1   # Evaluation of efficiency and sensitivity of 1D and 2D sample

2   # pooling strategies for diagnostic screening purposes

3

4   **Jasper Verwilt[1,2,3], Pieter Mestdagh[1,2,3,4], Jo Vandesompele[1,2,3,4]**

5   *1 OncoRNALab, Cancer Research Institute Ghent*

6   *2 Department of Biomolecular Medicine, Faculty of Medicine and Health Sciences,*

7   *Ghent University*

8   *3 Center for Medical Genetics, Ghent University*

9   *4 Biogazelle, Belgium*

10

11   # Abstract

12   As SARS-CoV-2 continues to spread around the world while the pandemic lasts,

13   testing facilities are forced to massively increment their testing capacities to handle

14   the increasing number of samples. While sample pooling methods have been

15   proposed or are effectively implemented in some labs, no systematic and large-scale

16   simulations have been performed using real-life quantitative data from testing

17   facilities. Here, we use anonymous data from 1632 positive cases to simulate and

18   compare 1D and 2D pooling strategies. We show that the choice of pooling method

19   and pool size is an intricate decision with a prevalence-dependent efficiency-

20   sensitivity trade-off.

21

22

23

24

Introduction

Massive screening of the population for SARS-CoV-2 infection is proposed as one of the key strategies in the global battle against the COVID-19 pandemic. Due to the immense number of samples that are analyzed during population screening, pooling of samples seems to be a valid strategy in overcoming shortages in reagents and increasing testing capacity.

Several pooling strategies in the frame of SARS-CoV-2 testing have been discussed in recent preprints. The most discussed method is referred to as one-time pooling[1–4]. In this strategy, the samples are pooled, pools are tested and only samples in positive pools are tested individually. A second popular approach is called sequential pooling[1,5] in which the samples are pooled, pools are tested and positive pools are split into two equally-sized sub-pools. These sub-pools are tested again and this process is repeated until the individual sample level is reached. The third method that has been explored is two-dimensional or 2D pooling[6]. This method organizes samples in a 2D matrix and then creates pools along the rows and columns of the matrix. The pools are tested and negative rows and columns are excluded from the matrix. Next, all remaining samples are tested individually. Finally, researchers have explored pooling strategies in which samples are assigned to multiple pools, but require only one round of testing. Given the composition and test results of the positive pools, the positive samples can immediately be identified without the need for further individual testing. P-BEST[7] and Tapestry[8] are examples of such strategies.

While attractive, pooling strategies come with inherent limitations. First, pooling dilutes each individual sample, possibly to such an extent that the viral RNA

50   becomes undetectable, resulting in false negative observations[9–11]. A second

51   limitation is that an increase in sample manipulations augments the risk of cross-

52   contamination and sample mix-ups, which can lead to false negatives and false

53   positives[6]. Other drawbacks are unique to specific strategies. The P-BEST pooling

54   protocol is very time consuming, even when using a pipetting robot[7]. The repeated

55   pooling method, on the other hand, suffers from a complicated re-pooling scheme[1].

56

57   Although the number of preprints and peer-reviewed publications on pooling

58   strategies for COVID-19 PCR-based testing has increased rapidly throughout the

59   pandemic, some important insights are still lacking. First of all, the proposed optimal

60   pooling strategy is often based on a binary classification of samples as either positive

61   or negative. However, this Boolean approach is not in accordance with the real-world

62   situation and does not allow for investigating the dilution effect of pooling. Second,

63   when using a quantitative representation of the viral loads, these values need to

64   reflect real-life data, as patients present a very wide range of viral loads. This is

65   reflected in the wide spectrum of Cq values reported by PCR-based tests[12]. Finally,

66   as pooling is most effective for population-wide screening (where a very low

67   prevalence is expected) it is important to determine the performance of strategies

68   when encountering a low fraction of positive samples.

69

70   Here, we evaluate one-time (or 1D) pooling and two-dimensional (2D) pooling (using

71   practical microtiter plate format pool sizes) as promising strategies for massive, low-

72   prevalence population screening using real-life RT-qPCR data from 1632 positive

73   samples.

74

Results

75

76 *Inverse relationship between efficiency and prevalence*

77 In order to evaluate the efficiency-gain of the adopted pooling strategies, we

78 calculated the number of tests that are needed to analyze all samples and divided

79 this number by the total number of samples. We calculated the median, minimum

80 and maximum prevalence-specific efficiency for each pooling strategy (Figure 1).

81 First, we observe an inverse relationship between efficiency and prevalence over the

82 evaluated prevalence range from 0.01% to 10% for all pooling strategies. Second,

83 there is no single most efficient strategy, because this depends on the prevalence

84 (notice crossings in Figure 1). Until a prevalence of 0.36%, 1x24 is the most efficient

85 strategy, from 0.40% to 2.51% 16x24 becomes the most efficient, from 2.82% to

86 4.47% the most efficient strategy is 12x24 and from 5.01% to 10% 8x12 is the most

87 efficient strategy. However, at high prevalence all strategies show similar efficiency

88 gain. Third, strategies employing a larger pool size display a higher efficiency when

89 the prevalence is low, but as the prevalence increases, there is a tipping point for

90 each strategy at which its smaller pool size variant becomes more efficient. At very

91 low prevalence, the efficiency of a 1 x n pool size becomes n and that of a m x n pool

92 strategy becomes $(m \times n) / (m + n)$. As a general trend, 2D pooling methods are less

93 sensitive to changes in prevalence in comparison with 1D pooling methods. We

94 conclude that the most efficient pool size very much depends on the prevalence, but

95 2D pooling methods generally are most efficient when prevalence is higher than

96 0.4%.

97

98 *Sensitivity decreases with lower prevalence*

99   The true performance of a pooling strategy cannot be evaluated by efficiency only.

100   Since the number of false negatives due to pooling is one of the main possible

101   drawbacks of sample pooling, it is necessary to take this into account when choosing

102   the optimal strategy. In this regard, we calculated the average sensitivity of the

103   different simulated settings (Figure 2, Supplemental Figure 3), which ranges from

104   0.636 to 0.968. For all of the pooling strategies—and irrespective of pool size—we

105   primarily see that sensitivity increases with increasing prevalence. There is a non-

106   linear relationship between sensitivity and prevalence. Furthermore, since prevalence

107   is linked with efficiency and as a result indirectly linked with sensitivity, we note that

108   an increase in efficiency comes with a decrease in sensitivity. We also observe that

109   at a prevalence lower than 1%, there is an increased variation in sensitivity for

110   different simulation cohorts. Thus, the increased efficiency at low prevalence comes

111   with a low and pool size-dependent problematically variable sensitivity.

112

113   *Sensitivity loss in function of viral load*

114   With the intention of determining the influence of the viral load and, as a proxy, Cq

115   value of the positive sample on the sensitivity associated with the pooling strategies,

116   we calculated the probabilities of a true-positive for each non-pooled original Cq

117   value (Supplemental Figure 2). We investigated how the sensitivity changes when

118   higher Cq values are progressively being included in the cohort, starting with

119   samples with highest viral load (lowest Cq value) (Figure 3, Supplemental Figure 4).

120   In the first place, we note that the Cq value at which sensitivity loss starts to occur

121   only depends on the 1D pool size or largest dimension of the 2D pool; i.e. Cq value

122   of 35 for 1x4, 34 for 1x8, 33.4 for 1x12 and 8x12, 33 for 1x16 and 12x16 and 32.4 for

123   1x24 and 16x24. These Cq values are (as expected) identical to

124    $37 - \log_2(largest\ pool\ size)$. Additionally, when Cq values larger than these cut-off

125   values are systematically included, the sensitivity drops exponentially. The rate at

126   which this reduction happens, decreases when prevalence increases. Finally, when

127   the prevalence is 10%, larger pool sizes result in a smaller reduction in sensitivity,

128   but for all other visualized prevalence values the sensitivity decreases with larger

129   pool size. Altogether, the extent to which low viral load samples contribute to the drop

130   in sensitivity depends on pool size and pooling strategy, although the sensitivity

131   decrease is most problematic when prevalence is low.

132

Discussion

134   Sample pooling strategies form an incredible asset in an attempt to increase

135   throughput in times when massive testing for COVID-19 would be needed. A plethora

136   of pooling strategies have been suggested, some more performant or practical than

137   others. 1D and 2D pooling methods were selected in this simulation study because

138   they are simple and quick to perform, have straightforward pipetting schemes and do

139   not require re-accessing the same sample more than twice. Pool sizes were selected

140   to be easily compatible with 96-well plates. Using a large real-life dataset of 1632

141   positive samples enabled us to simulate relevant settings and provide more accurate

142   outcomes in comparison with real wet-lab tests, which are rather limited in the

143   number of positive samples and may not cover the whole range of viral loads.

144   Because the original Cq distribution depends on the origin of the sample (hospital,

145   care center, …) and the stage of pandemic, our observations do as well. Firstly, our

146   results confirm the widely accepted idea that sample pooling methods show a higher

147   efficiency when prevalence is low[1–6,13] and that, for 1D and 2D pooling methods, as

148   prevalence increases, a threshold is reached after which smaller pool sizes become

149 more efficient[1,6]. However, appraising the performance of a pooling method

150 exclusively by its efficiency would ignore one of the major drawbacks of pooling: loss

151 of sensitivity due to dilution of the target. This issue becomes most pertinent when

152 the viral load is low[9–11,13,14]. Our results confirm that all tested pooling methods suffer

153 from false negatives, to a variable degree (Figure 2). This loss in sensitivity across all

154 prevalence conditions generally precludes use of pooling for diagnostic testing of

155 COVID-19 samples according to the U.S. Food and Drug Administration (FDA),

156 whereby only 1x16 and 1x24 strategies under high prevalence (≥10%) conditions

157 meet the minimal 95% sensitivity requirement[15]. When prevalence is high, the loss in

158 sensitivity for large pools is partly compensated by the fact that low viral load

159 samples can be 'rescued' by high viral load samples when present in the same pool.

160 (Figure 3). Intuitively, 2D pooling methods are especially vulnerable for false

161 negatives, as a high Cq sample would have to be 'rescued' in the corresponding row

162 and column pools, which is confirmed by our results. The influential role of

163 prevalence on efficiency as well as sensitivity presents an import challenge,

164 considering that, in order to make an informed decision on the preferential pooling

165 strategy, the prevalence has to be known. By nature, we cannot know the exact

166 prevalence before testing our samples, and as a result, the prevalence has to be

167 estimated. In general, we show that it is of extreme importance that an optimal

168 equilibrium between efficiency and sensitivity is achieved when deciding on the

169 pooling strategy and corresponding pool size.

170

171 Materials and Methods

172 *Patient samples*

173   Nasopharyngeal swabs were taken by a healthcare professional as a diagnostic test

174   for SARS-CoV-2, as part of the Belgian national testing platform. The individuals

175   were tested at nursing homes or in triage centers, between April 9th and June 7th. To

176   mimic low prevalence viral loads as much as possible, only batches of 94 patient

177   samples with fewer than 10 positives were included in this study. After additional

178   filtering as described in a further paragraph, this resulted in 113 928 patients in total,

179   of which 1632 positives (1.43%) with corrected Cq values ranging from 9.85 to 36.94

180   (median of 28.78) (Supplemental Figure 1).

181

182   *SARS-CoV-2 RT-qPCR test*

183   RNA extraction was performed using the Total RNA Purification Kit (Norgen Biotek

184   #24300) according to the manufacturer's instructions using 200 µl transport medium,

185   200 µl lysis buffer and 200 µl ethanol, with processing using a centrifuge (5810R with

186   rotor A-4-81, both from Eppendorf). RNA was eluted from the plates using 50 µl

187   elution buffer (nuclease-free water), resulting in approximately 45 µl eluate. RNA

188   extractions were simultaneously performed for 94 patient samples and 2 negative

189   controls (nuclease-free water). After addition of the lysis buffer, 4 µl of a proprietary

190   700 nucleotides spike-in control RNA (40 000 copies) and carrier RNA (200 ng of

191   yeast tRNA (Roche #10109517001) was added to all 96 wells from the plate). To the

192   eluate of one of the negative control wells, 7500 RNA copies of positive control RNA

193   (Synthetic SARS-CoV-2 RNA Control 2, Twist Biosciences #102024) were added.

194   Six µl of RNA eluate was used as input for a 20 µl RT-qPCR reaction in a CFX384

195   qPCR instrument using 10 µl iTaq one-step RT-qPCR mastermix (Bio-Rad

196   #1725141) according to the manufacturer's instructions, using 250 nM final

197   concentration of primers and 400 nM of hydrolysis probe. Primers and probes were

198 synthesized by Integrated DNA Technologies using clean-room GMP production. For

199 detection of the SARS-CoV-2 virus, the Charité E gene assay was used (FAM)[16]; for

200 the internal control, a proprietary hydrolysis probe assay (HEX) was used. Prior to

201 May 25th, 2 singleplex assays were performed; after May 25th, 1 duplex RT-qPCR

202 was performed (with $1/8^{th}$ of spike-in control RNA). Cq values were generated using

203 the FastFinder software v3.300.5 (UgenTec). Only batches were approved with a

204 clean negative control and a positive control in the expected range.

205

206 *Assembly of positive Cq value distribution*

207 First, a subset of full 96-well RNA plates is selected containing less than 10 positives.

208 The latter selection criterium is introduced to filter out plates originating from high

209 COVID-19 positive rate circumstances, such as hospitals, to avoid putative bias

210 towards high viral loads. Only plates are retained with a positive control value that is

211 within two standard deviations of the mean positive control value of all filtered plates.

212 Next, in order to correct the Cq values for inter-plate variation, the difference of the

213 average positive control Cq values per qPCR plate from the overall mean positive

214 control Cq is calculated. The Cq values in each qPCR plate are then corrected by the

215 qPCR plate-specific difference from the mean. Supplemental Figure 1 displays the

216 distribution of the corrected Cq values of the positive samples. Finally, by inspecting

217 this histogram, we dismiss all Cq values larger than 37 as noisy data, resulting in a

218 final set of 1632 positive Cq values.

219

220 *Simulation of 1D and 2D pooling strategies*

221 Simulations of are run using R 4.0.1. First, several cohorts of 100 000 patients are

222 repeatedly simulated with varying fractions f, defined by $f = 10^{-x}\ with\ x =$

223    $4, 3.95, ..., 1$, of positive cases (resulting in 61 cohorts of 0.01% to 10% prevalence).

224   This is done five times, resulting in five replicate cohort per prevalence (302 cohorts).

225   The Cq values of the positive samples are sampled with replacement from the set of

226   1632 positive Cq values. Next, the patients are randomly separated into pools

227   depending on the pooling strategy that is simulated. The pooling strategies that were

228   simulated are 1x4, 1x8, 1x12, 1x16, 1x24 (all 1D), and 8x12, 12x16 and 16x24. The

229   Cq value of the pool was calculated as follows:

230

$$c_{pool} = \log_2 P - \log_2 \sum_{i=1}^{p} 2^{-c_i}$$

231

232   With c_{pool} the Cq value of the pool, P the number of samples in the pool, p the

233   number of positive samples in the pool, $c_1, c_2, ..., c_p$ the Cq values of the positive

234   samples. If the Cq value of the pool is smaller than 37, it is classified as a positive

235   pool. For 1D pooling, only samples in positive pools are retained and the remaining

236   individual Cq values were checked to be positive. For 2D pooling, the Cq values of

237   the differently sized pools are checked simultaneously and the samples in negative

238   pools are removed, after which all Cq values of the remaining samples are checked

239   individually. Samples that were retained after the testing of the pools and had an

240   individual Cq lower than 37 are classified as positive, all other samples are classified

241   as negative.

242   The sensitivity is calculated as:

243

$$sensitivity = \frac{\# \; true \; positive \; samples}{\# \; true \; positive \; samples + \# \; false \; negative \; samples}$$

244

245 The analytical efficiency gain is calculated as:

246

$$efficiency\ gain = \frac{\#\ tests\ required\ for\ individual\ testing}{\#\ tests\ required\ for\ pooling\ strategy}$$

247

248 In all simulations, the number of tests required for individual testing is equal to the

249 number of samples.

250

Acknowledgements

252 We are grateful for the data from the Belgian federal taskforce for COVID-19 qPCR

253 testing.

254

Authors' contributions

256 Conceptualization: J.Va., P.M. and J.Ve.; Methodology: J.Va., P.M. and J.Ve.;

257 Software: J.Ve.; Formal Analysis: J.Ve.; Resources: J.Va. and P.M.; Data Curation:

258 J.Ve.; Writing – Original Draft: J.Va. and J.Ve.; Writing – Review & Editing; J.Va.,

259 P.M. and J.Ve.; Visualisation: J.Va. and J.Ve.; Supervision: J.Va and P.M.; Project

260 Administration: J.Va. and P.M.

261

Data availability

263 The code and Cq values are available on

264 https://github.com/jasperverwilt/covidpooling. The Cq values are available as an

265 RDML file in the RDMLdb database[17], under ID 2008AA74.

266

References

268   1.   Shani-Narkiss, H., Gilday, O. D., Yayon, N. & Landau, I. D. Efficient and

269   Practical Sample Pooling High-Throughput PCR Diagnosis of COVID-19.

270   *medRxiv* 2020.04.06.20052159 (2020). doi:10.1101/2020.04.06.20052159

271   2.   Guha, P., Guha, A. & Bandyopadhyay, T. Application of pooled testing in

272   screening and estimating the prevalence of Covid-19. *medRxiv*

273   2020.05.26.20113696 (2020). doi:10.1101/2020.05.26.20113696

274   3.   Adams, K. Expanding Covid-19 Testing: Mathematical Guidelines for the

275   Optimal Sample Pool Size Given Positive Test Rate. *medRxiv*

276   2020.05.21.20108522 (2020). doi:10.1101/2020.05.21.20108522

277   4.   Millioni, R. & Mortarino, C. Sequential informed pooling approach to detect

278   SARS-CoV2 infection. *medRxiv* 2020.04.24.20077966 (2020).

279   doi:10.1101/2020.04.24.20077966

280   5.   Bukhari, S. U. K., Khalid, S. S., Syed, A. & Shah, S. S. H. Smart Pooled

281   sample Testing for COVID-19: A Possible Solution for Sparsity of Test Kits.

282   *medRxiv* 2020.04.21.20044594 (2020). doi:10.1101/2020.04.21.20044594

283   6.   Sinnott-Armstrong, N., Klein, D. & Hickey, B. Evaluation of Group Testing for

284   SARS-CoV-2 RNA. *medRxiv* 2020.03.27.20043968 (2020).

285   doi:10.1101/2020.03.27.20043968

286   7.   Shental, N. *et al.* Efficient high throughput SARS-CoV-2 testing to detect

287   asymptomatic carriers. *medRxiv* 2020.04.14.20064618 (2020).

288   doi:10.1101/2020.04.14.20064618

289   8.   Ghosh, S. *et al.* Tapestry: A Single-Round Smart Pooling Technique for

290   COVID-19 Testing. *medRxiv* 2020.04.23.20077727 (2020).

291   doi:10.1101/2020.04.23.20077727

292   9.   Torres, I., Albert, E. & Navarro, D. Pooling of Nasopharyngeal Swab

293   Specimens for SARS-CoV-2 detection by RT-PCR. *medRxiv*

294   2020.04.22.20075598 (2020). doi:10.1101/2020.04.22.20075598

295 10. Yelin, I. *et al.* Evaluation of COVID-19 RT-qPCR test in multi-sample pools.

296   *Clin. Infect. Dis.* (2020). doi:https://doi.org/10.1093/cid/ciaa531

297 11. Gan, Y. *et al.* Sample Pooling as a Strategy of SARS-COV-2 Nucleic Acid

298   Screening Increases the False-negative Rate. *medRxiv* 2020.05.18.20106138

299   (2020). doi:10.1101/2020.05.18.20106138

300 12. Liu, Y. *et al.* Viral dynamics in mild and severe cases of COVID-19. *The Lancet*

301   *Infectious Diseases* **20**, 656–657 (2020).

302 13. Ben-Ami, R. *et al.* Pooled RNA extraction and PCR assay for efficient SARS-

303   CoV-2 detection. *medRxiv* 2020.04.17.20069062 (2020).

304   doi:10.1101/2020.04.17.20069062

305 14. Nguyen, N. T., Aprahamian, H., Bish, E. K. & Bish, D. R. A methodology for

306   deriving the sensitivity of pooled testing, based on viral load progression and

307   pooling dilution. *J. Transl. Med.* **17**, 252 (2019).

308 15. U.S. Food and Drug Administration. *Molecular Diagnostic Template for*

309   *Laboratories.* (2020).

310 16. Corman, V. M. *et al.* Detection of 2019 novel coronavirus (2019-nCoV) by real-

311   time RT-PCR. *Eurosurveillance* **25**, 2000045 (2020).

312 17. Ruijter, J. M. *et al.* RDML-Ninja and RDMLdb for standardized exchange of

313   qPCR data. *BMC Bioinformatics* **16**, (2015).

314

315

316 **Figures**
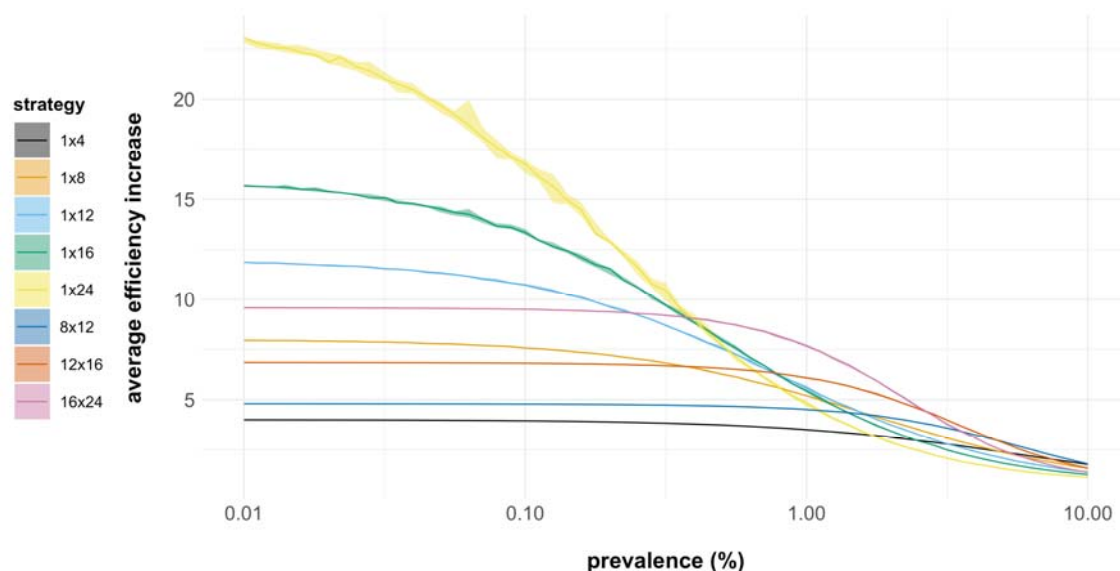
317

318 **Figure 1:** Average efficiency increase (number of samples divided by the number of tests performed)

319 over a 0.01 to 10% prevalence range for different pooling strategies. The averages are approximated

320 by the median over five replicate simulations of a cohort size of 100 000 patients each. The ranges

321 around the line indicate the minimum and maximum value over five replicate simulations.

322

323
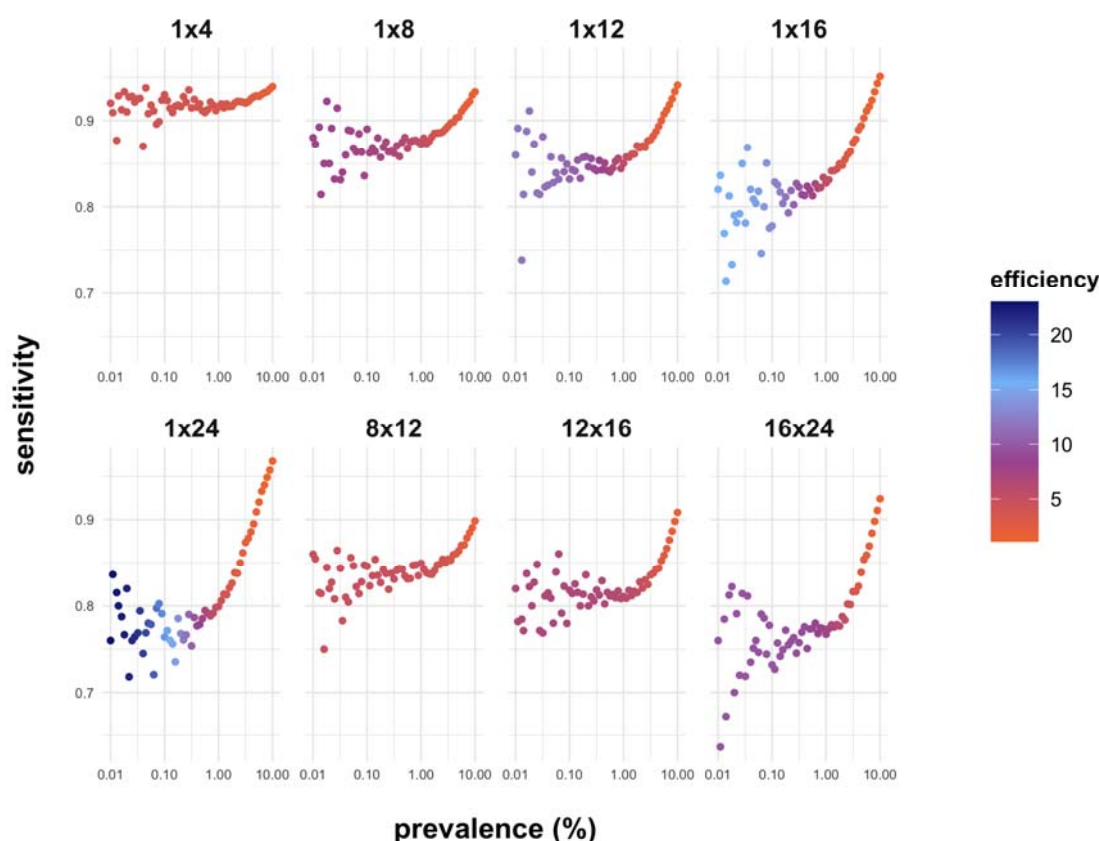
Figure 2: Sensitivity in function of the prevalence for different strategies. Each point is colored by its corresponding efficiency. Each data point illustrates the median value of five replicate simulations, each simulation starting from a set prevalence and pooling strategy for a testing cohort of 100 000 patients.
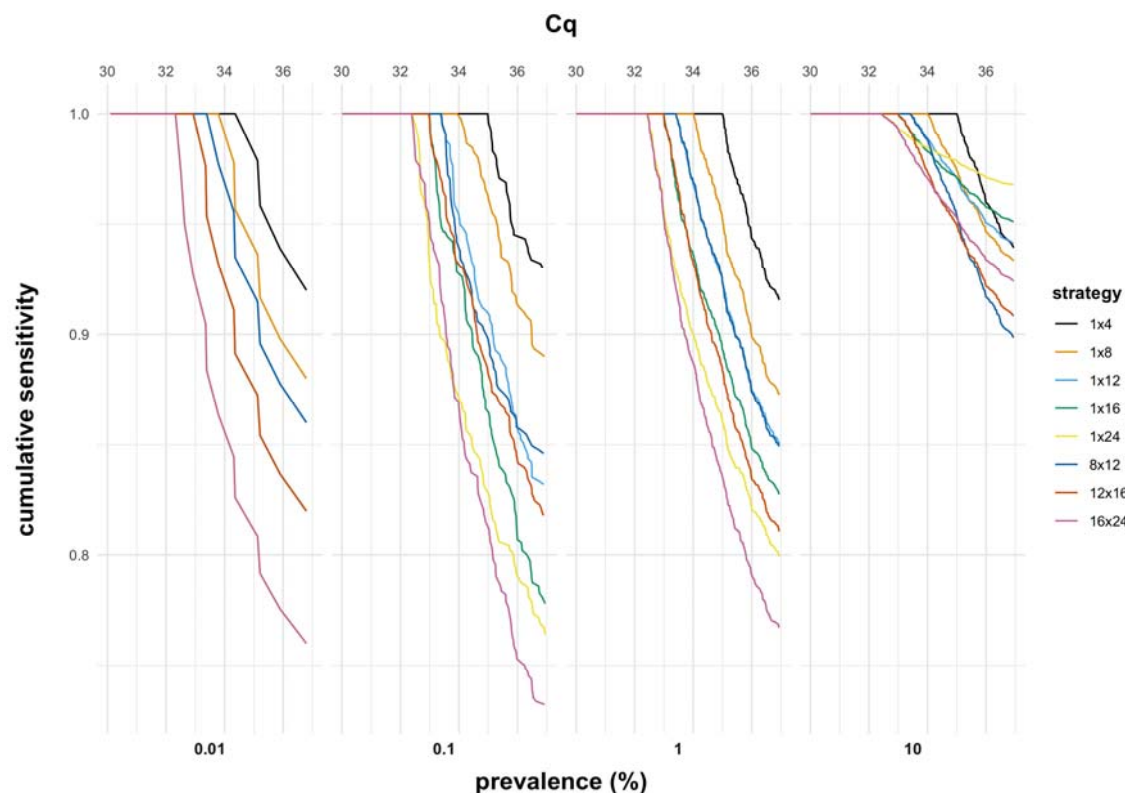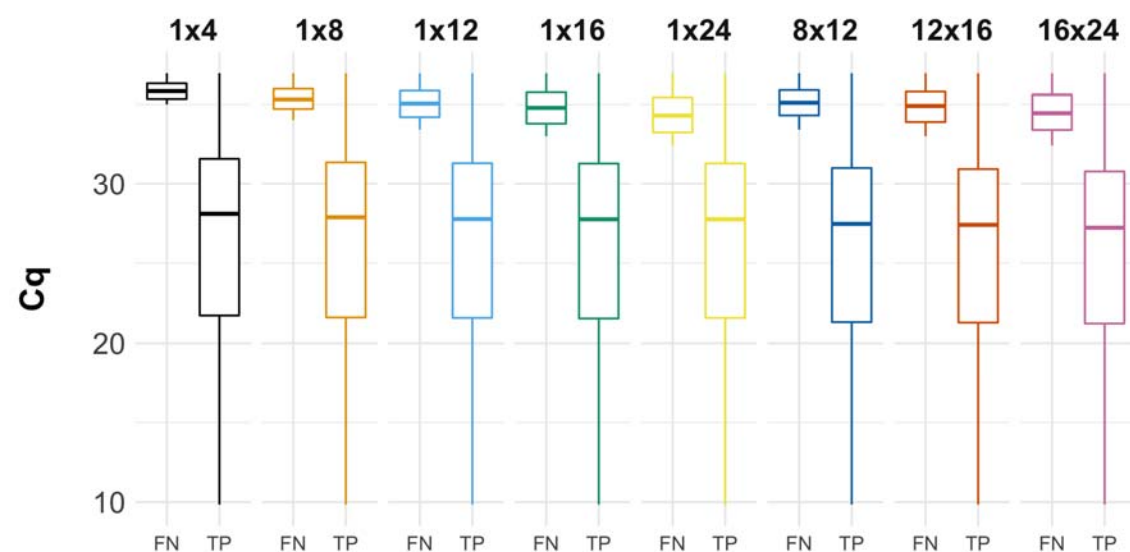
Figure 3: Cumulative sensitivities calculated for all Cq values up to a cut-off Cq for four prevalences. All true positives and false negatives over five replicate simulations for each sampled Cq value are counted and used for calculation of the sensitivity up to the cut-off Cq. The data is filtered for prevalences of 0.01%, 0.1%, 1% and 10% and grouped by prevalence.
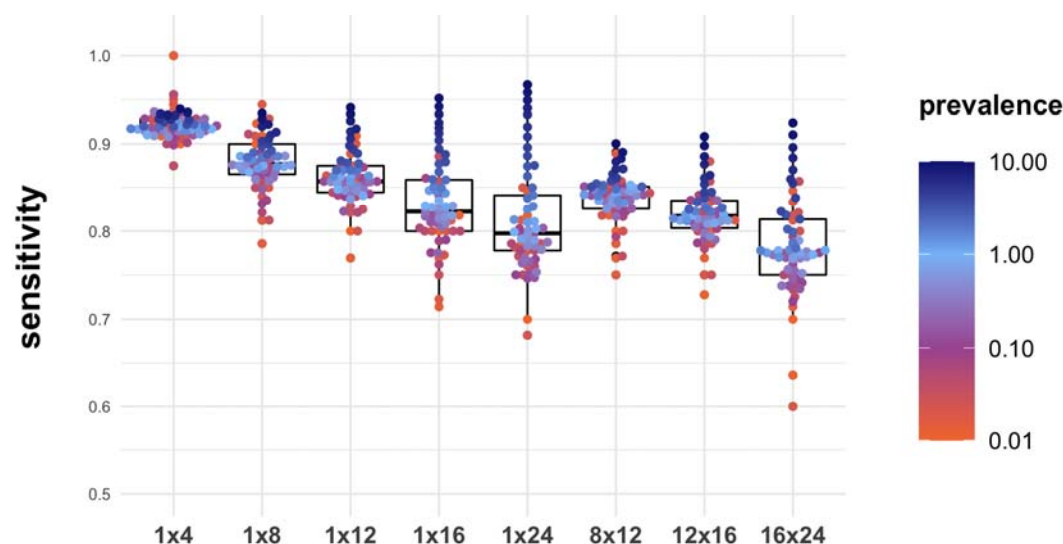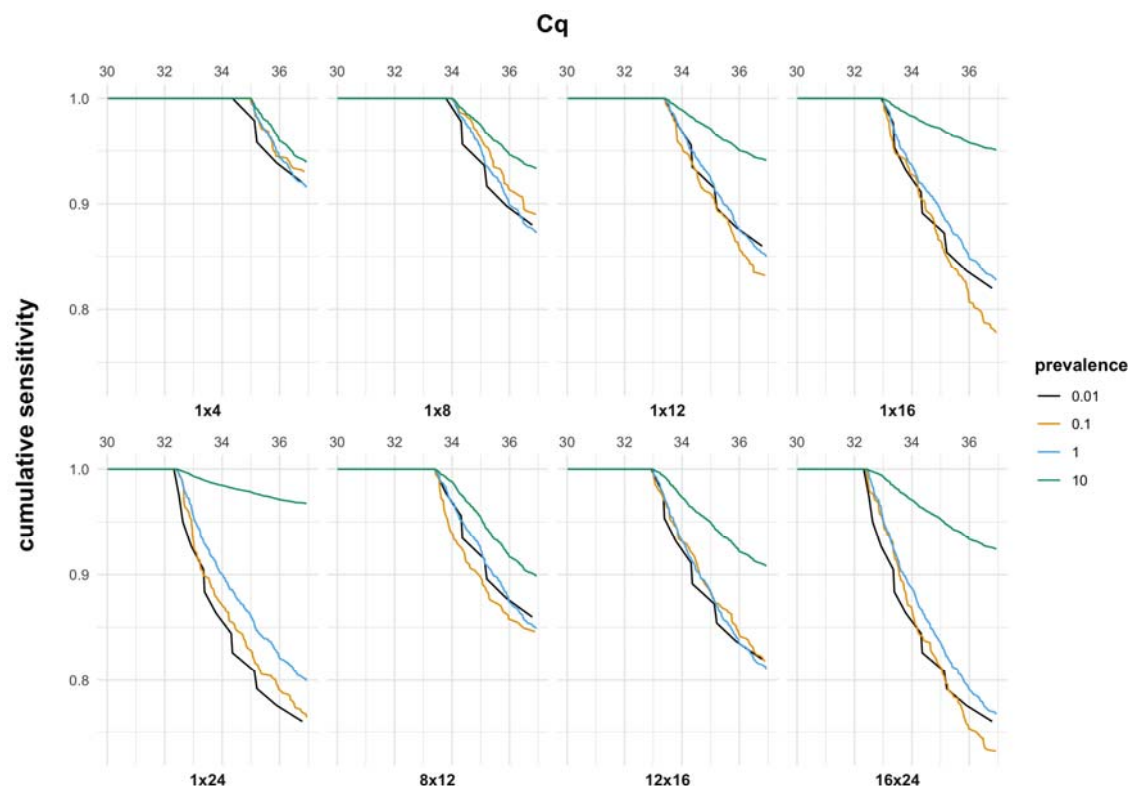
Supplemental Figures
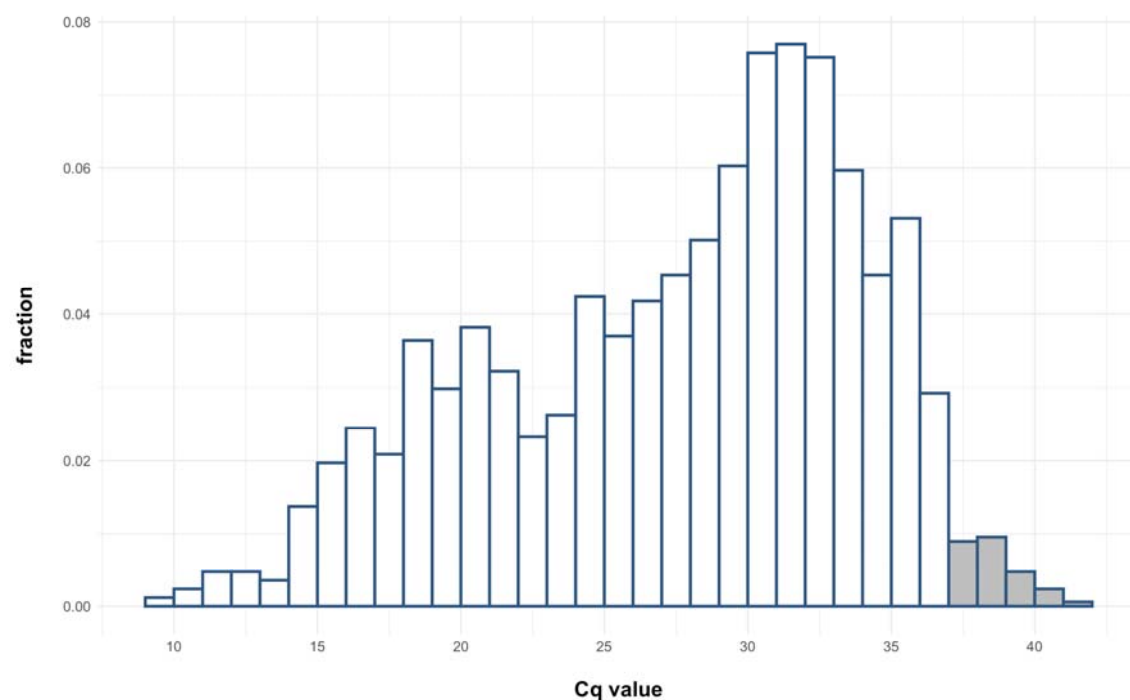
Supplemental Figure 1: The binned distribution of Cq values of all full 96-well RNA plates with good controls and less than 10 positives. The Cq values are corrected by calculating the difference between the mean positive control Cq value per 384-well qPCR plate from the overall mean of the positive control Cq values. Grey bars indicate the Cq values classified as noise.

345   **Supplemental Figure 2:** Cq value distributions for false negatives (FN) and true positives (TP) for

346   different pooling strategies. Each data point represents a Cq value from a replicate simulation with a

347   set prevalence.

348

349

350   **Supplemental Figure 3:** Distributions of the average sensitivity for the simulations in each group.

351   Each data point is the median sensitivity over five simulations.

352

353

Supplemental Figure 4: Cumulative sensitivities calculated for all Cq values up to a cut-off Cq for
each strategy. All true positives and false negatives over five replicate simulations for each sampled
Cq value are counted and used for calculation of the sensitivity up to the cut-off Cq. The data is filtered
for prevalences of 0.01%, 0.1%, 1% and 10% and colored accordingly.

358