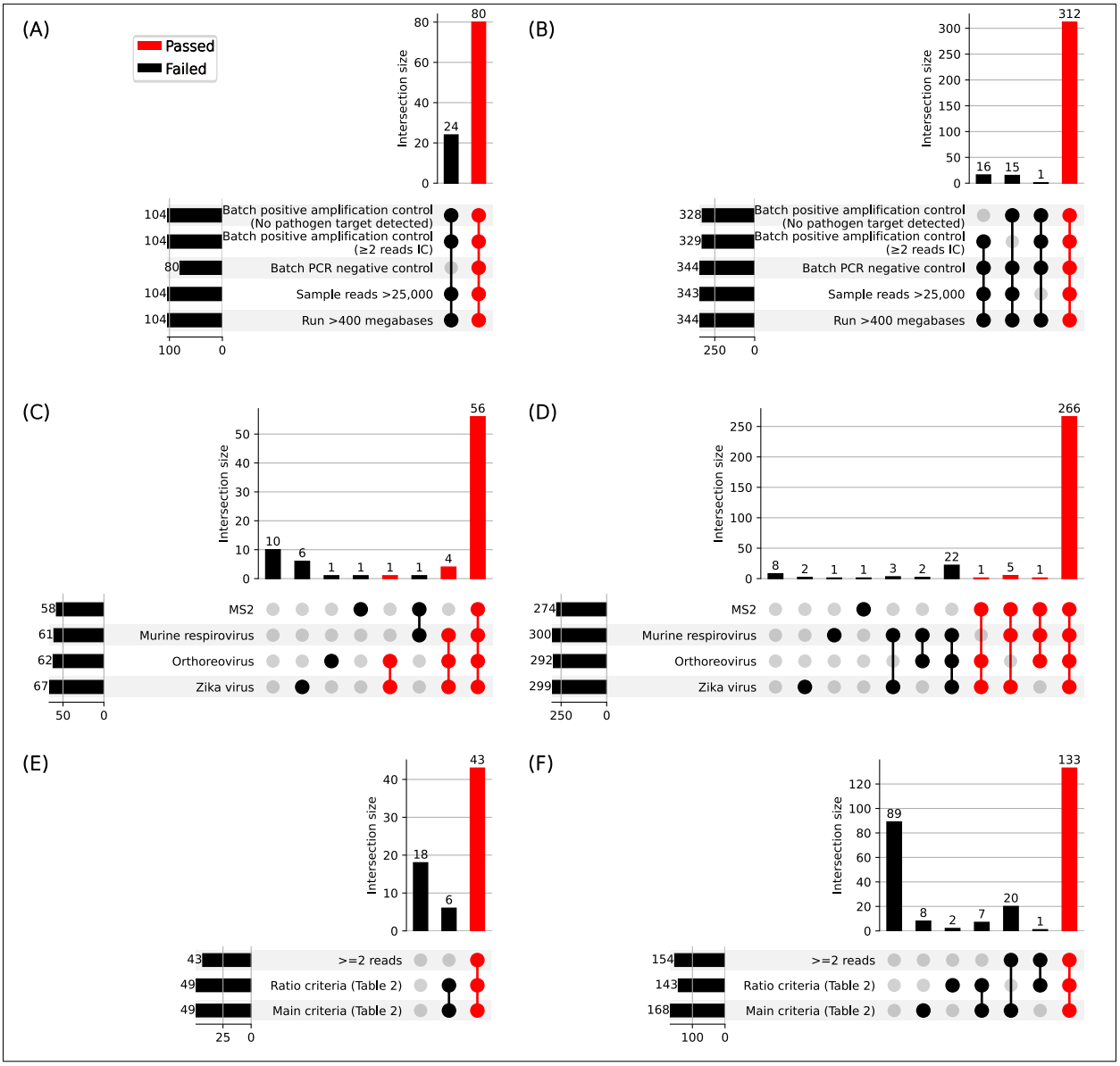


Figure 4 Positivity ratio criteria as developed in the derivation set (A,C) and tested in the validation set (B,D) Showing samples which passed sample QC thresholds (**Table 2**) in panels A, B and samples which failed sample QC thresholds in panels C, D. The sample QC pass (panels A, B) includes samples that failed batch controls, to provide context for barcode bleedover. In panel A classification of gold standard PCR-positives in the derivation set as true or false was based only on the number of reads and the main criteria (**Table 2**), ie did not include the ratio criteria shown by the red line, which was developed from the derivation set. In panel B, the classification of gold standard positives in the validation set does include the ratio criteria as shown by the red line. Individual thresholds shown in grey dashed lines.

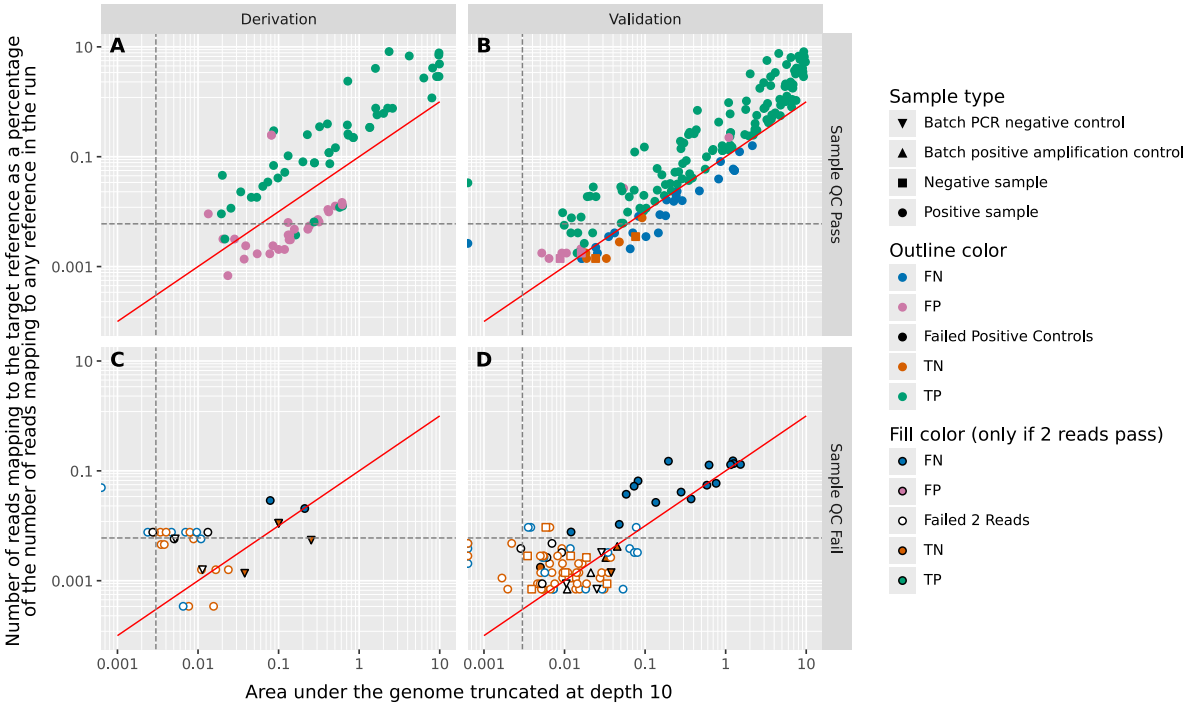


Figure 5 Validation set sample flow

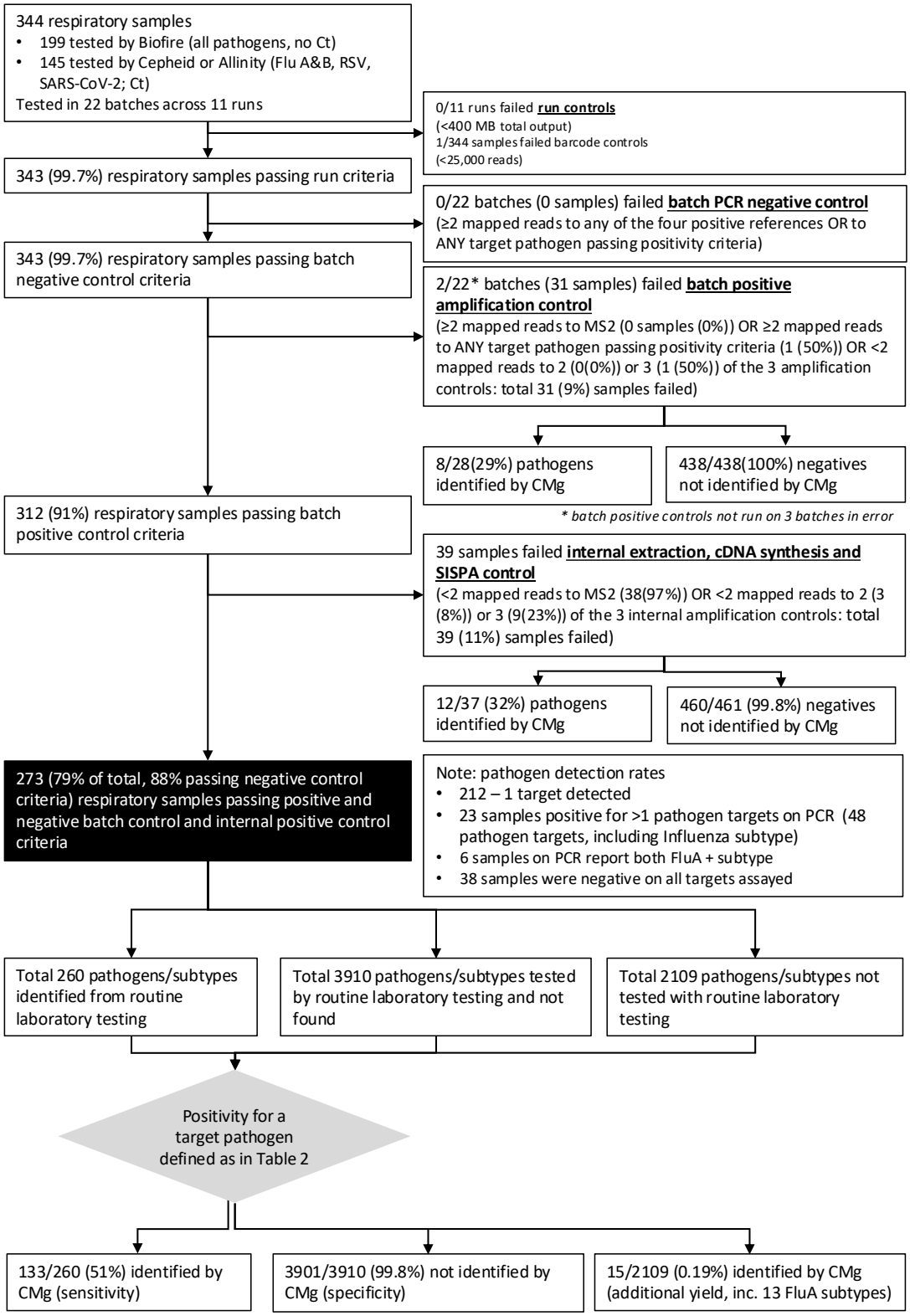


Figure 6 Pathogen-specific detections (A) and sensitivity (if >20 gold-standard detections) (B) in the derivation and validation sets. 95% confidence intervals calculated using Clopper-Pearson interval based on Beta distribution.

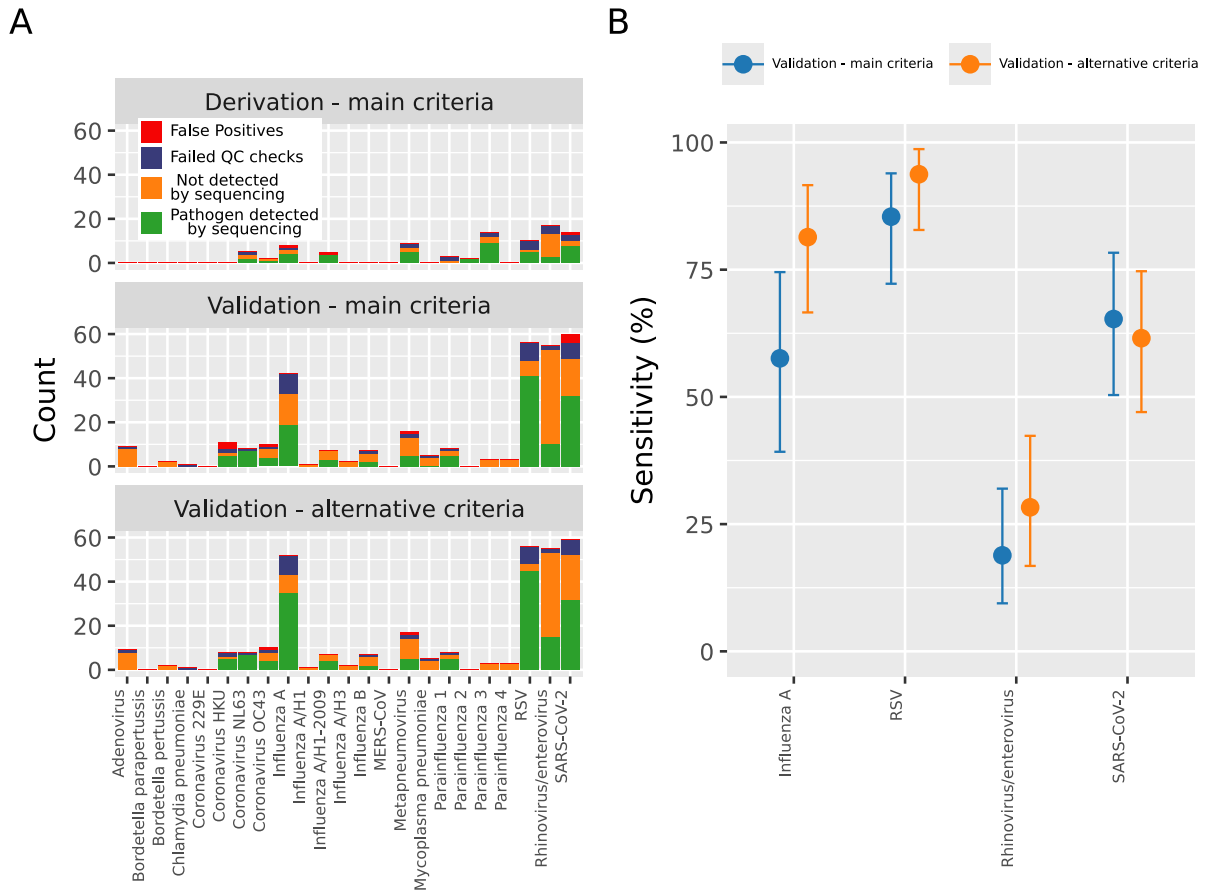
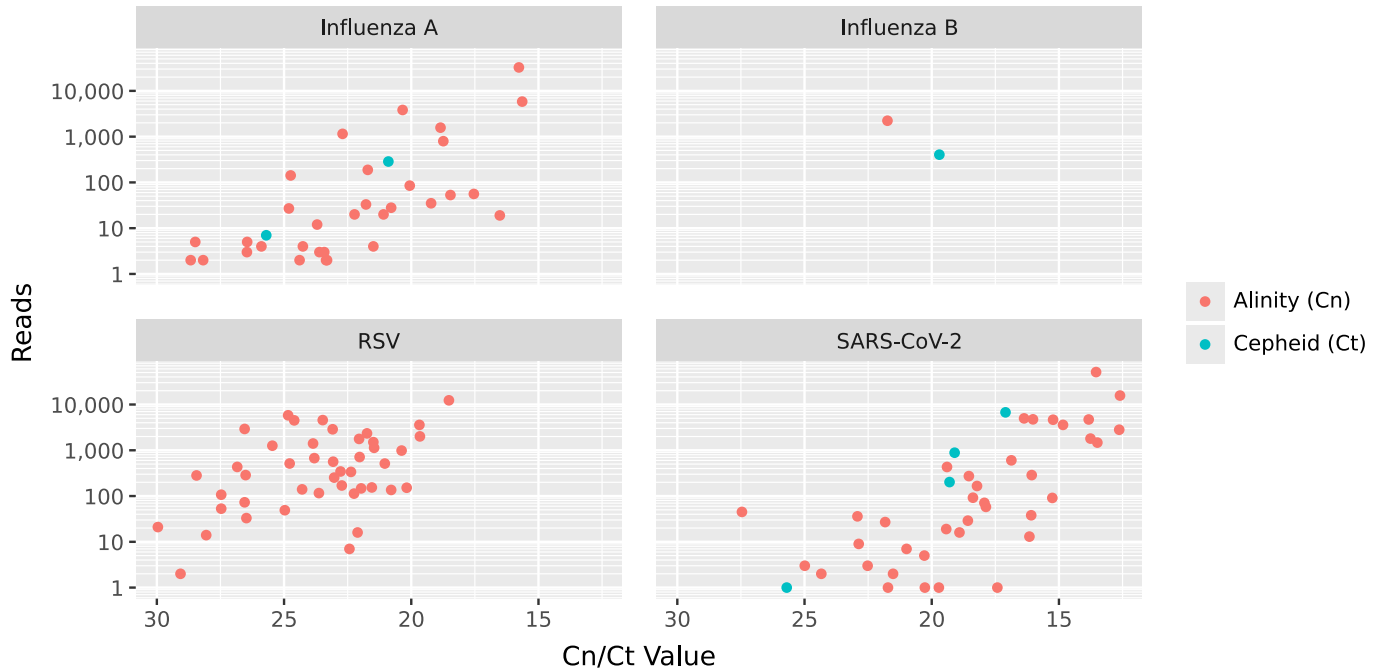
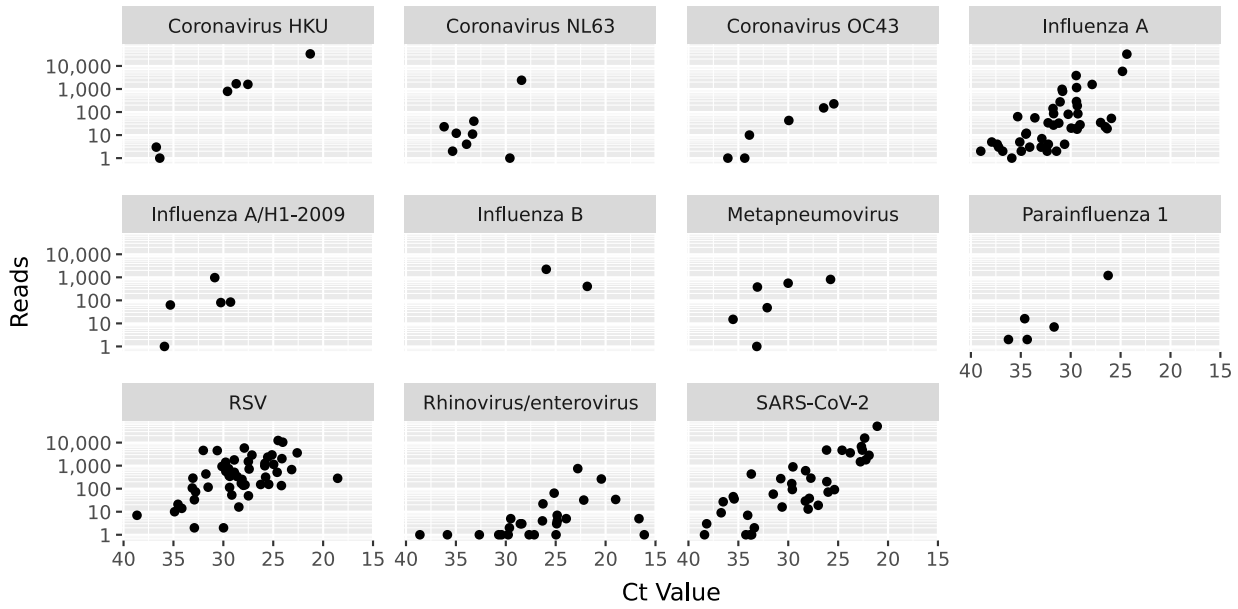


Figure 7 Association between number of mapped reads and Cn/Ct values from original assay (A) or study qPCR (B)

(A)



(B)