

1 **Title:** Use of compressed sensing to expedite high-throughput diagnostic testing for COVID-19 and beyond

2

3 **Authors:**

4 Kody A. Waldstein^{1,2†}, Jirong Yi^{3†}, Michael Myung Cho⁴, Raghu Mudumbai³, Xiaodong Wu³, Steven M.
5 Varga^{1,2,5‡}, and Weiyu Xu^{3‡*}

6 †‡Denotes equal contribution

7

8 **Affiliations:**

9 ¹Interdisciplinary Graduate Program in Immunology, University of Iowa, Iowa City, IA 52242, USA.

10 ²Department of Microbiology and Immunology, University of Iowa, Iowa City, IA 52242, USA.

11 ³Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA 52242, USA.

12 ⁴Department of Electrical and Computer Engineering, Penn State Behrend, Erie, PA 16563, USA.

13 ⁵Department of Pathology, University of Iowa, Iowa City, IA 52242, USA.

14

15 *To whom correspondence should be addressed: **Weiyu Xu**, weiyu-xu@uiowa.edu

16

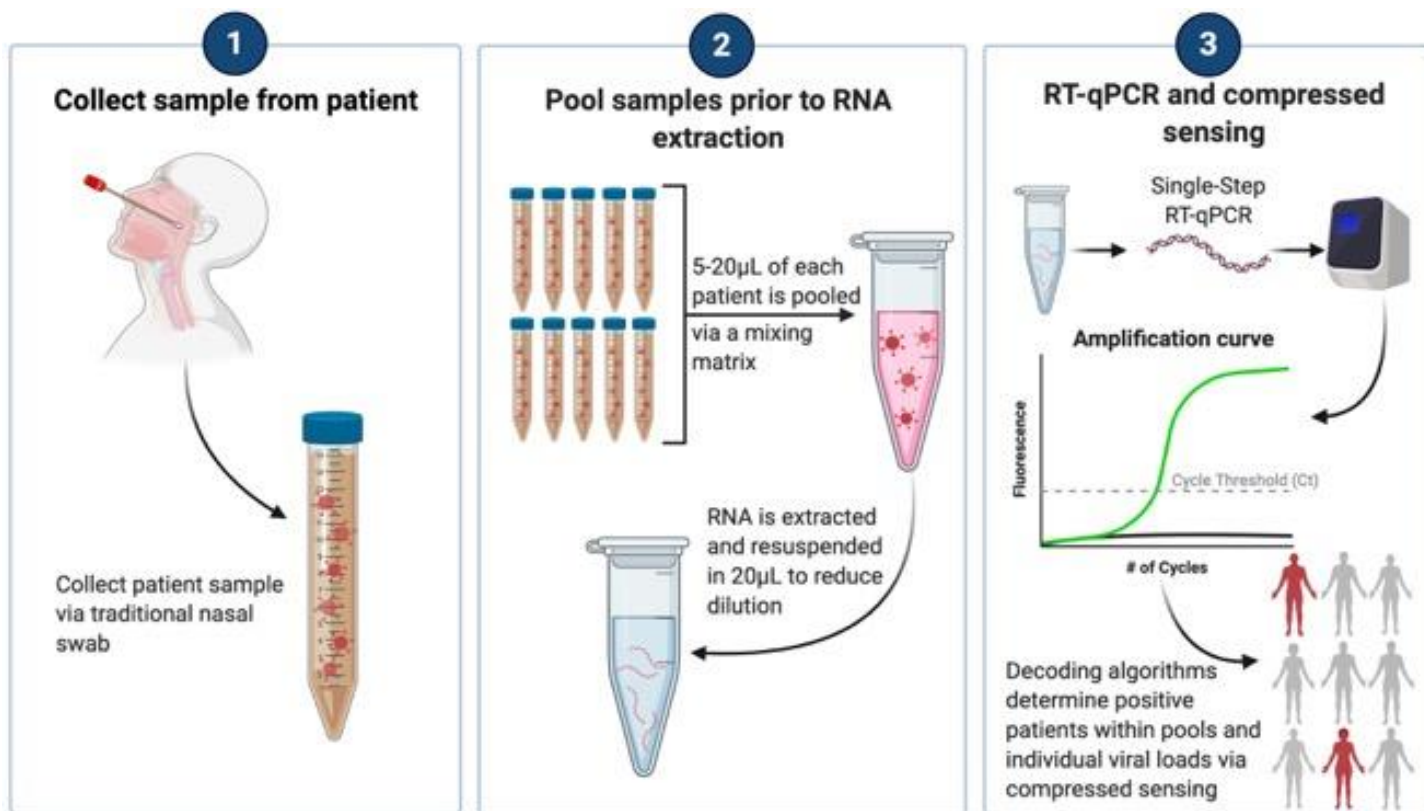
17

18 **Abstract**

19 The rapid spread of SARS-CoV-2 has placed a significant burden on public health systems to provide rapid and
20 accurate diagnostic testing highlighting the critical need for innovative testing approaches for future pandemics.
21 In this study, we present a novel sample pooling procedure based on compressed sensing theory to accurately
22 identify virally infected patients at high prevalence rates utilizing an innovative viral RNA extraction process to
23 minimize sample dilution. At prevalence rates ranging from 0-14.3%, the number of tests required to identify
24 the infection status of all patients was reduced by 75.6% as compared to conventional testing in primary human
25 SARS-CoV-2 nasopharyngeal swabs and a coronavirus model system. Additionally, our modified pooling and
26 RNA extraction process minimized sample dilution which remained constant as pool sizes increased. Our use of
27 compressed sensing can be adapted to a wide variety of diagnostic testing applications to increase throughput
28 for routine laboratory testing as well as a means to increase testing throughput to combat future pandemics.

29
30
31
32
33
34
35
36
37
38
39
40
41
42

43 **Graphical Abstract**



44

45

46

47

48

49

50

51

52

53

54

55

56

57

58 Introduction

59 The rapid community spread of SARS-CoV-2 has placed a significant burden on diagnostic testing and
60 public health to provide fast and accurate testing strategies. The number of COVID-19 tests being performed
61 each day has increased 8-fold since testing reagents became widely available with an average of over 1.5-2
62 million COVID-19 quantitative reverse transcription polymerase chain reaction (qRT-PCR) tests performed by
63 day in the United States alone (1-3). Additionally, multiple new and more infectious variants of COVID-19
64 have emerged worldwide harboring genetic mutations significant enough to evade recognition by host
65 antibodies causing some concern for current vaccine formulations (4-8). Testing and screening remains an
66 imperative safeguard to minimize spread, thus the development of innovative strategies and techniques to
67 increase testing capacity without reducing the accuracy and efficacy of testing is crucial.

68 A traditional method to increase testing capacity is by pooling samples as opposed to conducting
69 individualized testing, known as “group testing” (9-11). The principle is simple, if the prevalence rate is low
70 within the population, the majority of samples will inevitably test negative. In this scenario, a single negative
71 result indicates that all patients within that pool are also negative. However, the ability to accurately test using
72 this method diminishes quickly as the prevalence rate increases (12-15). Current CDC guidelines require
73 subsequent individual testing of all patients within a pool if the pool is positive (16). Worldwide SARS-CoV-2
74 prevalence rates continue to be >10% with a worldwide estimate of ~30% (17). These rates are well beyond the
75 capacity of traditional pooling methods as many pools will be positive requiring additional individual testing
76 and inevitably increasing the number of tests required. More sophisticated pooling efforts have arisen during the
77 pandemic though the testing models’ accuracy and effectiveness falls apart rapidly as the prevalence rate rises
78 and are thus not viable options for the current and future pandemics (13, 16, 18, 19).

79 In this study, we present a novel and innovative pooling protocol which utilizes mathematically-derived
80 mixing matrices and decoding algorithms to accurately identify positive patients within pools using the CDC-
81 approved range of positive Ct values at high prevalence rates. Additionally, we propose a new approach based
82 on compressed sensing theory for detection of viral load using pooled sample testing (20-22). We also employ

83 a modified RNA extraction process in which the patient swab samples are pooled prior to RNA extraction
84 allowing the sample to be concentrated thus minimizing sample dilution. This modified approach has shown
85 high accuracy and reproducibility at prevalence rates over 10% with large sample sizes using an experimental
86 mouse coronavirus, mouse hepatitis virus strain 1 (MHV-1) as well as human COVID-19 patient samples.

87

88 **Problem Formulation**

89 Notations: We use $[N]$ to denote the set $\{1, 2, \dots, N\}$, and \square_+^N to denote the set $[0, +\infty)^N$. We denote by $\Pr(E)$ the
90 probability of an event E , and use $\text{round}(x)$ to round x to the closest integer. The j -th element of a vector

91 $x \in \square_+^N$ is denoted by x_j or $(x)_j$. The support set or the set of indices corresponding to the nonzero elements of

92 a vector $p \in \square_+^n$ is denoted by $\text{supp}(p)$.

93 **Mixing matrix design**

94 *Parity check matrix and fixed dilution*

95 In this section, we introduce how the participation matrix P and the allocation matrix W are designed for
96 MHV-1 with small population size N , i.e., $N=7, 15$, and 31 . For a prevalence rate of 1%, there can be
97 approximately one infected sample for $N=7, 15$, and 31 . From information theory, we know that the parity
98 check matrices for Hamming codes can guarantee the correction of one error in codewords or the identification
99 of the parity check matrix column which corresponds to the error in the codewords (23). In the context of virus
100 testing, such parity check matrices can guarantee the identification of one positive from all the tested samples.
101 This exactly fits our need for a small population number with 1% prevalence, and we can use such parity
102 checking matrices as the participation matrices.

103 The construction of such parity check matrices can be described as follows. Suppose $P \in \{0, 1\}^{n \times N}$, then
104 we let $N = 2^n - 1$, and the columns of P are simply all the nonzero binary sequence of length n . As we consider
105 $N=7, 15$, and 31 , the corresponding participation matrices are shown in Figure 1 A-C.

The allocation matrix should be designed in correspondence with the practical clinic procedures for mixing the samples. As for the allocation matrix W for MHV-1 in our laboratory experiments, since we take 5 μL from each individual sample to form the sample pool which is then concentrated to a volume of 20 μL , this implies that the virus load for an individual sample in the mixing is $\frac{1}{4}$ of its original virus load. Thus, we can design the allocation matrix as follows:

$$W_{ij} = \begin{cases} 1/4, P_{ij} \neq 0 \\ 0, P_{ij} = 0 \end{cases}, i \in [n], j \in [N].$$

Bipartite graph matrix and equal partition

Though the parity check matrices of Hamming codes can be easily used as the participation matrix, it cannot scale up for high N or prevalence rates. This is because such parity check matrices can only guarantee the identification of one positive sample, while high N or prevalence rates can result in more than one positive sample in the population. Another consequence of a high N is the large number of nonzero elements in the participation matrix, which means high complexity during laboratory experiments. This motivates us to design participation matrices which can not only succeed in scenarios where more than 1 positive samples are present, but also have low complexity as indicated by the number of nonzero elements in the participation matrix. We propose to use the binary matrices constructed using a bipartite graph as the participation matrices (24, 25). For the COVID-19 experiments, we will use a well-designed binary matrix $P \in \{0,1\}^{16 \times 40}$ with each column having only 4 nonzero elements as shown in Figure 2.

For SARS-CoV-2 virus testing in our laboratory experiments, since equal volumes of samples participating in a particular pool are mixed together, and we did not perform sample concentration, the virus load for each individual sample in the mixture is actually scaled down by the number of participants. Thus, the allocation matrix can be designed as:

$$W_{ij} = \begin{cases} \frac{1}{N}, P_{ij} \neq 0 \\ \sum_{j=1}^N P_{ij} \\ 0, P_{ij} = 0 \end{cases}, i \in [n], j \in [N].$$

128 Since there is a one-to-one correspondence between the participation matrix P and the mixing matrix A ,
129 we will refer them alternatively in the subsequent sections without confusion.

130 *Mixing matrix and dilution upon adaptive requests*

131 Apart from the above pooling results with a prefixed mixing matrix, we can make requests for extra pooling
132 results adaptively according to the decoding results at each stage. The mixing matrices used in the adaptive
133 requests will depend on the specific decoding results, e.g., the determination of $P^{(i)} \in \mathbb{R}^{n_i \times N}$ cannot be
134 determined in advance. However, the corresponding allocation matrix will be designed according to the parity
135 check matrix for MHV-1 and the bipartite graph matrix for SARS-CoV-2 in our laboratory experiments.

136 *Sample pooling*

137 In many group testing processes, patient samples are pooled after RNA extraction or the total pool
138 volume dictates the RNA elution volume. In both cases, this means the fold dilution of each patient is dependent
139 on the total number of patients within a pool. Thus, as the number of patients pooled increases, the sample
140 become more dilute significantly increasing the probability of a false negative test result. This phenomenon has
141 required pools to remain small, usually under 5 patients per pool (14, 15, 26). To reduce the dilution effect of
142 pooling, a modified RNA isolation protocol was developed using TRIzol phenol/chloroform that can be more
143 broadly applied to RNA extraction kits and automated systems such as the KingFisher (27). With this method,
144 patient samples are pooled prior to RNA extraction. After the isopropanol precipitation and ethanol step, the
145 pelleted RNA can be significantly concentrated by reducing the final volume of water used to solubilize the
146 RNA thus minimizing the potential impact of sample dilution (**Fig. 2A**).

147 To test the dilution effect of traditional pooling on qRT-PCR Ct results as compared to our modified
148 RNA extraction protocol, we utilized the widely used murine coronavirus MHV-1 as a model system (28-33).
149 Using a MATLAB-derived computational script, we pseudo-randomly generated simulated patients based on a
150 Ct value range of 12-34 cycles. These experimental parameters were chosen from current CDC testing
151 guidelines and growing evidence that individuals with viral loads corresponding to a Ct value of 34 and above
152 are likely non-infectious and/or not reliable to diagnose positive patients (34-37). Additionally, in our hands, Ct

153 values greater than 36 are generated from MHV-1 and SARS-CoV-2 qRT-PCR reactions containing \approx 1-10
154 copies of the target gene and enter a realm where non-specific amplification and false positive rates increase.

155 Simulated patient samples were evaluated in qRT-PCR reactions as individuals to establish their ground
156 truth Ct values. The samples were subsequently individually mixed with viral transport media (VTM) to
157 generate dilutions of 8, 10, and 16-fold. The dilution was performed to simulate a situation where a single
158 patient within a pool is positive, and consequently, the addition of other negative patient samples contributes
159 solely to the dilution of the positive patient sample. RNA was extracted from each pool using TRIzol by either
160 the modified RNA extraction protocol or traditional group testing. An elution volume of 20 μ L was chosen to
161 allow a 5 μ L qRT-PCR test to be run in duplicate with 10 μ L remaining for a retest. (**Fig. 2A**).

162 As expected, samples pooled by traditional group testing exhibited a significant impact on the Ct value
163 resulting in signal dilution (**Fig. 2B**). However, the dilution effect was minimized or eliminated in the modified
164 RNA extraction protocol. (**Fig. 2B**). Importantly, the Δ Ct was consistent among all pools regardless of the
165 number of patients indicating the pool size could be significantly increased without causing further sample
166 dilution. One issue with increasing the number of patients within a pool is the corresponding increase in the
167 total volume of the pool. To reduce to total pool volume, we created pools by adding 5 μ L of sample from each
168 patient to the pool and eluting with 20 μ L resulting in a 1:4 dilution. This approach resulted in a significantly
169 smaller total pool volume with an average increase in Ct of 1.5 cycles with no correlation to the number of
170 patients within the pool (**Fig. 2B**).

171 These results suggest that the dilution caveat of traditional group testing can be minimized by
172 implementing our modified extraction protocol. Patient RNA samples can also be concurrently extracted
173 individually and banked if repeat testing is required. This approach provides a standard dilution effect that is
174 consistent regardless of either the pool size or the volume which significantly simplifies downstream
175 computation and decoding while reducing the chance of a false negative result.

176 *Virus load decoding with success certificate*

177 In this section, we describe a decoding algorithm which decodes each sample's viral load from testing
178 results of pooled samples. A unique feature of our decoding algorithm is the decoding success certificate it
179 provides: assuming that the testing results are accurate, we are guaranteed that the decoding results are the only
180 set of positive samples that fit the testing results.

181 We consider the problem of recovering a ground truth signal $x \in \mathbb{R}_+^N$ from its under-sampled
182 measurements. Given a mixing matrix $A \in \mathbb{R}^{n \times N}$ with $n < N$, suppose we have qualitative measurements
183 $p \in \{0,1\}^n$ and qualitative measurements $y \in \mathbb{R}_+^n$ for the n pools which are complicated functions of $x \in \mathbb{R}_+^N$, our
184 goal is to recover $x \in \mathbb{R}_+^N$ from $p \in \{0,1\}^n$ and $y \in \mathbb{R}_+^n$. More specifically, $p = h(f(Ax))$ where $h(\cdot): \mathbb{R}_+^n \rightarrow \{0,1\}^n$,
185 and $y = f(Ax)$ where $f(\cdot): \mathbb{R}_+^n \rightarrow \Omega$ and Ω is a set of valid Ct values.

186 In the qRT-PCR amplification and quantification process (2) for a mixture of multiple patient samples,
187 the quantitative relation between $b := Ax$ and Ct value y can be obtained via interpolation (38). This means the
188 function $f(\cdot)$ is the composite of the qRT-PCR amplification process and the interpolation operation. The Ct
189 value will be compared with a threshold value τ preset by the authority to determine the final status of the
190 mixture, and it varies under different scenarios. For the sake of reducing the false negative at the cost of more
191 later tests, the technician can be conservative enough to mark positive results for mixtures although they have
192 moderately large Ct values for which negative results can be assigned when the criterion is relaxed.

193 Our goal is to decode the status of x , i.e., positive (meaning that a sample is infected by virus) or
194 negative (meaning that a sample is not infected by virus) status for each sample, and the amount of virus in each
195 sample. We want to emphasize that in the virus testing practice, we will only have the Ct value data y , and the
196 qualitative data p which is obtained from the Ct value. The A_{ij} implies whether the sample j participates in the i -
197 th pooling test with $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, N$. Thus, if there is no error, a pool has positive results, i.e., $p_i = 1$
198 if and only if there is at least one positive element of x participating in the i -th pooling test. To achieve the
199 above goals, we apply techniques from compressed sensing to solve it, and we end up with solving under-
200 determined systems for x , i.e., $f^{-1}(y) = Ax$ where f^{-1} is the inverse function of f . The problem is usually solved by

201 $\min_x \|x\|_1$, such that $f^{-1}(y) = Ax$ under the assumption that x is sparse (20, 22). In virus testing, the Ct value is first
202 obtained from the qRT-PCR, and then used for interpolating the virus load $f^{-1}(y)$. This means the f^{-1} can be
203 treated as the interpolation procedure. We also consider $\min_x \|x\|_1$, such that $f^{-1}(y + \Delta y) = Ax$ where $\Delta y \in \mathbb{R}^n$
204 characterizes the noise occurring in the measurement of Ct values.

205 One difference between solving under-determined systems in compressed sensing and those in the virus
206 testing is that the values of N and n are small in the later, and large in the former. This subtle difference is
207 critical for successful recovery, and the commonly used L_1 minimization in compressed sensing may not be able
208 to recover x when N is small. Though the accuracy outcomes are favorable when N is large, this is not optimal
209 for reliability and keeping the complexity of mixing process low in clinical virus testing. (**Supplemental Fig.**
210 **1**). Thus, in this paper we will focus on the case where N is small.

211 *Compressed sensing decoding*

212 In this section, we present a novel algorithm for virus decoding (**Supplemental Fig. 2A**). Our proposed
213 algorithm consists of three components, i.e., a support set estimation component for qualitative decoding, a
214 quantitative decoding component which makes use of the results from the support set estimation component,
215 and an adaptive data requesting component which asks for more testing results for improve decoding
216 performance according to the qualitative and quantitative results.

217 In the support set estimation component, the goal is to give an initial estimate of the index sets of
218 positive samples, negative samples, and samples whose status cannot be determined, respectively. We propose
219 to solve a sequence of minimization and maximization pair for estimating an upper and a lower bound for each
220 element of $x \in \mathbb{R}_+^N$, i.e., for $i = 1, 2, \dots, N$, we solve

$$\begin{aligned} & \min_x x_i, \\ & \text{such that } L_j \leq (Ax)_j \leq U_j, j \in \text{supp}(p) \\ & x \geq 0, \end{aligned}$$

222 and

223

$$\begin{aligned} & \max_x x_i, \\ & \text{such that } L_j \leq (Ax)_j \leq U_j, j \in \text{supp}(p) \\ & x \geq 0, \end{aligned}$$

224

where $L_j = f^{-1}(y_j^{ub}) = f^{-1}(y_j + \eta)$ and $U_j = f^{-1}(y_j^{lb}) = f^{-1}(y_j - \eta)$ with $\eta > 0$ is a parameter characterizing the noise

225

in Ct value readings. We want to emphasize that in virus testing using qRT-PCR, a larger Ct value corresponds

226

to a smaller virus load (2). After we get the lower (upper) bound estimates $x_{lb}^* \in \mathbb{R}_+^N$ ($x_{ub}^* \in \mathbb{R}_+^N$), we compare

227

each of its element with a upper bound virus load threshold parameter $\varepsilon_{vlub} \in \mathbb{R}_+$ ($\varepsilon_{vllb} \in \mathbb{R}_+$). If $(x_{lb}^*)_i > \varepsilon_{vlub}$ or

228

$(x_{ub}^*)_i < \varepsilon_{vllb}$, we claim the i -th sample of x must be positive or negative. By repeating the comparison for each

229

$i \in [N]$, we can obtain index sets Pos and Neg which are the index sets of samples which must be positive and

230

negative, respectively. Finally, the index set of samples whose status cannot be determined can be obtained as

231

$U := [N] \setminus (Pos \cup Neg)$. The above algorithm is presented in Algorithm 2 (**Supplemental Fig. 2B**).

232

The set estimates Pos, Neg, U are then exploited in the quantitative decoding component whose core is

233

an exhaustive search algorithm. For the exhaustive search component, we solve a weighted least square for each

234

possible cardinality $k \in \{1, 2, \dots, |U|\}$ and for each possible support set $K \subseteq U$ with cardinality $|K| = k$, i.e.,

$$\begin{aligned} & \min_x \sum_{j \in \text{supp}(p)} \frac{(f^{-1}(y_j) - (Ax)_j)^2}{(f^{-1}(y_j))^2}, \\ & \text{such that } x_K > 0, x_{(Pos \cup U) \setminus K} = 0, x_{Neg} = 0. \end{aligned}$$

235

236

The main idea is to estimate a sample virus load $x \in \mathbb{R}_+^N$ such that the deviation between the estimated pool

237

virus load $(Ax)_j$ and the corresponding interpolated pool virus load $f^{-1}(y_j)$ is minimized. Due to the wide

238

range that the sample virus load can reside, i.e., from 10^{-6} to 10^6 , we normalize the deviation via a scaling

239

factor $\frac{1}{f^{-1}(y_j)}$. The algorithm is presented in supplemental files. Usually in practice, the combinatorial

240

characteristics of exhaustive search can bring high computational complexity and high accuracy. In our virus

241 testing problem, due to the small size of the problem, the exhaustive search can be a good option. Besides, the
242 support set estimation component can be used to further reduce the size of the combinatorial problem. Another
243 trick we use to reduce the computational complexity is that we try to find the sparsest solution. This is achieved
244 by finding the solution with the smallest support set such that the misfit between the estimated Ct value and the
245 measured Ct value is smaller than a given tolerance for all the observed positive pools.

246 In the data adaptive requesting component, based on the results from the support set estimation and the
247 quantitative decoding components, we design new pooling strategies for pooling samples. The extra pooled
248 testing results are obtained using individual samples whose status and virus load cannot be determined by
249 previous pooled testing results. The mixing matrices for pooling the undetermined individual samples can be
250 case-specific in practice. The algorithm is presented in Algorithm 3 (**Supplemental Fig. 2C**). Usually in
251 practice, the combinatorial characteristics of exhaustive search can bring high computational complexity though
252 it can achieve high accuracy for estimating x . In our virus testing problem, due to the small size of the problem,
253 the exhaustive search can be a good option. Besides, the Algorithm 2 can be used to further reduce the size of
254 the combinatorial problem. Another trick we use to reduce the computational complexity is that we try to find
255 the sparsest solution. This is achieved by finding the solution with the smallest support set such that the misfit
256 between the estimated Ct value and the measured Ct value is smaller than a given tolerance Δy for all the
257 observed positive pools.

258

259 **Results**

260 To demonstrate proof of concept, we began our initial experiments with the model coronavirus MHV-1
261 testing a range of experimental parameters (28-33). As in the pooling dilution effect experiments, a MATLAB-
262 based script was used to generate pseudorandom experimental parameters based on N total samples with a
263 prevalence rate of 1-10%. Samples were mixed together to form n different pools according to the participation
264 matrix in Figure 1. Total RNA was extracted from the generated pools utilizing our 1:4 modified pooling
265 technique (**Fig. 2A**). Total RNA isolated from sample pools was then amplified via qRT-PCR to generate a

266 numerical readout of cycle threshold values. To avoid accidental errors, for every group of N samples and a
267 given mixing matrix $A \in \mathbb{R}^{n \times N}$ (here n is just n_1 in Algorithm 1) experiments were duplicated.

268 In one of our experiments, pools 1, 2, 4, and 5 returned Ct values within the bounds to be considered
269 positive (**Table 1**). With this information alone, Algorithm 2 can decode the samples with
270 $Neg = \{3, 6, 8, 9, 12, 13, 14, 15, 16, 20, 21, 23, 24, 25, 27, 29\}$ as negative, and the rest of the samples are undetermined. This
271 means $U = \{1, 2, 4, 5, 7, 10, 11, 17, 18, 19, 22, 26, 28, 30, 31\}$, and $Pos = \emptyset$ (**Supplementary Table 1**). These sets are
272 consistent with the virus load decoded by exhaustive search in which the samples decoded by Algorithm 2 as
273 negative indeed have almost zero virus load, while those which are decoded as undetermined have virus loads
274 which are neither too big nor too small to be considered negative. However, from the decoding results from
275 Algorithm 3, we can see that apart from giving zero estimate for the virus load of samples specified by Neg , it
276 also estimates all samples from U , except sample 17, to have zero virus load. This can be validated with
277 request for one extra pooling test involving all the samples in U except 17.

278 After initial pooling and decoding, further pooling for confirmation testing may be required. We will
279 refer to the matrix in Figure 1C as $P^{(1)}$. From our decoding result, we request an additional pooling test ($P^{(2)}$)
280 since not all sample infection statuses can be determined with 100% certainty. Thus, we designed the mixing
281 matrix which pools all the samples that are highly likely false positive (**Supplementary Fig. 3**). Viral loads
282 which are very small in magnitude can be due to numerical error, and we can simply treat it as 0.

283 Overall, the infection status of 1325 unique experimentally generated samples were determined with
284 individual experimental prevalence rates ranging from 0-14.3% (**Table 2**). After a single round of testing, the
285 infection status of 97.4% of all samples was established with 100% certainty. One subsequent round of
286 verification testing identified the infection status of 98.9% samples with full certainty and 15 remaining samples
287 which required further testing to determine infection status with full certainty.

288 In total, 322 tests were required to identify all positive samples within the population of 1325 total
289 samples. This resulted in a 75.6% reduction in the total number of tests required as compared to individualized

290 testing. These experiments were repeated with similar parameters and results bringing the total number of
291 experimentally generated samples tested to 2650.

292 To validate our pooling and detection system, we obtained human patient RNA samples from the
293 University of Iowa diagnostic testing laboratory. Samples were provided as extracted RNA, thus our modified
294 RNA extraction protocol was not utilized and samples were mixed using traditional pooling (**Fig. 2A**). An
295 optimized participation matrix was generated to reflect the expected dilution effect (**Fig 1D**). Experimental
296 parameters were pseudo-randomly generated as previously described with a total N of 40 patients and a set
297 prevalence rate of 10%. The pooling results for one of two independent experiments is presented in table 3. For
298 both of the two runs, we requested extra pooling results for decoding, and thus required the generation of an
299 additional mixing matrix (**Supplementary Fig. 4**). Additional pooling results and individual patient viral loads
300 is shown in supplementary tables 2 and 3.

301 After one round of testing and compressed sensing decoding, 2 patients were identified and confirmed as
302 positive and 72 were confirmed as negative leaving 6 patients as likely positive. Two subsequent pools and four
303 individual confirmation tests provided adequate data points to determine the infection status of all patients with
304 100% certainty. 32 tests were required to determine the infection status of 92.5% of all patients. Additional
305 confirmatory testing brought the total tests performed to screen 80 patients to 38. This is a 52.5% reduction in
306 the number of tests needed as compared to current individual testing (**Table 2**).

307 **Discussion**

308 Together, our experimental data provides proof of concept and validates our compressed sensing pooling
309 system as an effective and reproducible method to greatly increase COVID-19 testing capacity while
310 simultaneously providing more diagnostic information by determining patient viral load. Using our novel
311 testing approach, we were able to identify positive samples with extreme accuracy at prevalence rates at 10% or
312 higher in both an MHV-1 coronavirus model system and human COVID-19 patient samples. This required
313 approximately one third as many tests as would be needed with current individual testing procedures.

314 Pooled testing is an effective approach to increase testing capacity and allow widespread screening to
315 occur and has been implemented with limited success during the COVID-19 pandemic (9-14, 26, 39-44).
316 However, current pooled testing efforts lose efficacy and precision at real world prevalence rates and ultimately
317 require substantial additional confirmation testing. In 2020 for the first time in the field, we proposed to use
318 compressed sensing techniques for quantitative virus testing with high prevalence, and computational
319 experiments validated the effectiveness of our method (42). Others such as Ghosh et al. and Shental et al.,
320 showed the superiority of compressed sensing virus testing technology using a non-adaptive approach though
321 their method could only succeed at low prevalence, e.g., less than 10% (43, 44). In contrast, our current work
322 uses an adaptive approach and can succeed at prevalence rates greater than 10% and utilizes a success
323 certificate to ensure results are accurate (**Section S1.3**). Additionally, one major caveat of pooled testing is
324 sample dilution and the increase of false negatives. To eliminate the pooling dilution effect, we utilized a
325 modified RNA extraction protocol which differs from current clinical diagnostic lab procedures by simply
326 concentrating the RNA to a set volume regardless of the patient input number (**Fig. 2A**). This standardizes the
327 dilution to an expected and reproducible ΔC_t from the ground truth value that does not change if the number of
328 patients within a pool increases (**Fig. 2B**). This protocol alone removes the risk of samples with low levels of
329 virus being diluted in a pool and being read as a false negative.

330 Our approach demonstrates an effective process to combat testing bottlenecks for future pandemics.
331 Many clinical testing labs currently utilize automated RNA extraction systems in which parameters can be
332 changed to fit our new protocols. Additionally, we have created a beta decoding software in which qPCR data
333 can be entered and the program will decode the data, identify positive patients, and generate additional pools for
334 further testing, if needed, all automatically (software code available upon request). Most importantly, the
335 application of our testing method is broad and can be applied to many testing applications within medicine and
336 beyond such as serum antibody testing, drug screening, avian influenza surveillance, water contamination
337 testing, etc. Our application of compressed sensing is perfectly positioned for testing applications such as these
338 as they are sparse by nature and require accurate results from many data points.

339 The emergence of new pathogens and deadly variants is ongoing and will continue to be a significant
340 threat to public health and humanity as a whole (4-8). Implementing a highly accurate pooled testing procedure
341 is absolutely critical to mitigating the spread of deadly viral pandemics such as COVID-19, thus saving lives
342 and decreasing the economic destruction from high mortality rates and widespread quarantines. Our use of
343 compressed sensing in pooled COVID-19 testing demonstrated high sensitivity in experimental infection
344 models with the model coronavirus MHV-1, as well as with primary human COVID-19 samples. The utilization
345 of compressed sensing theory in signal analysis is well established, but its use in the testing of physical
346 specimens has the potential to revolutionize how we provide accurate results when testing extremely large
347 numbers of samples. This will position healthcare professionals to rapidly respond to future pandemics by
348 identifying infected individuals early, minimizing spread, and thus saving lives.

349 **Materials and Methods**

350 *Generation of experimental parameters and positive MHV-1 samples*

351 We used a computer script to generate pseudorandom viral loads for each of N individual samples based
352 on an average prevalence rate of 5%, and positive patient Ct values in the range 12-34. The MHV-1 standard
353 curve was used to plot the generated sample Ct value (X) and interpolate the dilution of MHV-1 virus stock (Y)
354 required. According to these estimates Y , MHV-1 was diluted in viral transport media as in the CDC-approved
355 nasopharyngeal swab collection protocol. (34, 36).

356 *MHV-1 sample pooling*

357 5-20 μ L of generated MHV-1 samples were pooled together in equal volumes on ice as designated by
358 the appropriate mixing matrix. Negative samples were added as sterile viral transport media.

359 *Human patient sample pooling*

360 Human samples that were to be discarded were supplied as extracted RNA in 96-well plates from the
361 University of Iowa Diagnostic Testing Lab. Patients were identified as positive or negative with no information
362 on Ct number, viral load, or any patient identifiable information. 5 μ L of patient samples were pooled together
363 in equal volumes on ice as designated by the appropriate mixing matrix. The University of Iowa determined that
364 this project did not meet the regulatory definition of human subjects research and therefore IRB approval was
365 not required.

366 *Isolation of viral RNA*

367 Viral RNA was extracted via a modified TRIzol phenol/chloroform extraction protocol and can be
368 scaled as needed (**Fig 2**). A patient pool of 20 μ L total volume was mixed with 200 μ L TRIzol. The sample was
369 vortexed for 10 sec and incubated for 5 min at room temperature (RT). 40 μ L of chloroform was added,
370 vortexed for 10 sec, and incubated for 5 min at RT. The mixture was centrifuged at 12,000 x g for 10 min at
371 4°C. 100 μ L of the upper aqueous layer was transferred to a sterile 1.5 mL tube. 100 μ L of isopropanol
372 supplemented with 2 μ g glycogen was added, vortexed for 10 sec, and incubated for 5 min at RT. The pellet
373 was mixed with 180 μ L of 75% ethanol and resuspended by gentle inversion and centrifuged at 14,000 x g for
374 10 min at RT. The supernatant was aspirated and the pellet was air dried for 10 min in a sterile laminar flow
375 hood. The RNA pellet was resuspended in 20 μ L of RNase-free diethyl pyrocarbonate-treated H₂O and
376 incubated at 55°C for 5 min.

377 *qRT-PCR*

378 5 μ L of patient pools and samples were mixed with the GoTaq qRT-PCR master mix (Promega) and ran
379 in duplicate on a QuanStudio 3 thermocycler via the FAST qRT-PCR protocol as recommended by the CDC
380 (36). An MHV-1 virus stock or SARS-Cov-2 S protein containing plasmid of known concentrations were used
381 to generate a standard curve consisting of seven to ten 10-fold serial dilutions. The resulting amplification
382 curves were analyzed with AppliedBiosystems Design and Analysis 2.4.

383 Compressed Sensing Decoding

384 An optimization algorithm leveraging the non-negativity of viral loads was used to give an upper and
385 lower bound on the viral load for each sample. If the lower bound for a sample's viral load is not zero, we are
386 sure that that sample is positive; if the upper bound for a sample's viral load is equal to 0, we are sure that that
387 sample is negative. This identifies samples which are either definitely positive or definitely negative. For the
388 samples with ambiguous infection statuses, we perform exhaustive search for the smallest set of positive
389 samples (namely sparsest solution, having the smallest number of positive samples) fitting the observed viral
390 loads of these pools. The remaining samples were mixed together into a pooled sample to confirm that they are
391 indeed negative: if this pooled sample comes back positive, further testing will be necessary, but this is
392 statistically unlikely.

394 References

- 395 1. COVID-19 Response. COVID-19 Case Surveillance Public Data Access, Summary, and Limitations
396 (version date: November 30, 2020). *Centers for Disease Control and Prevention*.
- 397 2. Nolan T, Hands RE, and Bustin SA. Quantification of mRNA using real-time RT-PCR. *Nature*
398 *Protocols*. 2006;1(3):1559-82.
- 399 3. Daily State-by-State Testing Trends. *Johns Hopkins Coronavirus Resource Center*.
- 400 4. Greaney AJ, Loes AN, Crawford KHD, Starr TN, Malone KD, Chu HY, et al. Comprehensive mapping
401 of mutations to the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human
402 serum antibodies. *bioRxiv*. 2021:2020.12.31.425021.
- 403 5. McCarthy KR, Rennick LJ, Nambulli S, Robinson-McCarthy LR, Bain WG, Haidar G, et al. Natural
404 deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. *bioRxiv*.
405 2020:2020.11.19.389916.
- 406 6. Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, et al. Estimated
407 transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science*. 2021:eabg3055.
- 408 7. England PH. Investigation of novel SARS-CoV-2 variant: variant of concern 202012/01. *Public Health*
409 *England Briefing*. 2020.
- 410 8. Firestone MJ, Lorentz AJ, Wang X, Como-Sabetti K, Vetter S, Smith K, et al. First Identified Cases of
411 SARS-CoV-2 Variant B.1.1.7 in Minnesota - December 2020-January 2021. *MMWR Morb Mortal Wkly*
412 *Rep*. 2021;70(8):278-9.
- 413 9. Dorfman R. The Detection of Defective Members of Large Populations. *The Annals of Mathematical*
414 *Statistics*. 1943;14(4):436-40.
- 415 10. Arnold ME, Slomka MJ, Coward VJ, Mahmood S, Raleigh PJ, and Brown IH. Evaluation of the pooling
416 of swabs for real-time PCR detection of low titre shedding of low pathogenicity avian influenza in
417 turkeys. *Epidemiol Infect*. 2013;141(6):1286-97.
- 418 11. Taylor SM, Juliano JJ, Trottman PA, Griffin JB, Landis SH, Kitsa P, et al. High-Throughput Pooling
419 and Real-Time PCR-Based Strategy for Malaria Detection. *Journal of Clinical Microbiology*.
420 2010;48(2):512.
- 421 12. Morandi PA, Schockmel GA, Yerly S, Burgisser P, Erb P, Matter L, et al. Detection of human
422 immunodeficiency virus type 1 (HIV-1) RNA in pools of sera negative for antibodies to HIV-1 and
423 HIV-2. *J Clin Microbiol*. 1998;36(6):1534-8.

- 424 13. Mutesa L, Ndishimye P, Butera Y, Souopgui J, Uwineza A, Rutayisire R, et al. A pooled testing strategy
425 for identifying SARS-CoV-2 at low prevalence. *Nature*. 2020.
- 426 14. Pikovski A, and Bentele K. Pooling of coronavirus tests under unknown prevalence. *Epidemiology and*
427 *infection*. 2020;148:e183-e.
- 428 15. Cherif A, Grobe N, Wang X, and Kotanko P. Simulation of Pool Testing to Identify Patients With
429 Coronavirus Disease 2019 Under Conditions of Limited Test Availability. *JAMA Network Open*.
430 2020;3(6):e2013075-e.
- 431 16. Interim Guidance for Use of Pooling Procedures in SARS-CoV-2 Diagnostic, Screening, and
432 Surveillance Testing. *Centers for Disease Control and Prevention*. 2020;NCIRD Division of Viral
433 Diseases.
- 434 17. Louca S. COVID-19 prevalence in 161 countries and over time. *medRxiv*. 2020:2020.12.01.20241539.
- 435 18. Polage CR, Lee MJ, Hubbard C, Rehder C, Cardona D, Denny T, et al. Assessment of an Online Tool to
436 Simulate the Effect of Pooled Testing for SARS-CoV-2 Detection in Asymptomatic and Symptomatic
437 Populations. *JAMA Network Open*. 2020;3(12):e2031517-e.
- 438 19. Deka S, and Kalita D. Effectiveness of Sample Pooling Strategies for SARS-CoV-2 Mass Screening by
439 RT-PCR: A Scoping Review. *J Lab Physicians*. 2020;12(3):212-8.
- 440 20. Donoho DL. Compressed sensing. *IEEE Transactions on Information Theory*. 2006;52(4):1289-306.
- 441 21. Candes EJ, and Tao T. Decoding by linear programming. *IEEE Transactions on Information Theory*.
442 2005;51(12):4203-15.
- 443 22. Candes EJ, and Tao T. Near-Optimal Signal Recovery From Random Projections: Universal Encoding
444 Strategies? *IEEE Transactions on Information Theory*. 2006;52(12):5406-25.
- 445 23. Cover TM, and Thomas JA. *Elements of Information Theory, 2nd Edition*. Wiley; 2006.
- 446 24. Cho M, Vijay Mishra K, and Xu W. Computable performance guarantees for compressed sensing
447 matrices. *EURASIP Journal on Advances in Signal Processing*. 2018;2018(1):16.
- 448 25. Jafarpour S, Xu W, and Hassibi B. Efficient and Robust Compressed Sensing Using Optimized Expander
449 Graphs. *IEEE Transactions on Information Theory*. 2009;55(9):4299-308.
- 450 26. Praharaj I, Jain A, Singh M, Balakrishnan A, Dhodapkar R, Borkakoty B, et al. Pooled testing for
451 COVID-19 diagnosis by real-time RT-PCR: A multi-site comparative evaluation of 5- & 10-sample
452 pooling. *Indian J Med Res*. 2020;152(1 & 2):88-94.
- 453 27. KingFisher Flex User Manual *Thermo Fisher Scientific* Rev 1.2.
- 454 28. Hartwig SM, Holman KM, and Varga SM. Depletion of alveolar macrophages ameliorates virus-
455 induced disease following a pulmonary coronavirus infection. *PLoS One*. 2014;9(3):e90720.
- 456 29. Leibowitz JL, Srinivasa R, Williamson ST, Chua MM, Liu M, Wu S, et al. Genetic determinants of
457 mouse hepatitis virus strain 1 pneumovirulence. *J Virol*. 2010;84(18):9278-91.
- 458 30. De Albuquerque N, Baig E, Ma X, Zhang J, He W, Rowe A, et al. Murine hepatitis virus strain 1
459 produces a clinically relevant model of severe acute respiratory syndrome in A/J mice. *J Virol*.
460 2006;80(21):10382-94.
- 461 31. Khanolkar A, Hartwig SM, Haag BA, Meyerholz DK, Epping LL, Haring JS, et al. Protective and
462 pathologic roles of the immune response to mouse hepatitis virus type 1: implications for severe acute
463 respiratory syndrome. *J Virol*. 2009;83(18):9258-72.
- 464 32. Khanolkar A, Fulton RB, Epping LL, Pham NL, Tifrea D, Varga SM, et al. T cell epitope specificity and
465 pathogenesis of mouse hepatitis virus-1-induced disease in susceptible and resistant hosts. *J Immunol*.
466 2010;185(2):1132-41.
- 467 33. Khanolkar A, Hartwig SM, Haag BA, Meyerholz DK, Harty JT, and Varga SM. Toll-like receptor 4
468 deficiency increases disease and mortality after mouse hepatitis virus type 1 infection of susceptible
469 C3H mice. *J Virol*. 2009;83(17):8946-56.
- 470 34. Interim Guidelines for Collecting, Handling, and Testing Clinical Specimens from Persons for
471 Coronavirus Disease 2019 (COVID-19) (Version Date: November 5, 2020). *Centers for Disease Control*
472 *and Prevention*.
- 473 35. La Scola B, Le Bideau M, Andreani J, Hoang VT, Grimaldier C, Colson P, et al. Viral RNA load as
474 determined by cell culture as a management tool for discharge of SARS-CoV-2 patients from infectious
475 disease wards. *European Journal of Clinical Microbiology & Infectious Diseases*. 2020;39(6):1059-61.
- 476 36. CDC 2019-Novel Coronavirus (2019-nCoV) Real-Time RT-PCR Diagnostic Panel. CDC-006-00019,
477 Revision: 05. *Centers for Disease Control and Prevention*.
- 478 37. Tom MR, and Mina MJ. To Interpret the SARS-CoV-2 Test, Consider the Cycle Threshold Value. *Clin*
479 *Infect Dis*. 2020;71(16):2252-4.
- 480 38. Real-time PCR handbook. *Thermo Fisher Scientific*. 2014.

- 481 39. Alcoba-Florez J, Gil-Campesino H, García-Martínez de Artola D, Díez-Gil O, Valenzuela-Fernández A,
482 González-Montelongo R, et al. Increasing SARS-CoV-2 RT-qPCR testing capacity by sample pooling.
483 *Int J Infect Dis.* 2020:S1201-9712(20)32468-1.
- 484 40. Denny TN AL, Bonsignori M, et al. Implementation of a Pooled Surveillance Testing Program for
485 Asymptomatic SARS-CoV-2 Infections on a College Campus — Duke University, Durham, North
486 Carolina, August 2–October 11, 2020. *MMWR Morb Mortal Wkly Rep.* 2020;69:1743-7.
- 487 41. Jirong Yi MC, Xiaodong Wu, Weiyu Xu, Raghu Mudumbai. Error Correction Codes for COVID-19
488 Virus and Antibody Testing: Using Pooled Testing to Increase Test Reliability. *arXiv:200714919* 2020.
- 489 42. Jirong Yi RM, Weiyu Xu. Low-Cost and High-Throughput Testing of COVID-19 Viruses and
490 Antibodies via Compressed Sensing: System Concepts and Computational Experiments.
491 *arXiv:200405759.* 2020.
- 492 43. Shental N, Levy S, Wuvshet V, Skorniakov S, Shalem B, Ottolenghi A, et al. Efficient high-throughput
493 SARS-CoV-2 testing to detect asymptomatic carriers. *Science Advances.* 2020;6(37):eabc5961.
- 494 44. Ghosh S, Agarwal R, Rehan MA, Pathak S, Agarwal P, Gupta Y, et al. A Compressed Sensing
495 Approach to Pooled RT-PCR Testing for COVID-19 Detection. *IEEE Open Journal of Signal*
496 *Processing.* 2021:1-.
- 497

Funding: Research reported in this publication was supported by funds from the Iowa Institute of Artificial Intelligence (to WX), National Science Foundation Award #2031218 (to WX) and the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award number T32AI007485 (to KAW). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author contributions:

Conceptualization: RM, XW, SMV, WX

Methodology: KAW, JY, WX, MC

Investigation: KAW, JY

Funding acquisition: SMV, WX

Supervision: SMV, WX,

Writing – original draft: KAW

Writing – review & editing: KAW, JY, WX, SMV

Competing interests: The authors are coinventors of a pending patent covering the use of compressed sensing in diagnostic testing applications.

Data and materials availability: All data, code, and materials used in this study are available upon request.

Figure 1. Optimized group testing mixing matrix design. (A-C) Hamming code parity check pooling matrix design for $N=7, 15,$ and 31 . (A) $N=7$ numerical matrix with 3 pools (3×7). (B) $N=15$ numerical matrix with 4 pools (4×15). (C) $N=31$ pixel matrix with 5 pools (5×31). (D) Bipartite pooling matrix design optimized for high N and prevalence rates. $N=40$ pixel matrix with 16 pools (16×40). (A,B) 1 indicates patient is included in the pool. 0 indicates the patient is not included in the pool. (C,D) White pixel indicates patient included in pool. Black pixel indicates patient not included in pool.

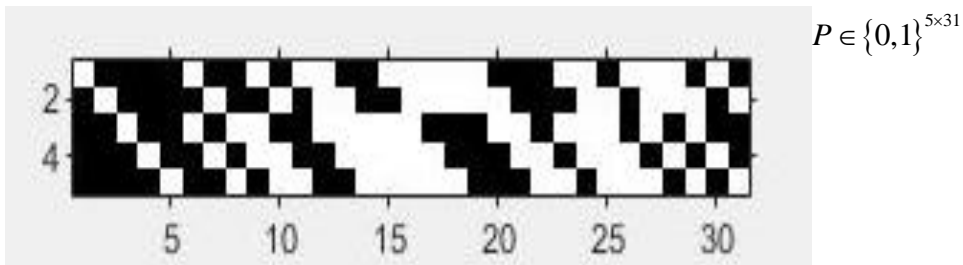
A)

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix} \in \{0,1\}^{3 \times 7}$$

B)

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} \in \{0,1\}^{4 \times 15}$$

C)



D)

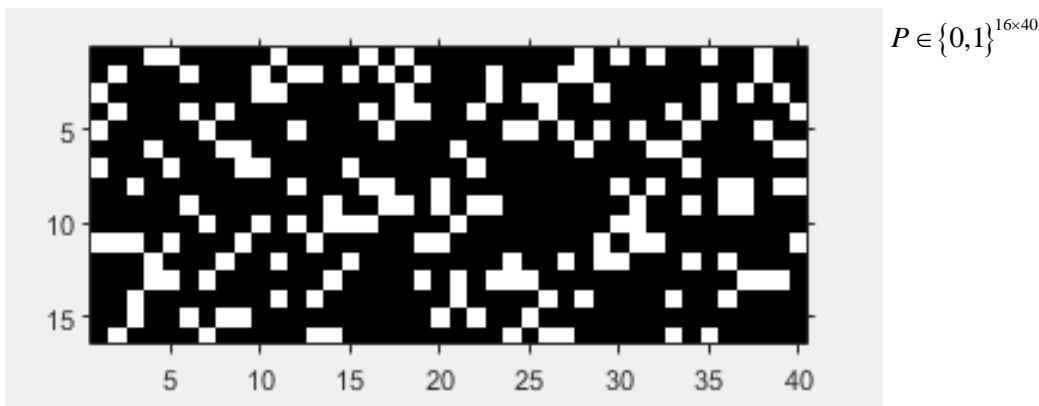
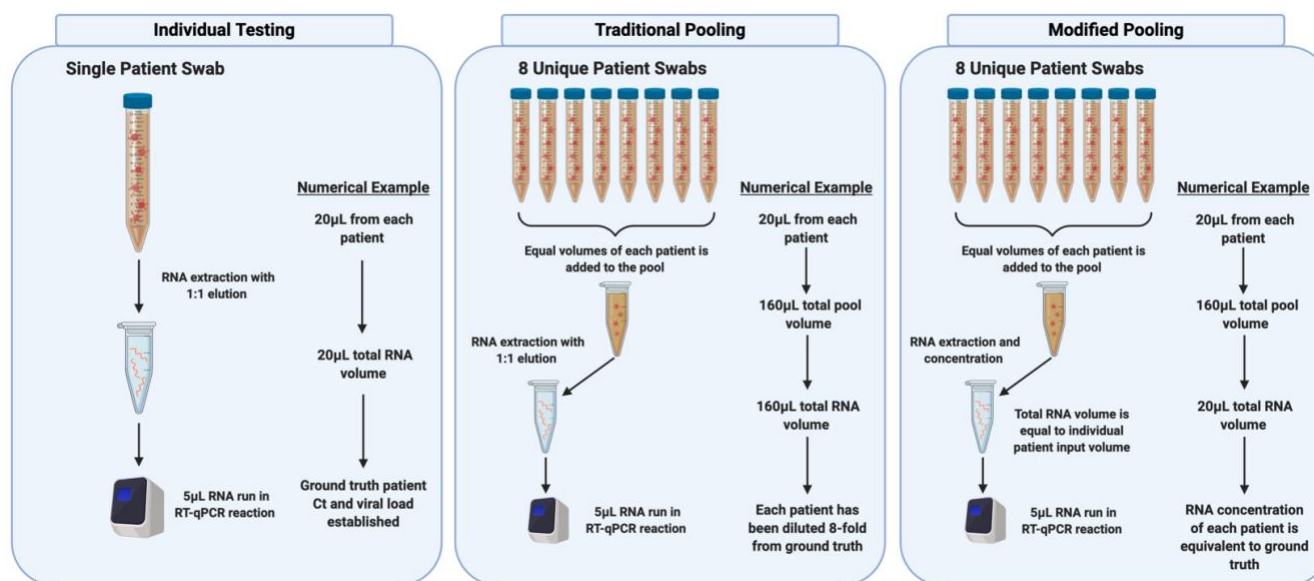
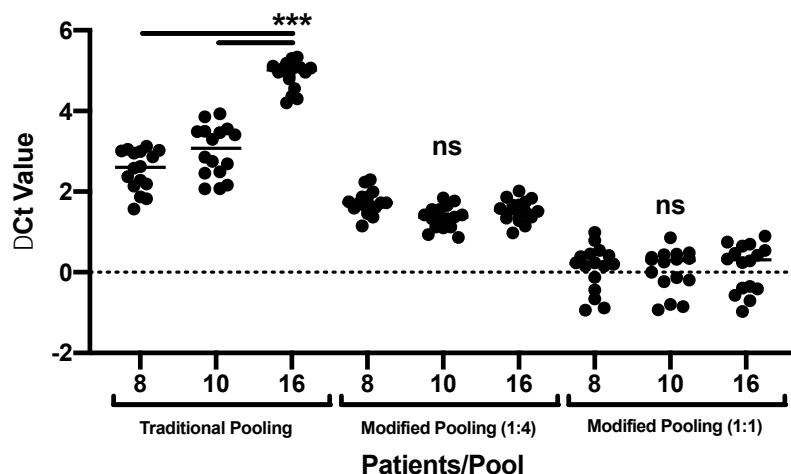


Figure 2. Modified pooling protocol eliminates dilution effect of group testing. (A) RNA extraction and qRT-PCR workflow in individual testing, traditional pooling (group testing), and the modified pooling protocol. Numerical examples are theoretical to display dilution effect and can be scaled to individual diagnostic testing facility protocols. (B) MHV-1 was used to generate individual samples of various viral loads (1×10^9 - 1×10^2 copy number/qRT-PCR reaction). qRT-PCR was performed on each samples to develop ground truth Ct values. Samples were then used in various pool sizes in traditional pooling and in the modified pooling protocol. Increases in sample Ct values from the ground truth values were calculated and plotted as Δ Ct Value.

A)



B)



Pool #	Status	Ct Duplicate 1	Ct Duplicate 2
Pool 1	Pos	35.068	34.234
Pool 2	Pos	35.107	34.526
Pool 3	Neg	NA	NA
Pool 4	Pos	35.021	34.697
Pool 5	Pos	34.031	34.123

Pool #	Status	Ct Duplicate 1	Ct Duplicate 2
Pool 1	Pos	33.92	32.961
Pool 2	Pos	20.68	20.909
Pool 3	Pos	34.065	36.231
Pool 4	Pos	26.562	27.051
Pool 5	Neg	NA	NA
Pool 6	Pos	26.719	26.864
Pool 7	Pos	19.977	20.386
Pool 8	Pos	27.063	27.756
Pool 9	Neg	NA	NA
Pool 10	Pos	20.636	20.945
Pool 11	Pos	27.574	27.266
Pool 12	Pos	20.196	20.925
Pool 13	Neg	NA	NA
Pool 14	Pos	32.336	32.185
Pool 15	Neg	NA	NA
Pool 16	Pos	33.133	32.97

Table 3. Compressed sensing decoded pooled testing significantly decreases the number of tests required to identify infected patients

	Total Experimental Samples Tested	Total Samples Pooled	Experimental Prevalence Rate	# of Patient Infection Status Identified			Tests Required	# Tests Saved Compared to Individual Testing
				Round 1	Round 2	Round 3		
MHV-1	175	7	0-14.3%	175 (100%)	NA	NA	75	100 (57.1%)
	375	75	0-6.7%	365 (97.3%)	375 (100%)	NA	103	272 (72.5%)
	775	217	0-6.45%	751 (96.9%)	761 (98.2%)	14 (100%)	144	631 (81.4%)
Total:	1325	299	0-14.3%	1291 (97.4%)	1311 (98.9%)	1325 (100%)	322	1003 (75.6%)
COVID-19	80	80	10%	74 (92.5%)	80 (100%)	NA	38	42 (52.5%)