

RESEARCH

Open Access



# Semi-quantitative group testing for efficient and accurate qPCR screening of pathogens with a wide range of loads

Ananthan Nambiar<sup>1\*†</sup>, Chao Pan<sup>2,3†</sup>, Vishal Rana<sup>2†</sup>, Mahdi Cheraghchi<sup>4</sup>, João Ribeiro<sup>5</sup>, Sergei Maslov<sup>1,3\*</sup> and Olgica Milenkovic<sup>2,3\*</sup>

<sup>†</sup>Ananthan Nambiar, Chao Pan and Vishal Rana have contributed equally and their names are listed in alphabetical order.

\*Correspondence:  
nambiar4@illinois.edu;  
maslov@illinois.edu;  
milenkovic@illinois.edu

<sup>1</sup>Department of Bioengineering,  
University of Illinois Urbana-  
Champaign, Urbana, IL, USA

<sup>2</sup>Department of Electrical  
and Computer Engineering,  
University of Illinois Urbana-  
Champaign, Urbana, IL, USA

<sup>3</sup>Center for Artificial Intelligence  
and Modeling, Carl R. Woese  
Institute for Genomic Biology,  
University of Illinois Urbana-  
Champaign, Urbana, IL, USA

<sup>4</sup>Department of Electrical  
Engineering and Computer  
Science, University of Michigan,  
Ann Arbor, MI, USA

<sup>5</sup>NOVA LINES and NOVA School  
of Science and Technology,  
Caparica, Portugal

## Abstract

**Background:** Pathogenic infections pose a significant threat to global health, affecting millions of people every year and presenting substantial challenges to healthcare systems worldwide. Efficient and timely testing plays a critical role in disease control and transmission prevention. Group testing is a well-established method for reducing the number of tests needed to screen large populations when the disease prevalence is low. However, it does not fully utilize the quantitative information provided by qPCR methods, nor is it able to accommodate a wide range of pathogen loads.

**Results:** To address these issues, we introduce a novel adaptive semi-quantitative group testing (SQGT) scheme to efficiently screen populations via two-stage qPCR testing. The SQGT method quantizes cycle threshold ( $C_t$ ) values into multiple bins, leveraging the information from the first stage of screening to improve the detection sensitivity. Dynamic  $C_t$  threshold adjustments mitigate dilution effects and enhance test accuracy. Comparisons with traditional binary outcome GT methods show that SQGT reduces the number of tests by 24% on the only complete real-world qPCR group testing dataset from Israel, while maintaining a negligible false negative rate.

**Conclusion:** In conclusion, our adaptive SQGT approach, utilizing qPCR data and dynamic threshold adjustments, offers a promising solution for efficient population screening. With a reduction in the number of tests and minimal false negatives, SQGT holds potential to enhance disease control and testing strategies on a global scale.

**Keywords:** Group testing, Pooled testing, Semiquantitative group testing, qPCR,  $C_t$  values, Viral load, COVID-19

## Background

Pathogenic infections in humans can cause a wide range of diseases, from mild ailments like the common cold or strep throat to more severe and life-threatening illnesses such as COVID-19, Ebola, and Tuberculosis [2, 4]. These diseases are spread through the proliferation of pathogens within the host and subsequent transmission to other susceptible individuals, often leading to an outbreak in a population. The amount of pathogen in



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

a host, typically referred to as the viral load in the case of viruses, is most frequently expressed in terms of the number of pathogen particles per milliliter of the collected fluid sample. It can vary significantly from the time of infection until recovery and can correlate with the severity of symptoms [18, 19, 35]. To quantify viral loads, the real-time reverse transcription-polymerase chain reaction (qPCR) method is widely used, which reports the number of amplification cycles before the amount of genetic material in the sample reaches a prescribed threshold for detection, known as the cycle threshold or  $C_t$  value.

Individual samples are usually tested using qPCR to monitor disease progression in patients, but when screening a population for infected individuals, it is more efficient to test large groups of samples simultaneously. Group testing (GT) is a strategy that involves pooling multiple samples prior to running qPCR tests, and subsequently detecting infected individuals in the groups based on the test results. This reduces the overall number of tests required while minimizing the false negative rate (FNR), which is critical in infectious disease screening methods, as undetected positive individuals can lead to the rapid spread of disease. Various GT strategies have been proposed in the past to increase the efficiency of wide-scale testing [14, 15, 24], which are implemented using adaptive or non-adaptive protocols. Adaptive testing allows for the sequential selection of groups, while non-adaptive testing requires the selection of all test groups at the same time.

The first known GT scheme, proposed by Dorfman [14], is an example of adaptive GT with binary outcomes (positive or negative), and is not designed to use the quantitative information about viral load. However, fully quantitative testing schemes, including compressive sensing [13, 20], are susceptible to measurement noise, require specialized pooling matrices, and come with performance guarantees only when the ratio of maximum to minimum viral load is confined to a relatively narrow interval [1]. This is not the case for many viruses, including SARS-CoV-2, where viral loads of patients may differ by multiple orders of magnitude [18]. Furthermore, the pooling of samples in both GT and compressive sensing methods leads to dilution, which can adversely impact the accuracy of test outcomes and cannot be directly addressed in a compressive sensing setting.

To address these limitations, we propose a new adaptive semi-quantitative group testing (SQGT) scheme that uses  $C_t$  values quantized into more than two bins in a structured way. In addition, our scheme combines test outcomes from two rounds to improve the likelihoods of subjects being labelled correctly. An idea for two rounds of testing with binary outcomes that is similar to our approach is array testing [6, 26]. The key differences between the aforementioned work and our method lie in the choice of the number of subjects in a test pool (fixed to  $\sqrt{n}$  in [6, 26], but optimized in our work), and the fact that we use nonbinary outcomes that reflect the actual recorded  $C_t$  values of the tested groups. Our work also addressed dilution effects via controlled changes in the threshold values. Since GT was used during the COVID-19 pandemic, multiple theoretical approaches mostly based on Dorfman's method have been developed [3, 33]. At the same time, several large-scale GT data sets containing  $C_t$  values in COVID-19 infected individuals have been generated and made publicly available [5, 12, 21]. Therefore we test our SQGT scheme on the only known complete real-world qPCR COVID-19 data from Israel [5] (complete in the sense that individuals involved in positive groups were

tested individually) and compare it Dorfman's method, showing an increase in testing efficiency. For example, for a population infection rate of 0.02, our SQGT method uses 24% fewer tests than the binary outcome Dorfman's GT method, while maintaining a negligible FNR compared to qPCR noise.

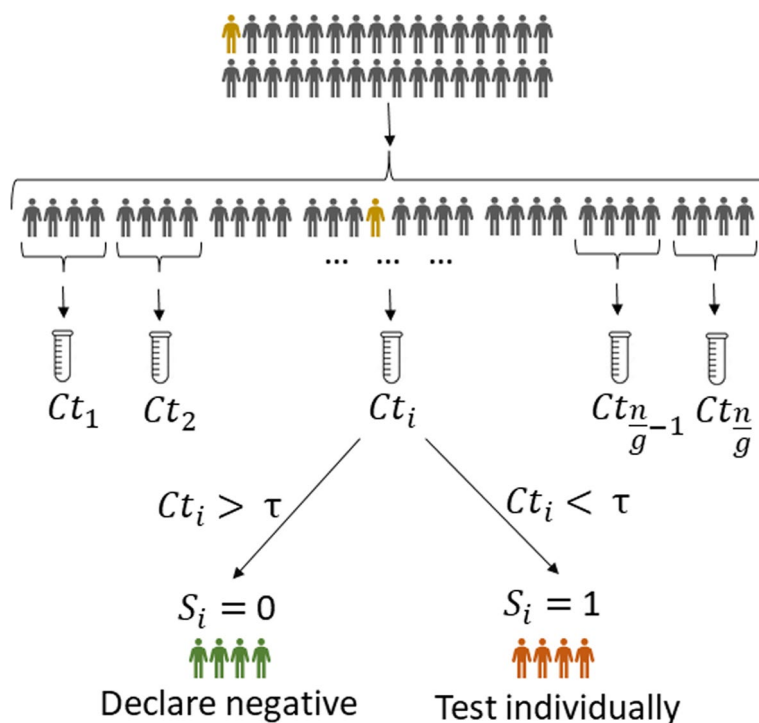
## Algorithms and results

### Basics of group testing

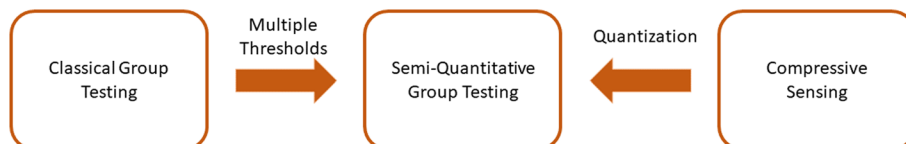
GT, in its most basic form, performs screening of a collection of potentially positive individuals by splitting them into test groups involving more than one individual so as to save on the total number of tests performed. The outcome of a group of test subjects is interpreted as follows: the result is declared positive (and denoted by 1) if at least one of the individuals in the tested group is infected; and, the test result is declared negative (and denoted by 0) if there are no infected individuals in the group. From a theoretical point of view, GT aims to find an optimal strategy for grouping individuals so that the number of binary tests needed to accurately identify all infected individuals is minimized. GT can be implemented using nonadaptive and adaptive approaches. Unlike adaptive GT, nonadaptive schemes require that all tests are performed simultaneously so that the outcome of one test cannot be used to inform the selection of individuals for another test. The first known GT scheme by Dorfman [14] is an example of adaptive screening since it involves two stages of testing, one of which isolates groups with infected individuals, and another one that identifies the actual infected individuals. In general, adaptive schemes use multiple stages of testing and different combinations of individuals to best inform the sequence of tests to be made. (The term "adaptive" is standard in the information theory community, although work in other research areas sometimes deviate from this terminology.) When specializing Dorfman's scheme for qPCR screening, the decision about positive and negative group labels is made based on  $Ct$  values (see Fig. 1).

Despite their widespread use, GT methods have notable shortcomings when used in systems that provide more quantitative information than a binary answer of the form "yes-no," such as is the case for qPCR screening. This motivates developing extensions of GT schemes that make use of the more quantitative information available from experiments. When all of the available quantitative information is used, the generalized GT scheme represents a form of compressive sensing (CS) [8, 11, 13]. However, CS-based schemes require the ratio of the maximum and minimum pathogen concentrations to be properly bounded [1]. This type of assumption does not hold for a large number of infectious diseases, including COVID-19, where the viral concentrations can vary over several orders of magnitude [18]. In the presence of infected individuals with widely different loads, CS approaches will mask individuals with low pathogen concentrations.

Here we propose a more structured approach to GT that straddles the classical Dorfman's scheme and fully quantitative CS approaches. Our SQGT scheme can be seen as a multi-threshold version of Dorfman's GT with two independently permuted groups of samples or a quantized version of adaptive CS (see Fig. 2). More details are provided in the following subsection.



**Fig. 1** Dorfman’s two-stage GT protocol. The test subjects are randomly partitioned into groups of optimized size  $g$  and tested as a group. All individuals in positive groups are subsequently tested individually. As before,  $Ct$  stands for the cycle threshold value of the group under consideration. Note that this GT protocol only uses a binary decision variable, yes (1) and no (0), for the case that  $Ct < \tau$  and  $Ct > \tau$ , respectively. The decision threshold  $\tau$  depends on the protocol used for qPCR

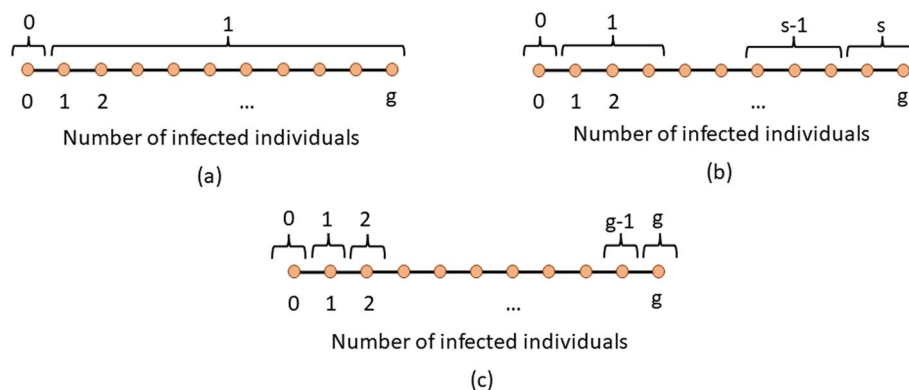


**Fig. 2** Semi-quantitative GT generalizes Dorfman’s GT by using more than one threshold and, like CS, uses information about the estimate of the total number of infected individuals, but with the numbers quantized according to predetermined cluster selections

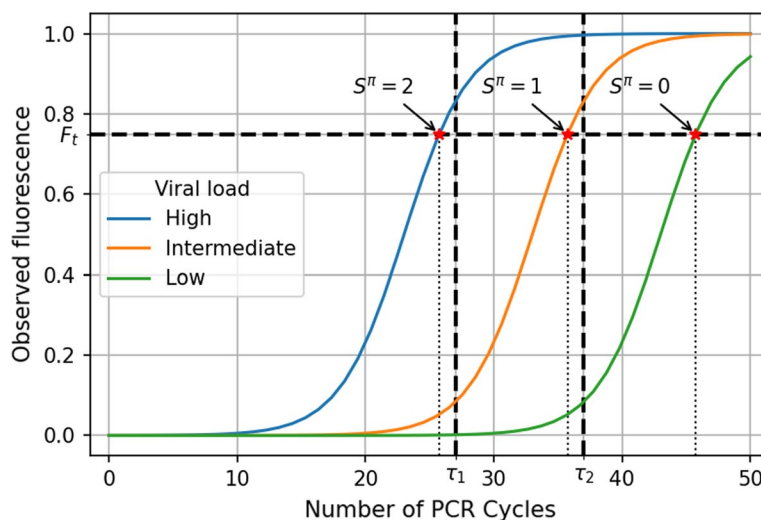
**Semi-quantitative group testing**

SQGT is a GT protocol that interprets test results as estimates of the number of infected individuals in each tested group. Broadly speaking, unlike Dorfman’s GT which generates binary responses (0, for a noninfected group, and 1 when at least one infected subject is present in the group, see Fig. 3a), SQGT produces answers of the form “between  $x$  and  $y$  infected individuals in the group” (see Fig. 3b). For qPCR experiments, the range of values for the number of infected individuals in the group may be estimated from the  $Ct$  value of the group.

For a general SQGT scheme, one seeks a collection of  $\geq 1$  measurement thresholds, such that the outcome of each test is an interval for the possible number of infected individuals, i.e., the outcome of an SQGT experiment specifies lower and upper bounds on the number of infected individuals in a group. If the thresholds are



**Fig. 3** GT interpreted through quantitative output quantization. The quantitative output corresponds to the actual number of infected individuals in a group. In (a), corresponding to Dorfman’s GT, the quantizer maps all outcomes involving more than one infected individual to a score 1. The score 0 indicates that there are no infected individuals in the group. In (b), corresponding to a general SQGT scheme, the quantizer is allowed to map any collection of outcomes to any choices of scores. This implies that the number of possible test results may be larger than two, but upper bounded by the size of the group  $g$ . The simplest version of SQGT based on a uniform quantizer is depicted in (c)



**Fig. 4** An example of qPCR amplification curves and two-threshold ( $\tau_1, \tau_2$ ) SQGT. The two thresholds apply to  $C_t$  values while the actual measurement corresponds to the intersection of the  $F_t$  line (the fluorescence threshold) and the amplification curve. For example, the left-most red star indicates the intersection of the high viral load amplification curve with  $F_t$  and the corresponding measurement falls into the quantization bin denoted by  $S^\pi = 2$

consecutive integers covering all possible options for the number of infected individuals in a group, the scheme reduces to additive (quantitative) GT [28, 31] (see Fig. 3c).

Although nonadaptive SQGT has been previously analyzed from an information-theoretic perspective [9, 16, 17], practical implementations for adaptive SQGT schemes are still lacking, especially in the context of qPCR testing. Our approach is the first adaptive SQGT scheme that is specifically designed for real-world qPCR testing. It operates directly on the  $C_t$  values and makes use of two thresholds,  $\tau_1$  and  $\tau_2$  (see Fig. 4). This choice for the number of thresholds balances the ease of implementation of a testing

scheme in a laboratory with the ability to use the quantitative information from a qPCR test more efficiently.<sup>1</sup>

The main idea behind our *Ct* value-based SQGT approach is to perform a two-stage SQGT protocol with randomly permuted groups of subjects and risk assessment based on the *Ct* values obtained after the first stage. More specifically, the scheme involves the following three steps:

- First, we create two separate, randomly permuted lists of  $n$  subjects. Each of these lists is then evenly divided into groups of a specified size,  $g$ , which are subsequently tested. It's important to underline that the ideal test group size,  $g$ , for our methodology may differ from that typically utilized in Dorfman's GT approach.
- Second, since GT inevitably leads to sample dilution, we adjust the *Ct* thresholds in the SQGT scheme to account for this effect. Note that each individual's sample contributes to two *Ct* values: one from the group they were initially part of in the first permuted list, and another from their group in the second permuted list. This dual-measurement system provides a way for cross-linking the results.
- Third, we examine the pair of *Ct* values associated with the individuals to stratify them into low-risk, medium-risk, and high-risk categories. Based on the risk category, the individuals are either immediately declared negative, or tested once again individually. Although the number of tests performed can be reduced by performing nonadaptive SQGT testing on all risky subjects (discussed in the Additional file 1: Section 1.4), for simplicity we opt for individual testing.

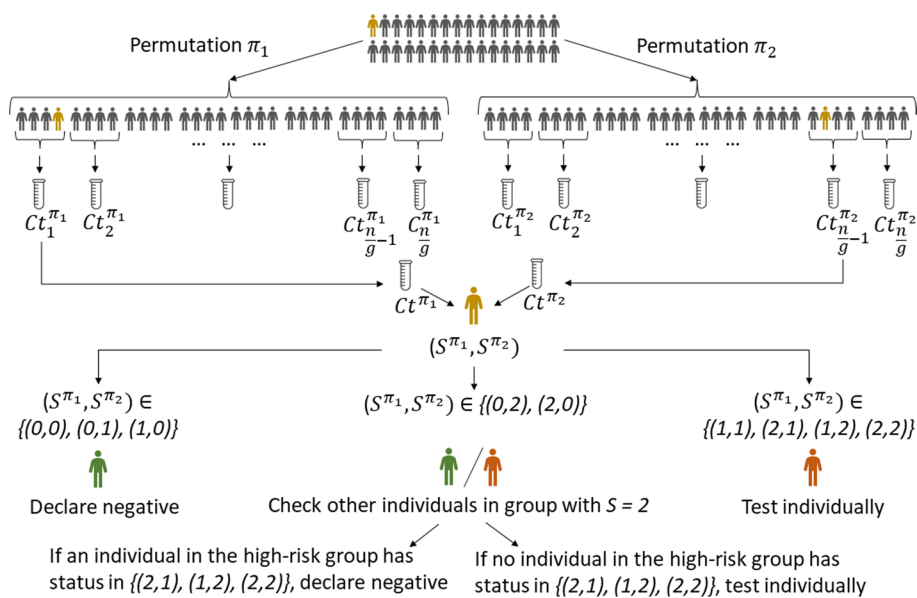
Next, we describe our scheme in detail. We consider a population of  $n$  individuals, arranged into groups of size  $g$ , and denote the fraction of infected individuals by  $p$ . Again, we only make use of two quantization thresholds, denoted by  $\tau_1$  and  $\tau_2$ . Our scheme consists of two stages.

In the first stage, we group the patient samples into groups of size  $g$ , ensuring that each individual contributes to two different groups. To achieve this, we use two random permutations,  $\pi_1$  and  $\pi_2$ , of the  $n$  individuals so that they appear in different random orders. Subsequently, the ordered lists are split into groups of  $g$  consecutive samples (for simplicity, we assume that  $n$  is a multiple of  $g$ ). The resulting groups are denoted by  $\gamma_1^{\pi_1}, \gamma_2^{\pi_1}, \dots, \gamma_{n/g}^{\pi_1}$  and  $\gamma_1^{\pi_2}, \gamma_2^{\pi_2}, \dots, \gamma_{n/g}^{\pi_2}$ . It is important to note that each individual belongs to two groups,  $\gamma_i^{\pi_1}$  and  $\gamma_j^{\pi_2}$  with  $i \in \{1, \dots, n/g\}$  and  $j \in \{1, \dots, n/g\}$ , where the two groups are created based on the two permuted lists. For both collections of groups, we perform separate qPCR experiments, denoting the outcomes as  $Ct_i^{\pi_1}$  and  $Ct_j^{\pi_2}$ , respectively. Then we quantize the *Ct* values into bins, and assign the test scores  $S_i^{\pi_1}$  for group  $\gamma_i^{\pi_1}$  and  $S_j^{\pi_2}$  for group  $\gamma_j^{\pi_2}$  using the threshold rule:

$$S^\pi = \begin{cases} 0, & \text{if } Ct^\pi \geq \tau_2; \\ 1, & \text{if } \tau_1 < Ct^\pi < \tau_2; \\ 2, & \text{if } Ct^\pi \leq \tau_1. \end{cases} \tag{1}$$

<sup>1</sup> We also observe in practice that using more than two thresholds leads to diminishing returns in the number of tests saved but significantly increasing the complexity of the scheme.





**Fig. 5** Our proposed two-stage SQGT scheme with two thresholds, as described in Eq. 1. The approach is to run two parallel rounds of Dorfman-like group tests. To assess if the individual marked in orange is infected, we test them in two different groups, and collect the scores  $(S^{\pi_1}, S^{\pi_2})$ . Based on this pair of scores, we decide if the individual marked in orange needs to be individually tested or not. See the text for more details

Consequently, each individual is labeled by a pair of test scores  $(S_i^{\pi_1}, S_j^{\pi_2})$ , representing the outcomes of the two group tests (for group  $\gamma_i^{\pi_1}$  and  $\gamma_j^{\pi_2}$ ) that the individual is involved in. We omit the subscripts  $i$  and  $j$  in the later context for simplicity of notation.

In the second stage, we classify individuals based on their scores  $(S^{\pi_1}, S^{\pi_2})$ . Individuals with scores  $\{(0,0), (0,1), (1,0)\}$  are deemed low-risk and declared negative. In particular, scores  $\{(0,1), (1,0)\}$  are declared to correspond to negative subjects because they were involved in a negative test group (score 0) and intermediate Ct value group (score 1). Subjects with scores  $\{(1,1), (2,1), (1,2), (2,2)\}$  are classified as high-risk and tested individually in a second stage of tests. For the remaining score pairs,  $\{(2,0), (0,2)\}$ , we proceed as follows: If the group with score 2 contains another individual with a score in  $\{(1,2), (2,1), (2,2)\}$ , we classify the first individual as negative; otherwise, we conduct an individual test. We chose this testing strategy because the scores 0 and 2 indicate two different test results for the same individual that may have been caused by an error and/or the presence of other infected individuals in the second pool. To determine whether the score 0 is due to a false negative result or the score 2 is due to the presence of other infected individuals in that pool, we need to examine the scores of other participants in the group who tested (highly) positive. This provides an intuitive motivation for the proposed strategy. Figure 5 illustrates the proposed two-stage SQGT scheme, while Fig. 1 depicts Dorfman’s GT scheme. Additional file 1: Sections 1.2 and 1.3 provide a detailed mathematical analysis of the various GT schemes discussed.

It is worth noting that conducting individual testing, as in the second stage of our SQGT scheme for the high-risk group, is suboptimal from the point of minimizing the number of tests. This issue is not limiting the application of the scheme since one can use a nonadaptive GT scheme in the second stage, thereby significantly reducing the number of second-stage tests. Since nonadaptive GT is conceptually more involved and

harder to implement in practice than the above procedure, pertinent explanations are delegated to Additional file 1: Section 1.4.

As we will demonstrate in the Results section, the proposed two-stage SQGT approach offers substantial reductions in the number of tests when compared to Dorfman-type tests. It remains to see if the reduction in the number of tests leads to undesirable increases in the FNR of the scheme. To address this question, we need to consider the influence of dilution effects on the test results and how one could adjust quantization thresholds to counter these effects.

### Dilution effects

In most experiments involving GT, the test samples come in specified unit concentrations that are equal across all test subjects. This means that a group sample involving  $g$  individuals will only use a fraction  $1/g$  of the unit sample from each individual. This inevitably leads to dilution of the group sample, the level of which depends on the number of infected individuals in that particular group. When there is only a small number of infected individuals in the group, the overall viral load of the group sample may be lower than the detection threshold, thereby leading to highly undesirable false negatives. False negative rate (FNR) is related to true positive rate (TPR) through  $FNR = 1 - TPR$ , and the TPR function is often referred to as the sensitivity function.

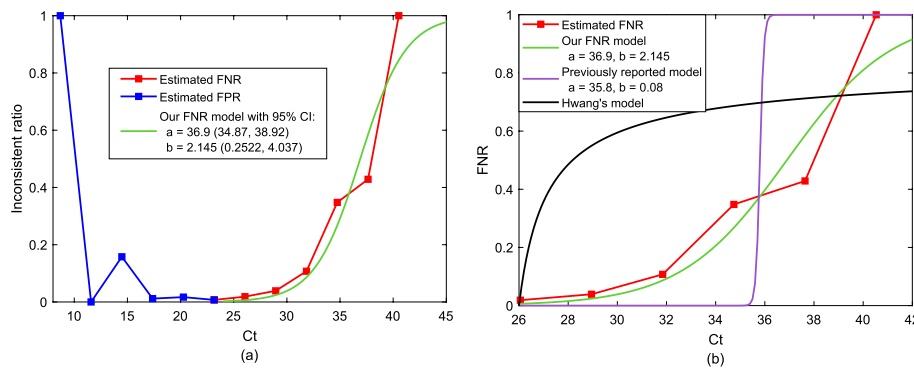
A mathematical model for dilution effects was first proposed in [23], which introduced a special TPR function  $TPR(p, g, d)$  of the form

$$\begin{aligned} TPR(p, g, d) &= \mathbb{P}(\text{test result is declared positive} | \text{there is at least 1 positive subject in the group}) \\ &= p \left[ 1 - (1 - p)^{g^d} \right]^{-1}. \end{aligned} \quad (2)$$

Here,  $p$  denotes the infection rate,  $g$  denotes the group size, and  $d$  denotes a parameter capturing the dilution level. When  $d = 0$ ,  $TPR(p, g, 0) = 1$ , indicating that there is no dilution; when  $d = 1$  and  $g$  is large,  $TPR(p, g, 1) \sim p$ , indicating that the sample is fully diluted and that the probability of correctly identifying a defective group is the same as the infection rate. More details on the TPR model for SQGT can be found in Additional file 1: Section 1.5.

Although the dilution model (2) is mathematically elegant and tractable for analysis, it provides a poor match for real-world measurements (see Fig. 6b). A more practical approach to quantifying dilution effects is to assess how dilution impacts the actual viral load in a group. The empirical studies [5, 7, 12, 25] consistently point out that the  $Ct$  values of groups tend to be higher than the  $Ct$  value of individual tests with high probability. This phenomenon is also due to dilution effects. Nevertheless, none of these works describe how to readjust the  $Ct$  value used for declaring positives in the presence of dilution. In the context of SQGT, this is an even more important issue as the increased  $Ct$  values can lead to degradation in the detection rate as well as a significantly increased number of measurements. This motivates exploring the relationship between the value of the  $Ct$  threshold used for an individual test and that used for a group test. For the worst-case scenario when there is only one infected individual in a group of size  $g$ , the group  $Ct$  value takes the form





**Fig. 6** FNR estimated from data reported in [5] and different FNR models fitted to the real-world experimental data. **a** We count the cases where the group test was positive but all subjects individually tested negative. The ratio of the number of these “inconsistent” tests and the total number of tests with the same Ct value is denoted as the “inconsistent ratio”. Specifically, we consider the right half of the curve ( $Ct > 25$ ) to be caused by the false negative results, which agrees with the intuition that the FNR increases as the Ct value increases. **b** We fit the FNR model from Eq. (4), and the ones from [23, 27] to the real-world experimental data. As it is apparent, the black and purple lines provide a poor fit to the data while our model (green line) with parameters ( $a = 36.9, b = 2.145$ ) represents a significantly more accurate fit

$$\begin{aligned}
 Ct &= -M \log_{10}(v/g) + B \\
 &= -M \log_{10}(v) + B + M \log_{10}(g),
 \end{aligned}
 \tag{3}$$

where  $v$  denotes the viral load of the infected individual, and  $M$  and  $B$  are positive values denoting the slope and the intercept for the PCR calibration curve [25]. The exact values of  $M$  and  $B$  need to be estimated from the experimental data. Equation (3) characterizes the relationship between the viral load and the Ct value, and it implies that compared to individual testing, the group Ct value will be higher by  $M \log_{10}(g)$ . The implication of this observation is that for GT, we need to increase the Ct thresholds by  $M \log_{10}(g)$ .

Elaborate regression models for dilution effects rely on knowledge of the underlying distribution of viral loads for positive and negative individuals [29]. In the absence of robust estimates for these distributions, we use a model that utilizes PCR calibration curves instead [25]

### Controlling and modelling FNRs of PCR tests

In order to quantify the trade-off between the FNR and the reduction in the number of group tests when using the proposed SQGT scheme, we express the FNR, an important metric with respect to test accuracy, as a function of the Ct value. For this purpose, we use the large-scale real-world GT dataset [5]. Our FNR model is based on the following “sigmoid” function,

$$FNR(Ct) = \left[ 1 + \exp\left(\frac{a - Ct}{b}\right) \right]^{-1},
 \tag{4}$$

where  $a, b$  are two tunable parameters that can be used to fit the measured/estimated FNRs. Note that similar ideas were also discussed in [27]; however, as may be seen from Fig. 6b, the FNR function ( $a = 35.8, b = 0.08$ ) proposed in [27] significantly deviates from real-world experimental data.

In practice, the values of FNRs are hard to estimate as this requires multiple tests of the same individual. In the GT context, there are two ways to estimate FNRs. The direct scenario is to compute FNRs by counting the instances when a group test was negative but at least one member from that group tested positive. However, in all experimental verification of GT protocols, individuals whose group tested negative are eliminated from future retesting. This renders the direct approach impossible to pursue in practice. The indirect approach is to count the cases where the group test was positive but all subjects individually tested negative. In this work, we follow the second approach to estimate the FNRs. The ratio of the number of these “inconsistent” tests and the total number of tests with the same  $Ct$  value is shown in Fig. 6a. Note that these results can correspond to either a false positive for the group test, or a false negative for one or more of the individual tests. Here we consider the right half of the curve ( $Ct > 25$ ) to be caused by the false negative results, which agrees with the intuition that the FNR increases as the  $Ct$  value increases. Our fitted FNR curve is shown in Fig. 6b, along with the estimated FNR curve from experimental results, and the models from [23, 27]. As it is apparent, the latter provides a poor fit to the data while our model with parameters ( $a = 36.9, b = 2.145$ ) represents a significantly more accurate fit.

The FNR shown in Fig. 6 corresponds to individual tests, for which we do not know the correct  $Ct$  values. Therefore, we shift the group test  $Ct$  values by  $M \log_{10}(g) = 2.895$  in Eq. (3) to estimate the individual  $Ct$  values. A detailed discussion of the data processing and FNR estimation pipeline is included in the “Methods” section.

#### Case study of the SQGT protocol applied to COVID-19 data

While the SQGT framework is broadly applicable to PCR-based pathogen screening, general data is usually limited for pathogens other than SARS-CoV-2. The COVID-19 pandemic has resulted in an unprecedented amount of publicly available qPCR test data, which motivates testing our SQGT framework on real-world SARS-CoV-2 data. Our reported results pertain to a set of 133, 816 SARS-CoV-2  $Ct$  values of qPCR tests performed in Israel between April and September 2020 as reported in [5]. This dataset is unique in so far that it reports the  $Ct$  values of group tests along with the  $Ct$  values of individual tests for subjects involved in positive group tests. Since both test results are needed for our analysis and modeling, it is the only currently available dataset for a study of this type. To explore a range of different infection scenarios without performing additional experiments, we simulated populations of 10,000 individuals of which a fraction  $p \in \{0.02, 0.05, 0.1\}$  was infected by the virus. The  $Ct$  values of the infected individuals were randomly sampled from the real-world dataset of [5], and converted into estimated viral loads using Eq. 5 (see also the “Methods” section). The viral loads of uninfected individuals were set to 0.

Following the SQGT scheme of Fig. 5, samples are partitioned into groups of  $g$  individuals whose viral loads were subsequently averaged and converted to  $Ct$  values as described in the “Methods” section (Eq. 6). Following standard diagnostic procedures, individuals were declared negative if their  $Ct$  values exceeded a threshold (in our case, set to 37 as suggested in [22]).

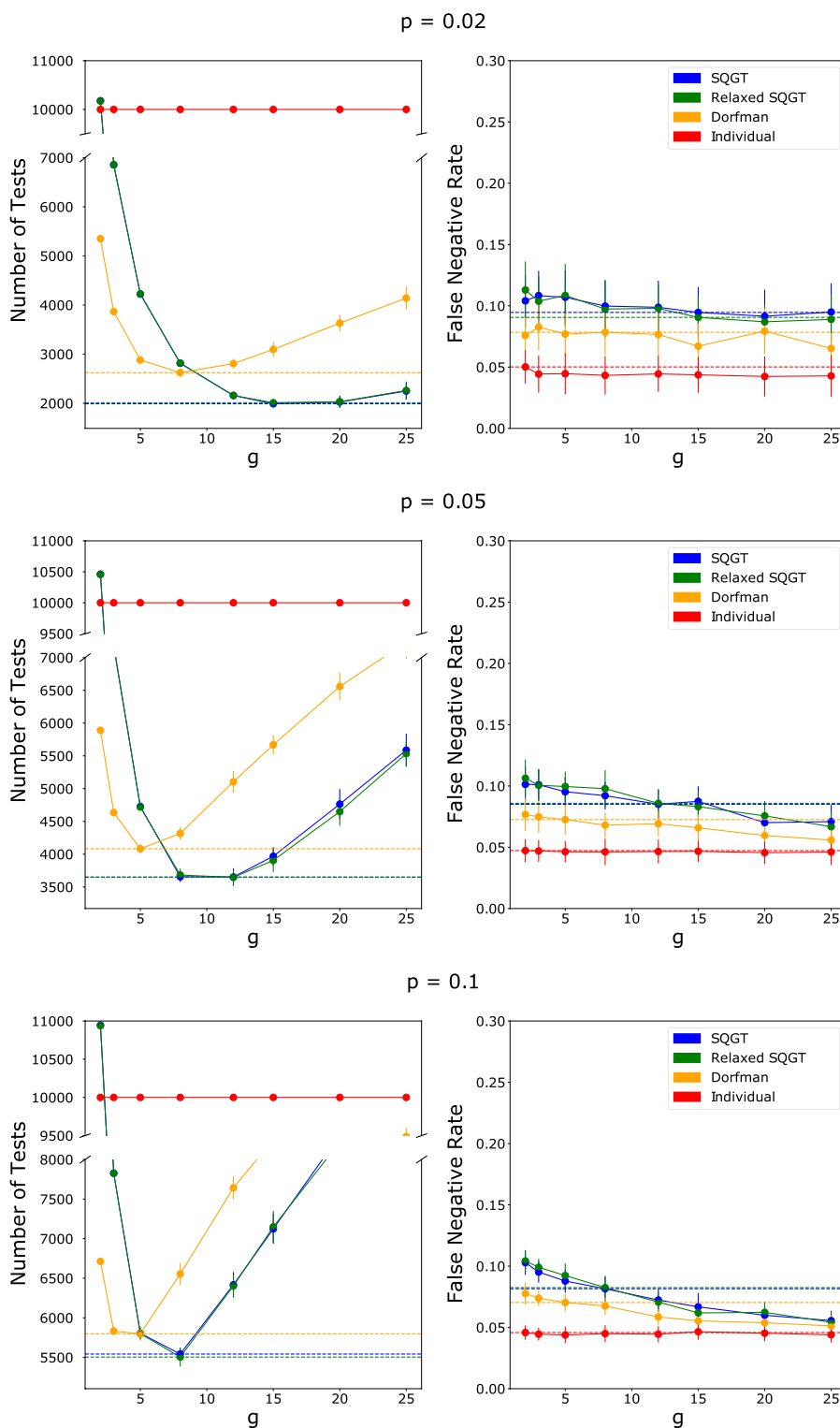
To analyze the magnitude of the savings in the number of tests required for the GT scheme compared to individual screening, independent of PCR assay noise, we ran

both Dorfman's GT and SQGT on the model data. The tests were performed under the assumption that qPCR assays are error-free. Additional file 1: Figure 1 shows these results for all three infection rates  $p$ . We performed a sweep of group sizes  $g$  for each value of  $p$  to identify their optimal values. While both GT schemes require significantly fewer than the 10,000 tests needed for individual testing, SQGT consistently outperforms Dorfman's GT for all three infection rate levels. In addition, Additional file 1: Figure 1 shows that the group-dependent thresholds help to avoid false negatives that would have occurred due to dilution effects, as expected.

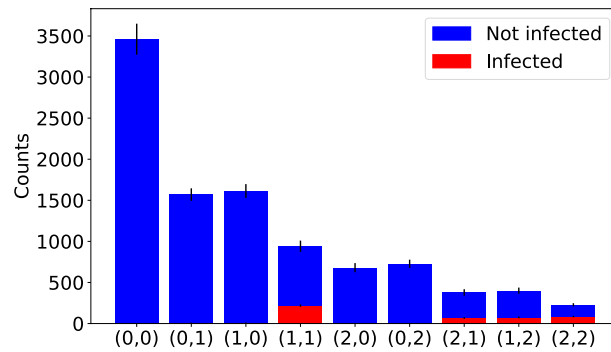
However, as noted in the previous section, qPCR assays are not error-free in practice, and as a result, the false negatives in GT schemes could be due to either dilution effects or qPCR noise. Therefore, we incorporated qPCR noise into our model to make it more realistic. This was done by including the empirically fitted FNR in Fig. 6 into the PCR assays in our model (see the "Methods" section for details). Figure 7 shows that while the noise has very limited effects on the number of tests required by each GT scheme, it does have the expected effect of increasing the FNR of both individual and group tests. For individual testing, the noise function we fit appears to correspond to an FNR of just under 0.05, which is comparable to the empirically determined values reported in [30] and [32]. The FNR values of both GT schemes are also consistently slightly higher than those of individual testing. To compare the FNR of SQGT and Dorfman's GT, we first identify the optimal group size for each scheme by picking the value  $g$  for which the scheme requires the least number of tests. When  $p = 0.02$ , the optimal value of  $g$  for SQGT was 15 with an average of 1989.8 tests required; at the same time, Dorfman's GT required 2623.6 tests for an optimal group size  $g = 8$ . These optimal group sizes correspond to FNRs of about 0.0946 for SQGT and 0.0784 for Dorfman's GT, respectively. When the infection rate is increased to 0.05, the optimal group sizes are smaller, with  $g = 12$  and  $g = 5$  for SQGT and Dorfman's GT, respectively. These group sizes correspond to 3651.7 tests with an FNR of 0.851 for SQGT and 4082.6 tests with an FNR of 0.726 for Dorfman's GT. Finally, at  $p = 0.1$  the optimal group size for SQGT was identified as  $g = 8$ , with 5,542.2 tests and an FNR of 0.815, while for Dorfman's GT the results indicated  $g = 5$ , with 5798.0 tests and an FNR of 0.703. The observed trend is that SQGT offers savings in the number of tests at the expense of a slight increase in FNR. It should also be noted that this increase is often within the error-bounds of the FNRs.

In addition, we tested a modified version of SQGT where individuals with a (2, 0) or (0, 2) result are declared negative without further testing. As shown in Fig. 7, this version of the SQGT method performs similarly to the regular SQGT. To investigate the reason behind this finding, we plotted the number of individuals for each possible outcome of the SQGT scheme for an infection rate of 0.04 and the corresponding optimal group size  $g = 12$ . As can be seen in Fig. 8, the (2, 0) and (0, 2) test results consist only of uninfected individuals. Therefore, it makes sense that declaring them negative without further testing has no effect on the FNR. For a mathematical analysis of the phenomena and related GT models, the reader is referred to Additional file 1: Section 1.2.

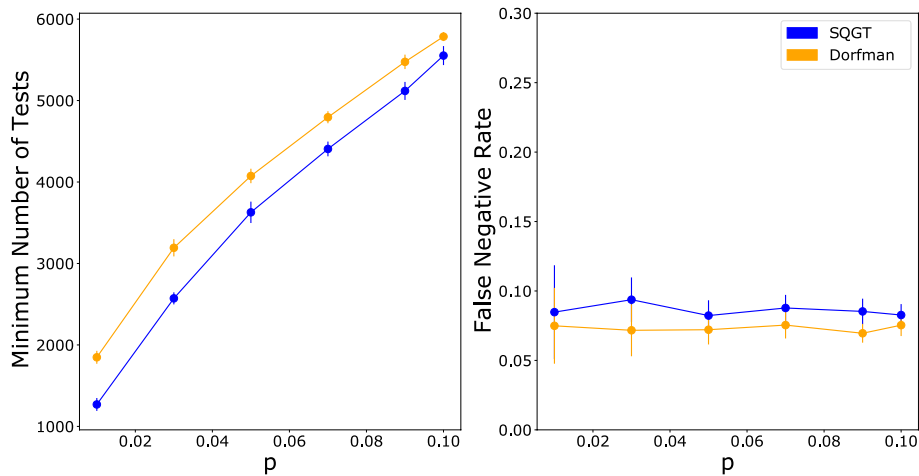
Finally, we examined how the number of tests required for the optimal group size varies over a wider range of infection rates, as shown in Fig. 9, alongside the corresponding FNRs. The figure shows that as the infection rate increases, the number of tests required for both GT schemes increases and the advantage of GT over individual testing



**Fig. 7** The number of tests used and the FNRs of the SQGT protocol (blue), Dorfman's GT (orange), and individual testing (red) for infection rates  $p \in \{0.02, 0.05, 0.1\}$ . The dashed lines show the number of tests and FNRs for the optimal group size (i.e., the group size that minimizes the number of tests needed) for each scheme. The optimal group size is strongly influenced by the scheme used for testing, and one cannot expect the same group size to be optimal for Dorfman's scheme and the proposed SQGT scheme



**Fig. 8** The number of individuals with each possible outcome for the pair of test results in the SQGT scheme. The number of infected individuals is shown in red, while the number of healthy (uninfected) individuals is shown in blue



**Fig. 9** The optimal number of tests used in Dorfman’s GT (orange) and SQGT (blue) versus the infection rate,  $p$  (left panel), and the corresponding FNRs (right panel). Optimal refers to the smallest number of tests possible under all possible choices of group sizes  $g$

decreases. This is a property that has been already established in the past for Dorfman’s scheme [14]. The values of the parameter  $p$  are chosen to reflect the different infection rates observed during the Covid-19 pandemics. For values of  $p$  greater than 0.38, the most effective (non-adaptive) testing method is individual testing. In general, the larger the value of  $p$ , the smaller the savings in the number of tests compared to individual screening. In addition, the figure shows that SQGT for PCR screening always saves more tests than Dorfman’s scheme with only a small increase in FNR (within the margin of error of Dorfman’s FNR). We repeated this experiment with viral loads drawn from a different dataset by [25] and obtained similar results as shown in Additional file 1: Figure 2.

### Discussion

We introduced the concept of Semi-Quantitative Group Testing (SQGT) as an extension of traditional GT methods, with a specific focus on qPCR-based pathogen screening. GT methods, in their classical form, are based on binary test outcomes (positive or negative) and are effective for identifying infected individuals in

a cost-efficient manner. However, they fail to utilize the full quantitative information provided by qPCR assays, which can lead to suboptimal performance in scenarios with widely varying pathogen concentrations.

SQGT addresses this limitation by interpreting test results as estimates of the number of infected individuals in each group. The proposed SQGT scheme utilizes two quantization thresholds to categorize qPCR results into different risk categories, allowing for a more refined analysis of the infection status within each group. By employing random permutations and two-stage testing, SQGT can reduce the number of tests needed while still maintaining a high level of test accuracy.

The study also addressed the issue of dilution effects in GT protocols, which can lead to false negatives in qPCR-based testing. To mitigate this problem, we incorporated group size-dependent thresholds in the SQGT framework, adjusting for the dilution effect and improving the overall accuracy of the test results.

Through extensive simulations and analysis using real-world qPCR data from SARS-CoV-2 testing, we demonstrated that SQGT outperforms traditional GT schemes (such as Dorfman's GT) in terms of test efficiency while maintaining a comparable or slightly higher FNR. For example, for a population infection rate of  $p = 0.02$ , our conceptually simple SQGT method uses 24% fewer tests than the binary outcome Dorfman's GT method, while maintaining a negligible FNR compared to qPCR noise. In conclusion, SQGT provides substantial reductions in the number of tests required for pathogen screening, making it a promising approach for large-scale population testing, especially during pandemics or outbreaks.

It is important to note that the proposed SQGT scheme is tailored specifically for qPCR testing and it involves two stages of testing, as originally suggested by Dorfman's scheme. The two stages are crucial for adaptive screening which informs the tests in the second stage based on the results in the first stage. Nonadaptive testing scheme, on the other hand, would result in potentially smaller delays of the test results but would require significantly more tests. They are also often too complicated to implement in practice as they require combinatorial sample mixing and decoding.

Additionally, our studies were performed under two assumptions, error-free qPCR assays, and qPCR assays with a sigmoidal model of false negatives as a function of  $Ct$  values. The incorporation of qPCR assay noise into the simulations led to a slight increase in FNRs, highlighting the need for careful consideration of assay accuracy for a broader range of practical pathogen detection schemes.

For other pathogens and datasets, our SQGT scheme can be modified as needed by combining adaptive and nonadaptive test schemes, including more than two thresholds, and integrating a specialized technique for identifying "heavy hitters" (i.e., individuals with very high viral loads). These approaches are mathematically analyzed in the Additional file 1: Section 1.3.

An alternative approach to our choice of permutations for the first stage of the SQGT scheme is to use combinatorial designs [10], which would ensure that no two individuals participate in more than one test group together. We choose random permutations because they allow for tractable mathematical analysis (see Additional file 1). Another approach could be to use array-based sample pooling [26]. In this

setting, group sizes are limited to a value that is suboptimal for Dorfman's scheme and its modifications, especially in the presence of dilution effects.

## Methods

### Data

The real-world COVID-19 GT data [5] used in this paper contains 133, 816 samples collected between April and September 2020 in Israel and tested experimentally via Dorfman's pooling. The original data contains the following information for each individual sample:

- Sample id: A unique id for tracking the sample;
- Month: Information about the month when the sample is collected;
- Group id: An id indicating which group an individual sample belongs to in the test scheme. Samples within the same group share the same group id, and the test groups are of size 5 and 8;
- Result: Final test result for a sample (positive/negative);
- Sample viral  $Ct$ :  $Ct$  value of an individual test. Note that this value is not available when the group test involving the sample is negative;
- Group viral  $Ct$ :  $Ct$  value of the group to which the individual sample belongs to;
- Sample human  $Ct$ :  $Ct$  value of an individual test for amplifying the human ERV-3 [34] gene. This  $Ct$  value lying below a predetermined threshold serves as an internal control for whether a test was successful or not;
- Group human  $Ct$ :  $Ct$  value of the group test used for amplifying the human ERV-3 gene.

As pointed out in the Results section, there are some experimental inconsistencies between the results of the group tests and the individual tests. Specifically, in 70 out of 1887 positive tests, the results of the group tests were positive while all individuals within the groups tested negative. These results can be explained as false positives for the group test, or as false negatives for the individual tests. We used this information to model the FNR of the dataset as described in our Results section. Note that for simplicity we assume that there is only one positive individual sample within the group when a false negative result is recorded, as this is the most probable scenario. We hence use  $(\text{Group test } Ct - M \log_{10}(g))$  as the estimated  $Ct$  value for the individual test in the presence of a false negative, where  $g$  as before denotes the group size, while  $M \log_{10}(g) = 2.895$ . Our fitted model shown in Fig. 6a is obtained through the MATLAB `fit` function.

### Modelling COVID-19 group testing schemes

#### Modelling PCR tests

When modelling an individual test, individual  $i$  with a viral load  $v_i$  will have

$$Ct_i = -M \log_{10}(v_i) + B. \quad (5)$$

The values for  $M$  and  $B$  are set based on a previously established calibration curve [25]. Then given a threshold  $Ct_I$ , an individual  $i$  is considered positive for the virus if  $Ct_i < \tau_{In}$ . In our simulations we use  $\tau_{In} = 36$ .



To model a pooled test, the viral loads of individuals in a group are averaged and plugged into Eq. 6 to determine the  $Ct$  for the group. That is, for group  $j$  with individuals  $\{1, 2, \dots, g\}$

$$Ct_j = -M \log_{10} \left( \frac{1}{g} \sum_{i=1}^g v_{ji} \right) + B. \quad (6)$$

These group  $Ct$ s can then be used for different GT schemes as described in the “Algorithms and results” section.

#### Including PCR noise into models

Since PCR tests are not error-free, we also include some noise into the tests based on the FNR function

$$FNR(Ct) = \left[ 1 + \exp \left( \frac{a - Ct}{b} \right) \right]^{-1}, \quad (7)$$

where  $b$  is empirically determined to be 2.145 as discussed in the “Algorithms and results” section and  $a$  is the threshold used for the PCR test. To include this noise into our PCR simulations, we use the following procedure:

```

if test is individual then
     $Ct \leftarrow -M \log_{10}(v_i) + B$ 
else
     $Ct \leftarrow -M \log_{10} \left( \frac{1}{g} \sum_{i=1}^g v_{ji} \right) + B.$ 
end if
result  $\leftarrow$  Scheme( $Ct$ )
if truth == positive then
    if Bernoulli( $p = FNR(Ct)$ ) then
        result  $\leftarrow$  negative
    end if
end if

```

First, the  $Ct$  value of a test is calculated using Eq. 5 or 6. If the ground truth of the test is that it is positive, it is converted into a negative (no infected individuals) with probability  $FNR(Ct)$ . Otherwise, the result of the test is left as determined by the testing scheme.

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12859-024-05798-3>.

**Additional file 1.** The Supplementary Information file includes a supplementary figure showing results from error-free PCR simulations and the formal analysis performed in this study.

#### Acknowledgements

We thank Professor Nigel Goldenfeld and Dr. Rayan Gabrys for useful discussions.

**Author contributions**

Conceptualization, SM, and OM; investigation, AN, CP, and VR; formal analysis, AN, CP, VR, MC, JR, SM, OM; writing—original draft, AN, CP, VR, MC, JR, SM, and OM; writing—review and editing, AN, CP, VR, SM, and OM. All contributing authors are listed in alphabetical order.

**Funding**

The work was supported by NSF Grants 2107344 and 2107345.

**Availability of data and materials**

The notebooks and data used for simulations are available at <https://github.com/maslov-group/SQGT>.

**Declarations****Ethics approval and consent to participate**

Although we performed an analysis of real data involving human subjects, we did not collect the data ourselves or perform the actual experiments. The rules and protocols used adhere to the regulations in the countries in which the data was acquired. The data from [5] was approved by the Hadassah Medical Center Institutional Review Board with a waiver from the need for informed consent. The anonymized data from [25] was cleared under paragraph 25 of the Berlin Landeskrankenhausgesetz.

**Consent for publication**

Not applicable.

**Conflict of interest**

The authors declare no conflict of interest.

Received: 12 August 2023 Accepted: 26 April 2024

Published online: 17 May 2024

**References**

1. Aeron S, Saligrama V, Zhao M. Information theoretic bounds for compressed sensing. *IEEE Trans Inf Theory*. 2010;56(10):5111–30.
2. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Introduction to pathogens. In: *Molecular biology of the cell*, 4th edn. Garland Science; 2002.
3. Alcoba-Florez J, Gil-Campesino H, de Artola DG-M, Díez-Gil O, Valenzuela-Fernández A, González-Montelongo R, Ciuffreda L, Flores C. Increasing SARS-CoV-2 RT-qPCR testing capacity by sample pooling. *Int J Infect Dis*. 2021;103:19–22.
4. Baker RE, Mahmud AS, Miller IF, Rajeev M, Rasambainarivo F, Rice BL, Takahashi S, Tatem AJ, Wagner CE, Wang L-F, et al. Infectious disease in an era of global change. *Nat Rev Microbiol*. 2022;20(4):193–205.
5. Barak N, Ben-Ami R, Sido T, Perri A, Shtoyer A, Rivkin M, Licht T, Peretz A, Magenheim J, Fogel I, et al. Lessons from applied large-scale pooling of 133,816 SARS-CoV-2 RT-PCR tests. *Sci Transl Med*. 2021;13(589): eabf2823.
6. Berger T, Mandell J W. Bounds on the efficiency of two-stage group testing. In: *Codes, curves, and signals: common threads in communications*. Springer; 1998. pp. 213–232.
7. Brault V, Mallein B, Rupprecht J-F. Group testing as a strategy for covid-19 epidemiological monitoring and community surveillance. *PLoS Comput Biol*. 2021;17(3): e1008726.
8. Candes EJ, Romberg JK, Tao T. Stable signal recovery from incomplete and inaccurate measurements. *Commun Pure Appl Math J Issued Courant Inst Math Sci*. 2006;59(8):1207–23.
9. Cheraghchi M, Gabrys R, Milenkovic O. Semiquantitative group testing in at most two rounds. In: 2021 IEEE international symposium on information theory (ISIT). IEEE; 2021. pp. 1973–1978.
10. Colbourn CJ. *CRC handbook of combinatorial designs*. Boca Raton: CRC Press; 2010.
11. Dai W, Milenkovic O. Subspace pursuit for compressive sensing signal reconstruction. *IEEE Trans Inf Theory*. 2009;55(5):2230–49.
12. de Salazar A, Aguilera A, Trastoy R, Fuentes A, Alados JC, Causse M, Galán JC, Moreno A, Trigo M, Pérez-Ruiz M, et al. Sample pooling for SARS-CoV-2 RT-PCR screening. *Clin Microbiol Infect*. 2020;26(12):1687–e1.
13. Donoho DL. Compressed sensing. *IEEE Trans Inf Theory*. 2006;52(4):1289–306.
14. Dorfman R. The detection of defective members of large populations. *Ann Math Stat*. 1943;14(4):436–40.
15. Eberhardt JN, Breuckmann NP, Eberhardt CS. Multi-stage group testing improves efficiency of large-scale covid-19 screening. *J Clin Virol*. 2020;128: 104382.
16. Emad A, Milenkovic O. Semiquantitative group testing. *IEEE Trans Inf Theory*. 2014;60(8):4614–36.
17. Emad A, Milenkovic O. Code construction and decoding algorithms for semi-quantitative group testing with nonuniform thresholds. *IEEE Trans Inf Theory*. 2016;62(4):1674–87.
18. Fajnzylber J, Regan J, Coxen K, Corry H, Wong C, Rosenthal A, Worrall D, Giguelf F, Piechocka-Trocha A, Atyeo C, et al. Sars-cov-2 viral load is associated with increased disease severity and mortality. *Nat Commun*. 2020;11(1):1–9.
19. Fraser C, Hollingsworth TD, Chapman R, de Wolf F, Hanage WP. Variation in hiv-1 set-point viral load: epidemiological analysis and an evolutionary hypothesis. *Proc Natl Acad Sci*. 2007;104(44):17441–6.
20. Ghosh S, Agarwal R, Rehan MA, Pathak S, Agarwal P, Gupta Y, Consul S, Gupta N, Goenka R, Rajwade A, et al. A compressed sensing approach to pooled RT-PCR testing for covid-19 detection. *IEEE Open J Signal Process*. 2021;2:248–64.

21. Hogan CA, Sahoo MK, Pinsky BA. Sample pooling as a strategy to detect community transmission of sars-cov-2. *JAMA*. 2020;323(19):1967–9.
22. Hu X, Xing Y, Jia J, Ni W, Liang J, Zhao D, Song X, Gao R, Jiang F. Factors associated with negative conversion of viral rna in patients hospitalized with covid-19. *Sci Total Environ*. 2020;728: 138812.
23. Hwang FK. Group testing with a dilution effect. *Biometrika*. 1976;63(3):671–80.
24. Indyk P, Ngo HQ, Rudra A. Efficiently decodable non-adaptive group testing. In: Proceedings of the twenty-first annual ACM-SIAM symposium on discrete algorithms. SIAM; 2010. pp. 1126–1142.
25. Jones TC, Biele G, Mühlemann B, Veith T, Schneider J, Beheim-Schwarzbach J, Bleicker T, Tesch J, Schmidt ML, Sander LE, et al. Estimating infectiousness throughout SARS-COV-2 infection course. *Science*. 2021;373(6551): eabi5273.
26. Kim H-Y, Hudgens MG, Dreyfuss JM, Westreich DJ, Pilcher CD. Comparison of group testing algorithms for case identification in the presence of test error. *Biometrics*. 2007;63(4):1152–63.
27. Lin Y, Ren Y, Wan J, Cashore M, Wan J, Zhang Y, Frazier P, Zhou E. Group testing enables asymptomatic screening for covid-19 mitigation: feasibility and optimal pool size selection with dilution effects. *arXiv preprint arXiv:2008.06642*; 2020.
28. Lindstrom B. Determining subsets by unramified experiments. In: A survey of statistical design and linear models; 1975.
29. McMahan CS, Tebbs JM, Bilder CR. Regression models for group testing data with pool dilution effects. *Biostatistics*. 2013;14(2):284–98.
30. Takahashi H, Ichinose N, Okada Y. False-negative rate of sars-cov-2 rt-pcr tests and its relationship to test timing and illness severity: a case series. *IDCases*. 2022;28: e01496.
31. Wolf J. Born again group testing: multiaccess communications. *IEEE Trans Inf Theory*. 1985;31(2):185–91.
32. Woloshin S, Patel N, Kesselheim AS. False negative tests for sars-cov-2 infection—challenges and implications. *N Engl J Med*. 2020;383(6): e38.
33. Yelin I, Aharony N, Tamar ES, Argoetti A, Messer E, Berenbaum D, Shafran E, Kuzli A, Gandali N, Shkedi O, et al. Evaluation of covid-19 RT-qPCR test in multi sample pools. *Clin Infect Dis*. 2020;71(16):2073–8.
34. Yuan CC, Miley W, Waters D. A quantification of human cells using an ERV-3 real time PCR assay. *J Virol Methods*. 2001;91(2):109–17.
35. Zheng S, Fan J, Yu F, Feng B, Lou B, Zou Q, Xie G, Lin S, Wang R, Yang X, et al. Viral load dynamics and disease severity in patients infected with sars-cov-2 in Zhejiang province, China, January–March 2020: retrospective cohort study. *bmj*. 2020;369: m1443.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.