

## Swab pooling for large-scale RT-qPCR screening of SARS-CoV-2

Ana Paula Christoff PhD<sup>□1</sup>, Giuliano Netto Flores Cruz BPharm<sup>□1</sup>, Aline Fernanda Rodrigues Sereia PhD<sup>1</sup>, Dellyana Rodrigues Boberg PhD<sup>1</sup>, Daniela Carolina de Bastiani BSc<sup>1</sup>, Laís Eiko Yamanaka MSc<sup>1</sup>, Gislaine Fongaro PhD<sup>2</sup>, Patrícia Hermes Stoco PhD<sup>3</sup>, Maria Luiza Bazzo PhD<sup>4</sup>, Edmundo Carlos Grisard PhD<sup>3</sup>, Camila Hernandes PhD<sup>5</sup>, Luiz Felipe Valter de Oliveira PhD<sup>1\*</sup>

<sup>1</sup>BiomeHub Biotechnologies, Florianópolis-SC, Brasil

<sup>2</sup>Laboratório de Virologia Aplicada, Departamento de Microbiologia, Imunologia e Parasitologia, Universidade Federal de Santa Catarina (UFSC), Florianópolis-SC, Brazil.

<sup>3</sup>Laboratório de Protozoologia, Departamento de Microbiologia, Imunologia e Parasitologia, Universidade Federal de Santa Catarina (UFSC), Florianópolis-SC, Brazil.

<sup>4</sup>Laboratório de Biologia Molecular, Microbiologia e Sorologia – LBMMS, Universidade Federal de Santa Catarina (UFSC), Florianópolis-SC, Brazil.

<sup>5</sup>Hospital Israelita Albert Einstein (HIAE), São Paulo, SP, Brasil

<sup>□</sup>These authors contributed equally.

**\*Correspondence author:** Luiz Felipe Valter de Oliveira, BiomeHub Biotechnologies. Av. Luiz Boiteux Piazza, 1302, Sapiens Parque, Canasvieiras, Florianópolis, SC, Brazil

## ABSTRACT

Pool testing has been proposed as an alternative for large-scale SARS-CoV-2 screening. However, dilution factors proportional to the number of pooled samples have been a source of major concern regarding its diagnostic performance. Further, sample pooling can lead to increased laboratory workload and operational complexity. Therefore, pooling strategies that minimize sample dilution, loss of sensitivity, and laboratory overload are needed to allow reliable and large-scale screenings of SARS-CoV-2. Here, we describe a pooling procedure in which nasopharyngeal swabs are pooled together at the time of sample collection (swab pooling), decreasing laboratory manipulation and minimizing dilution of the viral RNA present in the samples. Paired analysis of pooled and individual samples from 613 patients revealed 94 positive individual tests. Having individual testing as a reference, no false-positives or false-negatives were observed for swab pooling. A Bayesian model estimated a sensitivity of 99% (Cr.I. 96.9% to 100%) and a specificity of 99.8% (Cr.I. 99.4% to 100%) for the swab pooling procedure. Data from additional 18,922 patients screened with swab pooling were included for further quantitative analysis. Mean Cq differences between individual and corresponding pool samples ranged from 0.1 Cq (Cr.I. -0.98 to 1.17) to 2.09 Cq (Cr.I. 1.24 to 2.94). Overall, 19,535 asymptomatic and presymptomatic patients were screened using 4,400 RT-qPCR assays, resulting in 246 positive patients (positivity rate 1.26%). This corresponds to an increase of 4.4 times in laboratory capacity and a reduction of 77% in required tests. Finally, these data demonstrate that swab pooling can significantly minimize sample dilution and sensitivity issues commonly seen in its traditional counterpart. Therefore, swab pooling represents a major alternative for reliable and large-scale screening of SARS-CoV-2 in low prevalence populations.

## INTRODUCTION

The COVID-19 pandemic, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has dramatically impacted public health worldwide in this year of 2020<sup>1,2</sup>. Rapid identification and isolation of infected individuals is essential, but it can be particularly challenging given the infectious potential of both asymptomatic and presymptomatic cases<sup>3,4</sup>. In this scenario, massive population SARS-CoV-2 testing is an urgent need to allow the isolation of infected individuals and, ultimately, the pandemic control.

The most sensitive and recommended test is based on the RT-qPCR method, which detects an active infection through the identification of the viral RNA in nasopharyngeal samples<sup>5-8</sup>. However, several limitations have hampered large-scale population screenings using RT-qPCR, mainly related to the worldwide shortage of supplies and their relatively high cost. To overcome these limitations and scale-up testing capability, some research groups have proposed pooling samples for testing, in which several individuals are simultaneously analyzed using a single test<sup>9-15</sup>.

In the many ways that pool testing was proposed so far, all of them are based on individual sample mixing by the laboratory (sample pooling). This procedure involves substantial sample manipulation, leading to operational challenges and, more importantly, to substantial dilution of viral RNA present in any of the pool samples. Such a dilution effect directly impacts the analytical sensitivity of the RT-qPCR assay, potentially leading to reduced diagnostic sensitivity<sup>9</sup>. Here, we describe a pooling procedure in which nasopharyngeal swabs are pooled together at the time of sample collection (swab pooling), decreasing laboratory manipulation and minimizing dilution of the viral RNA present in the sample.

## **METHODS**

### **Study design**

A retrospective study was performed using de-identified results from nasopharyngeal samples subjected to RT-qPCR-based SARS-CoV-2 testing from May 5<sup>th</sup> to July 31<sup>st</sup>, 2020. Sample collection was performed focusing on low-prevalence COVID-19 populations (*e.g.* asymptomatic or presymptomatic individuals). A total of 45 pool samples and their 613 corresponding individual samples were analyzed in parallel to assess correspondence between individual and pool qualitative results, *i.e.*, individual samples were tested regardless of the results from their corresponding pools. Further, 18,922 additional patients were tested using swab pooling, corresponding to 1,344 pools. Among these, only positive pools had their respective individual samples tested. Individuals from negative pools were considered negative for SARS-CoV-2 RNA detection. Comparison of paired cycle quantification (Cq) values from all positive samples obtained was carried out to assess potential quantitative biases due to swab pooling. This study was approved by the Hospital Israelita Albert Einstein Ethics Committee (number 36371220.6.0000.0071). The patient informed consent was waived off by the ethics committee as the research was performed on de-identified, anonymised samples.

### **Specimen collection and swab pooling for SARS-CoV-2 screening**

Nasopharyngeal samples were collected by trained healthcare professionals using nylon flocked swab, stored in tubes containing sterile saline solution and submitted to laboratory processing within a maximum of 48h after sample collection.

For the swab pooling method, two swabs were collected from the same individual (Figure 1). The first swab, collected through one nostril, was stored in an individual tube containing 3 mL of saline solution. A second swab, collected from the second nostril, was stored in a pool tube containing 5 mL of saline solution. As a general rule, each pool tube was allowed to contain up to 16 swabs, from 16 different patients, collected apart within a maximum of 1h. In this way, the pooling of swabs is performed at the time of sample collection, dismissing further manipulation, mixing and dilution of the samples by the

laboratory. When a given pool tested positive, all the corresponding individual samples were also tested to identify the infected patients. If a pool yielded a negative result, all individuals within that pool were considered negative for SARS-CoV-2 RNA detection.

### **SARS-CoV-2 RT-qPCR detection**

RNA isolation was performed from nasopharyngeal samples using guanidine thiocyanate lysis solution followed by magnetic beads capture and purification (BiomeHub, Brazil). Samples were eluted in 40  $\mu$ L of RNase-free water. RNA reverse transcription (RT) was performed using SupesScript<sup>TM</sup> IV (Invitrogen, USA) and random hexamers, according to the manufacturer instructions.

Real-Time PCR detection of SARS-CoV-2 was performed using the following genetic markers: a region of the gene encoding the viral envelope protein (*E*) with P1 probe and the RNA-dependent RNA polymerase gene (*RdRp*) with P2 probe for discriminatory assay, as described in the Charité-Berlin protocol<sup>8,16</sup>. Also, data from the detection of the surface glycoprotein gene (*S*) using SYBR Green intercalating fluorophore as previously described<sup>17</sup> were included. Amplifications below cycle quantification (Cq) 40 were considered positive. Amplifications were performed in 7500 Fast, QuantStudio 6 Pro Real Time PCR (Applied Biosystems, USA), or in a CFX 384 (BioRad, USA). Cycle quantification values from the RT-qPCR amplifications were used for data analysis.

### **Statistical analysis**

All statistical analyses were performed using R statistical software (v. 3.6.3)<sup>18</sup>. Data wrangling and visualization were performed using the tidyverse package suite (v. 1.3.0)<sup>19</sup>. Modeling was performed using the brms R package (v. 2.12.0) and the Stan probabilistic programming language (v. 2.19.1)<sup>20,21</sup>. Additional R packages included ggpubr (v. 0.2.5), RColorBrewer (v. 1.1.2), binom (v. 1.1.1), and patchwork (v. 0.0.1)<sup>22-25</sup>.

Concordance between pool and individual tests was determined by considering their corresponding qualitative results, *i.e.*, a test was considered concordant if the individual result matched

the result from its corresponding pool. Among positive tests, we quantified the mean Cq difference between individual samples and their corresponding pools. We employed a Bayesian hierarchical model with patient-specific intercepts as follows:

$$\begin{aligned}
 Cq_i &\sim N(\mu_i, \sigma^2) \\
 \mu_i &= \alpha + \alpha_{patient[i]} + \beta_{pool} * Pool_i + \beta_E * E_i + \beta_{RdRp} * RdRp_i \\
 &\quad + \beta_{Pool:E} * Pool_i * E_i + \beta_{Pool:RdRp} * Pool_i * RdRp_i \\
 \alpha_{patient} &\sim N(0, \sigma_\alpha) \\
 \alpha &\sim N(25, 3) \\
 \beta. &\sim N(0, 5) \\
 \sigma_\alpha &\sim Exponential(1) \\
 \sigma &\sim Exponential(1)
 \end{aligned} \tag{1}$$

where  $Pool_i$  is an indicator variable which equals 1 when the  $i^{th}$  observation is from a pool sample and 0 otherwise. The  $E_i$  and  $RdRp_i$  variables adjust for variation in the genetic marker used for the RT-qPCR assay. In these settings, the population-level intercept  $\alpha$  represents the average Cq value for an individual test using the  $S$  gene, whereas the patient-specific intercept  $\alpha_{patient[i]}$  accounts for patient-to-patient variability. The  $\beta.$  coefficients allow quantification of mean Cq differences between individual and pool tests across different genetic markers. We set weakly informative priors for all parameters. Results were reported as posterior means and 95% credible intervals.

For the sample dilution in swab pooling experiment, the same model as in (1) was employed, except that varying intercepts varied with inoculating samples instead of with patients; also, only  $E$  and  $RdRp$  genes were used. Credible intervals for proportions were obtained using a Beta(1, 1) prior for the binomial likelihood. Observed correlations were reported as Spearman's rank correlation as well as Pearson's correlation coefficient.

## RESULTS

### Sample dilution in swab pooling - Proof of concept

In a laboratory experiment, 16 positive nasopharyngeal samples were selected as inoculating samples to be mixed in equal volumes into 16 negative pool samples as well as into 16 negative individual samples, according to a dilution factor of 1.67. This dilution factor corresponds to volumes between samples collected in swab pooling tubes (with 5 mL saline solution) and samples collected in individual tubes (with 3 mL saline solution). For this paired experiment, we observed a mean Cq difference between pool and individual samples of 0.42 Cq (95% Cr.I. -0.22 to 1.09) for the *E* gene and 0.6 Cq (95% Cr.I. -0.05 to 1.24) for the *RdRP* gene (Figure 2A).

In Figure 2B we show the expected slopes ( $\Delta Cq$ ) in RT-qPCR amplifications with variable amplification efficiencies and considering different dilution scenarios. In sample pooling, the dilution factor is equal to the number of individual samples within a pool, yielding expected mean Cq differences of at least 3.32 Cq and as high as 7.37 Cq depending on the number of pooled samples and the amplification efficiency. For swab pooling, on the other hand, the dilution factor is kept fixed at 1.67 so that the expected variation due to dilution alone is constrained between 0.73 and 1.08 Cq.

### Paired analysis of pool and individual tests

To investigate any loss of diagnostic sensitivity due to swab pooling, we analyzed individual and pool samples from 613 patients regardless of their pool results (*i.e.*, positive and negative pools). All the individual and pool samples were analyzed in parallel resulting in 94 positive individual tests and 20 positive pools (Figure 3A). Among the 20 positive pools, at least one individual sample in each pool tested positive for SARS-CoV-2. Positive patients per pool varied from 1 to 11 (Figure 3B). We observed no clear evidence of correlation between the pool Cq values and the number of positive samples within each pool (Figure 3C). Paired comparisons of the pool and their respective individual Cq values can be visualized in Figure 3D. Further analysis of Cq variation is performed in the next section.

Qualitatively, we did not observe any positive individual test paired with a negative pool, *i.e.*, no false-negatives due to swab pooling. In fact, we observed complete agreement (100%) between qualitative results from the pool and individual paired samples. Hence, we employed a simple beta-binomial model with flat priors on performance estimates that would otherwise reach 100%. Having individual testing as a reference, the current data supports a sensitivity of 99% (95% Cr.I. 96.9% to 100%) and a specificity of 99.8% (95% Cr.I. 99.4% to 100%) for the swab pooling procedure, indicating evidence of strong similarity in diagnostic performance.

### **Large-scale screening for SARS-CoV-2 using swab pooling**

To investigate any biases in quantitative results, we included data from additional 1,344 pools and their respective individual tests. In total, 19,535 patients (1,389 pools) were screened using the swab pooling method herein described. Considering all combined results, we observed 246 positive patients for SARS-CoV-2 distributed in 163 pools, resulting in a positivity rate of 1.26%. For 12 pools (0.86%), amplification of both *E* and *RdRp* genes was detected but no associated positive individual sample was identified. In such cases, a new sample collection was requested by the laboratory.

Among the 163 positive pools, 100 (61.3%) contained exactly 16 pooled swabs (Figure 4A). Also, 104 pools (63.8%) corresponded to exactly one positive individual test each. Over 81% of positive pools presented at most 3 correspondent positive individual tests. Some Cq values above 40 were individually inspected and considered positive for 3 pools and 3 individual samples in the *E* gene (41.06, 40.58, 40.58 for pools and 41.93, 41.61, 40.99 for individuals) given that their respective *RdRp* gene amplification was also positive but with lower Cqs (34.58 to 37.52). Additionally, one pool sample with a Cq value of 44.79 for the *RdRp* gene was considered positive as its *E* gene counterpart showed a Cq value of 34.83.

Correlation between Cq values from individual tests and their corresponding pools was strongest for pools associated with one or two positive samples, seemingly diminishing with the increase in the number of positive samples within the pools (Figure 4B). To estimate the Cq variation due to swab



pooling, we assigned to each patient the Cq value from their individual test and the Cq value from their respective pool. Using a hierarchical model with patient-specific intercepts, we estimated the mean Cq difference between individual tests and their corresponding pools for each genetic marker (Figure 5). For the *S* gene (94 patients), the mean Cq difference was estimated to be 0.1 Cq (95% Cr.I. -0.98 to 1.17). Differences for the *E* and *RdRp* genes (152 patients) were estimated to be 1.8 Cq (95% Cr.I. 0.93 to 2.66) and 2.09 Cq (95% Cr.I. 1.24 to 2.94), respectively.

## DISCUSSION

Pool testing has gained importance to fight the COVID-19 pandemic, as challenges involving cost and logistics are at the core of shared struggles to promote large-scale screenings worldwide<sup>10</sup>. Traditional pooling methods proposed<sup>9-15</sup> rely on the combination of multiple individual samples prior to RNA extraction or RT-qPCR, leading to a dilution factor directly related to the number of samples in the pool<sup>9,11-15</sup>. This dilution effect has been of major concern over the diagnostic performance of pool testing procedures<sup>26</sup>. Here, we report a pooling strategy that readily minimizes such an effect and enables large-scale screening for SARS-CoV-2.

Assessing data from 19,535 screened patients, swab pooling and individual testing showed hardly distinguishable performances both qualitatively and quantitatively. With complete agreement between paired qualitative results from 613 patients, the presented data indicates evidence of strong similarity in diagnostic sensitivity and specificity. We did not observe a clear correlation between pool Cq values and number of positive individuals within the pools, as previously suggested considering other pooling methods<sup>14</sup>. Also, the correlation between Cq values from individual tests and their corresponding pools seems to be stronger for pools with no more than two positive individuals. This is mainly reflecting the potentially wide range of individual Cq values for samples composing a single pool.

Although we do use a larger volume for swab pooling (5 mL of saline solution versus 3 mL in individual tubes), the corresponding dilution factor of 1.67 will lead to an expected mean increase of 1.08 Cq even under sub-optimal amplification efficiencies. In a laboratory-controlled experiment, we did not

detect clear differences due to dilution alone, with point estimates from 0.43 to 0.61 Cq. In practice, observational data from 246 positive patients generated point estimates of mean Cq differences between individual tests and their corresponding pools ranging from 0.1 to 2.09 Cq. While such values are hardly significant in terms of analytical sensitivity, the expected counterparts for traditional pooling would range from 3.3 to 5 Cq under optimal amplification conditions. This range corresponds to dilution factors from 10 to 32, when equivolumetric pools from 10 to 32 samples, respectively, are formed post-collection by the laboratory as traditionally proposed<sup>9,11,13,15</sup>. In a worst-case scenario for swab pooling, a mean Cq difference of 2.94 Cq (*RdRp* gene, upper limit of 95% credible interval) would still be considerably lower than the expected differences for sample pooling with 10 samples and perfect amplification efficiency. Nonetheless, there is always a limitation towards samples with Cq's higher than 35, in which case mean differences as small as 1 Cq could still result in false-negative tests regardless of the pooling strategy.

Operationally, the major difference between swab pooling and traditional methods regards sample collection: while in swab pooling we combine multiple swabs in the same tube at the time of sample collection, traditional strategies pool equal volumes from individually collected samples. Beyond dilution, the latter methodology adds complexity to laboratory operations and may lead to increased workload to already saturated laboratory facilities. Traditional pooling requires significant sample manipulation with a risk of contamination and even sample exchange during the laborious pooling process.

On the other hand, collecting two swabs from the same patient can be operationally trivial. While one swab goes into the pooling tube, the other one will only be processed by the laboratory if the pool tests positive. A critical step, this sample collection process can still represent an important limitation of swab pooling as it can cause variation between pooled and individual swabs. In this study, we detected 12 pools with positive results but no positive associated individual test. Of these, 8 pools were associated with two specific collection events (4 pools collected each day). Thus, it is likely that such inconsistencies are attributable to the sample collection process. Still, these cases represented 0.86% of all 1,389 tested pools. Notably, the proper training of sample collection staff represents a cheaper and easier-to-implement alternative to increased laboratory complexity. Any laboratory capable of routine processing

of diagnostic samples for SARS-CoV-2 can also perform swab pool analysis using the same detection methods and infrastructure already in use.

Using swab pooling during sample collection, laboratories in which traditional pooling is currently unfeasible become readily able to contribute to large-scale screenings. Swab pooling, therefore, represents a gain in operational performance for reliable testing of SARS-CoV-2 at scale. As it is well-known, however, any pooling strategy only boosts testing capability for low positivity rates<sup>27</sup>. Swab pooling does not address this matter and is, therefore, suitable for screening populations with low expected prevalence of COVID-19.

The data in the present study comes from the application of swab pooling in asymptomatic or presymptomatic populations, yielding a 1.26% positivity rate. The proposed method was used with pools containing a majority of 16 individuals, but the optimum pool size can be determined by each laboratory during internal validation. Pools with 8, 10, 16, or even 32 swabs may be desirable depending on local epidemiological status and target populations. Upon validation, swab pooling may be applied to any reasonable pool size traditionally proposed to optimize testing scale. Here, over 19,500 patients were screened using approximately 4,400 RT-qPCR assays, corresponding to an increase of 4.4 times in laboratory capacity and a reduction of 77% in the total of required tests.

Finally, identification of infected patients is essential to contain the spread of SARS-CoV-2. This has been hampered by the fact that several people carrying the virus remain asymptomatic or presymptomatic<sup>4,28</sup>. Thus, massive and sensitive testing of asymptomatic and presymptomatic individuals is of utmost importance to fight the COVID-19 pandemic, especially at the moment in which the world attempts to resume economic and social activities.

## **CONCLUSION**

Pool testing is a major alternative for large-scale screening of SARS-CoV-2 in low prevalence populations. Here, we demonstrate that the swab pooling minimizes sample dilution, can be as sensitive as individual testing and reduces laboratory workload. A total of 77% of tests were saved in the screening of 19,535 asymptomatic or presymptomatic patients.

## **ACKNOWLEDGEMENTS**

We would like to thank all BiomeHub, HIAE, and UFSC staff who were involved in all stages of sample collection, laboratory processing, and results discussion. We are grateful for their countless efforts to persist in high-quality research during such difficult times for science in our country, especially during the COVID-19 pandemic. Conflict of interest disclosures: all authors from BiomeHub are currently full-time employees of this research and consulting company specialized in microbiome biotechnologies.

BiomeHub funded the study design, analysis and data submission for publication.

## REFERENCES

1. Yang P, Wang X. COVID-19: a new challenge for human beings. *Cell Mol Immunol.* 2020;17(5):555-557. doi:10.1038/s41423-020-0407-x
2. Lai C-C, Shih T-P, Ko W-C, Tang H-J, Hsueh P-R. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and corona virus disease-2019 (COVID-19): the epidemic and the challenges. *Int J Antimicrob Ag.* 2020;55(3):105924. doi:10.1016/j.ijantimicag.2020.105924
3. He X, Lau EHY, Wu P, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat Med.* 2020;26(5):672-675. doi:10.1038/s41591-020-0869-5
4. Lee S, Kim T, Lee E, et al. Clinical Course and Molecular Viral Shedding Among Asymptomatic and Symptomatic Patients With SARS-CoV-2 Infection in a Community Treatment Center in the Republic of Korea. *Jama Intern Med.* 2020;180(11). doi:10.1001/jamainternmed.2020.3862
5. Vogels CBF, Brito AF, Wyllie AL, et al. Analytical sensitivity and efficiency comparisons of SARS-CoV-2 RT-qPCR primer-probe sets. *Nat Microbiol.* Published online 2020:1-7. doi:10.1038/s41564-020-0761-6
6. Sethuraman N, Jeremiah SS, Ryo A. Interpreting Diagnostic Tests for SARS-CoV-2. *Jama.* 2020;323(22):2249-2251. doi:10.1001/jama.2020.8259
7. Caruana G, Croxatto A, Coste AT, et al. Diagnostic strategies for SARS-CoV-2 infection and interpretation of microbiological results. *Clin Microbiol Infec.* 2020;26(9):1178-1182. doi:10.1016/j.cmi.2020.06.019
8. World Health Organization. Laboratory testing for coronavirus disease 2019 (COVID-19) in suspected human cases: interim guidance, 2 March 2020. *World Health Organization.* Published online March 2, 2020. <https://apps.who.int/iris/handle/10665/331329>
9. Yelin I, Aharony N, Tamar ES, et al. Evaluation of COVID-19 RT-qPCR test in multi-sample pools. *Clin Infect Dis.* Published online 2020. doi:10.1093/cid/ciaa531
10. Lohse S, Pfuhl T, Berkó-Göttel B, et al. Pooling of samples for testing for SARS-CoV-2 in asymptomatic people. *Lancet Infect Dis.* Published online 2020. doi:10.1016/s1473-3099(20)30362-5
11. Ben-Ami R, Klochendler A, Seidel M, et al. Large-scale implementation of pooled RNA extraction and RT-PCR for SARS-CoV-2 detection. *Clin Microbiol Infec.* 2020;26(9):1248-1253. doi:10.1016/j.cmi.2020.06.009
12. Hogan CA, Sahoo MK, Pinsky BA. Sample Pooling as a Strategy to Detect Community Transmission of SARS-CoV-2. *Jama.* 2020;323(19):1967-1969. doi:10.1001/jama.2020.5445
13. Torres I, Albert E, Navarro D. Pooling of nasopharyngeal swab specimens for SARS-CoV-2 detection by RT-qPCR. *J Med Virol.* Published online 2020. doi:10.1002/jmv.25971

14. Gupta E, Padhi A, Khodare A, et al. Pooled RNA sample reverse transcriptase real time PCR assay for SARS CoV-2 infection: A reliable, faster and economical method. *Plos One*. 2020;15(7):e0236859. doi:10.1371/journal.pone.0236859
15. Perchetti GA, Sullivan K-W, Pepper G, et al. Pooling of SARS-CoV-2 samples to increase molecular testing throughput. *J Clin Virol*. 2020;131:104570. doi:10.1016/j.jcv.2020.104570
16. Corman VM, Landt O, Kaiser M, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Eurosurveillance*. 2020;25(3):2000045. doi:10.2807/1560-7917.es.2020.25.3.2000045
17. Chan JF-W, Yuan S, Kok K-H, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet*. 2020;395(10223):514-523. doi:10.1016/s0140-6736(20)30154-9
18. Team RC. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2019. <http://www.r-project.org>
19. Wickham H, Averick M, Bryan J, et al. Welcome to the Tidyverse. *J Open Source Softw*. 2019;4(43):1686. doi:10.21105/joss.01686
20. Bürkner P. brms: An R Package for Bayesian Multilevel Models Using Stan. *J Stat Softw*. 2017;80(1). doi:10.18637/jss.v080.i01
21. Carpenter B, Gelman A, Hoffman M, et al. Stan: A Probabilistic Programming Language. *J Stat Softw*. 2017;76(1). doi:10.18637/jss.v076.i01
22. Kassambara A. *Ggpubr: "ggplot2" Based Publication Ready Plots. R Package Version 0.4.0.*; 2020. <https://CRAN.R-project.org/package=ggpubr>
23. Neuwirth E. *RColorBrewer: ColorBrewer Palettes. R Package Version 1.1-2.*; 2014. <https://CRAN.R-project.org/package=RColorBrewer>
24. Dorai-Raj S. *Binom: Binomial Confidence Intervals For Several Parameterizations. R Package Version 1.1-1.*; 2014. <https://CRAN.R-project.org/package=binom>
25. Pedersen TL. *Patchwork: The Composer of Plots. R Package Version 1.0.1.*; 2020. <https://CRAN.R-project.org/package=patchwork>
26. Cherif A, Grobe N, Wang X, Kotanko P. Simulation of Pool Testing to Identify Patients With Coronavirus Disease 2019 Under Conditions of Limited Test Availability. *Jama Netw Open*. 2020;3(6):e2013075. doi:10.1001/jamanetworkopen.2020.13075
27. Eberhardt JN, Breuckmann NP, Eberhardt CS. Multi-Stage Group Testing Improves Efficiency of Large-Scale COVID-19 Screening. *J Clin Virol*. 2020;128:104382. doi:10.1016/j.jcv.2020.104382
28. Moghadas SM, Fitzpatrick MC, Sah P, et al. The implications of silent transmission for the control of COVID-19 outbreaks. *Proc National Acad Sci*. 2020;117(30):17513-17515. doi:10.1073/pnas.2008373117

29. Ramakers C, Ruijter JM, Deprez RHL, Moorman AFM. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci Lett*. 2003;339(1):62-66. doi:10.1016/s0304-3940(02)01423-4
30. Livak KJ, Schmittgen TD. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the  $2^{-\Delta\Delta CT}$  Method. *Methods*. 2001;25(4):402-408. doi:10.1006/meth.2001.1262
31. Kralik P, Ricchi M. A Basic Guide to Real Time PCR in Microbial Diagnostics: Definitions, Parameters, and Everything. *Front Microbiol*. 2017;8:108. doi:10.3389/fmicb.2017.00108

## FIGURE LEGENDS

**Figure 1. Molecular screening for SARS-CoV-2 through swab pooling method.** 1) From each individual, two swabs are collected. One nasopharynx swab is collected from one nostril and stored in an individual 3 mL tube. Then, another swab is collected through the other nostril and stored in a 5 mL tube containing up to 15 other individuals (pool tube). 2) In the laboratory, the RNA from pooled swabs is extracted, and SARS-CoV-2 detection is performed using RT-qPCR. If a given tested pool presents a positive result, all corresponding individual samples are then processed to identify infected patients.

**Figure 2. Swab pooling and dilution effects.** (A) In a laboratory controlled experiment, 16 positive samples were inoculated in negative individual and pool samples maintaining the dilution factor of 1.67 between individual (3 mL) and swab pooling (5 mL) samples. Mean Cq differences between individual and pool samples estimated for *E* and *RdRp* genes along with 95% credibility intervals are shown in the top-left corner of the graphs. (B) Expected Cq variations ( $\Delta Cq$ ) for sample pooling methods with 10, 16 or 32 samples compared to swab pooling in different amplification efficiencies. Expected  $\Delta Cq$  was calculated using  $\text{Efficiency}^{\text{slope}} = \text{dilution factor}^{29-31}$ . In swab pooling, the number of samples are not related to the dilution factor.

**Figure 3. Paired analysis of pool and individual tests.** (A) Results for samples analyzed in parallel as pools and as individual tests. (B) Total number of positive samples for each positive pool. (C) Correlation between the number of positive samples within a given pool and the corresponding pool Cq (*R*: Pearson's corr. coefficient;  $\rho$ : Spearman's rank corr. coefficient). Points were colored by the total number of individuals within the pool (pool size). (D) Cycle quantification values for positive pools and corresponding positive individual samples.

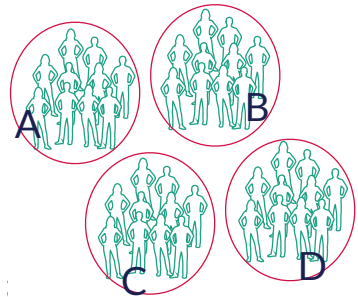
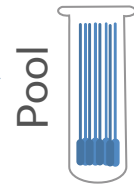
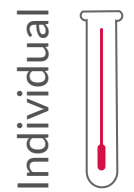
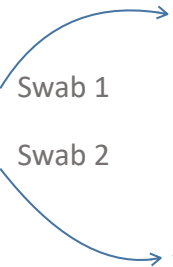
**Figure 4. Large-scale molecular screening of SARS-CoV-2 with swab pooling.** (A) Number of positive pools related to the original pool size and the number of positive samples within the pool. (B)



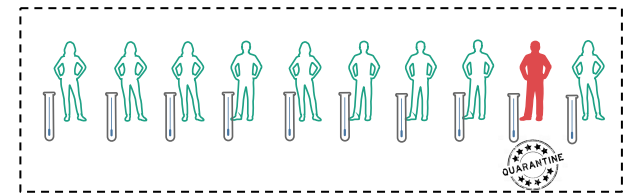
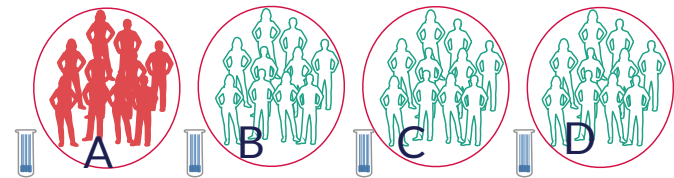
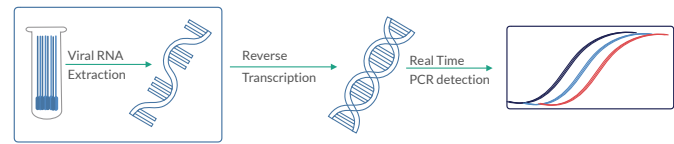
Correlation between Cq values from pools and their respective individual samples stratified by the number of positive samples within the pool (1, 2, and 3 or more positive individual samples). Point color represents the pool size (from less than 10 to 16 individuals). Cq values for the three marker genes tested were included. Correlation coefficients are presented in the figure ( $R$ : Pearson's corr. coefficient;  $\rho$ : Spearman's rank corr. coefficient).

**Figure 5. Paired cycle quantification comparison and estimated mean differences for pool or individual samples.** Cq values from individual tests and corresponding pools were assigned to each patient. A hierarchical model with patient-specific intercepts was used to estimate mean variations between individual and pool samples across varying genetic markers. Estimates for mean Cq differences along with 95% credible intervals are shown in the top-left corner of each graph.

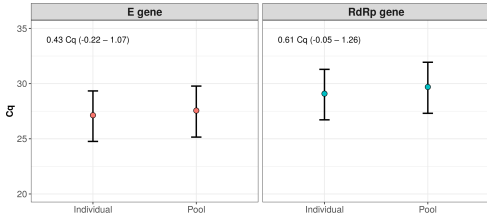
1



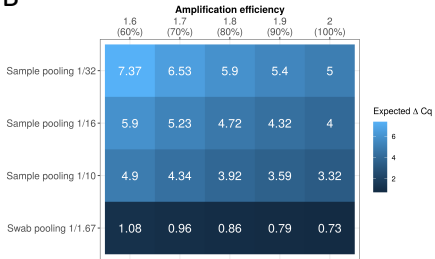
2



A



B



A

45 POOLS

613 INDIVIDUALS

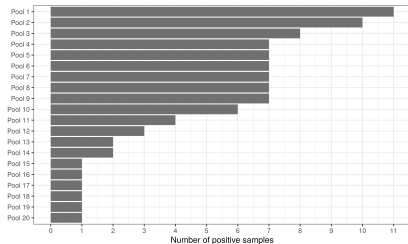
20 Positive Pools

94 Positive Individuals

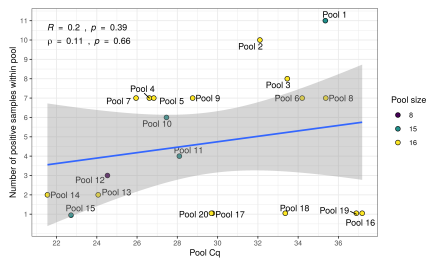
25 Negative Pools

519 Negative Individuals

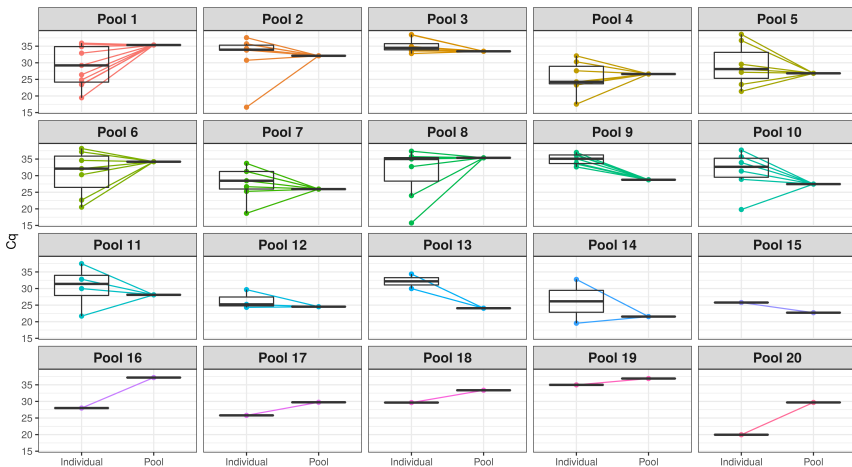
B



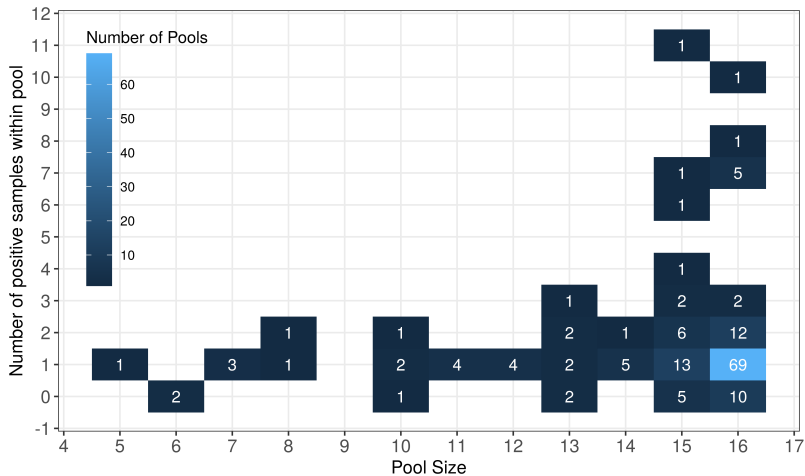
C



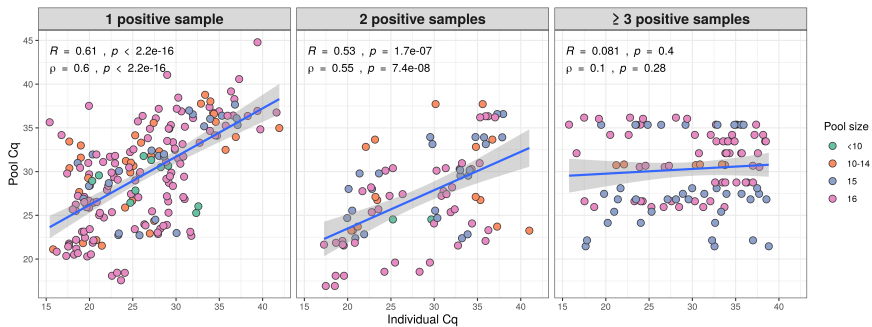
D



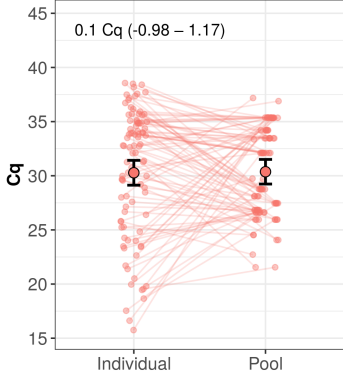
A



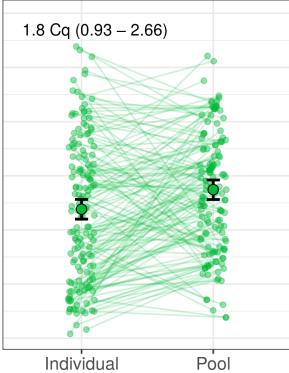
B



### S gene



### E gene



### RdRp gene

