Clinical validation of a novel metagenomic nanopore sequencing method for detecting viral respiratory pathogens: diagnostic accuracy study

Nicholas D. Sanderson*[#1,2], Kate E. 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

Corresponding author

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

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

Abstract

Background

Clinical metagenomics (CMg) offers high-throughput respiratory pathogen detection with a wider range than targeted, probe-dependent diagnostics. Sequencing cost and the challenges of high host biomass in non-invasive samples are barriers to the use of CMg in high-throughput respiratory pathogen detection.

Methods

We optimised a long-read sequencing workflow to detect RNA viruses in nasopharyngeal swabs, employing pathogen enrichment and ONT sequencing. As a pre-requisite for agnostic pathogen detection, we first derived quality control (QC) criteria and diagnostic thresholds against a gold-standard comprising 23 pathogen targets detected by routine multiplex PCR. We validated this workflow using 344 prospectively collected upper respiratory tract samples submitted for routine testing.

Findings

Using pre-defined QC and positivity criteria, the workflow's sensitivity versus PCR was 51% (95%CI: 45%-57%) (133/260 positive targets detected) (ranging from 19%-85% across pathogens with >20 gold- standard detections), and specificity 99.8% (95%CI: 99.6%-99.9%) (3836/3845 negative targets not detected). Sensitivity improved to 58% using post-hoc optimised thresholds, 61% only considering RNA pathogens, 70% excluding rhinovirus/enterovirus and 83% excluding samples with qPCR Ct values ≥35. Read crossover from multiplex sequencing contributed most (7/9) false-positives: only 2 plausible additional pathogens were identified (rhinovirus and coronavirus OC43). 41 respiratory syncytial virus (RSV), 13 influenza A and 10 rhinovirus/enterovirus were successfully sub-typed by sequencing. Multiplexed nanopore sequencing costs were £112/sample.

Interpretation

Although CMg has substantial diagnostic potential, this validation study demonstrates the technical limitations of current metagenomic sequencing methods applied to viral detection in upper respiratory tract samples with high host and low pathogen biomass. Its greater sensitivity at higher viral loads

demonstrates the importance of identifying the most appropriate use cases to maximise its utility and value.

Funding: National Institute for Health and Care Research Oxford Biomedical Research Centre.

Research in context

Evidence before this study

Clinical metagenomics (CMg) provides the potential for rapid respiratory pathogen detection with a wider range than targeted, probe-dependent diagnostics. Searching PubMed for studies published between Jan 1 2018 and December 1 2025 using the terms ‘respiratory tract infection’ AND ‘metagenomics’ AND ‘diagnosis’, with no language restrictions yielded 734 items. We identified 15 clinical studies assessing the diagnostic performance of a metagenomic workflow for respiratory samples. Only 3 of these included upper respiratory tract samples (nasopharyngeal swabs); the remaining studies exclusively investigated invasive samples from the lower respiratory tract (bronchioalveolar lavage fluid). One of the three relevant studies assessed detection of viral pathogens in 83 specimens from critical care patients as part of pan-microbial nanopore sequencing metagenomic assay; only 2 samples included were nasopharyngeal samples. The second validated a high-depth short-read CMg workflow in 191 samples (85% nasopharyngeal), reporting 93.6% sensitivity and 98.8% specificity across eight respiratory pathogens detected by respiratory PCR. This workflow required substantial equipment and sequencing costs. A third study validated a nanopore sequencing CMg workflow, with lower laboratory footprint and sequencing costs than the second study. In 359 nasopharyngeal samples, they reported 61% sensitivity and 100% specificity against four respiratory viral targets detected by PCR. The costs of CMg mean that routine deployment as a high-throughput test with broad use will require demonstrating good performance across a range of common respiratory pathogens (such as can be assayed with commercially available extended multiplex PCR testing) and in upper respiratory tract samples.

Added value of this study

This prospective study is the first to evaluate nanopore sequencing for detecting a broad range of common respiratory viruses in nasopharyngeal samples. We validated a CMg workflow employing a modernised SISPA step for viral amplification, in 344 prospectively collected non-invasive samples, from which 19 different respiratory pathogens were detected by routine PCR testing. Using pre-defined bioinformatic thresholds overall sensitivity was 51% compared with PCR. This could be improved to 83% by limiting analysis to RNA viruses with Ct <35 and using post-hoc exploratory bioinformatic criteria. Sequencing costs were 45% lower than short-read sequencing.

Implications of all the available evidence

Our results demonstrate rigorous bioinformatic thresholds are essential in multiplexed CMg sequencing to reduce false-positive detections, a particular danger with imperfect barcode de-multiplexing. However these thresholds impair true-positive detection in samples with a high ratio of host-to-pathogen biomass. Further research should focus on identifying in which samples and clinical settings CMg can offer greater value in patient care.

Introduction

Respiratory tract infections remain a leading cause of illness with substantial impact on healthcare systems.[1] Diagnostic testing for respiratory infections continues to pose major challenges.[2,3] Much routine testing remains targeted,[4] testing for either a single pathogen, or configured as 'syndromic' panels testing pre-defined pathogen arrays.[5,6] While sensitive, these assays are constrained in scope,[7] are vulnerable to naturally occurring variations in the assay target,[8,9] and may fail to distinguish between important species or subtypes.

Clinical metagenomics (CMg) has emerged as a promising modality for rapid, untargeted and simultaneous pathogen detection directly from clinical samples. Studies reporting CMg use in critical care have focussed on lower respiratory tract specimens such as bronchoalveolar lavage (BAL),[10,11]

97 reporting performance for bacterial detection and antimicrobial resistance profiling.[12,13] However, use
98 can extend to viral and fungal infections.[14]

99 Technical challenges limit broader routine application of respiratory CMg. Non-invasive naso-
100 pharyngeal swabs contain high ratios of host to pathogen biomass, making accurate detection of signal
101 from noise problematic: some CMg workflows count only a single pathogen read as detection.[15,16]
102 Dominant human nucleic acid requires depletion strategies to increase relative pathogen abundance.[16]
103 Short-read sequencing to very high depths can overcome these challenges, but at a high price:
104 sequencing costs of a well-validated workflow were $260USD/test (estimated £198GBP on 2025
105 average exchange rate).[17] Pathogen enrichment strategies include targeted baits, or untargeted
106 amplification (sequence independent, single primer amplification (SISPA), or phi29 polymerase).[18,19]
107 Further, contaminating material and barcode/index misclassification can all produce false-positive
108 results,[20] requiring optimised laboratory practices and bioinformatic processes.[21]

109 Whilst ultimately the CMg goal is agnostic pathogen detection, validation of sensitivity and specificity
110 against known pathogen targets is necessary (but not sufficient) to assess performance. We therefore
111 undertook a clinical evaluation of an end-to-end workflow for detecting common RNA respiratory
112 viruses from nasopharyngeal samples. The workflow combines lower-cost virus enrichment and non-
113 target depletion approaches, and Oxford Nanopore Technologies (ONT) sequencing. We used a
114 derivation sample set to develop bioinformatic thresholds, and then evaluated their performance in an
115 independent validation sample set, using commercial multiplex PCR assays as the comparator to
116 quantify sensitivity and specificity, and employing multiple laboratory and bioinformatic quality
117 controls.

118 **Methods**

---

119 Respiratory samples for pathogen detection

120 Nasopharyngeal samples included in this study were submitted in viral transport medium for testing in
121 the Microbiology Laboratory, Oxford University Hospitals NHS Foundation Trust (Oxford, UK). PCR

122 was performed with three platforms: two four-target assays (Alinity Resp-4-plex assay (Abbott, USA),
123 or Xpert Xpress CoV-2/Flu/RSV plus (Cepheid, USA)), or a 23-target assay (BIOFIRE FilmArray
124 Respiratory 2.1 panel (Biomerieux, France)) (**Tab.S1**), following manufacturer's instructions. The
125 four-target assays (covering influenza A/B, respiratory syncytial virus (RSV), and severe acute
126 respiratory syndrome coronavirus 2 (SARS-CoV-2)) were performed routinely, while 23-target PCR
127 testing was undertaken for samples from immunocompromised patients or those admitted to critical
128 care. Samples surplus to routine testing were retrieved. Inclusion criteria were: (i) nose and throat
129 swabs, (ii) <7 days since patient sampling, iii) pathogen detected at cycle threshold (Ct) or cycle
130 number (Cn) value ≤30 for Resp-4-Plex and Xpert assays or sampled tested by BIOFIRE FilmArray
131 Respiratory 2.1 Panel (no Ct/Cn available). For approximately every eight positive samples, one
132 FilmArray negative sample was collected (**Tab.1(a)**).

133 Procedures

134 A derivation set was used to define bioinformatic criteria for pathogen detection by sequencing and
135 thresholds for quality control (QC) for both positive and negative controls, criteria which were then
136 fixed and tested using the validation set, with all validation assays and sequencing conducted by
137 laboratory staff blinded to validation study results throughout. Validation sample size was determined
138 by the duration of the winter 24/25 season and available staff.

139 Sample preparation and sequencing followed a single optimised laboratory method in the validation
140 set, while derivation sample preparation and sequencing used a similar near final method (**Fig.1,**
141 **supplemental methods**). Briefly, this comprised host depletion by low-speed centrifugation and
142 nuclease treatment with M-SAN HQ (ArcticZymes), followed by viral precipitation with Intact Virus
143 Precipitation Reagent (Thermo Fisher Scientific). Viral material was retrieved by high-speed
144 centrifugation. RNA was extracted using Direct-zol RNA Microprep kit (Zymo Research), and
145 underwent cDNA synthesis and SISPA. DNA libraries were prepared using Rapid Barcoding Kit and
146 sequenced on R10.4.1 flow cells on GridION Mk1 (ONT).

147 Samples were sequenced in batches with batch positive and negative controls, and internal
148 amplification controls in each sample (**Tab.2**). Samples within both the derivation and validation sets
149 were randomly mixed across sequencing runs. MS2 was added to each sample as a control of nucleic
150 acid extraction and subsequent cDNA synthesis and SISPA. Three viruses were added to each sample
151 following nucleic acid extraction as internal controls (ICs) of cDNA synthesis and SISPA. Each batch
152 of 16 samples was accompanied by a positive amplification control consisting of the same 3 spiked-in
153 viruses in a negative (no template) control. During cDNA synthesis, 3µl of single-stranded (ss) cDNA
154 was reserved for quantification by qPCR (TaqMan Microbe Detection Assay, Thermo Fisher
155 Scientific) targeting the respiratory virus(es) detected on gold-standard testing

156 As the necessary first step, we targeted identification of 23 known pathogens/pathogen subtypes
157 comprising the diagnostic targets of the commercial multiplex PCR platform (**Tab.S1**). To test the
158 ability of the CMg workflow to reproduce PCR results at the highest possible resolution,[22] pathogen
159 detection was based on mapping reads to a curated database of matching known respiratory pathogen
160 genome sequences (**Tab.S2**) using minimap2. Although the CMg workflow was focussed on RNA
161 virus detection, we evaluated performance with respect to all FilmArray panel targets. We evaluated
162 28 bioinformatic metrics against true-positives and true-negatives to optimise detection (**Tab.S3**;**S4**),
163 individually and in combination. Because each sample had a positive or negative result from PCR
164 gold-standard testing for multiple pathogens which were or were not be detected by metagenomic
165 sequencing, each sample contributed multiple observations to the analysis (i.e. each pathogen tested
166 for using gold-standard PCR contributed one observation with the outcome being presence or absence
167 of that pathogen on sequencing results). The analysis workflow is provided on
168 https://github.com/oxfordmmm/AD_analysis.

169 Results

170 In total 382 respiratory samples were retrieved, divided temporally into separate sample sets: a) a
171 derivation (n=38 clinical samples, tested repeatedly to produce n=104 individual sequences); and b) a
172 validation set (n=344 clinical samples, each tested once to produce n=344 individual sequences).

173 ## Derivation Set: Defining thresholds for quality control

174 The derivation set included 104 sequencing results from 38 samples, all with pathogens detected on
175 multiplex PCR testing (**Tab.1**, **Fig.2**). Total data obtained per run and reads per sample (**Fig.S1**) were
176 used to set heuristic thresholds for sequencing run quality checks (**Tab.2**).

177 ## Derivation Set: Improved base-calling accuracy and read thresholds reduce
178 false-positives from barcode crossover

179 In the derivation set, barcode crossover following standard high accuracy base-calling (HAC) led to
180 sequencing reads being mis-attributed to another sample within a multiplexed pool (**Fig.S2**). The super
181 accuracy base-calling (SUP) model demonstrated improved performance, reducing median reads
182 assigned to unused barcodes from 0.1% (0.08%-0.11%) to 0.01% (0.01%-0.013%), and the number of
183 sequences with ≥1 read mapping to a pathogen for which the sample was PCR-negative from 63/104
184 (62%) to 34/104 (33%). All 34 false-positives in SUP-called sequences occurred on sequencing runs
185 which included another PCR-positive sample with a high number of reads for the pathogen detected,
186 and were based on single read in 12/104 (12%) sequences. We therefore implemented a "≥2 reads"
187 threshold for target pathogen detection (**Tab.2**). We found 56/104 (54%) derivation set sequences with
188 ≥2 reads mapped to the pathogen identified by routine PCR testing and 18/104 (17%) with ≥2 reads
189 mapping to a pathogen for which the sample was PCR-negative.

190 This "≥2 reads" quality threshold was also applied to sequencing batch positive and negative controls,
191 and ICs (**Tab.2**). 2/6 (33%) derivation sequencing runs, containing 24/104 (23%) sequences, failed
192 sequencing batch negative control thresholds (i.e. had ≥2 reads mapping to MS2 or one the three viral
193 internal amplification controls). Among runs passing sequencing batch controls, 19/80 (24%)
194 sequences failed ICs: 8/19 (42%) failed to detect two IC viruses and 11/19 (58%) failed to detect all
195 three IC viruses (**Fig.3C**). In total, 61/104 (59%) derivation set sequences passed all negative and
196 positive batch control and ICs and were used to establish bioinformatic thresholds for pathogen
197 detection.

198 ## Derivation Set: Defining bioinformatic thresholds for pathogen detection

199 Within 61 derivation sequences passing QC, 67 pathogen targets were detected by routine PCR (i.e.
200 gold-standard positives), and 1154 pathogen targets were not detected by PCR (i.e. gold-standard
201 negatives) (**Tab.1**). We evaluated 28 bioinformatic metrics against true-positives and true-negatives to
202 optimise detection (**Tab.S3**;**S4**), individually and in combination (**supplemental methods**). Many
203 metrics and combinations had very similar performance (**Tab.S3;S4;S5**), and were highly collinear
204 (**Fig.S3**). Maximising Youden's index, and taking into account sequencing performance aspects
205 (**Tab.S3**), we chose a final combination of three metrics encompassing coverage breadth, depth and
206 the proportion of mapped reads, to address known sources of error in metagenomics.[23,24] Specifically,
207 a pathogen detection in sequence data required a) the percentage of the reference genome with
208 coverage of ≥1 read was >0.25%, or b) the mean read depth (truncated at a maximum of 10 at all
209 positions) across the entire genome (denoted "area under the genome") was >0.003, or c) the number
210 of reads in the sample mapping to the target reference genome as a percentage of pathogen reads on
211 the run mapping to any pathogen reference genome was >0.006% (**Tab.2**).

212 Combining these three "OR" criteria with the "AND" criterion for ≥2 target pathogen reads led to
213 28/1929 false-positives including all sequences in the derivation set (including those not passing
214 batch/internal QC control), and 4/1154 false-positives after restricting to sequences that passed QC
215 checks. False-positives had proportionately lower "area under the genome" compared with the
216 percentage of reads in the sample mapping to the target reference genome as a percentage of reads on
217 the run mapping to any pathogen reference genome (**Fig.4A**). We therefore added a third "AND"
218 "ratio criterion" for positivity, specifically that the ratio between these criteria should be >0.1 (**Tab.2,**
219 **Fig.4A**).

220 Hence the final pathogen detection criteria determined were: (percentage of reference genome with ≥1
221 read >0.25% OR "area under the genome" >0.003 OR percentage of reads mapping to reference
222 genome>0.006%)) AND ≥2 target pathogen reads AND "ratio criterion">0.1 (**Tab.2**).

223 43/67 (64%) PCR-positives in the derivation set were detected by these three criteria (**Fig.3E**): among
224 24 PCR-positives classified negative, 18/67 (27%) had no reads mapped to the target pathogen, and
225 6/67 (9%) had one read mapped to the target pathogen (failed the ≥2 reads threshold). 1151/1154
226 (99.7%) of the pathogens tested for, but negative by multiplex PCR, were not detected by sequencing.
227 The three false-positive target detections, occurring in 2 of the 38 original samples, were SARS-CoV2
228 (two reads) and influenza A, further subtyped as H1-2009 (five reads; counted as two target
229 detections).

230 Validation Set: Diagnostic performance

231 To validate the performance of the derived thresholds, we used 344 independent upper respiratory
232 tract samples with known multiplex PCR results (**Tab.1a**) from patients with a wide range of ages and
233 inpatient and emergency locations (**Tab.1b)**. Samples were positive for diverse pathogens by PCR,
234 dominated by RNA viruses with a small number of DNA viruses and bacteria (**Fig.2**).

235 Samples were extracted in 22 batches and sequenced on 11 runs (**Fig.5**). 2/22 (9%) batches
236 (containing 31/344 (9%) samples) failed batch QC, with one batch amplification control sample failing
237 all three IC viruses (<2 mapped reads), and one batch amplification control sample failing due to
238 detection of reads from a target pathogen (**Fig.5**). 39 samples failed internal extraction or amplification
239 QC, 38/39 (97%) failing to detect the MS2 extraction and amplification control; in addition, 2/39 (8%)
240 did not detect two IC viruses and 9/39 (23%) failed to detect any IC viruses (**Fig.3D**). Overall,
241 273/344 (79%) respiratory samples in the validation set passed negative and positive control QC
242 criteria.

243 In these 273 samples, there were 260 pathogen gold-standard target detections by multiplex PCR, with
244 212 (78%) samples positive for a single pathogen, 21 (8%) positive for two pathogen targets (42
245 targets in total), 2 (1%) positive for 3 pathogen targets (6 targets in total) (**Tab.1**). Within these
246 detections, six (2%) samples were positive for influenza A with additional detection of the H1 subtype
247 (counted as 2 detections, following the FilmArray panel). 38/273 (14%) samples passing QC were
248 PCR-negative for all pathogens.

249 133/260 (51%, 95%CI: 45-57%) PCR-positive results (i.e. gold-standard positives) were also detected
250 by sequencing (**Tab.3**, **Figs3D, 4B**). Of samples with multiple pathogens targets found by PCR, 1/21
251 (4.8%) had two pathogens targets detected by sequencing, while 0/2 (0%) had three pathogen targets
252 detected by sequencing. There were 127 false-negative results: 89 (34% of the 260 total results) had
253 no reads mapped to the target pathogen, 17 (7% of the total) had only a single mapped read and
254 therefore fell below the threshold designed to minimise barcode crossover, 20 (8%) had ≥2 reads and
255 met at least one of the three main criteria but failed the ratio criteria (as defined in **Tab.2**), and 1
256 (0.4%) had ≥2 mapped reads and achieved the ratio criteria but did not meet any of the three main
257 criteria (**Figs.3D, 4B**).

258 3901/3910 (99.8%) targets negative by PCR (as tested in the original assays) were also not detected
259 from sequencing (**Tab.3**); 38 samples were PCR-negative for all pathogens, of which 36 (95%) were
260 also negative by sequencing. There were nine false-positive pathogen detections by sequencing, all
261 potentially attributable to barcode crossover (**Fig.4B**): specifically a) four SARS-CoV2 detections with
262 two reads each, all on a single run that contained a sample with early cycle SARS-CoV2 detection on
263 PCR (Resp-4-Plex Cn=13.54; 51,158 reads); b) four coronavirus detections (3 HCoV-HKU1, 1
264 HCoV-OC43) with 2-38 reads from one run (Run 11) that contained a sample with 33,347 reads
265 mapping to HCoV-HKU1 (PCR by FilmArray, no Cn/Ct reported); and c) one metapneumovirus
266 detection with 79 reads from a run including two other metapneumovirus-positive samples with 48 and
267 810 reads (PCR by FilmArray).

268 The ratio criterion performed poorly in the validation set compared with the derivation set, resulting in
269 20 false-negatives solely from applying this criterion (**Figs.3D, 4B**). We therefore compared the main
270 and ratio criteria (**Tab.2**) with an alternative criterion which had relatively similar performance to
271 these criteria in the derivation set (**Tabs.S3, S5**). This criterion defined detection by the number of
272 reads in the sample mapping to the target reference genome as a percentage of reads on the run
273 mapping to the specific target genome (rather than all reference genomes) being >0.037%. Replacing
274 the main and ratio criteria with this alternative criterion (but retaining the ≥2 reads criterion) increased
275 the number of samples passing batch controls from 260 to 274, increased sensitivity to 58% (95% CI

276 52-64%; 159/274), and improved specificity to 99.9% (95% CI 99.8-100%; 4015/4017) with only 2
277 false-positive detections (**Tab.3, Fig.S4**), the metapneumovirus and coronavirus OC43 detections.

278 Performance by pathogen

279 While overall sensitivity using pre-specified criteria was 51%, assay performance varied substantially
280 by pathogen. Sequencing detected fourteen different RNA viruses in samples passing QC, but no
281 bacteria or DNA viruses. *Chlamydia pneumoniae* was detected (3 reads) in one sample whose positive
282 controls failed (**Fig.6A**). Among pathogen targets with more than 20 PCR-positive results, sensitivities
283 using pre-specified criteria were: RSV 85% (95% CI: 75-95%); SARS-CoV-2 65% (95% CI: 53-
284 78%); influenza A 58% (95% CI: 40-75%), and rhinovirus/enterovirus 19% (95% CI: 8-29%), with
285 small improvements using the alternative criteria for all except SARS-CoV-2 (**Fig.6B**).

286 The sample workflow was designed to optimise RNA pathogen detection. Excluding DNA pathogen
287 targets (adenovirus and bacteria) from the evaluation modestly increased sensitivity using the
288 alternative criteria to 61% (159/260) (95% CI 55-67%) (**Tab.3**). Enterovirus and rhinovirus infections
289 were frequently at low viral loads (**Fig.S5B**), of unclear significance;[25] removing these two targets
290 increased overall sensitivity using the alternative criteria further to 70% (145/207) (95% CI 63%-76%)
291 (**Tab.3**).

292 CMg provides greater taxonomic resolution and additional target detection

293 Ten rhinovirus/enterovirus FilmArray positive samples were further identified by sequencing as
294 Rhinovirus B3 (n=6), Rhinovirus C (n=1), Rhinovirus A89 (n=1), and Enterovirus A (n=2). All 41
295 samples with RSV detected by sequencing were further identified as either RSV-A (n=22) or RSV-B
296 (n=19). In addition, 13 samples with influenza A detected on 4-plex PCR were subtyped with
297 sequencing: 12 as influenza A/H1-2009 and one as influenza A/H3.

298 We detected two pathogens (rhinovirus and coronavirus OC43) by CMg sequencing that were not
299 assayed in the original PCR test in 2/111 (2%) samples originally tested by 4-plex. These were not on
300 runs with other strong positives, representing plausible additional detection.

301 ## Performance by viral load (Cn,Ct)

302 We compared PCR Cn/Ct values from the original Resp-4-Plex/Xpert platforms with Ct values from
303 study-specific qPCR following ss cDNA synthesis (**Fig.S5**). Study qPCR Ct values were consistently
304 5-10 cycles higher than original Cn/Ct values, potentially due to loss of material during processing
305 and/or to the use of different assays.

306 Samples with higher Ct values from either assay generated fewer sequencing reads as expected
307 (**Fig.7**). When using the alternative detection criteria and restricting analysis to samples with
308 detectable pathogen by qPCR, sensitivity increased to 75% (95% CI, 68-81%), and increased further
309 to 83% (95% CI, 76-88%) when restricting to qPCR Ct <35 (**Tab.3**).

310 ## Sequencing assay costs

311 The sequencing and consumables costs were calculated based on multiplexing 32 clinical samples on
312 each sequencing run at £102.41/sample (**Tab.S6**). Sequencing platform costs added a further
313 £9.61/sample. The analysis pipeline can be performed using GridION onboard computing capacity; no
314 additional server or compute costs were therefore incurred. In addition to £112.02 sequencing and
315 infrastructure costs, each sample required 70 minutes of labour.

316 # Discussion

---

317 Here, we developed and assessed the performance of an end-to-end metagenomic workflow to detect a
318 range of RNA respiratory viruses. By applying stringent quality thresholds, we achieved high
319 specificity both for target pathogens (99.8%) and at an individual sample level (>96%). Sensitivity
320 using pre-defined thresholds in a large validation set was modest compared to multiplex PCR, being
321 51% overall, ranging from 85% for RSV to 19% for rhino/enterovirus. This could be improved to 83%
322 by limiting analysis to RNA viruses with Ct <35 and using post-hoc exploratory bioinformatic criteria.
323 The final cost was £112/sample, achieved by 32-multiplexing across one flow cell. While cost
324 compares favourably with Filmarray, and the sequencing cost was 45% less than an Illumina-based

325 respiratory virus sequencing assay,[17] the hands-on and labour time of the assay was substantial
326 (estimated 70 minutes/sample).

327 The ultimate goal of the assay was agnostic pathogen detection: however, a necessary first step was
328 reliable detection of known targets from routine testing. We therefore developed a mapping-based
329 analysis pipeline, the most sensitive approach to demonstrate performance against known targets.[22]
330 We validated this workflow against PCR targets using prospectively collected samples including 19/23
331 multiplex-PCR targets, substantially broader than the 4-8 targets validated in other RNA virus CMg
332 workflows to date.[16,17]

333 The varying sensitivities obtained highlight challenges of CMg for detecting infections with high host
334 and low pathogen biomass (as from nasopharyngeal swabs). Despite human cell depletion and RNA
335 amplification, only 66% of samples yielded sequence reads for the detected PCR target, and 20% of
336 gold-standard detections were not found by study qPCR. Such relatively low yield of pathogen reads
337 has also been observed in other respiratory CMg workflows. One workflow required only 3 non-
338 overlapping pathogen reads from a minimum of 5,000,000 per sample,[17] while two studies reported a
339 single positive read on ONT sequencing as the threshold for virus detection.[11,16] Our results
340 demonstrate that the consequence of relaxed thresholds is false-positive detections, a particular danger
341 with imperfect barcode de-multiplexing.

342 The assay performed particularly poorly for rhinovirus/enterovirus. Genome size (7kb vs 13-30kb
343 other viruses) may contribute to this poor sensitivity (**Tab.S2**). Additionally, differences between
344 study sequences and the references used for mapping (**Tab.S2**) were greatest for
345 rhinovirus/enterovirus (**Fig.S6**), suggesting that species diversity might also affect performance.

346 Our findings contrast with a recent UK respiratory CMg study, reporting sensitivity for viral detection
347 of 89% (95%CI: 71-98),[14] likely for multiple reasons. Alcolea-Medina et al sampled patients in critical
348 care, with >85% of viral detections from lower respiratory tract samples, where common respiratory
349 viruses have higher viral loads.[26,27] Moreover, sparse host material in BAL samples allowed
350 investigators to omit host depletion steps.[14] Finally, the study multiplexed ≤5 samples, and 66% of

included samples had no respiratory viruses detected by PCR testing. The lower risk of barcode cross-over allowed only a single read to indicate pathogen detection,[11] albeit at greater sequencing costs.

RNA extraction followed by SISPA, as employed here, has high hands-on time and requires more specialised laboratory practitioners than commercial multiplex-PCRs. These technical aspects and the modest sensitivity of the workflow mean that it will not be scalable for routine diagnosis. However, metagenomic workflows using established laboratory methods and limited specialised equipment may be valuable for pandemic preparedness, since RNA viruses are likely candidates for emerging pathogens.[28–30]

CMg undoubtedly has diagnostic potential. However, with a laboratory workflow optimised for RNA extraction, and bioinformatics optimised to stringent recognition of sample and barcode cross-over, this study demonstrates the technical limitations of high-throughput metagenomic sequencing for routine viral diagnostics in upper respiratory tract samples with high host and low pathogen biomass. Its improved performance at higher viral loads demonstrates the importance of identifying the most valuable use cases for CMg.

Role of the Funding source

The study funder of this study had no role in the study design, data collection, analysis or interpretation, or in the writing of this report.

Data Sharing

Deidentified patient data is included in supplementary material. All sequence data are available at the ENA (project number PRJEB102181).

Code Availability

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

374 ## References

375 1 UKHSA. Infectious diseases impacting England, 2025 report.
376 https://assets.publishing.service.gov.uk/media/67fe0e08712bf73dea13541d/Infectious_dise
377 ases_impacting_England_2025_report_2.pdf.

378 2 Stojanovic Z, Gonçalves-Carvalho F, Marín A, *et al.* Advances in diagnostic tools for
379 respiratory tract infections: from tuberculosis to COVID-19 - changing paradigms? *ERJ*
380 *Open Res* 2022; **8**: 00113–2022.

381 3 Charlton CL, Babady E, Ginocchio CC, *et al.* Practical Guidance for Clinical Microbiology
382 Laboratories: Viruses Causing Acute Respiratory Tract Infections. *Clin Microbiol Rev*
383 2019; **32**: e00042-18.

384 4 UK Health Security Agency. UK Standards for Microbiology Investigations. 2025.
385 https://www.rcpath.org/profession/publications/standards-for-microbiology-
386 investigations.html (accessed Feb 2, 2026).

387 5 Yanagihara K. The role of molecular diagnosis in acute respiratory tract infection. *Respir*
388 *Investig* 2019; **57**: 511.

389 6 Evans SE, Jennerich AL, Azar MM, *et al.* Nucleic Acid-based Testing for Noninfluenza
390 Viral Pathogens in Adults with Suspected Community-acquired Pneumonia. An Official
391 American Thoracic Society Clinical Practice Guideline. *Am J Respir Crit Care Med* 2021;
392 **203**: 1070–87.

393 7 Bouzid D, Casalino E, Mullaert J, *et al.* Added value of rapid respiratory syndromic testing
394 at point of care versus central laboratory testing: a controlled clinical trial. *J Antimicrob*
395 *Chemother* 2021; **76**: iii20–7.

396 8 Blais J, Lavoie SB, Giroux S, *et al.* Risk of Misdiagnosis Due to Allele Dropout and False-
397 Positive PCR Artifacts in Molecular Diagnostics: Analysis of 30,769 Genotypes. *J Mol*
398 *Diagn JMD* 2015; **17**: 505–14.

399 9 Castañeda-Mogollón D, Kamaliddin C, Oberding L, *et al.* A metagenomics workflow for
400 SARS-CoV-2 identification, co-pathogen detection, and overall diversity. *J Clin Virol Off*
401 *Publ Pan Am Soc Clin Virol* 2021; **145**: 105025.

402 10 Heitz M, Levrat A, Lazarevic V, *et al.* Metagenomics for the microbiological diagnosis of
403 hospital-acquired pneumonia and ventilator-associated pneumonia (HAP/VAP) in intensive
404 care unit (ICU): a proof-of-concept study. *Respir Res* 2023; **24**: 285.

405 11 Charalampous T, Alcolea-Medina A, Snell LB, *et al.* Routine Metagenomics Service for
406 ICU Patients with Respiratory Infection. *Am J Respir Crit Care Med* 2024; **209**: 164–74.

407 12 Hsu W-Y, Kao T-W, Cho H-C, *et al.* Performance of a hybrid capture-based target
408 enrichment next-generation sequencing for the identification of respiratory pathogens and
409 resistance-associated genes in patients with severe pneumonia. *Microbiol Spectr* 2025; **13**:
410 e0213024.

411 13 Chen H, Huang Q, Wu W, *et al.* Assessment and clinical utility of metagenomic next-
412 generation sequencing for suspected lower respiratory tract infections. *Eur J Med Res*
413 2024; **29**: 213.

414 14 Alcolea-Medina A, Snell LB, Humayun G, *et al.* Rapid pan-microbial metagenomics for
415 pathogen detection and personalised therapy in the intensive care unit: a single-centre
416 prospective observational study. *Lancet Microbe* 2025; **6**.
417 DOI:10.1016/j.lanmic.2025.101174.

418 15 Alcolea-Medina A, Alder C, Snell LB, *et al.* Unified metagenomic method for rapid
419 detection of microorganisms in clinical samples. *Commun Med* 2024; **4**: 135.

420 16 Gauthier NPG, Chan W, Locher K, *et al.* Validation of an Automated, End-to-End
421 Metagenomic Sequencing Assay for Agnostic Detection of Respiratory Viruses. *J Infect*
422 *Dis* 2024; **230**: e1245–53.

423 17 Tan JK, Servellita V, Stryke D, *et al.* Laboratory validation of a clinical metagenomic next-
424 generation sequencing assay for respiratory virus detection and discovery. *Nat Commun*
425 2024; **15**: 9016.

426 18 Yin Y, Zhu P, Guo Y, *et al.* Enhancing lower respiratory tract infection diagnosis:
427 implementation and clinical assessment of multiplex PCR-based and hybrid capture-based
428 targeted next-generation sequencing. *EBioMedicine* 2024; **107**: 105307.

429 19 Guo Y, Li Z, Li L, *et al.* A dual-process of targeted and unbiased Nanopore sequencing
430 enables accurate and rapid diagnosis of lower respiratory infections. *EBioMedicine* 2023;
431 **98**: 104858.

432 20 Diao Z, Zhang Y, Chen Y, *et al.* Assessing the Quality of Metagenomic Next-Generation
433 Sequencing for Pathogen Detection in Lower Respiratory Infections. *Clin Chem* 2023; **69**:
434 1038–49.

435 21 Torres Montaguth OE, Buddle S, Morfopoulou S, Breuer J. Clinical metagenomics for
436 diagnosis and surveillance of viral pathogens. *Nat Rev Microbiol* 2025; : 1–15.

437 22 Marić J, Križanović K, Riondet S, Nagarajan N, Šikić M. Comparative analysis of
438 metagenomic classifiers for long-read sequencing datasets. *BMC Bioinformatics* 2024; **25**:
439 15.

440 23 Cane J, Sanderson N, Barnett S, *et al.* Nanopore sequencing of influenza A and B in
441 Oxfordshire and the United Kingdom, 2022–23. *J Infect* 2024; **88**: 106164.

442 24 Street TL, Bejon P, Leach L, Oakley S, Young BC, Sanderson ND. Nanopore adaptive
443 sampling for bacterial identification from periprosthetic joint replacement tissue. *Microb*
444 *Genomics* 2025; **11**. DOI:10.1099/mgen.0.001507.

445 25 Wright PF, Deatly AM, Karron RA, *et al.* Comparison of Results of Detection of
446 Rhinovirus by PCR and Viral Culture in Human Nasal Wash Specimens from Subjects
447 with and without Clinical Symptoms of Respiratory Illness. *J Clin Microbiol* 2007; **45**:
448 2126–9.

26 Bouzid D, Hingrat QL, Salipante F, *et al.* Agreement of respiratory viruses' detection between nasopharyngeal swab and bronchoalveolar lavage in adults admitted for pneumonia: a retrospective study. *Clin Microbiol Infect* 2023; **29**: 942.e1-942.e6.

27 Lin C, Xiang J, Yan M, Li H, Huang S, Shen C. Comparison of throat swabs and sputum specimens for viral nucleic acid detection in 52 cases of novel coronavirus (SARS-Cov-2)-infected pneumonia (COVID-19). *Clin Chem Lab Med* 2020; **58**: 1089–94.

28 Ameni G, Zewude A, Tulu B, *et al.* A Narrative Review on the Pandemic Zoonotic RNA Virus Infections Occurred During the Last 25 Years. *J Epidemiol Glob Health* 2024; **14**: 1397–412.

29 Carrasco-Hernandez R, Jácome R, López Vidal Y, Ponce de León S. Are RNA Viruses Candidate Agents for the Next Global Pandemic? A Review. *ILAR J* 2017; **58**: 343–58.

30 Garcia-Blanco MA, Ooi EE, Sessions OM. RNA Viruses, Pandemics and Anticipatory Preparedness. *Viruses* 2022; **14**: 2176.

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. Accessed and verified the data: N.D.S and B.C.Y. Drafted the work or substantively revised it: N.D.S, P.B., A.S.W, B.C.Y. All authors have had full access to the study data and accept responsibility to submit for publication

Acknowledgements

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.

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

489 **(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)

490 * tests for influenza A/B, RSV and SARS-CoV-2 (**Tab.S1**).
491 ** tests for adenovirus, coronaviruses (229E, HKU1, NL63, OC43, SARS-CoV-2, MERS-CoV),
492 human metapneumovirus, human rhino/enterovirus, influenza viruses (including subtypes influenza A,
493 A/H1, A/H3, A/H1-2009), influenza B virus, parainfluenza viruses (1-4), respiratory syncytial virus,
494 *Bordatella parapertussis*, *Bordetella pertussis*, *Chlamydia pneumoniae* and *Mycoplasma pneumoniae*
495 (**Tab.S1**).
496 † Original derivation samples were split and sequenced multiple times. Some sequenced samples
497 included pooled samples tested with multiple platforms, see ‡. For the validation set each sample was
498 sequenced once only.
499 †† Samples that passed run, batch and sample internal controls, see **Tab.2**.
500 ‡ One derivation set sample was pooled and tested using both Resp-4-plex and FilmArray; this was
501 sequenced 3 times.
502 ‡‡ Includes influenza subtypes as separate pathogen targets.
503 ° Included in error

504 **(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)

505 * missing values from anonymised trial samples. Percentages of non-missing values.
506 ** DNA target or multiple pathogens identified on routine testing and insufficient sample to run qPCR
507 for all (see **supplemental methods**)
508 Note: showing n (%) or median (IQR). QC=quality control. See **Fig.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.

Type	Threshold for failure (any)	Applied to	Rationale
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.
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 **Fig.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 (**Tab.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 **Tab.1**.

Figure 1 Clinical metagenomic workflow overview

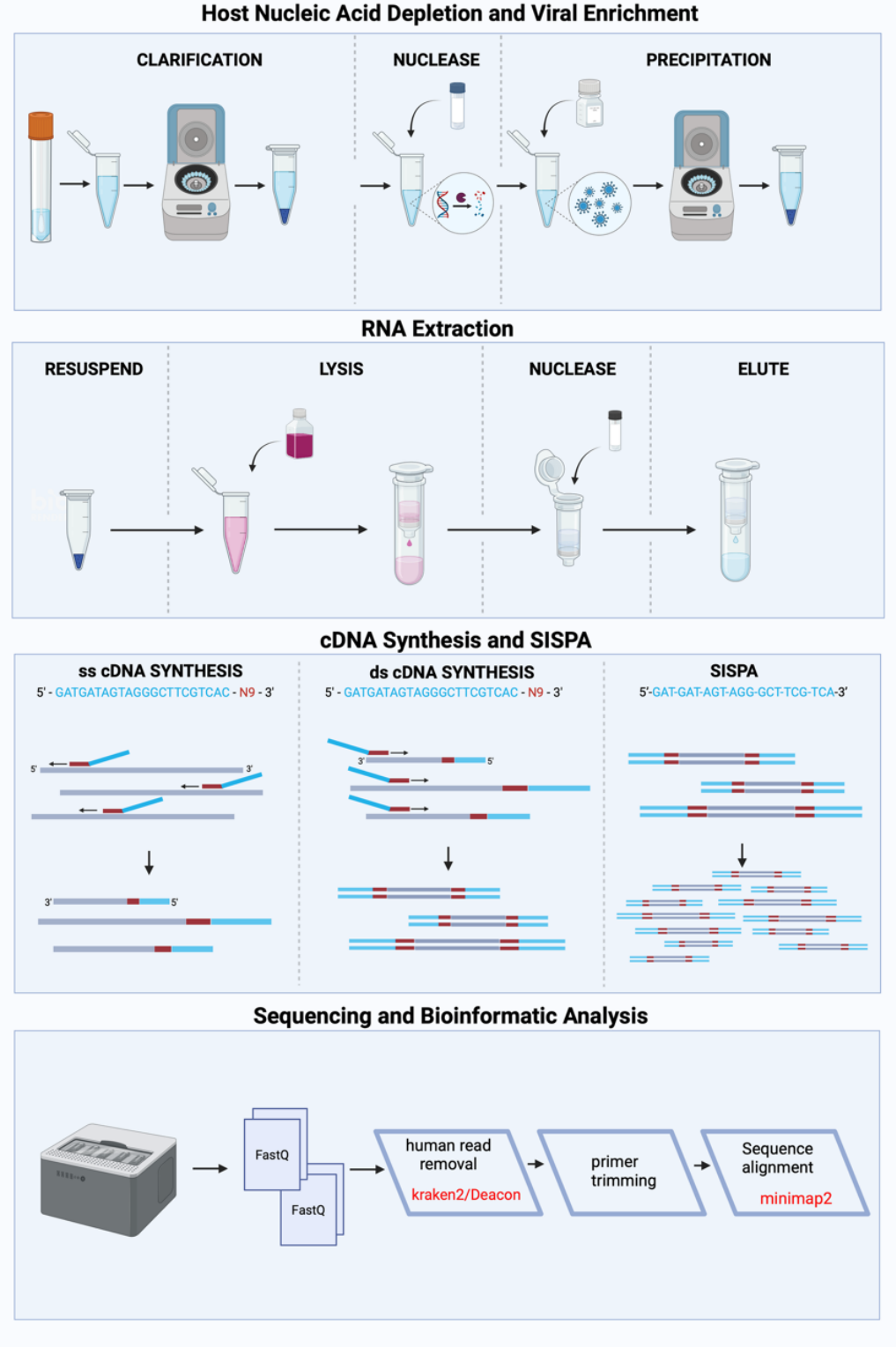


Figure 2 Pathogens identified (A, B) and not identified (C,D) on routine diagnostic testing in (A, C) derivation and (B, D) validation sets. Each sample could be positive for multiple pathogens, see **Table 1**.

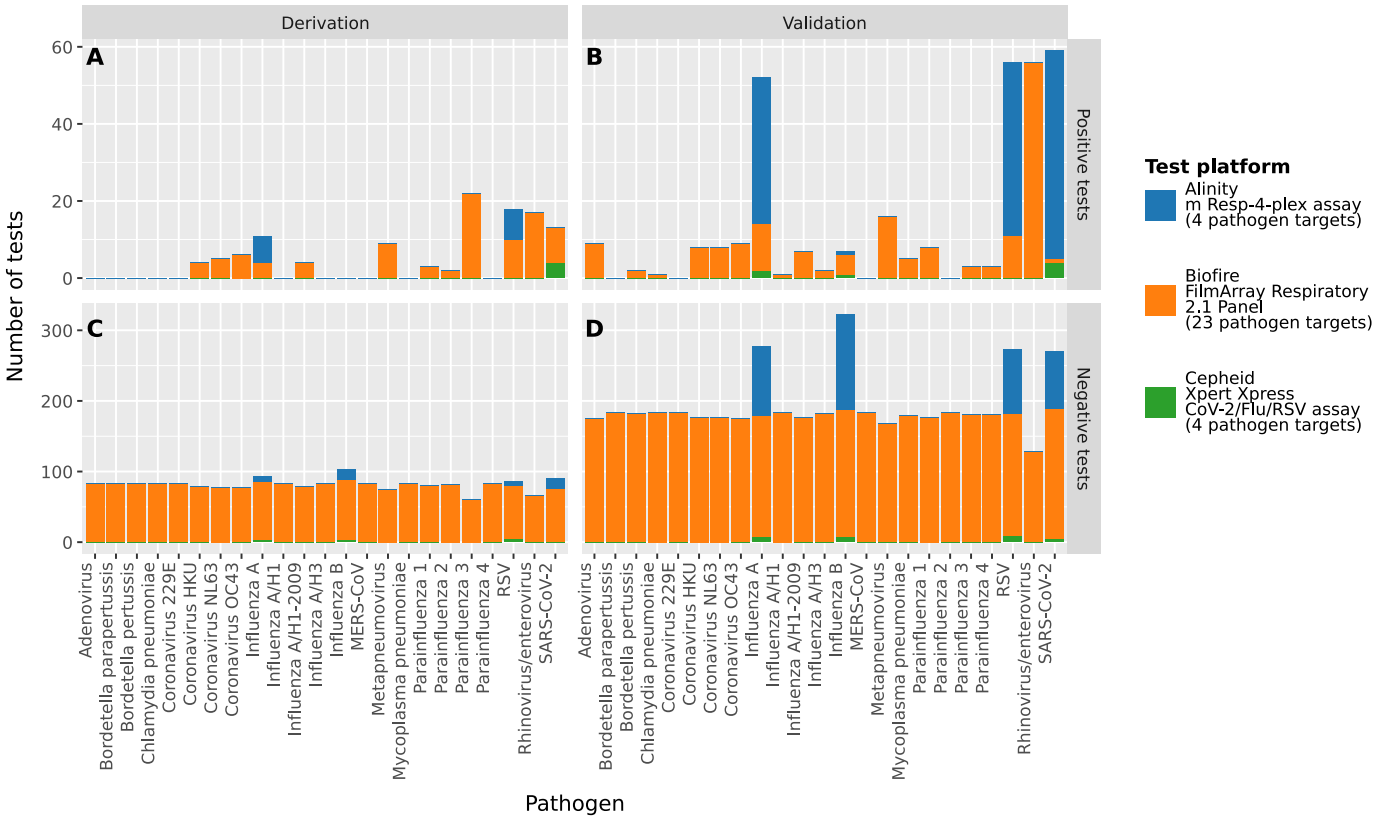


Figure 3 Batch control (A, B) and positive control (C, D) passed per sample and target detection (E, F) criteria passed for (A, C, E) derivation and (B, D, F) validation sets. Red indicates passed and black failed. RNA from murine respirovirus, orthoreovirus and Zika virus were spiked as cDNA synthesis 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.

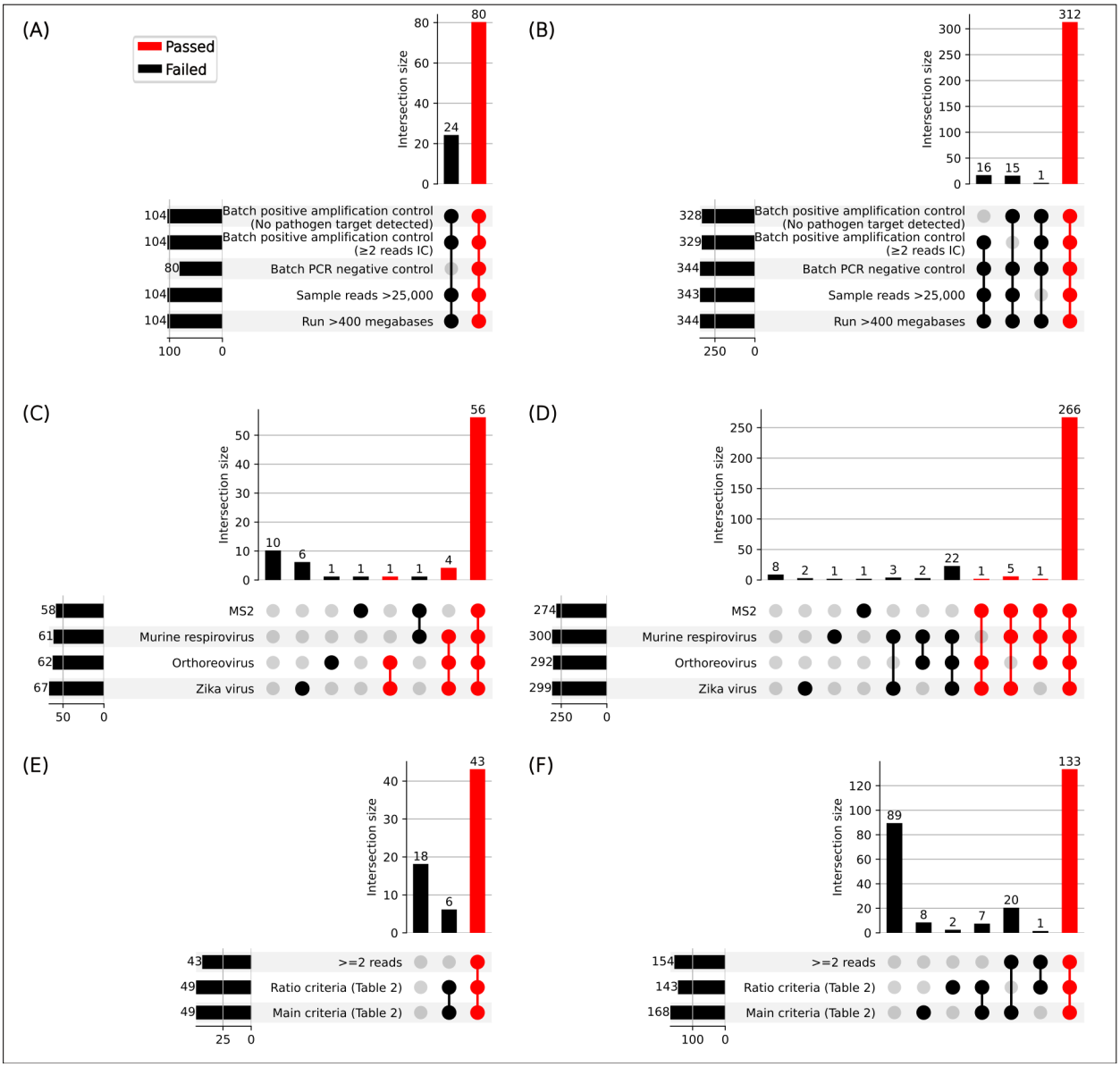


Figure 4 Positivity ratio criteria as developed in the derivation set (A,C) and tested in the validation set (B,D) Showing samples which passed sample QC thresholds (**Tab.2**) in panels A, B and samples which failed sample QC thresholds in panels C, D. The sample QC pass (panels A, B) includes samples that failed batch controls, to provide context for barcode crossover. 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 (**Tab.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.

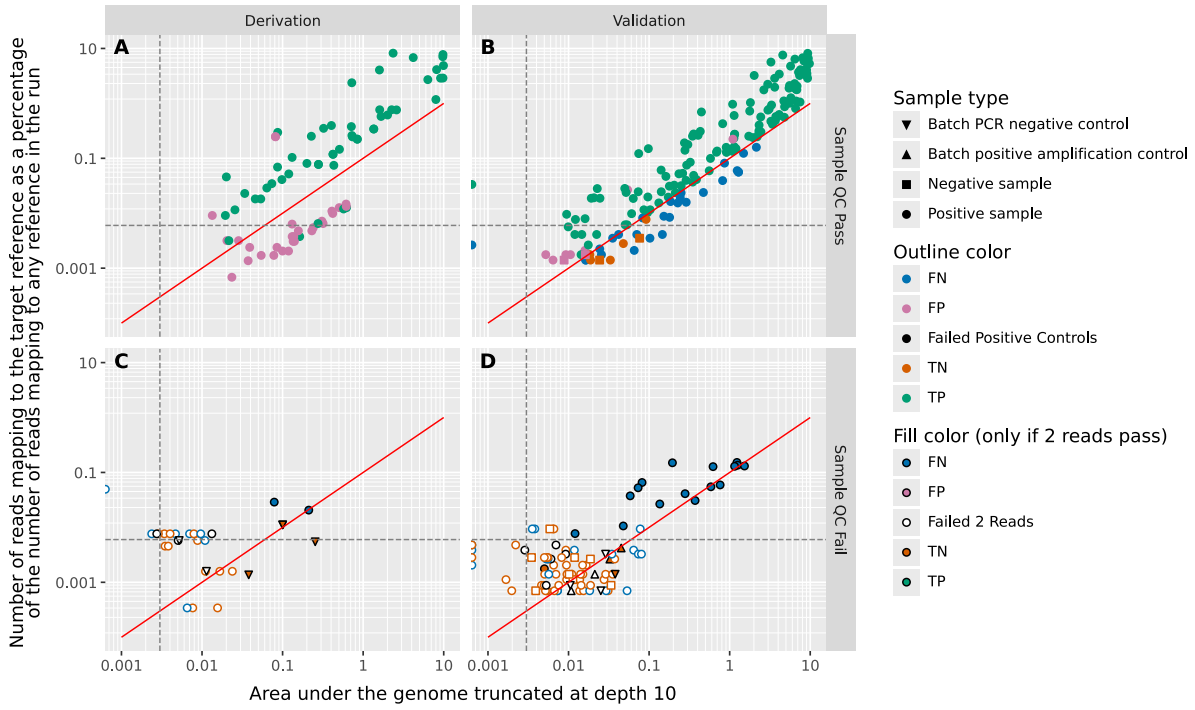


Figure 5 Validation set sample flow

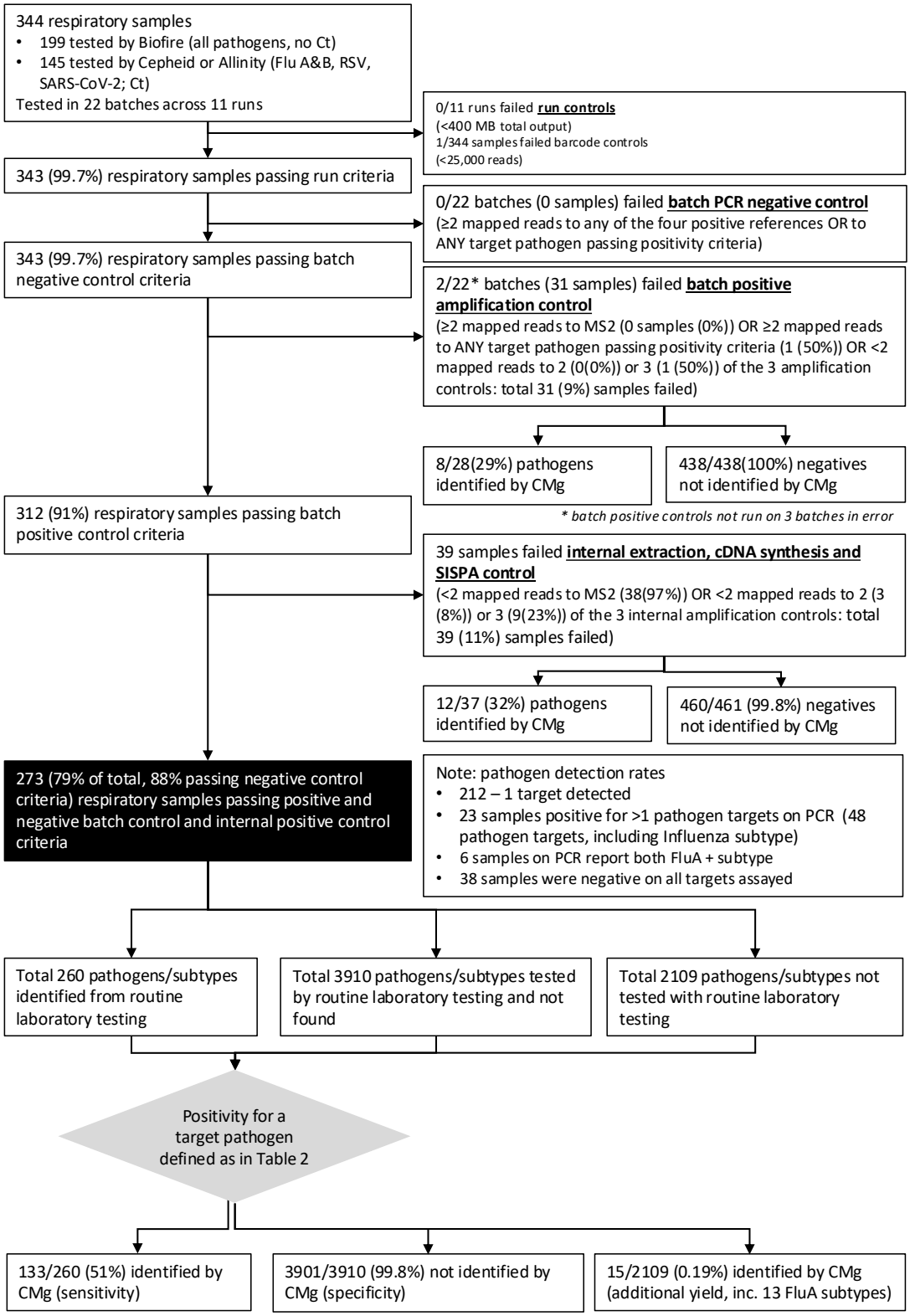


Figure 6 Pathogen-specific detections (A) and sensitivity (if >20 gold-standard detections) (B) in the derivation and validation sets. 95% confidence intervals calculated using Clopper-Pearson interval based on Beta distribution.

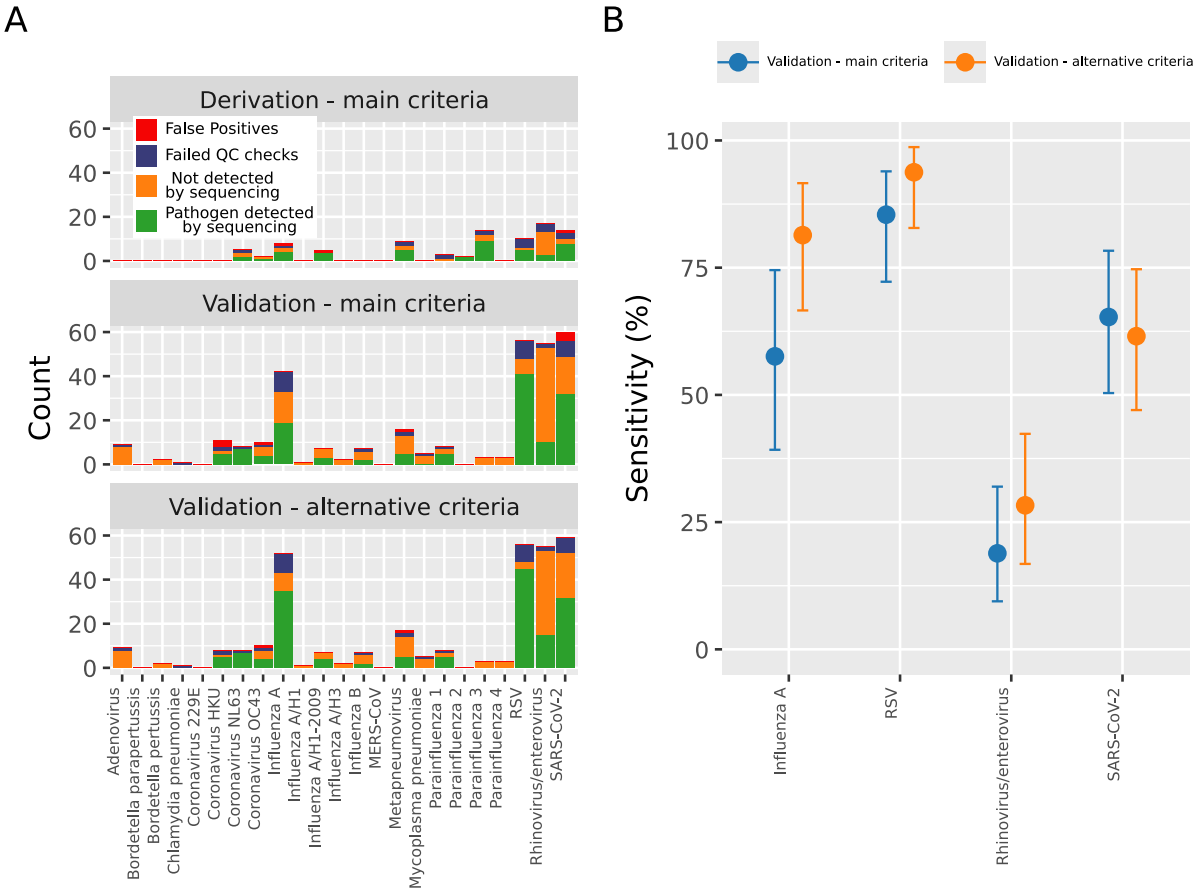
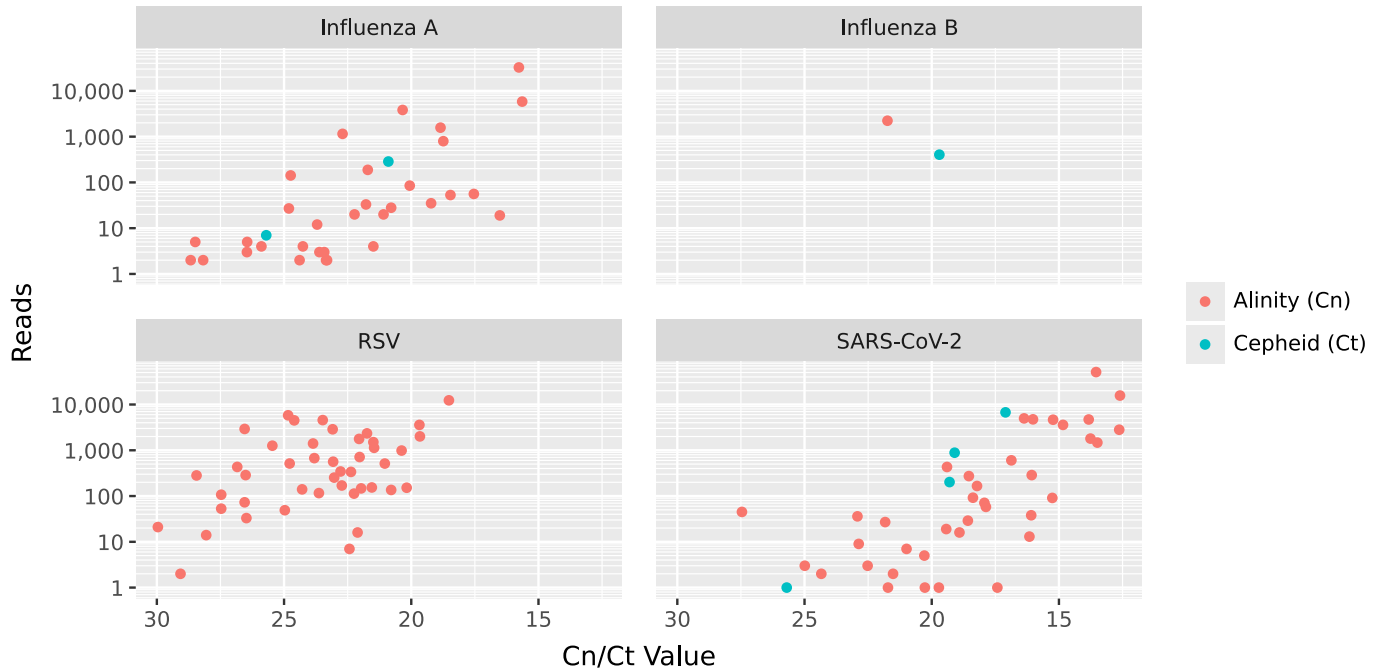
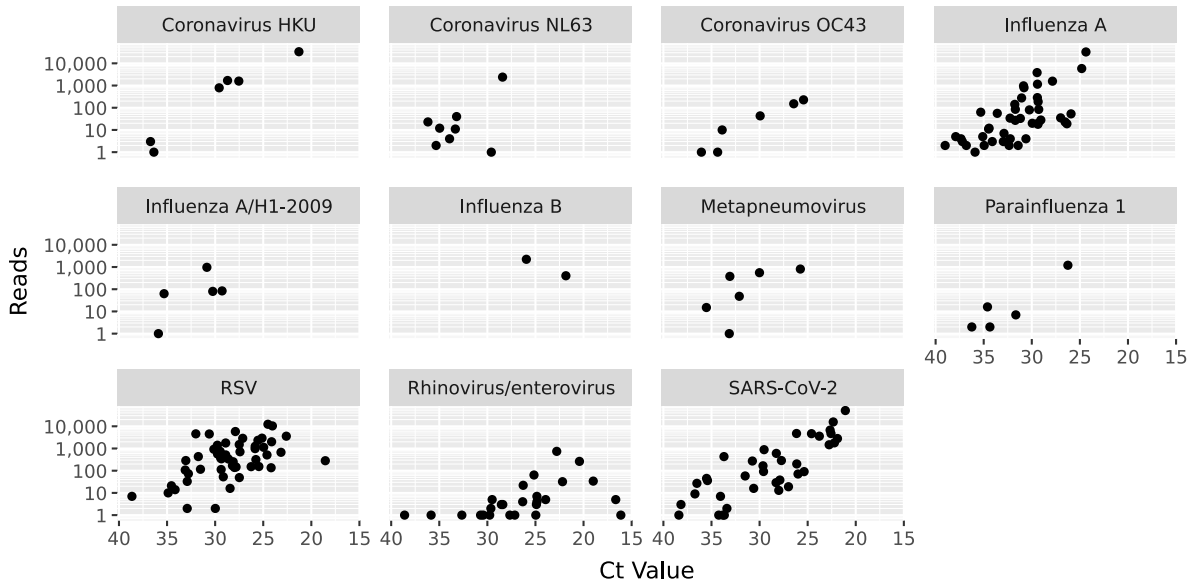


Figure 7 Association between number of mapped reads and Cn/Ct values from original assay (A) or study qPCR (B)

(A)



(B)