Implementation of an efficient sample pooling strategy for high throughput diagnostic testing of Severe Acute Respiratory Syndrome Corona Virus-2

By

Ugochukwu Jacob, Anyaneji (217080488)

A Dissertation by manuscripts submitted in fulfilment of the requirement of the degree of Master of Medical Science (MMedSc)

Department of Virology, School of Laboratory Medicine and Medical Sciences, College of Health Sciences, University of KwaZulu-Natal, Durban, South Africa

Supervisors:

Prof Tulio de Oliveira

Prof Francesco Petruccione

Co-supervisors:

Dr. Jennifer Giandhari, Dr Richard Lessells, Dr Lavanya Singh (KRISP, UKZN).

PREFACE

The work described in this dissertation was carried out at the KwaZulu-Natal Rescarch Innovation Sequencing Platform (KRISP), Nelson R Mandela School of Medicine, University of KwaZulu-Natal, Durban, South Africa, from February 2020 to September 2021, under the supervision of Prof Tulio de Oliveira and Prof Francesco Petruccione.

This work has not been submitted in any form for any degree or diploma to any tertiary institution. Where use has been made of the work of others, it is duly acknowledged in the text.

Ugochukwu Jacob Anyaneji [redacted] Date 19/01/2022

(217080488)

Professor Tulio de Oliveira

Supervisor [redacted] Date 19/01/2022

Professor Francesco Petruccione

Supervisor [redacted] Date 19/01/2022

DECLARATION

I, Mr Ugochukwu Jacob Anyaneji, declare as follows:

1. That the work described in this dissertation has not been submitted to UKZN or other tertiary institutions for purposes of obtaining an academic qualification, whether by myself or any other party.

2. This dissertation does not contain another person's writing, data, pictures, or other information unless specifically acknowledged as being sourced from other persons or researchers. Where other written sources have been quoted then:

• Their words have been re-written, but the general information attributed to them has been

referenced.

• Where their exact words have been used, then it has been appropriately referenced in the reference section.

Signed…  Date…19/01/2022……..

PUBLICATIONS

Lavanya Singh and Ugochukwu J. Anyaneji: Contributed equally to the publication of the article from this project and co-shared first authorship. Author contribution are as stated below in the published paper.

- Lavanya Singh[1,*,#], **Ugochukwu J. Anyaneji[1,#]**, Wilfred Ndifon[2,*], Neil Turok[3], Stacey A. Mattison[1], Richard Lessells[1], Ilya Sinayskiy[4, 5], Emmanuel J. San[1], Houriiyah Tegally[1], Shaun Barnett[6], Trevor Lorimer[6], Francesco Petruccione[4, 5], Tulio de Oliveira[1,*] 2021. **Implementation of an efficient SARS-CoV-2 specimen pooling strategy for high throughput diagnostic testing**. *Scientific Reports,* 11**,** 17793.

  Author contributions: LS, UJA, WN and NT designed the study; LS, UJA and SAM co-ordinated the experiments, performed validation of the pooling experiments, data analysis and interpretation; LS and UJA wrote the manuscript; WN, NT and RL contributed to editing the manuscript; WN and NT contributed to the theory; IS, SB, TL contributed to the development and implementation of the robotics automation workflow; HT and EJS developed and implemented the scripts used for the automated web applications; FP and TdO supervised the project; all authors reviewed and approved the final manuscript.

Other publications and my contributions related to the project is stated below

- VOGELS, C. B. F., BREBAN, M. I., OTT, I. M., ALPERT, T., PETRONE, M. E., WATKINS, A. E., KALINICH, C. C., EARNEST, R., ROTHMAN, J. E., GOES DE JESUS, J., MORALES CLARO, I., MAGALHÃES FERREIRA, G., CRISPIM, M. A. E., BRAZIL, U. K. C. G. N., SINGH, L., TEGALLY, H., **ANYANEJI, U. J**., NETWORK FOR GENOMIC SURVEILLANCE IN SOUTH, A., HODCROFT, E. B., MASON, C. E., KHULLAR, G., METTI, J., DUDLEY, J. T., MACKAY, M. J., NASH, M., WANG, J., LIU, C., HUI, P., MURPHY, S., NEAL, C., LASZLO, E., LANDRY, M. L., MUYOMBWE, A., DOWNING, R., RAZEQ, J., DE OLIVEIRA, T., FARIA, N. R., SABINO, E. C., NEHER, R. A., FAUVER, J. R. & GRUBAUGH, N. D. 2021. **Multiplex qPCR discriminates variants of concern to enhance global surveillance of SARS-CoV-2**. *PLOS Biology,* 19**,** e3001236.
- Wilkinson, E., M. Giovanetti, H. Tegally, J. E. San, R. Lessels, D. Cuadros, D. P. Martin, A.-R. N. Zekri, A. K. Sangare, A.-S. Ouedraogo, A. K. Sesay, A. Hammami, A. A. Amuri, A. Sayed, A. Rebai, A. Elargoubi, A. J. Trotter, A. K. Keita, A. A. Sall, A. Kone, A. Souissi, A. V. Gutierrez, A. J. Page, A. Iranzadeh, A. Lambisia, A. Sylverken, A. Ibrahimi, B. Dhaala, B. Kouriba, B. Kleinhans, C. Brook, C. Williamson, C. B. Pratt, C. G. Akoua-Koffi, C. N. Agoti, C. M. Morang'a, D. J. Nokes, D. J. Bridges, D. L. Bugembe, D. Baker, D. Doolabh, D. Ssemwanga, D. Tshabuila, D. Bassirou, D. S. Y. Amuzu, D. Goedhals, D. Maruapula, E. Foster-Nyarko, E. K. Lusamaki, E. Simulundu, E. Moraa, E. N. Ngabana, E. E. Fahime, E. Jacob, E. Lokilo, E. Mukantwari, E. Belarbi, E. Simon-Loriere, E. A. Anoh, F. Leendertz, F. Ajili, F. Wasfi, F. T. Takawira, F. Derrar, F. Bouzid, F. M. Muyembe, F. Tanser, G. K. Mbunsu, G. Thilliez, G. Kay, G. Githinji, G. v. Zyl, G. A. Awandare, G. Schubert, G. P. Maphalala, H. C. Ranaivoson, H. Lemriss, H. Abe, H. H. Karray, H. Nansumba, H. A. Elgahzaly, H. Gumbo, I. Smeti, I. B. Ayed, I. B.-B. Boubaker, I. Gaaloul, I. Gazy, I. Ssewanyana, J. B. Lekana-Douk, J.-C. C. Makangara, J.-J. M. Tamfum, J.-M. Heraud, J. G. Shaffer, J. Giandhari, J. Li, J. Yasuda, J. Q. Mends, J. Kiconco, J. Morobe, J. N. Nkengasong, J. O. Gyapong, J. T. Kayiwa, J. A. Edwards, J. Gyamfi, J. Farah, J. M. Ngoi, J. Namulondo, J. C. Andeko, J. J. Lutwama, J. O'Grady, K. A. Tumedi, K. M. Said, K. Hae-Young, K. O. Duedu, L. Belyamani, L. Singh, L.

d. O. Martins, M. Mine, M. Ramuth, M. Mastouri, M. Aouni, M. e. Hefnawi, M. I. Matsheka, M. Kebabonye, M. Turki, M. M. Nyaga, M. Mareka, M. Damaris, M. Cotten, M. W. Mburu, M. Mpina, M. Owusu, M. R. Wiley, M. A. Ali, M. Abouelhoda, M. G. Seadawy, M. K. Khalifa, M. Sekhele, M. Ouadghiri, M. Mwenda, M. Allam, M. V. T. Phan, N. Abid, N. Touil, N. Kharrat, N. Ismael, N. Mabunda, N.-y. Hsiao, N. B. Silochi, N. Saasa, N. Mulder, P. Combe, P. Semanda, P. E. Oluniyi, P. Arnaldo, P. K. Quashie, P. A. Bester, P. Dussart, P. K. Mbala, P. Kaleebu, R. Ayivor-Djanie, R. Njouom, R. O. Phillips, R. Gorman, R. A. Kingsley, R. A. A. Carr, S. E. Kabbaj, S. Gargouri, S. Masmoudi, S. Kassim, S. Trabelsi, S. Kammoun, S. Lemriss, S. H. Agwa, S. Calvignac-Spencer, S. Doumbia, S. M. Mandanda, S. Aryeetey, S. S. Ahmed, S. Moyo, S. Gaseitsiwe, S. Lekana-Douki, S. Prosolek, S. Ouangraoua, S. A. Mundeke, S. Rudder, S. Panji, S. Pillay, S. Engelbrecht, S. Nabadda, S. Behillil, S. L. Budiaki, S. van der Werf, T. Mashe, T. Aanniz, T. Mohale, T. Le-Viet, T. Schindler, **U. J. Anyaneji**, U. Ramphal, V. Fonseca, V. Enouf, V. Gorova, W. H. Roshdy, W. K. Ampofo, W. Preiser, W. T. Choga, Y. Bediako, Y. K. Tebeje, Y. Naidoo, Z. R. de Laurent, S. K. Tessema and T. de Oliveira (2021). **"A year of genomic surveillance reveals how the SARS-CoV-2 pandemic unfolded in Africa."** 2021.2005.2012.21257080.

DEDICATION

This research project is dedicated to:

- God almighty, my source of inspiration and strength
- My dear mother; Mrs Martha Anyaneji
- My brother turned father; Odinakachukwu Moses Nwokediegwu
- Friends, family and every well-wisher

ACKNOWLEDGEMENTS

I am grateful to God for His guidance, divine inspiration, and wisdom that brought me to this point in life.

My sincere and heartfelt appreciation goes to my supervisors, Professors Tulio de Oliveira and Francesco Petruccione, who double as both academic and social fathers.

Special thanks to my co-supervisors, Dr Jennifer Giandhari, Dr Richard Lessells, Dr Lavanya Singh, for constructive criticism and guidance throughout this work.

I wish to specially express my profound gratitude to my family for their love and support throughout my study.

I heartily appreciate all members of the innovative platform-KRISPers!

TABLE OF CONTENTS

LIST OF FIGURES

ABBREVIATIONS

CDC	Centre for Disease Control
COVID-19	Coronavirus disease 2019
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
MERS	Middle East Respiratory Syndrome
PCR	Polymerase Chain Reaction
RT-PCR	Real-Time Polymerase Chain Reaction
LFA	Lateral Flow Assay
Ct	Cycle threshold
CST	Community Screening and Testing
SOFA	Sequential Organ Failure Assessment
UTRs	Untranslated regions
ORF1a/1b	Open reading flame 1a/1b
NTD	N-Terminal domain
RBD	Receptor-binding domain
CPD	Cytoplasmic domain
APCs	Apoptotic cells
DCs	Dendritic cells
ACE2	Angiotensin converting enzyme 2
SGTF	Spike gene target failure
API	Application programming interface

ABSTRACT

Early identification and isolation of SARS-CoV-2 infected individuals, predominantly asymptomatic, is crucial for the containment of the disease. Such surveillance strategies can only be achieved through mass and regular testing. However, the high cost of PCR tests, currently the standard gold test, is a significant challenge in low-income countries. Group testing, otherwise known as specimen pooling, offers a viable alternative in cost reduction and skill requirements.

Most research on specimen pooling focuses on reducing the number of tests required to pinpoint a positive specimen in a group of pools. This study explored a unique strategy that utilizes the least possible number of tests to identify positive samples in the various pools without requiring a second round of testing. This conserves resources and reduces turnaround time, which indirectly impacts the high transmissibility of the virus.

We implemented a proof of concept pooling strategy using 3-d and 4-d hypercubes to test a professional sports team in South Africa and show that infected samples can be reliably detected in groups of 27 and 81, with minimal loss of assay sensitivity for samples with individual Ct values up to 32. We also report on our automation of sample pooling, using a liquid-handling robot and an automated web interface to identify positive samples. We conclude that hypercube pooling allows for the reliable RT-PCR detection of SARS-CoV-2 infection at a significantly lower cost than lateral flow antigen tests.

CHAPTER ONE

Introduction and literature review

1.1.1 Introduction

The end of 2019 marked the beginning of a new normal for social activities following the identification of a cluster of pneumonia cases observed in Chinese hospital admissions. Concerns regarding the strange outbreak spurred scientists to search for the possible cause. This led to the identification of the new coronavirus that has continued to plague human activities globally. A throat swab sample study conducted by the Chinese Centre for Disease Control (CDC) revealed that a novel coronavirus, now known as (server acute respiratory syndrome coronavirus) SARS-CoV-2, was responsible for the outbreak (Zhu et al., 2020). As of October 2021, more than 200 million people have been infected due to the rapid spread of the virus (WHO, 2021). The emergence of the virus was believed to have a zoonotic origin from a wet market in Wuhan city of Hubei province in China. Human-to-human transmission of the virus was later identified to contribute significantly to the high infection rate (Li et al., 2020a). In China, for instance, a rapid increase in cases was seen in less than a month and a further spread to other provinces in a short timeframe and consequently extending to international and intercontinental spread (Wu et al., 2020).

The first case of SARS- CoV- 2 infection in South Africa (SA) was reported on 5 March 2020 (NICD, 2020b) and spontaneously increased with a substantial burden on the national public health system. The outbreak's impact spread across all human activities and has posed an unusual burden on the global public health system as has never been seen in recent centuries. To this end, the international community has been exploring various options to effectively mitigate the impacts of the pandemic and possibly halt the spontaneous infection growth.

Several policy proposals targeted at suppressing the spread of the virus have supported the implementation of population-wide testing to monitor infection and detect outbreaks quickly (Majid et al., 2020). However, such testing is expensive and requires skilled personnel, making it unfeasible in most, if not all, low-income countries. In the United States, an acute shortage of testing reagents was previously reported, leading to the stopping of testing in certain places, resulting in an exponential increase in infection rate (Abdalhamid et al., 2020). The limited availability of the resources needed for wider population testing has remained a challenge. This highlights the need for a more efficient and cost-effective testing strategy to achieve an optimal outcome.

Several diagnostic approaches for detecting SARS-CoV-2 have been developed since the beginning of the outbreak, each with different detection times (Carter et al., 2020). Most commonly available tests are dominant on molecular assays, such as reverse transcription-polymerase chain reaction (RT-PCR). However, other techniques such as isothermal nucleic acid amplification assays, serological and

immunological assays for anti-SARS-CoV-2 antibody, hybridization microarray assays, and chest CT scans are also being explored. The gold standard PCR test has many limitations, such as multiple temperature changes in each cycle, high cost of reagents and laborious instrumental work, and days to get the results. At the same time, the serological assays are functionally limited by low sensitivity and specificity (Majumder and Minko, 2021a). Thus, increasing the need for developing rapid, less laborious and cost-effective diagnostic methods, which can directly detect the presence of SARS-CoV-2 virus in clinical samples in the shortest possible time. Specimen pooling offers a viable alternative to drive the benefits of mass population testing and is highly effective in different population-based screening scenarios. A study aimed at boosting testing efficiency showed that a pooling strategy could preserve testing resources and time with an increase in testing capacity of 69% for an incidence rate of SARS-CoV-2 infection of 10% or less (Cabrera Alvargonzalez et al., 2020). The pooling approach groups and tests samples drawn from multiple individuals as a single specimen. If the test result is negative, everyone in the group is considered negative; if it returns positive, at least one individual is infected with SARS-CoV-2. Further individual testing of constituent samples of the positive group is carried out to identify the positive ones (Dorfman, 1943).

Community screening and testing (CST) was among the unique and early responses to the pandemic in SA, which almost overwhelmed the country and its medical infrastructure. This strategy saw the mobilization of 28 000 health care workers who screened over 7 million individuals (more than 1 in 10 South Africans). As the cumulative number of cases increased exponentially due to the virus's rapid spread through the country, the national laboratories became overwhelmed with the demand for testing (Bhekisisa, 2020). This resulted in severe backlogs, with delayed turnaround times, which further contributed to the spread of the virus. Front-line staff shortages compounded this problem due to suspected/confirmed SARS-CoV-2 infection and the shortage of testing kits required to make accurate diagnoses. The testing demand decreased in SA, but as the country eventually moved to lockdown level 1, a second wave of infection appeared inevitable, as this was the trend witnessed in other countries of the world. Therefore, new strategies needed to be developed to facilitate quick and accurate testing and release of results to contain the virus and maintain the integrity of our health systems (Mallapaty, 2020).

In light of limited resources, many nations are turning to pooled sample testing (Ghosh et al., 2020), which offers a viable alternative and efficient use of scarce testing kits. Pooled testing, also known as group testing, is a procedure that involves combining individual samples drawn from multiple patients into pools and testing it as a single sample (Bilder and Tebbs, 2012). With the first and the most widely used form of pooling known as "Dorfman testing", all samples of a negative pool are considered negative. In contrast, the positive pool is retested individually to identify the positive samples (Dorfman, 1943b), thus reducing the associated cost of resources and time needed to test all samples individually. Various forms of pooling have been developed from the foremost Dorfman method ranging from adaptive and non-adaptive pooling. The adaptive process generally involves two-round testing, which

allows the detection of the positive pools and further retesting to identify the positive samples. In the non-adaptive methods, advanced improvement has been made to address the time consuming two round testing. Many of the methods in use deploy a single round of testing and a detection algorithm to identify the positive samples, thereby improving turnaround time, conservation of testing kits and cost reduction (Furstenau et al., 2020).

We explored a simplified multi-dimensional pooling method developed by Wilfred Ndifon and Neil Turok, which is very efficient for low prevalence (Mutesa et al., 2020a). By optimizing the pool size, the method identifies infected individuals in fewer tests than the standard procedure. Best results are obtained for relatively low prevalence. Field tests have already been reported and have attracted the attention of the international media. We estimate that using this approach for community screening in South Africa could save more than half a million tests, significantly reducing the cost of testing. This research focused on reducing the number of tests needed to pinpoint a positive case in a group. In the context of coronavirus and the associated high transmissibility, a shorter turnaround time is crucial to reduce infection spread. The proposed technique minimises cost and number of tests and has been improved to increase throughput in the shortest time possible.

1.1.2 Hypercube pooling strategy

The concept of the hypercube pooling follows a geometric algorithm in which each sample in the group to be tested is represented by a set of N points on a cubic lattice in D dimensions, organized in the form of a hypercube with L points on a side. Instead of testing individual sample, each gets divided into equal D subsamples according to N points on each L planar sides. They are recombined by slicing the hypercube into L planar slices, perpendicular to one of the principal directions on the lattice. each set of slice consist of subsamples in each of the D principal direction and are pooled and tested in parallel in a round of slicing tests. If there is one positive sample, then one slice out of the L slices, in each of the D directions, will yield a positive result. That slice indicates the coordinate of the point corresponding to the infected individual, along the associated principal direction.

Groups of 27 or 81 samples are sub-pooled into 9 and 12 slices respectively, each consisting of 9 or 27 subsamples respectively, resulting in each sample being represented in 3 or 4 different slices. Thus, the test results for the slices can be used to infer which sample is positive, based on its consistent detection within each of the hypercube slices, without an individual test ever being required. The multidimensional pooling approach takes advantage low positivity rate and is very efficient when the group pools are negative. The criteria for the choice of group size depends on positivity rate. For a pool size of 27 and 81, prevalence is estimated at a range of 2.3-0.78% and 0.78-0.26% respectively (Mutesa, 2020).

1.1.3 Epidemiology of SARS-CoV-2

Following the initial report of SARS-CoV-2 in China in December 2019, the disease spread rapidly, and the number of infections increased exponentially. On January 13 2020, the World Health Organisation (WHO) confirmed the first case of infection outside of China in Thailand (WHO, 2020), and within months, the disease had spread to all continents except Antarctica. By 15 May 2020, there were 4,621,410 cases and 308,542 related deaths (Giandhari et al., 2020) worldwide, involving almost every country in the world. The first death reported in China was of a 61-year-old man from Wuhan, and the first death outside of China was recorded in the Philipines (skynews, 2021). At the time of writing, statistics online show more than 170 million people have been infected globally, with more than three million deaths (worldometer, 2021). Much of Africa was spared the impact of the virus in the early stages. However, the first case of infection on the continent was recorded in Nigeria in February 2020, and in early March, the first case in South Africa was reported (NICD, 2020a).

South Africa has recorded the highest number of cases in Africa. As of September 2021, more than 2 million infected cases and 84 152 deaths (NICD, 2021) had been reported. The pandemic has been extremely difficult for most African countries and other developing countries to contend with, given the high cost of the available tests for diagnosing the infection (Reddy et al., 2020). Following the confirmation of the first case in SA, decisive action was implemented by the government, including many travel restrictions and largescale testing. However, these efforts were hampered by restricted population testing due to scarcity of resources and the high cost of testing.

1.1.4 Transmission dynamics

Initially, zoonotic transmission was considered the sole cause of infection since initial cases were epidemiologically linked to the wet market in China. However, the number of people who developed the disease without a history of exposure to the market or other people with any respiratory disease and occurrences in hospital settings increased, suggesting a person-to-person spread of the virus (Robyn et al., 2020). The precise transmission mode is unknown, but as common with other respiratory viruses, directly or indirectly, droplet-borne infection through formites is modelled as the dominant transmission mode (Faridi et al., 2020). Considering the basic reproduction rate (Ro), the possible number of secondary infections likely to emanate from exposure to a primary case in a susceptible population was estimated to be within the range of 2 to 3 (Covid-19 National Emergency Response Center et al., 2020). This is indicative of the risk of an epidemic spread.

Another factor determining infectivity is the secondary attack rate, the probability of infection occurring in a specific group of susceptible people exposed to a primary case, which was high among family members and close contacts. This points to the impact of social behaviour on transmissibility. A study on the effect of undocumented infection on viral dissemination postulated that undocumented infection

accounts for 79% of documented cases owning to high transmissibility during disease onset or asymptomatic phase (Li et al., 2020b). These could be significantly averted by widespread early and regular testing of the vulnerable populations. The unique features of SARS-CoV- 2 support the short incubation period and time it takes for the epidemic to double in size, which has been described in the literature. A study suggested 3-7 days as the doubling time while the incubation period was estimated to be 4-6 days (Park et al., 2020); these factors, together with limited testing, contribute to rapid infection spread.

Following the first case report in SA, a genomic epidemiological study on early transmission of SARS-CoV-2 revealed a heterogeneous introduction and transmission patterns, which resulted in early high death cases and a fast infection growth (Giandhari et al., 2020). The multiple introductions may have been augmented by restrictions of testing to recent travellers to at-risk countries. Substantial exportation of presymptomatic, asymptomatic cases and the absence of large-scale public health interventions on both population and personal levels was identified to exacerbate the exponential spread of SARS-CoV-2.

COVID 19 pandemic has overwhelmingly surpassed that of SARS-CoV-2 and MERS in terms of the number of laboratory-confirmed cases and spatial range of the epidemic areas (Deng and Peng, 2020). Thus, informing the need for a readied mitigations intervention and future occurrences.

1.1.5 Structure of SARS-CoV-2

Coronaviruses are a diverse group of highly transmissible and pathogenic viruses capable of infecting a wide variety of hosts. The name is derived from the Latin root word ‘corona’, meaning crown, which refers to their unique appearance under the electron microscope. They appear as round particles with a rim of projections resembling the solar corona. They are enveloped, positive-sense, single-stranded RNA viruses of ~30kb. Coronavirus belongs to the family Coronaviridae, which produces mild to severe respiratory diseases in humans (Cui et al., 2019).

In recent times, three notable species have emerged and have been responsible for disease outbreaks, beginning with the severe acute respiratory syndrome coronavirus (SARS–CoV) in 2002, followed by the Middle East respiratory syndrome coronavirus (MERS–CoV) in 2012 (Dhar Chowdhury and Oommen, 2020), and now the severe acute respiratory syndrome coronavirus 2 (SARS–CoV2) named due to homology (~80%) to SARS-CoV (Ksiazek et al., 2003). As shown in figure 1.1, SARS-CoV-2 consists of four structural proteins designated as; Spike (S), membrane (M), envelop (E), and nucleocapsid (N).

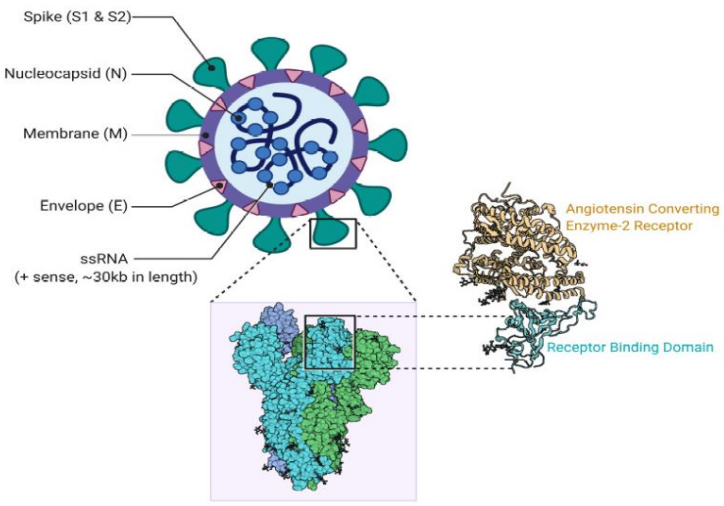


***Figure 0.1.1**: SARS- CoV 2 Structure by Rohan Bir Singh (https://www.ncbi.nlm.nih.gov/books/NBK554776/figure/article-52171.image.f3/)*

The Spike proteins of SARS-CoV-2 have been described as different from other viruses, given the presence of multiple cleavage sites for various proteases. This unique feature is likely responsible for their high and broad infectivity to many host organisms. Another characteristic of SARS-CoV-2 that distinguishes it from other coronaviruses is the existence of the furin cleavage site ("RPPA" sequence) at the S1/S2 site (figure 1), which may confer high pathogenicity to the virus (Coutard et al., 2020).

1.1.6 Genomic organisation

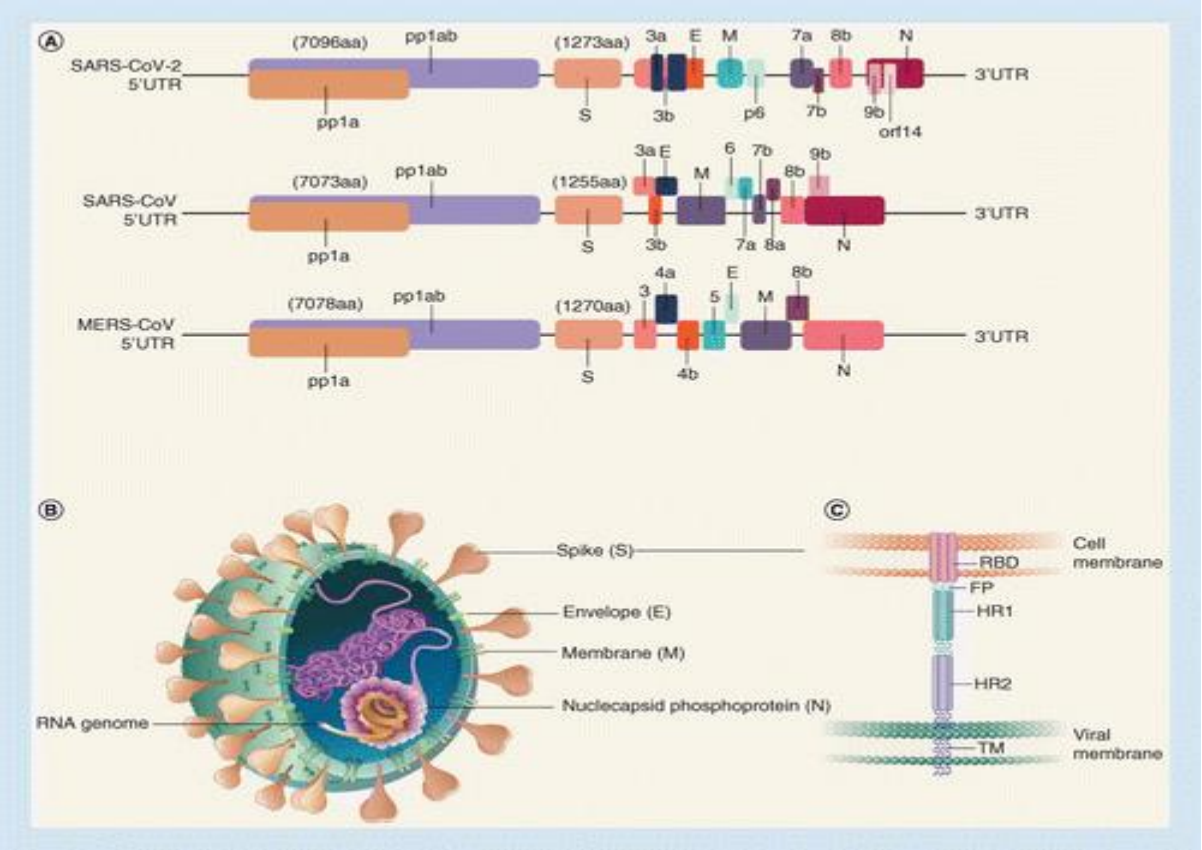


Figure 0.2: *Diagramatic representation of SARS-CoV-2 genome organisation relative to other coronaviruses (Taken from https://www.futuremedicine.com/cms/10.2217/fmb-2020-0063/asset/images/medium/figure1.gif)*

Bioinformatic data reveals that SARS-CoV-2 shares close to 80% and 50% genome identity with other beta coronaviruses such as SARS-CoV and MERS, respectively (Lu et al., 2020). The single-stranded RNA genome consists of 29891 nucleotides in size and encodes 9860 amino acids. Consistent with other βCoVs, the SARS-CoV-2 genome contains two flanking untranslated regions (UTRs) and a single long open reading frame encoding a polyprotein. The genome order of arrangement follows a 5′ to 3′ sequence including replicase (ORF1a/ORF1b), spike (S), envelope (E), membrane (M), and nucleocapsid (N). It lacks the hemagglutinin-esterase gene, which is characteristically found in lineage A β-CoVs. Additionally, seven putative ORFs encode accessory proteins interspersed between structural genes (Chan et al., 2020).

The majority of SARS-CoV-2 proteins shares a similar length and amino acid identity with the corresponding proteins in SARS-CoV except for the S protein, which differs significantly (Lu et al., 2020). Spike is composed of a transmembrane trimetric glycoprotein protruding from the viral surface, which determines the diversity of coronaviruses and host tropism (Bosch et al., 2003a). Its glycoprotein

consists of two functional subunits. The S1 subunit contains a signal peptide, followed by an N-terminal domain (NTD) and receptor-binding domain (RBD). In contrast, the S2 subunit contains conserved fusion peptide (FP), heptad repeat (HR) 1 and 2, transmembrane domain (Reddy et al.), and cytoplasmic domain (CP) (Chan et al., 2020).

A comparative analysis of SARS-CoV-2 S protein reveals its distinctiveness from the S proteins of most members in the subgenus Sarbecovirus, sharing amino acid sequence similarities of 76.7–77.0% with SARS-CoVs from civets and humans, 75–97.7% with bat coronaviruses in the same subgenus and 90.7–92.6% with pangolin coronaviruses (Zhou et al., 2020). In the receptor-binding domain, there is a 73% similarity in amino acids. Another important genomic feature observed in the S protein is the insertion of four amino acid residues in the junction of S1 and S2 subunits (Coutard et al., 2020). Such integration provides a polybasic cleavage site for furin and other proteases not found in other coronaviruses. This feature may be associated with the high pathogenicity of SARS-CoV-2, which allows it to bind to different host receptors.

A structural study suggested that the furin cleavage site can reduce the stability of S protein and provide the conformational adaption needed to bind the RBD to its receptor (Wrobel et al., 2020). However, whether or not the association of the furin cleavage site to the high transmissibility of SARS-CoV-2 as a gain function is yet to be demonstrated, thus necessitating further investigation. Another notable distinction is found in the ORF8 of SARS-CoV-2, which encodes a novel protein with 40% amino acid identity with the ORF8 of SARS-CoV. Unlike SARS-CoV, the ORF8 of SARS-CoV-2 lacks a motif that triggers intercellular stress pathways (Chan et al., 2020).

1.1.7 Evolution of SARS-CoV-2 variants

The emergence of SARS-CoV-2 variants poses a threat to the containment of the pandemic. Since the onset of the pandemic, several variants of concern (Alpa, Beta, Delta and Gamma) have been identified in different countries, including the UK, South Africa, Brazil and India, respectively, as the origin of detection of the VOC. These variants are associated with increased transmissibility, increased risk of reinfection and reduced vaccine efficacy (Tao et al., 2021). Several factors ranging from error-prone polymerase fidelity of viral replicating enzymes, host antiviral enzymes and environmental mutagens drives the rapid evolution of SARS-CoV-2 (Duffy et al., 2008), which ushers in a new and more complex phase of the pandemic. These variants are currently only detectable by genomic surveillance tools, which are expensive and not readily available (de Oliveira et al., 2021b).

Similar to host adaptation for survival, rapid genetic evolution occurs in viruses to enable them to adapt to host resistances. With SARS-CoV-2, several evolutionary changes, including mutations in the major functional proteins, have been reported since the outbreak. To explore the genetic variation of different SARS-CoV-2 strains, the Chinese national centre for bioinformatics aligned 77,801 genome sequences

of SARS-CoV-2 detected globally and identified 15,018 mutations, including 14,824 single-nucleotide polymorphisms (Zhao et al., 2020a). For instance, four amino acid alterations were identified in the S protein proximal to the receptor-binding interface. However, whether or not they affect the binding of the virus to host receptors is yet unknown. Another alteration, D614G has been identified in the S1 subunit and appears to be more frequently distributed in S variant sites. Following the identification, another variant with G614 in S protein was reported and dominantly circulating globally (Korber et al., 2020).

High viral load was observed in patients infected with the G614 variants, but clinical data suggested no disease severity in the D614G variant. In an in vitro experiment model, pseudotyped viruses carrying S proteins with the G614 variant were shown to generate more infectious titres than the original D614G, suggesting the alteration may have increased the infectivity of SARS-CoV-2. A phylogenetic analysis using the amino acids sequences of ORF1 a/b and the four structural genes of SARS-CoV-2 clustered with lineage B of βCoVs (Paraskevis et al., 2020). These SARS-related coronaviruses have been consistently found in horseshoe bats found in China and likely suggest a possible origin from bats. However, it has not been ascertained whether an intermediate or amplification animal host infected with SARS-CoV-2 exists in transmission to humans.

1.1.8 Pathogenesis of SARS-COV-2

Clinical observations indicate that although all population ages are susceptible to SARS-CoV-2, the older population is most seriously affected. Clinical manifestations vary, but the older age group (<60 years above) and people with comorbidities show high disease severity. In contrast, the young population and children have mild or no symptoms (Wu and McGoogan, 2020). Fever, fatigue, and dry cough have been observed as the most common symptoms among patients. Less observed symptoms included sputum production, headache, haemoptysis, diarrhoea, anorexia, sore throat, chest pain, chills, nausea, and vomiting in studies of patients in China (Chen et al., 2020).

Data from a Chinese hospital used to study risk factors associated with in-hospital death revealed older age, higher sequential organ failure assessment (SOFA) score, and d-dimer >1 μg/mL to be consistent in multi-variable analysis. For the univariable analysis, coronary artery disease, diabetes, and hypertension were also considered risk factors (Zhou et al., 2020a). Most people present with disease symptoms between 1-15 days (most commonly around day five; thereafter, dyspnoea and pneumonia develop within a median time of eight days from the time of illness onset) (Guan et al., 2020).

Regarding disease severity of Covid-19, several research interests has delved into the relationship between severity of disease and gender. A case study of the Chinese series showed an equal number of cases between males and females; however, the data suggested that men suffered more than women, similar to observations from other countries (Jin et al., 2020). A study found that men older than 68

years had a higher risk of respiratory failure, acute cardiac injury, and heart failure that led to death, regardless of a history of cardiovascular disease. Also, adverse outcomes of COVID-19 were associated with comorbidities such as hypertension, cardiovascular disease, and lung disease. These conditions are more prevalent in men; the study suggests they are correlated with smoking and drinking alcohol. Immunological differences associated with sex differences are also suggested to explain the severity differences (Hall et al., 2020). Further, a study that explored factors influencing the adoption of non-pharmaceutical measures to prevent infection spread found that women are 50% more likely to implement the measures than men, which may partly impact disease severity (Moran and Del Valle, 2016).

Since the beginning of the outbreak, clinical observations recorded in children have continued to reflect a small prevalence percentage. In reference to data from the Chinese centre for disease control and prevention from February 2020, children younger than ten years and within the 11-19 age group occupied 1% of recorded cases (Wu and McGoogan, 2020). Given that the paediatric population represents 20% of the total population, this indicates a low prevalence among children. However, considering the global problem of limited testing capacities, this may not be a true reflection of the actual prevalence in the paediatric population.

A confounding factor was that schools in China were closed for most of the epidemic, which coincided with the Chinese New Year holidays and may have influenced the results. This may have contributed to less exposure of children and consequently impacted the case number. In Italy, data of Covid-19 patients within the age range of 8 to 18 years represented 1.2% of the population, and the fatality rate was similar to that in China (Giorgi Rossi et al., 2020). Similar observations were also recorded in the US, Korea and other countries in the study, suggesting fewer symptomatic manifestations in children than adults.

A patient history record for children recorded that only 73% developed fever, cough, or shortness of breath. This is comparable to 93% of adults reported in the same time frame, between the ages of 18 and 64 years. The estimated hospitalization rate for children aged one to 17 was 14%, at most (Yuki et al., 2020). Despite the overall favourable outcome for the paediatric population, many deaths were also reported in the US and other countries.

1.1.9 Life cycle and receptor use

A five-step process involved in the life cycle of SARS-CoV-2 with its host has been described and includes; attachment, penetration, biosynthesis, maturation and release (Figure 1.3). Upon adherence to the host receptor interface, the virus gains entrance (penetration) into the host cell through endocytosis or fusion of viral and host membranes. Thereafter, viral contents are released and the RNA integrated into the nucleus for replication. The intermediate viral mRNA synthesises new viral particles

(biosynthesis), which undergoes maturation and release (Bosch et al., 2003b). S proteins are actively involved in viral binding to host receptors. The S1 subunit of S protein is responsible for binding to receptors, while the S2 subunits mediate the fusion of viral and cellular membranes (Walls et al., 2020). Angiotensin-converting enzyme 2 (ACE 2) has been identified as a functional receptor for SARS-CoV-2 and has been shown to bind effectively to S proteins. ACE 2 is highly expressed in the lungs, heart, ileum, kidney and bladder (Li et al., 2003). The abundance of ACE 2 in lung epithelial cells likely explains the high pathogenicity observed in the lung. Whether or not SARS-CoV-2 binds to another receptor requires further investigation.

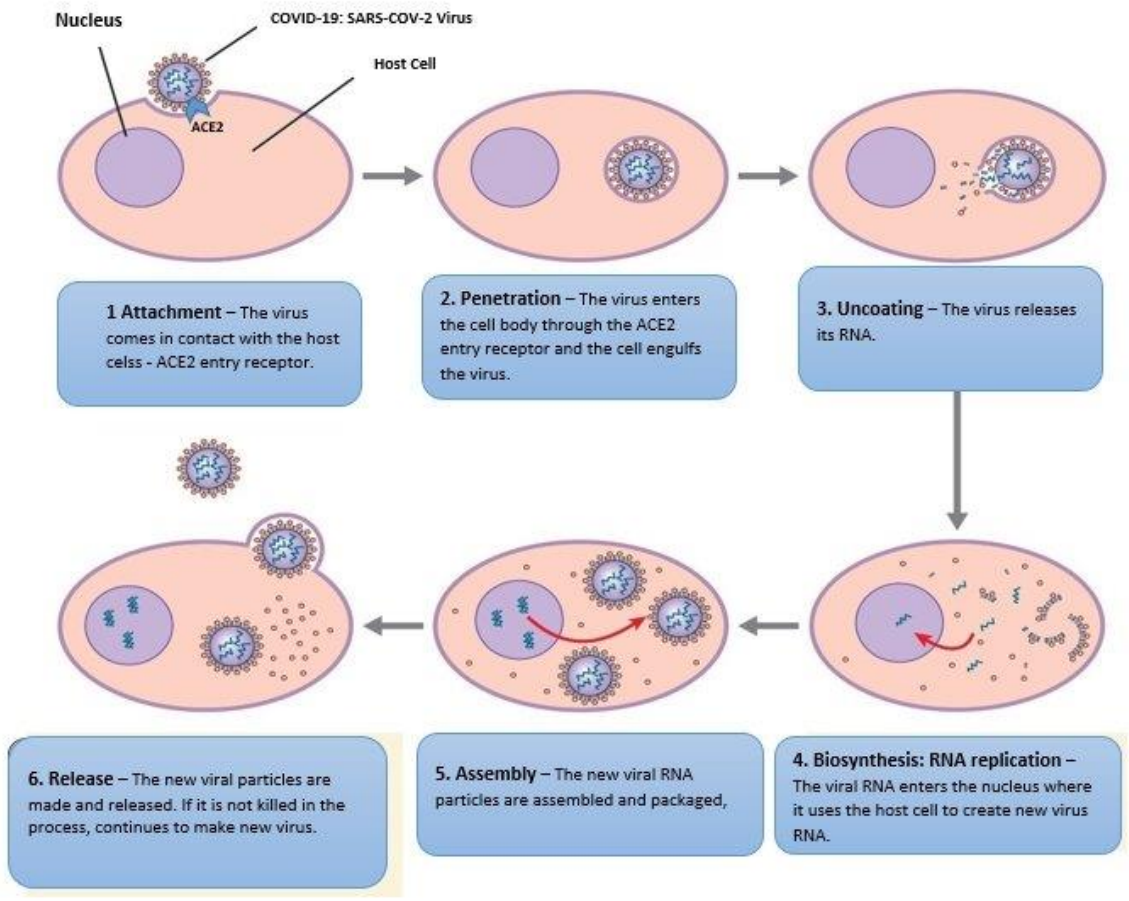


Figure 0.3 *Image showing the life cycle and host receptor for SARS-CoV-2 (Dr. Amit Ray, https://amitray.com/life-cycle-of-coronavirus-covid-19-the-sars-cov-2-virus/)*

Similar to other coronaviruses, proteolytic processing of S protein is essential for the endocytic process of SARS-CoV-2. A two-step sequential protease cleavage model was used to describe the protease cleavage activities of S proteins with host receptors (Belouzard et al., 2009). The process consists of cleavage at the S1/S2 site for priming and another cleavage for activation at the S′2 site, a position adjacent to a fusion peptide within the S_2 subunit. After the cleavage at the S1/S2 cleavage site, a non-covalent bond exists between S_1 and S_2 subunits. The distal S_1 subunit contributes to stabilising the membrane-anchored S_2 subunit at the prefusion state, while the cleavage at the S′2 site presumably

activates the spike for membrane fusion via irreversible, conformational changes. This form of cleavage activity is unusual among viruses and enables the cleaving of different proteases. The existence of the furin cleavage site at the S1/S2 site is another unique characteristic of SARS-CoV-2, which allows for cleavage during biosynthesis instead of direct incorporation of S protein observed during assembling of viral particles in SARS-CoV (Walls et al., 2020). Again, this feature likely increases the pathogenicity of SARS-CoV-2. It has also been reported that SARS-CoV-2 can recognize other non-human ACE 2 receptors with varied use efficiency (Zhao et al., 2020b), thus indicating different host susceptibilities to SARS-CoV-2.

The pathogenic manifestation of SARS-CoV-2 ranges from mild symptoms to severe respiratory failure. On binding to respiratory tract epithelial cells, they start to replicate and migrate down to the airways and enter the alveolar cells of the lungs. Rapid replication of the virus in the lungs potentially triggers strong immune responses with an accompanying cascade of cytokine storms (CS). CS causes acute respiratory disease syndrome and has been identified as a potent cause of death. Patients aged 60 years and above with pre-existing diseases have a high risk of developing respiratory disease and are unlikely to survive (Mehta et al., 2020). Histopathological changes are primarily observed in the lungs and result in bilateral diffused alveolar damage, hyaline membrane formation, desquamation of pneumocytes and fibrin deposits in lungs of severe COVID-19 patients. An immunohistochemical study identified SARS-CoV-2 antigen in the upper airway, bronchiolar epithelium and submucosal gland epithelium, type I and type II pneumocytes, alveolar macrophages and hyaline membranes in the lungs (Martines et al., 2020).

The airway consists of three cell types that mediate its immunity, including epithelial cells, dendritic cells (DCs), which are located below the epithelium, while the macrophages are positioned on the apical side. DCs and macrophages provide innate immunity against viruses until adaptive immunity is developed. T-cell mediated immune response against SARS-CoV-2 has been reported in which presentation of viral particles are initiated by DCs and macrophages (Yoshikawa et al., 2009). The question of how SARS-CoV-2 enters apoptotic cells remains unclear, although DCs and macrophages can phagocytize virus-infected apoptotic cells.

Following phagocytosis, these antigen-presenting cells move to the draining lymph nodes with viral particles, consequently activating T cells. $CD4^+$ and $CD8^+$ T cells play an essential role. $CD4^+$ T cells activate B cells to promote virus-specific antibody production, while $CD8^+$ T cells mediate the killing of infected cells. In addition to respiratory symptoms, thrombosis and pulmonary embolism have been identified in severe disease cases, thus supporting the elevated d-dimer levels equally observed in severe disease (Wang et al., 2018). Endothelium functions in promoting vasodilation, fibrinolysis, and anti-aggregation. Given that endothelium plays a significant role in thrombotic regulation, hypercoagulable profiles seen in severe diseases likely indicate significant endothelial injury. Endothelial cells also

express ACE2. Endothelial cells form one-third of lung cells, and microvascular permeability resulting from the endothelial injury may facilitate viral invasion.

1.1.10 The current state of SARS-CoV-2 control and infection management

To contain further spreading of COVID-19, several control measures, including non-pharmaceutical activities, quarantine, social distancing, and travel restrictions, each with varying degrees of impact, were instituted by governments of various countries. A study that used the deterministic SEIR model estimated that travel restrictions led to 91% case reductions in seven days compared to a baseline scenario with no restriction (Tian et al., 2020). A similar study found that travel restrictions delayed the epidemic in Wuhan for 2.91 days and 3-5 days (Chinazzi et al., 2020). Considering the simulating effect of pre-symptomatic transmission of COVID-19, a study reported the impact of quarantine depends on when it is implemented and the proportion of the quarantine. Therefore, the study recommended a 63% quarantine rate and prolonged control measures throughout the outbreak to avert the epidemic (Xiong and Yan, 2020).

A modelling study that simulated the impact of various control measures found that workplace social distancing was more effective in reducing the viral spread than school closure. Further, the study demonstrated that a combination strategy of case isolation, close contact quarantining, school closure, and workplace social distancing was most effective in reducing the outbreak in size (Xiong and Yan, 2020). Overall, the effectiveness of the control measures may be significantly hampered by asymptomatic and pre-symptomatic transmissions, which can only be overcome through early and regular testing.

Despite several ongoing research and clinical trials on drugs to treat COVID-19, no clinically approved targeted therapy has been achieved. Management efforts, including early diagnosis, immediate patient isolation, and protective conditions to prevent infection spread, effectively contain the virus. Typical treatment for COVID-19 includes general supportive care, respiratory support, and nutritional support (Majumder and Minko, 2021b). Research efforts on drugs are hampered mainly by the high cost and time associated with developing a new drug (Pushpakom et al., 2019). Repurposing existing drugs based on the relatedness of the disease mechanism to the target drug improve the cost and time parameters. Research targets for developing therapeutic drugs for COVID-19 have focused on either the blockage of viral entry mechanisms to host cells or the suppression or complete interference of one of the replication processes (Correia, 2010). However, the low stability associated with the target drug components before delivery to the target site and the loss of binding specificity is potent limitations to the research approach (Rogers et al., 2020).

Vaccines, on the other hand, are biological preparations from altered forms of disease agents capable of stimulating the immune system to fight the pathogen and at the same time create a memory that helps

the immune system to respond more robustly to the repetition of the same or a similar disease (Graham, 2013). Vaccines differ from therapeutic drugs in their ability to keep the memory of previous encounters with the pathogen. They consist of essential components of the disease agent that serve as an antigen. Various vaccines consist of either protein, nucleic acids or a synthetic substitute of the disease agent capable of stimulating immune response but not inducing the disease (Majumder and Minko, 2021b). Existing forms of vaccines are being explored and repurposed for use against SARS-CoV-2 and have shown positive results. If successful, it is expected that a combination of personal defence and epidemic control measures will lead to complete elimination or extermination of the virus.

1.1.11 SARS-CoV-2 testing during the pandemic

Following the identification of the virus, scientists were able to sequence and make available the viral genome. This enabled the development of Reverse Transcription—Polymerase Chain Reaction (RT-PCR) assays, which are now the gold standard for molecular diagnosis of the respiratory syndrome COVID19. RT-PCR test involves isolation of viral RNA from the clinical sample followed by generation of complementary DNA (cDNA) by RNA-dependent DNA polymerase reaction. Next, cDNA is converted to double-stranded DNA (dsDNA) through PCR amplification. The amplification of this sample is performed until the viral cDNA is detected by a fluorescent or electrical signal that allows for the real-time quantification of the virus (VanGuilder et al., 2008). Because of the rising number of cases and rapid spread of the virus, the world has been facing a shortage of RT-PCR supplies, hampering the widescale testing containment strategies of the public health system.

Specimen pooling involves combining specimens into groups and processing them as a single specimen. As already mentioned, this technique was proposed by Dorfman to screen recruits in the 1940s. It has undergone several modifications and better improved for applications in HIV and hepatitis screening in blood banks. Research on the technique has focused on reducing the number of tests needed to pinpoint a positive case in a group. In the context of coronavirus and the associated high transmissibility, a shorter turnaround time is crucial to reduce infection spread. This study aimed to improve the technique to reduce cost and number of tests and increase throughput in the shortest time possible.

Pooling has gained global application, with many laboratories exploring testing alternatives to ensure optimal use of resources and efficiency. A recent study demonstrated an eight per cent increase in testing efficiency and an eight per cent cost reduction (Shental et al., 2020) using a single round of testing. Another study reported a method that creates multiple pools from samples followed by a single round of RT-qPCR and an application that produces results with the viral load of each sample (Ghosh et al., 2020). This not only increases testing capacity and efficiency but, at the same time, provides more detailed information than routine PCR testing. Though the single round method appears to have many complex steps, it is advantageous over the two-round pooling. It addresses the dilemma of performing

RNA extraction before pooling in both the first and second-round testing. To this end, a robust and simplified method is needed to achieve a population-wide screening as institutional and organizational activities resumes.

1.1.12 Aim and Objectives

This study aims to refine the pooling method and apply it to the South African context to investigate whether it could speed up and reduce the cost of testing of SARS-CoV-2. A successful study could streamline large scale testing of the whole population.

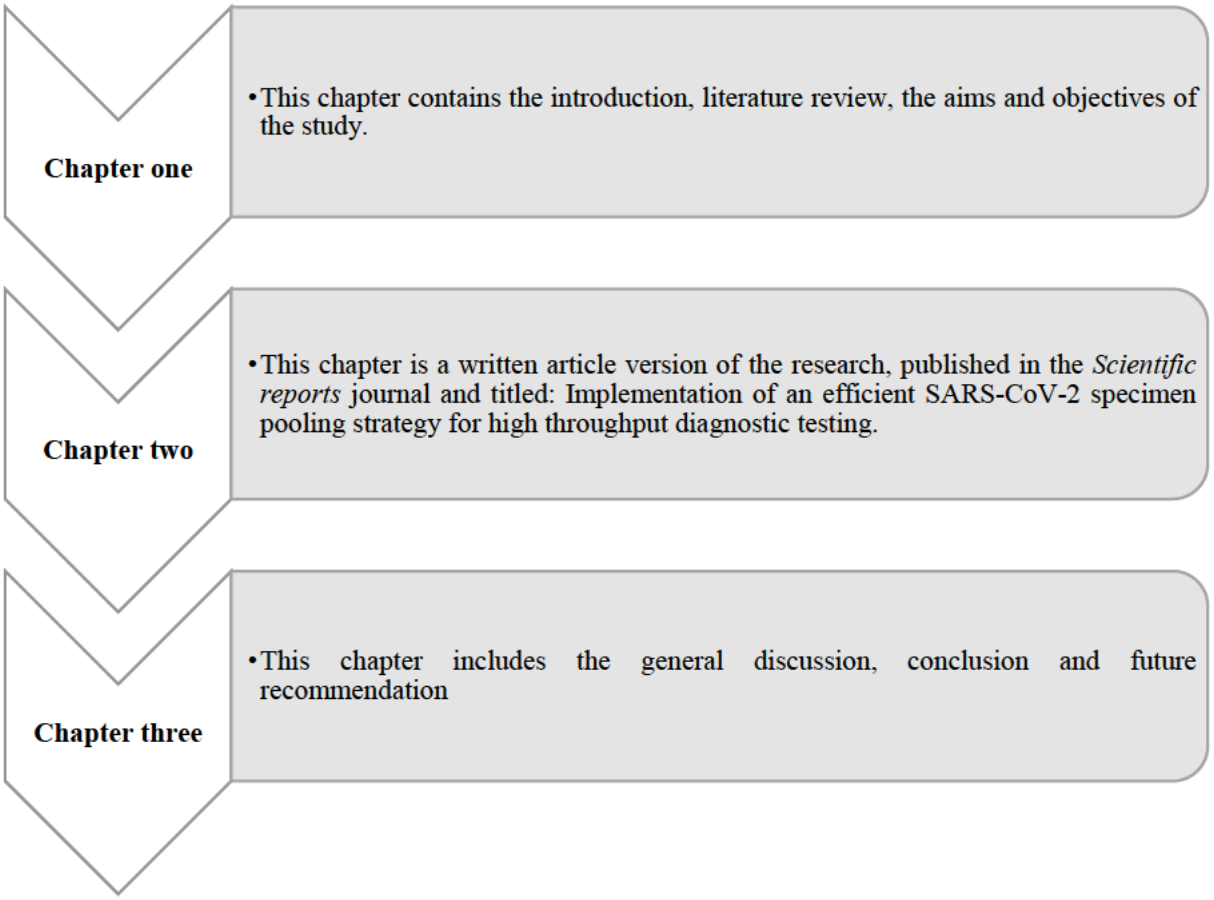
Specifically, the objectives are to:

1. Develop a SARS-CoV-2 qPCR pooling protocol for diagnostic testing
2. Validate this method using locally available samples
3. Improve the technique for application in the South African context (based on different prevalence)

The method was evaluated in two scenarios using pool sizes of 27 and 81 to estimate the most appropriate cost-effectiveness ratio of positive vs negative. The experiment was repeated several times to evaluate the accuracy and check for loss of sensitivity.

1.1.13 Dissertation outline

This dissertation comprises three chapters with an overall aim of detailing the research findings on implementing specimen pooling for high throughput diagnosis of SARS-CoV-2.



Bridging Chapter one and Chapter two

Chapter one gives a general overview of the role of mass screening using specimen pooling to control the SARS-CoV-2 pandemic. The review further describes the epidemiology, life cycle, structure and general management of the virus. Specimen pooling offers a viable alternative to the standard gold test for achieving the benefits of mass testing in a pandemic scenario. Hence the optimization of the method and its application is an urgent necessity of which this study addresses. Chapter two provides detailed review and study outcomes from implementing the hypercube-based pooling method described in the dissertation. Positive samples were reliably detected in groups of 27 and 81, with minimal loss of assay sensitivity for samples with individual Ct values of up to 32.

Results from the study confirm our method saves time and cost as compared to other diagnostic approaches. This Chapter is published in the journal *Scientific Reports*.

Singh, L., Anyaneji, U.J., Ndifon, W. *et al.* Implementation of an efficient SARS-CoV-2 specimen pooling strategy for high throughput diagnostic testing. *Sci Rep* **11,** 17793 (2021). https://doi.org/10.1038/s41598-021-96934-z

DOI: https://doi.org/10.1038/s41598-021-96934-z

CHAPTER TWO

2.1. Published article

Implementation of an efficient SARS-CoV-2 specimen pooling strategy for high throughput diagnostic testing

Lavanya Singh[1,*,#], Ugochukwu J. Anyaneji[1,#], Wilfred Ndifon[2,*], Neil Turok[3], Stacey A. Mattison[1], Richard Lessells[1], Ilya Sinayskiy[4, 5], Emmanuel J. San[1], Houriiyah Tegally[1], Shaun Barnett[6], Trevor Lorimer[6], Francesco Petruccione[4, 5], Tulio de Oliveira[1,*]

[1]KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), School of Laboratory Medicine & Medical Sciences, University of KwaZulu-Natal, Durban, South Africa

[2]African Institute for Mathematical Sciences, The Next Einstein Initiative, Rwanda

[3]Higgs Centre for Theoretical Physics, School of Physics and Astronomy, University of Edinburgh, United Kingdom

[4]School of Chemistry and Physics, University of Kwa-Zulu Natal, Westville, South Africa

[5]National Institute for Theoretical and Computational Sciences (NITheCS), KwaZulu-Natal, South Africa

[6]Discipline of Electrical, Electronic and Computer Engineering, University of KwaZulu-Natal, Durban, South Africa

[*]Correspondence: Lavanya Singh (singhl@ukzn.ac.za), Wilfred Ndifon (wndifon@nexteinstein.org), Tulio de Oliveira (tuliodna@gmail.com)

[#]These authors contributed equally

Abstract

The rapid identification and isolation of infected individuals remains a key strategy for controlling the spread of SARS-CoV-2. Frequent testing of populations to detect infection early in asymptomatic or presymptomatic individuals can be a powerful tool for intercepting transmission, especially when the viral prevalence is low. However, RT-PCR testing – the gold standard of SARS-CoV-2 diagnosis – is expensive, making regular testing of every individual unfeasible. Sample pooling is one approach to lowering costs. By combining samples and testing them in groups the number of tests required is reduced, substantially lowering costs. Here we report on the implementation of pooling strategies using 3-d and 4-d hypercubes to test a professional sports team in South Africa. We have shown that infected samples can be reliably detected in groups of 27 and 81, with minimal loss of assay sensitivity for samples with individual Ct values of up to 32. We report on the automation of sample pooling, using a liquid-handling robot and an automated web interface to identify positive samples. We conclude that hypercube pooling allows for the reliable RT-PCR detection of SARS-CoV-2 infection, at significantly lower costs than lateral flow antigen tests.

Keywords

SARS-CoV-2, RT-PCR, pooled testing, high-throughput

Introduction

A novel coronavirus, SARS-CoV-2, emerged at the end of 2019 in the city of Wuhan, China. The highly transmissible nature of SARS-CoV-2 (Hu et al., 2020) has resulted in a pandemic which continues to persist. South Africa (SA) and other African countries are currently facing a resurgence or “third wave” of infections, which in some countries is more severe than

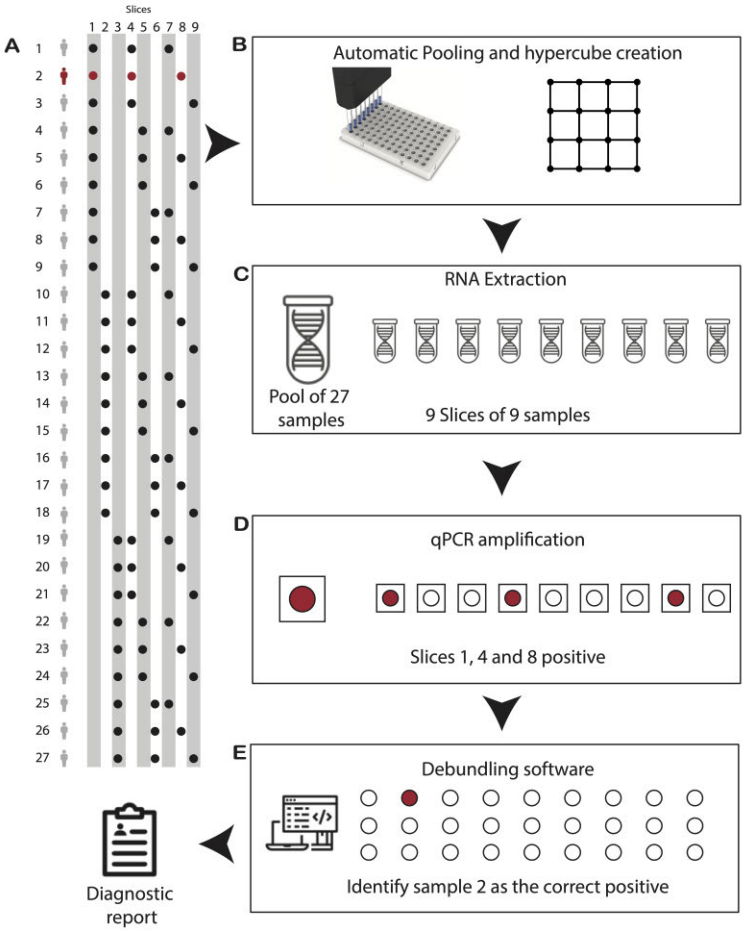
previously experienced.(Long et al., 2020, Tegally et al., 2020a, Vahidy et al., 2020) Although there are approved vaccines against SARS-CoV-2, these are not yet available in amounts sufficient to control the pandemic. To make matters worse, there are new SARS-CoV-2 variants (which are typically identified using genomics surveillance tools (de Oliveira et al., 2021a, Rao et al., 2020)) that have been identified in SA,(Diseases, 2020) which may be associated with higher transmissibility and hence a more rapid spread of the virus. There is therefore an urgent need for more efficient population screening and the isolation of infected individuals to reduce the transmission of SARS-CoV-2.

Presymptomatic or asymptomatic individuals, who are infectious viral carriers,(Barak et al., 2020) are the hidden drivers of the pandemic. They represent an estimated proportion ranging from 18 – 81% of infections,(Nikolai et al., 2020) thereby posing a major challenge to the containment of SARS-CoV-2. If such individuals can be efficiently detected through frequent, repeated population testing at scale and thereby enabled to isolate before they infect others, the spread of the virus can be prevented.[8] Therefore, efficient and affordable, high throughput SARS-CoV-2 testing is highly desirable as a means of controlling the pandemic.

Reverse-transcription real-time polymerase chain reaction (RT-PCR) testing is the gold-standard technology used for SARS-CoV-2 diagnosis. This test can cost up to US$ 56 (ZAR 850) per test, making high-throughput RT-PCR testing of every individual impractical. Sample pooling offers an attractive solution. By combining samples and testing them together, instead of performing individual tests, one can significantly reduce the number of tests and the associated labour and consumable costs. This method was first proposed by Dorfman (Dorfman, 1943) in 1943. Hypercube pooling, developed by Mutesa *et al* (2020),(Mutesa et al., 2020b) requires even fewer tests and yields greater cost savings. For

example, at viral prevalences p<0.05%, hypercube pooling yields a 100-fold cost-reduction as opposed to a 22-fold for Dorfman's algorithm. Hypercube pooling therefore offers a highly affordable means of testing large numbers of samples. We also describe a quantitative cost comparison with lateral flow antigen (Bendavid et al.) tests, finding hypercube pooling to have a significant cost advantage at low viral prevalence (Supplementary Information - Cost comparison of hypercube-based pooled tests vs. lateral flow antigen tests**)**.

Hypercube pooling has been eloquently explained in the literature.(Mutesa et al., 2020b) Briefly, the samples to be tested are divided into equally sized subsamples which are pooled together according to a mathematical algorithm. Here we report on pooled testing methods which can uniquely identify infected samples among groups of 27 and 81 samples (using 3- and 4-dimensional hypercubes, respectively), in far fewer than 27 or 81 tests. If the group is negative, then all individual samples in that group are deemed negative. If a group is positive, then its sub-samples are recombined in the form of slices within the hypercube corresponding to different overlapping sub-pools. Groups of 27 or 81 samples are sub-pooled into 9 and 12 slices respectively, each consisting of 9 or 27 subsamples respectively. Each individual sample is then represented in 3 or 4 different slices respectively, and the test results for the slices can be used to infer which sample is positive, based on its consistent detection within each of the hypercube slices, without an individual test ever being required. The slicing patterns for these groups are shown in the Supplementary Information Tables S1 and S2 and the complete workflows are presented in Figure. 1.


A
Slices
1 2 3 4 5 6 7 8 9
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
B
Automatic Pooling and hypercube creation
C
RNA Extraction
Pool of 27 samples
9 Slices of 9 samples
D
qPCR amplification
Slices 1, 4 and 8 positive
E
Debundling software
Identify sample 2 as the correct positive
Diagnostic report

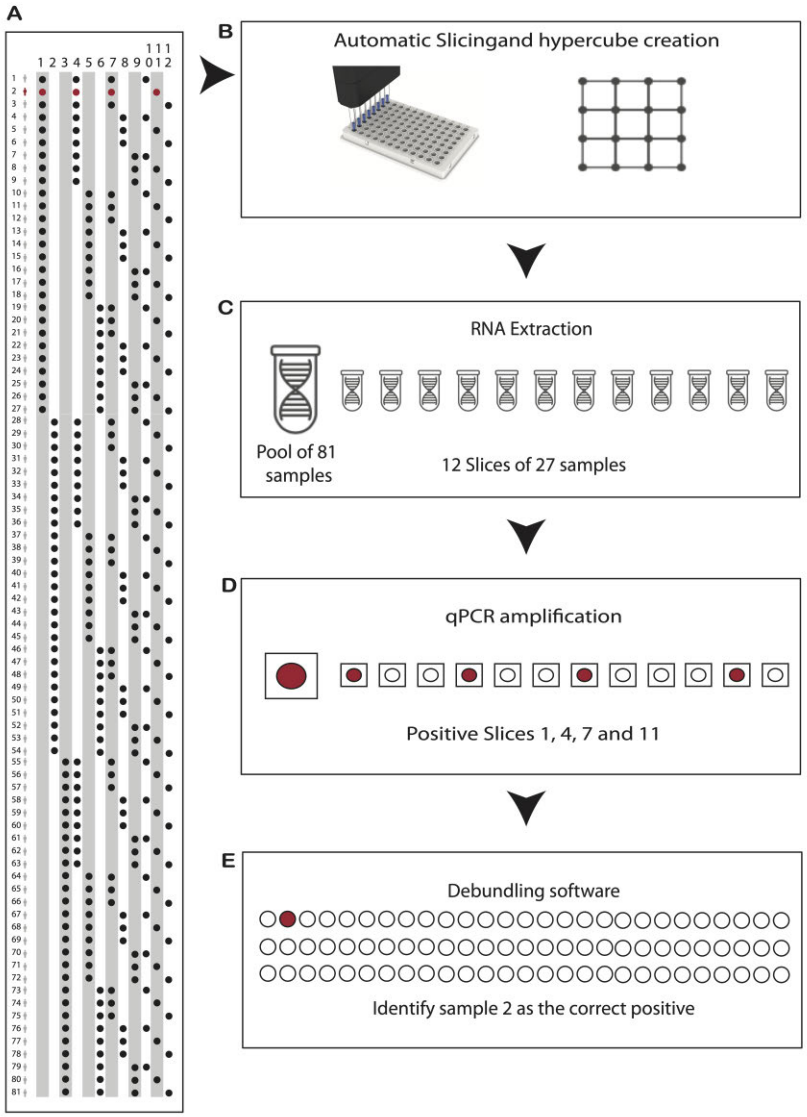


Figure. 1. Examples of 3-dimensional (top) and 4-dimensional (bottom) hypercube workflows; *(A) one positive sample (red, sample 2) is contained in a group of 27 samples or, respectively, 81 samples; (B) 9 (resp. 12) slice pools, each consisting of 9 (resp. 27) sub-samples, are created using a liquid handling robot; (C) RNA extraction is done on the group pool of 27 (resp. 81) samples (1 reaction) and 9 (resp. 12) slice pools (9 or 12 reactions); (D) qPCR results are positive for the group of 27 (resp. 81) samples and the respective slices 1, 4 and 8 (resp. 1, 4, 7, and 11); (E) De-bundling software is used for the automatic determination of the positive sample in the hypercube and for generating individual diagnostic reports.*

Although the pooling strategy described above offers considerable benefits over individual testing, it does involve tedious and repetitive pipetting steps, which can be cumbersome

when performed manually and can raise the risk of human error. For example, a 3-d hypercube involves 108 precise pipetting steps in different combinations, *i.e.*, a group-level pool of 27 sub-samples plus 9 slice-level sub-pools of 9 sub-samples each. The 4-d hypercube involves 405 pipetting steps, *i.e.*, the first group-level pool of 81 sub-samples plus 12 slice-level sub-pools of 27 sub-samples each. To reduce these intensive demands on laboratory staff as well as the risk of human error, we have implemented an automated pipeline using a liquid-handling robot, Opentrons OT-2. This open source robotic platform lends itself to scalability and rapid deployment. The OT2 is modularly constructed and utilises widely available non-proprietary electronic components that are neither difficult nor expensive to repair or replace. Flexibility is offered through the Python Application Programming Interface (API), which simplifies custom protocol development. The additional laboratory automation described here, refers to the automatic determination of the individual positive result(s) and the de-bundling (or ungrouping) of pooled results into individual results using an automated web interface. The advantages of automated sample pooling include increased human-resource time (which can be better allocated for data analysis, experimental setup, etc.), the elimination of the possibility of human error (due to fatigue or distraction, for example), the reduction of costs associated with manual labour, and the flexibility to easily replicate or modify protocols based on the starting number of samples and desired pooling strategy.

Results

3-d 27-sample hypercube experiments:

The results of three 3-d hypercube experiments (Experiments A – C), in which a single positive sample is diluted up to 27-fold, are shown in Figure 2. The plots are based on the Ct scores obtained for each of the three target genes (Supplementary Information Table S3).

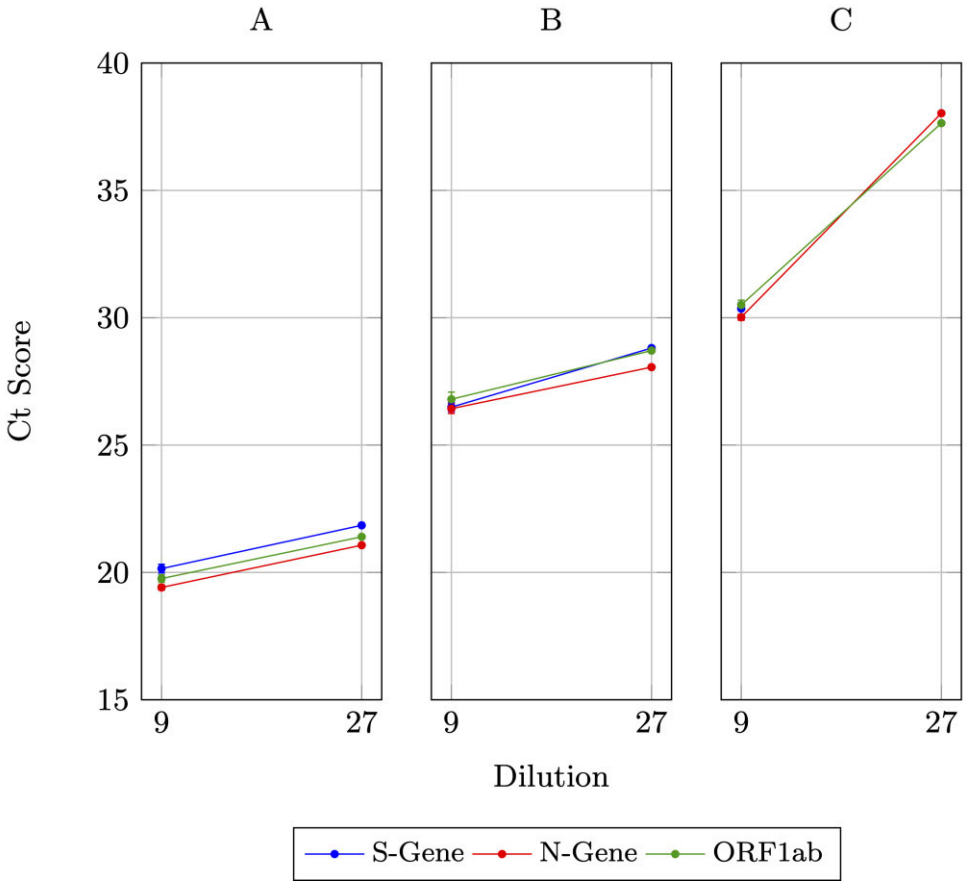


Figure 2. Detection of a single positive sample after 9-fold and 27-fold dilution, in 3-d hypercube pooling. Error bars are shown for the 3 positive pools, where the positive sample is diluted 9-fold.

The results show that the positive samples for each experiment can be successfully amplified and detected in the pool containing 27 samples and in each of the positive slices containing 9 samples each.

4-d 81-sample hypercube experiments:

The results of eight 4-d hypercube experiments (Experiments A – H), in which a single positive sample is diluted up to 81-fold, are shown in Figure. 3. The plots are based on the Ct scores obtained for each of the three target genes (Supplementary Information Table S4).

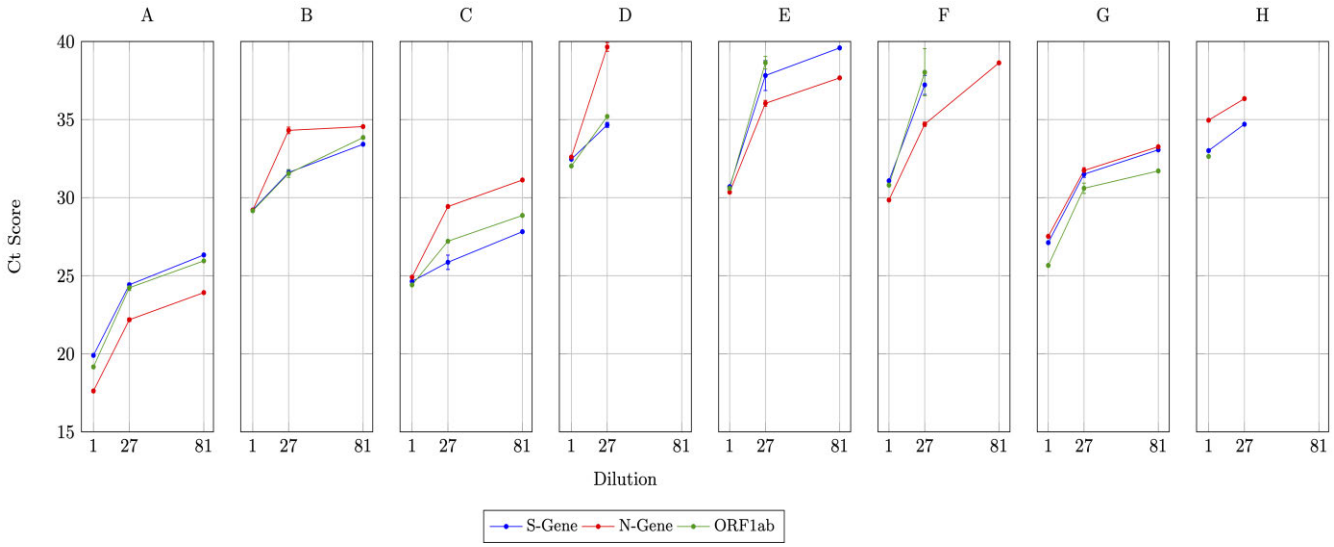


Figure. 3. Detection of a single positive sample out of 81 for each of eight 4-d hypercube experiments (A – H). Ct scores are shown for: a) the undiluted sample, b) the mean and standard errors for the four positive slice pools which is diluted 27-fold, and c) the hypercube pool in which the single positive is diluted 81-fold.

The operational sensitivity of the pooled test procedure is decreased for samples with an initial (undiluted) Ct value exceeding 32 as indicated by experiments D and H, where amplification in the 81-pool samples was not successful (none of the target genes were detected).

Efficiency – number of tests saved using the pooling method:

We received approximately 121 samples per week over an 8 week period, for testing (ie. 968 individual samples in total). Each week, three groups of either 27 or 81 pooled tests were performed (ie. 24 sets of pooled tests in total). The starting point for each set of pooled tests is a single, group-level test, in which either 27 or 81 samples are pooled and then tested. Slice-level tests are conducted only if the group-level test is positive. Since all 24 group-level tests were negative, a greater than 40 (i.e. 986 divided by 24)-fold reduction in the number of tests used was achieved compared to individual testing.

Laboratory automation:

We created an automated and easy-to-use web interface to assist in the de-bundling of pooled testing results and the identification of the positive sample(s) found in a pool. The de-bundling module requires as input the analysis file generated from the pooled testing results and a list of sample identifiers. It produces a de-bundling report, which can be easily integrated into a laboratory information management system (LIMS) such as LabWare 7 (used in our laboratory), for individual result reporting.

To infer the positive sample, the user specifies the pool size, i.e. 27 or 81, the number of positive slices and their respective slice numbers. The application then uses this information to determine and output individual positive test results. Both applications, freely available at http://krisp2.ukzn.ac.za:8080/rega-pooling/pooling (Supplementary Information Figure. S3), were used for all the analysis and reporting of results described in this paper.

Discussion

Diagnostic detection of SARS-CoV-2 is crucial for managing the pandemic.(Mujawar et al., 2020, Paliwal et al., 2020) The early identification and isolation of SARS-CoV-2 infected individuals remain a key strategy to interrupt community transmission. In this pilot study, we

have successfully implemented hypercube pooling for frequent testing of a professional sports team. We show that this procedure reliably detects a single positive sample in a group of 81, provided the starting sample has a Ct ≤ 32. Whilst there is a need for further evaluation, particularly with samples with a wider range of Ct values and pools with multiple positive samples for different viral prevalence scenarios, these results indicate that the pooling strategy is indeed a promising approach for cost-efficient RT-PCR testing.

Other studies have demonstrated the successful implementation of pooled testing using groups of up to 5 and 8 samples each.(Barak et al., 2020) To our knowledge, this is the first study that has implemented the SARS-CoV-2 hypercube-based pooled testing strategy, using group sizes as large as 81.

We selected for analysis samples with Ct values that are typical for the population under study. For comparison, we also selected samples with Ct values that are one standard deviation away from the mean Ct value. We used historical data from the population of samples tested routinely in KZN to determine the distribution of Ct values present at various percentile ranges (Supplementary Information Figure. S1). The asymmetrical (left-skewed) distribution of the data may be attributed to the population of samples that was used to generate the curve i.e. symptomatic people with active disease (and therefore lower Ct scores) presenting at health facilities.

The retention of the sensitivity of the pooled test procedure as the hypercube dimension size (and consequently the dilution) increases is an obvious concern.

Our results demonstrated no loss of assay sensitivity for samples with an initial (undiluted) Ct value ≤ 32 (Figure. 3). Samples with higher Ct values would typically have a lower viral load, which may be deemed as clinically and / or epidemiologically insignificant, since the

infection is not likely to be contagious (a cut-off Ct > 30 can be associated with non-infectious samples).(Jefferson et al., 2020)

Propagating SARS-CoV-2 from clinical samples can also be used as a valuable proxy for infectiousness. There are however, conflicting reports on the cut-off Ct value at which the virus is not cultivable, which can range anywhere from 30 to >35.(Singanayagam et al., 2020, Young et al., 2020) In any event, high Ct values are expected to be associated with low infectivity;(Bullard et al., 2020) the maximum viral load occurs during the onset of symptoms, when the risk of transmission is highest.(Cevik et al., 2020a) In respiratory samples, the viral load is highest during the initial stage of infection (patients in the early stages of infection usually have Ct values of 20 – 30 or less(Young et al., 2020)), and reaches a peak in the second week, and then falls.(Kevadiya et al., 2021a) The pooling strategy we have implemented is therefore most beneficial for detecting asymptomatic or presymptomatic people who are on the viral load incline. In addition, the effects of the loss of assay sensitivity can be mitigated by implementing more frequent testing – a solution which could be more easily afforded given the demonstrated cost savings of the pooled testing strategy.

There have also been reports of inherent inaccuracy due to higher false negative rates associated with the pooled testing method.(Daon et al., 2020) Our method however, includes many consistency checks. For example, the detection of a single positive pool (corresponding to a hypercube slice) when all other pools are negative is indicative of a testing error.(Mutesa et al., 2020b) This is especially beneficial when compared to approaches such as binary pooled testing(Hwang, 1972) which relies on repeated, sequential testing of the positive sample, and a single negative test can prematurely terminate the sequence of tests leading to a false negative result. Another cause of false

negative results is sequence variation at primer or probe binding sites on the viral RNA.(Reijns et al., 2020a) In South Africa, a new SARS-CoV-2 lineage (501Y.V2) characterised by various mutations in the spike (S) gene has been reported.(Tegally et al., 2020a) However, there was no apparent loss of sensitivity for the S-gene in the RT-PCR assay used in this study.

We have developed and successfully implemented a pilot SARS-CoV-2 pooled testing strategy for a prominent South African rugby team. The successful implementation of pooled testing requires that 3 important criteria are met *viz*. (i) efficiency, (ii) sensitivity and (iii) operational feasibility (Barak et al., 2020). Our study was successful in satisfying each of these criteria. Firstly, we achieved a >40-fold gain in testing efficiency, and a corresponding reduction in cost, compared to individual testing. The cost reduction is most impressive when all individuals are negative, because 27 or 81 different samples can be determined to be negative by using just one qPCR reaction, in a 3-d or 4-d hypercube, respectively. This was the case with the screening of a South African rugby team at the KRISP laboratory, where, during the first 8 weeks, all of the samples were tested using hypercubes of sizes 27 or 81, and found to be negative. Only from the 9th week was an additional round of slice testing required. Secondly, no significant loss of sensitivity for samples with appreciable viral loads ($Ct \leq 32$) was observed. Thirdly, we validated our pooling workflow, *i.e.,* by using liquid-handling robots to automated sample pooling and software applications to automate the identification of positive samples.

There are more conventional methods for high-throughput diagnostic testing which are available. For example, SARS-CoV-2 lateral flow antigen (Bendavid et al.) tests are point-of-care rapid tests that can be used for large-scale screening. Although LFAs may be appealing because they are cheap, since they do not require a laboratory with specialized

equipment or personnel and can provide results in 15 – 30 minutes, they do present some limitations; the sensitivity of this type of technology can vary considerably around an average sensitivity of 56.2% (95% CI 29.5 to 79.8%), (Dinnes et al., 2020) thereby decreasing its utility in screening certain populations such as health-care workers and other front-line personnel. We have performed a detailed cost comparison of the hypercube-based pooled testing strategy compared to the LFAs. We estimate that at prevalences below 0.43%, it costs over 6 times more to achieve the reliable detection (ie. a detection probability of at least 99.9%) of SARS-CoV-2 infectious individuals by using LFAs compared to hypercube testing (Supplementary Information - Cost comparison of hypercube-based pooled tests vs. lateral flow antigen tests). Despite their higher cost for reliable detection, LFAs may be used to complement hypercube-based testing, particularly when quick results are needed and the lower sensitivity of the assay is not considered problematic.

The pooled testing strategy and downstream automation protocol described in this study have together led to a significant reduction in cost, kit usage and turnaround time in our laboratory. While we demonstrated its successful application for frequent testing of a professional sports team, this approach can potentially be applied for screening other low prevalence, asymptomatic population groups. Further evaluation of this approach in different populations and epidemic settings is desirable.

References

1. Hu, B., Guo, H., Zhou, P. & Shi, Z.-L. Characteristics of SARS-CoV-2 and COVID-19. *Nature Reviews Microbiology*, 1-14 (2020).

2. Long, S. W. *et al.* Molecular architecture of early dissemination and massive second wave of the SARS-CoV-2 virus in a major metropolitan area. *MBio* **11** (2020).

3. Tegally, H. *et al.* Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv* (2020).

4. Vahidy, F. S. *et al.* Characteristics and outcomes of COVID-19 patients during initial peak and resurgence in the Houston metropolitan area. *Jama* **324**, 998-1000 (2020).

5. de Oliveira, T. *et al.* A novel variant of interest of SARS-CoV-2 with multiple spike mutations detected through travel surveillance in Africa. *medRxiv* (2021).

6. Rao, K. *et al.* Review on Newly Identified Coronavirus and its Genomic Organization. (2020).

7. Diseases, N. I. f. C. *The emergence of a mutated SARS-COV-2 lineage in South Africa*, <https://www.nicd.ac.za/the-emergence-of-a-mutated-sars-cov-2-lineage-in-south-africa/> (2020).

8. Barak, N. *et al.* Lessons from applied large-scale pooling of 133,816 SARS-CoV-2 RT-PCR tests. *medRxiv* (2020).

9. Nikolai, L. A., Meyer, C. G., Kremsner, P. G. & Velavan, T. P. Asymptomatic SARS Coronavirus 2 infection: Invisible yet invincible. *International Journal of Infectious Diseases* (2020).

10. Dorfman, R. The detection of defective members of large populations. *The Annals of Mathematical Statistics* **14**, 436-440 (1943).

11. Mutesa, L. *et al.* A pooled testing strategy for identifying SARS-CoV-2 at low prevalence. *Nature*, 1-8 (2020).

12. Mujawar, M. A. *et al.* Nano-enabled biosensing systems for intelligent healthcare: towards COVID-19 management. *Materials Today Chemistry* **17**, 100306 (2020).

13. Paliwal, P. *et al.* Grand challenges in bio-nanotechnology to manage the COVID-19 pandemic. *Frontiers in Nanotechnology* **2**, 5 (2020).

14. Jefferson, T., Spencer, E., Brassey, J. & Heneghan, C. Viral cultures for COVID-19 infectivity assessment. Systematic review. *MedRxiv* (2020).

15. Singanayagam, A. *et al.* Duration of infectiousness and correlation with RT-PCR cycle threshold values in cases of COVID-19, England, January to May 2020. *Eurosurveillance* **25**, 2001483 (2020).

16. Young, B. E. *et al.* Epidemiologic features and clinical course of patients infected with SARS-CoV-2 in Singapore. *Jama* **323**, 1488-1494 (2020).

17. Bullard, J. *et al.* Poliquin, Predicting infectious Severe Acute Respiratory Syndrome Coronavirus 2 from diagnostic samples. *Clin. Infect. Dis* (2020).

18. Cevik, M., Kuppalli, K., Kindrachuk, J. & Peiris, M. Virology, transmission, and pathogenesis of SARS-CoV-2. *bmj* **371** (2020).

19. Kevadiya, B. D. *et al.* Diagnostics for SARS-CoV-2 infections. *Nature Materials*, 1-13 (2021).

20. Daon, Y., Huppert, A. & Obolski, U. Inflated false-negative rates in pooled RT-PCR tests of SARS-CoV-2. *medRxiv* (2020).

21. Hwang, F. K. A method for detecting all defective members in a population by group testing. *Journal of the American Statistical Association* **67**, 605-608 (1972).

22. Reijns, M. A. *et al.* A sensitive and affordable multiplex RT-qPCR assay for SARS-CoV-2 detection. *PLoS biology* **18**, e3001030 (2020).

23. Dinnes, J. *et al.* Rapid, point-of-care antigen and molecular-based tests for diagnosis of SARS-CoV-2 infection. *Cochrane Database of Systematic Reviews* (2020).

Online Methods

SARS-CoV-2 clinical diagnostics:

The standard protocol used for SARS-CoV-2 detection at the KRISP Laboratory, University of KwaZulu-Natal, is to extract viral ribonucleic acid (RNA) from nasopharyngeal and / or oropharyngeal swab samples. This is followed by RT-PCR (TaqPath COVID-19 CE-IVD RT-PCR kit, ThermoFisher Scientific, MA, USA) for the detection of three SARS-CoV-2 target genes (viz. S-gene, N-gene and ORF1ab). A sample is considered "positive" if at least 2 out of the 3 target genes are amplified with an above-background fluorescence signal, cycle threshold, Ct, $\leq$ 40 PCR cycles. A sample is considered "inconclusive" if 1 out of the 3 target genes is positive (Ct $\leq$ 40) and should then be repeat tested. A sample is considered "negative" if none of the 3 target genes is amplified. Swab samples collected in viral transport medium (VTM) for routine SARS-CoV-2 surveillance in KwaZulu-Natal, South Africa were used in this study. All experiments were conducted in accordance with relevant local guidelines and regulations.

Historical data from a population of approximately 1200 samples tested routinely in KZN was used to determine the distribution of Ct values present at various percentile ranges (Supplementary Information Figure. S1). Positive samples used for each of the experiments were chosen based on their respective Ct scores to represent different quantiles of the Ct value distribution.

Pooling experiments were conducted to verify the successful detection of a positive sample embedded inside: (Covid-19 National Emergency Response Center, 2020.) a 3-d hypercube containing 26 other, negative samples; and (2) a 4-d hypercube containing 80 other, negative samples.

27-sample, 3-d hypercube experiments

Three positive samples were used in three independent pooling tests to evaluate proof-of-concept and determine whether these samples would still result in successful PCR amplification when diluted 27- (in group-level pools) and 9- (in slice-level pools) fold using 3-d hypercubes.

First, for each positive sample, a group-level pool containing 27 samples was created. Each positive sample was combined with 26 negative swab samples as follows: sample tubes were labelled from 1 to 27. Samples were added to each of the labelled tubes, with the positive sample inserted blindly, i.e. in a position of the hypercube not known to the person performing the pooling experiments or the analysis of results. Each sample was vortexed briefly and 200 μl was manually pipetted into a separate collection tube containing the pooled sample. The collection tube (with a total volume of 5400 μl) was then briefly vortexed and 200 μl was used for paramagnetic bead-based RNA extraction using the Chemagic360 automated extraction system (PerkinElmer Inc., Cat.# CMG-1049, MA, USA). The purified nucleic acid was eluted in 100 μl of elution buffer. 10 μl of RNA was then used in the RT-PCR reaction (TaqPath™ COVID-19 CE-IVD RT-PCR Kit, Thermo Fisher Scientific).

Group-level pools that were positive were sub-pooled into 9 slice-level pools, containing 9 samples each (Supplementary Information - Table S1). Each sample found in the positive, group-level pool was represented in 3 different slice-level pools.

81-sample, 4-d hypercube experiments

As in the 3-d experiments, positive samples were chosen to represent the different quantiles of the Ct value distribution that is typical of our study population (Supplementary Information Figure. S1).

One positive swab sample was combined with 80 negative swab samples in each of eight experiments (Experiments A – H). For each experiment, sample tubes were labelled from 1 to 81, and each of the respective positive samples was inserted blindly into a position within the 81-sample hypercube. Each sample was vortexed briefly and 200 μl was added to a separate collection tube containing the pooled sample. The collection tube (with a total volume of 16 200 μl for each experiment) was then briefly vortexed and 200 μl was used for RNA extraction and subsequent RT-PCR, as described above. For each experiment, the 81-sample pool was further sub-pooled into 12 slices containing 27 samples each (Supplementary Information - Table S2). Each individual sample was therefore represented in 4 different slices in each experiment.

Based on the consistent amplification of the positive sample across the respective slices in each of the 3-d and 4-d hypercube experiments, the location of the positive sample could be deduced.

Laboratory robotics:

The robotics work flow is summarised in the following steps (Supplementary Information Figure. S2):

1. A new 200 µl tip is collected by the single-channel 300 µl electronic pipette.
2. The electronic pipette draws 200 µl from the sample rack for the current slice iteration.

3. The electronic pipette dispenses the sample in the relevant sample sub-pool.
4. The tip is discarded into the robot's trash.

The programming codes for the 3-d and 4-d pooling experiments using the OT-2 liquid handling robot is available at https://github.com/krisp-kwazulu-natal/efficient-SARS-CoV-2-pooled-testing-strategy-code

Software for the de-bundling and inference of positive results

Custom scripts were implemented in Python for the de-bundling and inference of positive results from a pool of 27 or 81 samples. When inferring positive results, our scripts take into account degenerates and can infer more than one positive sample per pool at low prevalence. For ease of use, we developed a user-friendly, web-based front end using the Java Web Toolkit (Jwt), an open-source, robust Java web front end (GUI; http://webtoolkit.eu/jwt) library. Minimal user interaction is required as user interaction is limited to uploading the files to be processed or providing pool details in the case of inference. A report detailing the de-bundling results is provided to the user once processing of the data is complete. Our de-bundling and inference software is freely available and can be accessed at the URL: http://krisp2.ukzn.ac.za:8080/rega-pooling/pooling. The programming scripts for the software we developed are available at https://github.com/krisp-kwazulu-natal/efficient-SARS-CoV-2-pooled-testing-strategy-code

Acknowledgements

This work is based upon research supported by the South African Research Chair Initiative of the Department of Science and Technology and National Research Foundation. TdO and FP want to thank the South African Department of Science and Innovation (DSI) for funding the research proposal on Massive Screening for Covid-19 using Pooled Testing (Grant number: DSI/CON 0000/2020).

Author information

Affiliations

KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), School of Laboratory Medicine & Medical Sciences, University of KwaZulu-Natal, Durban, South Africa

Lavanya Singh, Ugochukwu J. Anyaneji, Stacey A. Mattison, Houriiyah Tegally, Emmanuel J. San, Richard Lessells, Tulio de Oliveira

African Institute for Mathematical Sciences, The Next Einstein Initiative, Rwanda

Wilfred Ndifon

Higgs Centre for Theoretical Physics, School of Physics and Astronomy, University of Edinburgh, United Kingdom

Neil Turok

School of Chemistry and Physics, University of Kwa-Zulu Natal, University Road, Chiltern Hill, Westville, South Africa

Francesco Petruccione, Ilya Sinayskiy

National Institute for Theoretical and Computational Sciences (NITheCS), KwaZulu-Natal, South Africa

Francesco Petruccione, Ilya Sinayskiy

Department of Electrical, Electronic and Computer Engineering, University of KwaZulu-Natal, Durban, South Africa

Trevor Lorimer, Shaun Barnett

Author contributions

LS, UJA, WN and NT designed the study; LS, UJA and SAM co-ordinated the experiments, performed validation of the pooling experiments, data analysis and interpretation; LS and UJA wrote the manuscript; WN, NT and RL contributed to editing the manuscript; WN and NT contributed to the theory; IS, SB, TL contributed to the development and implementation of the robotics automation workflow; HT and EJS developed and implemented the scripts used for the automated web applications; FP and TdO supervised the project; all authors reviewed and approved the final manuscript.

Ethics declarations

We used de-identified remnant nasopharyngeal and oropharyngeal swab samples from patients tested positive for SARS-CoV-2 by RT–qPCR from public health laboratories in South Africa. The project was approved by University of KwaZulu-Natal Biomedical Research Ethics Committee (Protocol reference no. BREC/00001195/2020; project title: COVID-19 transmission and natural history in KwaZulu-Natal, South Africa: epidemiological investigation to guide prevention and clinical care).

Competing interests

The authors declare no competing interests.

CHAPTER THREE

3.1. General discussion, conclusion and future recommendation.

Rapid diagnosis of SARS-CoV-2 is crucial to evaluate the disease spread and contain the effects of the pandemic. Several factors, including the emergence of new variants, the role of transmission before symptoms, and presymptomatic/asymptomatic individuals, sustain outbreaks, thus making early detection vital. Given the many limitations around the standard gold test, a pertinent need is to consider applying a strategy that could help conserve scarce reagents and reduce other constraints associated with the individual PCR test. For this purpose, we implemented a prove of concept sample pooling strategy using 3-d and 4-d hypercubes to test samples from a professional sports team in South Africa and show that infected samples can be reliably detected in groups of 27 and 81, with minimal loss of assay sensitivity for samples with individual Ct values up to 32. We reported on the automation of sample pooling, using a liquid-handling robot and an automated web interface to identify positive samples. We conclude that hypercube pooling allows for the reliable RT-PCR detection of SARS-CoV-2 infection at a significantly lower cost than lateral flow antigen (Bendavid et al. 2020) tests.

Implementing our pooling strategy is highly beneficial for early detection of infection at the initial stages and detects samples within the detection limit threshold values. Studies reveal that maximum viral loads are reached at the onset of symptoms, with an associated increase in the risk of transmission (Cevik et al., 2020b). Samples from the respiratory system have also been reported to reach the highest viral load during the early stages of infection and present Ct values ranging from 20-30 or less, followed by low viral load (Kevadiya et al., 2021b). More frequent testing may be implemented to mitigate the effect of loss of assay sensitivity, which is more easily compensated by the cost affordability of the entire pooling strategy. (Shental et al., 2020) reported the efficiency gains of sample group testing using pooling strategy and, as supported by the US food and drug administration, published guidelines for sample pooling. A unique advantage of our method is that it is a nonadaptive single-stage method. Positive samples are detected using a single round of testing, which reduces the overall test time and does not require storing samples for a second round of testing. The pooling method is equally highly attractive for diagnostic laboratories, given the logical simplicity augmented by the automated pipetting component.

Our method incorporates many consistency checks that addressee inherent inaccuracy from higher false negatives, likely resulting from dilution. For example, detecting a single positive slice pool with the other negative slices indicates a mistake in testing. This is quite beneficial compared to binary testing methods that solely rely on repeated testing of the positive pools in subsequent rounds of testing, which can incorrectly obstruct the identification of the positive samples where there are false-negative results. False-negative results may also arise from sequence variation at primer or probe binding sites on the viral RNA (Reijns et al., 2020b). A similar instance has been observed from a new SARS-CoV-2 variant (501Y.V2) identified in South Africa, which is commonly characterised by various mutations

in the spike (S) gene (Tegally et al., 2020b). However, there was no apparent loss of sensitivity for the S-gene in the RT-PCR assay used in this study.

Our pooling method is more affordable than other conventional methods for large-scale screening, such as lateral flow assays, which are considered cheap and do not require highly skilled personnel. It has reliable sensitivity and a wider range of applications in different populations contexts. The sensitivity of LFAs has considerable variation averaged at 56.2% (95% CI 29.5 to 79.8%) (Lee et al., 2020), which reduces their applicability in different population contexts (such as health-care workers and other front-line personnel). A comparative study of LFAs to molecular genetic assays found that the sensitivity of LFAs was significantly lower in patients presenting to the hospital but higher in patients with at least seven days of symptoms and those with CRP levels >100 mg/L upon presentation (Ong et al., 2020). The gross dependence of LFAs on the production of antibodies allows a significant time for viral spread and consequently annuls the benefits of early detection of infection. We analysed a cost comparison of the hypercube-based pooled testing strategy compared to the LFAs and found that, at prevalence below 0.43%, it costs over six times more to achieve reliable detection of SARS-CoV-2 infectious individuals using LFAs compared to hypercube testing (assuming a detection probability of 99.9%).

Our method can be easily adapted to work on any standard liquid-dispensing robot, which offers increased scalability and rapid deployment. This is essential to reduce or eliminate human error that may occur from the tedious manual pipetting. Another benefit of our method is the automatic determination of the individual positive result(s) and the de-bundling (or ungrouping) pooled results into individual results using an automated web interface. Our method offers advantages such as increased human-resource time (which can be better allocated for data analysis, experimental setup, etc.), elimination of the possibility of human error (due to fatigue or distraction, for example), reducing costs associated with manual labour, and the flexibility to easily replicate or modify protocols based on the starting number of samples and desired pooling strategy.

As the pandemic continues, the influence of host selection pressure, environmental pressure, and therapeutic pressure influence the evolution of newer variants. Processes underlying viral evolution results in changes in viral RNA sequences for assay binding sites. These new SARS-CoV-2 variants may have increased transmissibility and the potential to evade immune responses. This, therefore, warrants the need for improved surveillance of the different circulating lineages. In light of this, we validated an assay specifically targeting the mutations in the emerging variants of concern, as shown in chapter three of this study. This method enables timely detection and efficient surveillance of the new variants. Further to expanding testing efforts to accommodate the detection of new viral variants, it is worthwhile to explore the utilization of the assay in the pooling strategy.

Our study on assay for detection and discrimination of new variants revealed that the gold standard Taqpath PCR assay was only able to detect the B.1.1.7 (also 501Y.V1) variant, first detected in the

United Kingdom, due to a key deletion in these viruses, spike Δ69–70, which causes a "spike gene target failure" (SGTF) result(Vogels et al., 2021). However, an SGTF result is not definitive for B.1.1.7, and this assay cannot detect other variants of concern (VOC) that lack spike Δ69–70, such as B.1.351 (also 501Y.V2), detected in South Africa, and P.1 (also 501Y.V3), recently detected in Brazil. We identified a deletion in the ORF1a gene (ORF1a Δ3675–3677) in all three variants, which has not yet been widely detected in other SARS-CoV-2 lineages. Using ORF1a Δ3675–3677 as the primary target and spike Δ69–70 to differentiate, we designed and validated an open-source PCR assay to detect SARS-CoV-2 VOC. Our assay can be rapidly deployed in combination with our pooling method in laboratories worldwide to enhance surveillance for the local emergence and spread of B.1.1.7, B.1.351, and P.1. (Vogels et al., 2021).

To allow for an increase in testing efficiency, which is sequel to minimizing the number of pools required to screen a given population, designs should vary according to the prevalence rate. Pooling designs generally differ by the number of pools and by the number of samples per pool. Specifically, as the prevalence rate in population rises, more pools are required to correctly identify all positive carriers in a single testing round. Similarly, the required number of samples per pool decreases with an increase in the positivity rate. Conversely, where there are low case rates, more efficient pooling designs can be used, i.e., fewer pools are required to test the same number of individuals while the number of samples per pool increases. As the pandemic continues to spread, the infection rates within a given population change continuously. To provide a close to optimal pooling design, it is ideal to make a rough estimate of the current positivity in the population. In general, this rate can be estimated based on the carrier rates observed during the last few days for a given population since they typically vary over time. Further studies are therefore necessary to ascertain the consistency of our observation in different population positivity rates.

In summary, this study explored a novel method for efficient SARS-CoV-2 diagnostic testing that can significantly increase throughput and testing resources and be readily implemented in clinical diagnostic laboratories.

3.2. Conclusion

The hypercube pooling strategy offers an attractive compromise between minimizing the total number of tests to reduce costs and maximizing the speed of testing to reduce the spread of the virus. We have demonstrated its viability for group sizes up to 81 samples, showing that significant cost savings can be achieved. We have quantified the loss of sensitivity due to dilution and discussed several ways to mitigate it—for example, through frequently repeated group tests. These strategies could enable the use of larger pool sizes, bringing even more significant cost savings at low prevalence. The most striking aspect of our approach is how rapidly the cost of testing a population can reduce, pooled test sensitivity permitting, as the viral prevalence decreases. This should incentivise decision-makers to act firmly to

drive the prevalence down through mass screening, contact tracing and isolation. As the viral prevalence is reduced, all aspects of this strategy become more and more affordable

3.3. Recommendation

During a rapidly changing epidemic, testing strategies will need to adapt to potential increases in the positive test rate. Group testing of pooled specimens also requires highly sensitive assays to low positive samples that are not missed. Therefore, strategies must be employed to closely monitor the use of pooling as the positive rate of test specimens increases in an outbreak. Additionally, the impact of different extraction methods on the recovery of RNA and overall test sensitivity need to be evaluated. Therefore, laboratories must perform their own validation pool studies for kits used for each RNA extraction and amplification based on the prevalence rate of COVID-19 in their own region. The use of saliva samples, which have been reported to be more sensitive than nasopharyngeal swabs, needs further exploration. In conclusion, this study showed that pooling is an effective approach to expand the impact of limited test resources and reagents during specific stages of an infectious disease outbreak.

REFERENCES

ABDALHAMID, B., BILDER, C. R., MCCUTCHEN, E. L., HINRICHS, S. H., KOEPSELL, S. A. & IWEN, P. C. 2020. Assessment of Specimen Pooling to Conserve SARS CoV-2 Testing Resources. *American Journal of Clinical Pathology,* 153**,** 715-718.

BARAK, N., BEN-AMI, R., SIDO, T., PERRI, A., SHTOYER, A., RIVKIN, M., LICHT, T., PERETZ, A., MAGENHEIM, J. & FOGEL, I. 2020. Lessons from applied large-scale pooling of 133,816 SARS-CoV-2 RT-PCR tests. *medRxiv*.

BELOUZARD, S., CHU, V. C. & WHITTAKER, G. R. 2009. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proceedings of the National Academy of Sciences of the United States of America,* 106**,** 5871-5876.

BENDAVID, E., MULANEY, B., SOOD, N., SHAH, S., LING, E., BROMLEY-DULFANO, R., LAI, C., WEISSBERG, Z., SAAVEDRA, R., TEDROW, J., TVERSKY, D., BOGAN, A., KUPIEC, T., EICHNER, D., GUPTA, R., IOANNIDIS, J. & BHATTACHARYA, J. 2020. COVID-19 Antibody Seroprevalence in Santa Clara County, California. *medRxiv*.

BHEKISISA. 2020. National COVID-19 testing backlog means patients wait up to 10 days for results. *Bhekisisa Centre for Health Journalism*.

BILDER, C. R. & TEBBS, J. M. 2012. Pooled-testing procedures for screening high volume clinical specimens in heterogeneous populations. *Statistics in medicine,* 31**,** 3261-3268.

BOSCH, B. J., VAN DER ZEE, R., DE HAAN, C. A. M. & ROTTIER, P. J. M. 2003a. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *Journal of virology,* 77**,** 8801-8811.

BOSCH, B. J., VAN DER ZEE, R., DE HAAN, C. A. M. & ROTTIER, P. J. M. 2003b. The Coronavirus Spike Protein Is a Class I Virus Fusion Protein: Structural and Functional Characterization of the Fusion Core Complex. 77**,** 8801-8811.

BULLARD, J., DUST, K., FUNK, D., STRONG, J., ALEXANDER, D., GARNETT, L., BOODMAN, C., BELLO, A., HEDLEY, A. & SCHIFFMAN, Z. 2020. Poliquin, Predicting infectious Severe Acute Respiratory Syndrome Coronavirus 2 from diagnostic samples. *Clin. Infect. Dis*.

CABRERA ALVARGONZALEZ, J. J., REY CAO, S., PÉREZ CASTRO, S., MARTINEZ LAMAS, L., CORES CALVO, O., TORRES PIÑON, J., PORTEIRO FRESCO, J., GARCIA COMESAÑA, J. & REGUEIRO GARCIA, B. 2020. Pooling for SARS-CoV-2 control in care institutions. *BMC Infectious Diseases,* 20**,** 745.

CARTER, L. J., GARNER, L. V., SMOOT, J. W., LI, Y., ZHOU, Q., SAVESON, C. J., SASSO, J. M., GREGG, A. C., SOARES, D. J., BESKID, T. R., JERVEY, S. R. & LIU, C. 2020. Assay Techniques and Test Development for COVID-19 Diagnosis. *ACS Central Science,* 6**,** 591-605.

CEVIK, M., KUPPALLI, K., KINDRACHUK, J. & PEIRIS, M. 2020a. Virology, transmission, and pathogenesis of SARS-CoV-2. *bmj,* 371.

CEVIK, M., KUPPALLI, K., KINDRACHUK, J. & PEIRIS, M. 2020b. Virology, transmission, and pathogenesis of SARS-CoV-2. 371**,** m3862.

CHAN, J. F.-W., KOK, K.-H., ZHU, Z., CHU, H., TO, K. K.-W., YUAN, S. & YUEN, K.-Y. 2020. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerging microbes & infections,* 9**,** 221-236.

CHINAZZI, M., DAVIS, J. T., AJELLI, M., GIOANNINI, C., LITVINOVA, M., MERLER, S., PASTORE, Y. P. A., MU, K., ROSSI, L., SUN, K., VIBOUD, C., XIONG, X., YU, H., HALLORAN, M. E., LONGINI, I. M., JR. & VESPIGNANI, A. 2020. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science,* 368**,** 395-400.

CORREIA, I. R. 2010. Stability of IgG isotypes in serum. *mAbs,* 2**,** 221-232.

COUTARD, B., VALLE, C., DE LAMBALLERIE, X., CANARD, B., SEIDAH, N. G. & DECROLY, E. 2020. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Research,* 176**,** 104742.

COVID-19 NATIONAL EMERGENCY RESPONSE CENTER, E., CASE MANAGEMENT TEAM, K. C. F. D. C. & PREVENTION 2020. Coronavirus Disease-19: Summary of 2,370 Contact Investigations of the

First 30 Cases in the Republic of Korea. *Osong public health and research perspectives,* 11**,** 81-84.

DAON, Y., HUPPERT, A. & OBOLSKI, U. 2020. Inflated false-negative rates in pooled RT-PCR tests of SARS-CoV-2. *medRxiv*.

DE OLIVEIRA, T., LUTUCUTA, S., NKENGASONG, J., MORAIS, J., PAIXÃO, J. P., NETO, Z., AFONSO, P., MIRANDA, J., DAVID, K. & INGLÊS, L. 2021a. A novel variant of interest of SARS-CoV-2 with multiple spike mutations detected through travel surveillance in Africa. *medRxiv*.

DE OLIVEIRA, T., LUTUCUTA, S., NKENGASONG, J., MORAIS, J., PAIXÃO, J. P., NETO, Z., AFONSO, P., MIRANDA, J., DAVID, K., INGLÊS, L., RAISA RIVAS CARRALERO, A. P. A. P., FREITAS, H. R., MUFINDA, F., TESSEMA, S. K., TEGALLY, H., SAN, E. J., WILKINSON, E., GIANDHARI, J., PILLAY, S., GIOVANETTI, M., NAIDOO, Y., KATZOURAKIS, A., GHAFARI, M., SINGH, L., TSHIABUILA, D., MARTIN, D. & LESSELLS, R. J. 2021b. A novel variant of interest of SARS-CoV-2 with multiple spike mutations detected through travel surveillance in Africa. 2021.03.30.21254323.

DENG, S.-Q. & PENG, H.-J. 2020. Characteristics of and Public Health Responses to the Coronavirus Disease 2019 Outbreak in China. 9**,** 575.

DINNES, J., DEEKS, J. J., ADRIANO, A., BERHANE, S., DAVENPORT, C., DITTRICH, S., EMPERADOR, D., TAKWOINGI, Y., CUNNINGHAM, J. & BEESE, S. 2020. Rapid, point-of-care antigen and molecular-based tests for diagnosis of SARS-CoV-2 infection. *Cochrane Database of Systematic Reviews*.

DISEASES, N. I. F. C. 2020. *The emergence of a mutated SARS-COV-2 lineage in South Africa* [Online]. Available: https://www.nicd.ac.za/the-emergence-of-a-mutated-sars-cov-2-lineage-in-south-africa/ [Accessed 21 December 2020].

DORFMAN, R. 1943. The Detection of Defective Members of Large Populations. *The Annals of Mathematical Statistics,* 14**,** 436-440.

DUFFY, S., SHACKELTON, L. A. & HOLMES, E. C. 2008. Rates of evolutionary change in viruses: patterns and determinants. *Nature Reviews Genetics,* 9**,** 267-276.

FARIDI, S., NIAZI, S., SADEGHI, K., NADDAFI, K., YAVARIAN, J., SHAMSIPOUR, M., JANDAGHI, N. Z. S., SADEGHNIIAT, K., NABIZADEH, R., YUNESIAN, M., MOMENIHA, F., MOKAMEL, A., HASSANVAND, M. S. & MOKHTARIAZAD, T. 2020. A field indoor air measurement of SARS-CoV-2 in the patient rooms of the largest hospital in Iran. *Science of The Total Environment,* 725**,** 138401.

FURSTENAU, T. N., COCKING, J. H., HEPP, C. M. & FOFANOV, V. Y. 2020. Sample pooling methods for efficient pathogen screening: Practical implications. *PLOS ONE,* 15**,** e0236849.

GHOSH, S., RAJWADE, A., KRISHNA, S., GOPALKRISHNAN, N., SCHAUS, T. E., CHAKRAVARTHY, A., VARAHAN, S., APPU, V., RAMAKRISHNAN, R., CH, S., JINDAL, M., BHUPATHI, V., GUPTA, A., JAIN, A., AGARWAL, R., PATHAK, S., REHAN, M. A., CONSUL, S., GUPTA, Y., GUPTA, N., AGARWAL, P., GOYAL, R., SAGAR, V., RAMAKRISHNAN, U., KRISHNA, S., YIN, P., PALAKODETI, D. & GOPALKRISHNAN, M. 2020. Tapestry: A Single-Round Smart Pooling Technique for COVID-19 Testing. 2020.04.23.20077727.

GIANDHARI, J., PILLAY, S., WILKINSON, E., TEGALLY, H., SINAYSKIY, I., SCHULD, M., LOURENCO, J., CHIMUKANGARA, B., LESSELLS, R., MOOSA, Y., GAZY, I., FISH, M., SINGH, L., KHANYILE, K. S., FONSECA, V., GIOVANETTI, M., ALCANTARA, L. C., PETRUCCIONE, F. & DE OLIVEIRA, T. 2020. Early transmission of SARS-CoV-2 in South Africa: An epidemiological and phylogenetic report. 2020.05.29.20116376.

GIORGI ROSSI, P., MARINO, M., FORMISANO, D., VENTURELLI, F., VICENTINI, M., GRILLI, R. & REGGIO EMILIA, C.-W. G. 2020. Characteristics and outcomes of a cohort of COVID-19 patients in the Province of Reggio Emilia, Italy. *PloS one,* 15**,** e0238281-e0238281.

GRAHAM, B. S. 2013. Advances in antiviral vaccine development. *Immunol Rev,* 255**,** 230-42.

GUAN, W.-J., NI, Z.-Y., HU, Y., LIANG, W.-H., OU, C.-Q., HE, J.-X., LIU, L., SHAN, H., LEI, C.-L., HUI, D. S. C., DU, B., LI, L.-J., ZENG, G., YUEN, K.-Y., CHEN, R.-C., TANG, C.-L., WANG, T., CHEN, P.-Y., XIANG, J., LI, S.-Y., WANG, J.-L., LIANG, Z.-J., PENG, Y.-X., WEI, L., LIU, Y., HU, Y.-H., PENG, P.,

WANG, J.-M., LIU, J.-Y., CHEN, Z., LI, G., ZHENG, Z.-J., QIU, S.-Q., LUO, J., YE, C.-J., ZHU, S.-Y. & ZHONG, N.-S. 2020. Clinical Characteristics of Coronavirus Disease 2019 in China. 382**,** 1708-1720.

HALL, K. S., SAMARI, G., GARBERS, S., CASEY, S. E., DIALLO, D. D., ORCUTT, M., MORESKY, R. T., MARTINEZ, M. E. & MCGOVERN, T. 2020. Centring sexual and reproductive health and justice in the global COVID-19 response. *Lancet (London, England),* 395**,** 1175-1177.

HU, B., GUO, H., ZHOU, P. & SHI, Z.-L. 2020. Characteristics of SARS-CoV-2 and COVID-19. *Nature Reviews Microbiology***,** 1-14.

HWANG, F. K. 1972. A method for detecting all defective members in a population by group testing. *Journal of the American Statistical Association,* 67**,** 605-608.

JEFFERSON, T., SPENCER, E., BRASSEY, J. & HENEGHAN, C. 2020. Viral cultures for COVID-19 infectivity assessment. Systematic review. *MedRxiv*.

JIN, J.-M., BAI, P., HE, W., WU, F., LIU, X.-F., HAN, D.-M., LIU, S. & YANG, J.-K. 2020. Gender Differences in Patients With COVID-19: Focus on Severity and Mortality. *Frontiers in public health,* 8**,** 152-152.

KEVADIYA, B. D., MACHHI, J., HERSKOVITZ, J., OLEYNIKOV, M. D., BLOMBERG, W. R., BAJWA, N., SONI, D., DAS, S., HASAN, M. & PATEL, M. 2021a. Diagnostics for SARS-CoV-2 infections. *Nature Materials***,** 1-13.

KEVADIYA, B. D., MACHHI, J., HERSKOVITZ, J., OLEYNIKOV, M. D., BLOMBERG, W. R., BAJWA, N., SONI, D., DAS, S., HASAN, M., PATEL, M., SENAN, A. M., GORANTLA, S., MCMILLAN, J., EDAGWA, B., EISENBERG, R., GURUMURTHY, C. B., REID, S. P. M., PUNYADEERA, C., CHANG, L. & GENDELMAN, H. E. 2021b. Diagnostics for SARS-CoV-2 infections. *Nature Materials,* 20**,** 593-605.

KORBER, B., FISCHER, W. M., GNANAKARAN, S., YOON, H., THEILER, J., ABFALTERER, W., HENGARTNER, N., GIORGI, E. E., BHATTACHARYA, T., FOLEY, B., HASTIE, K. M., PARKER, M. D., PARTRIDGE, D. G., EVANS, C. M., FREEMAN, T. M., DE SILVA, T. I., ANGYAL, A., BROWN, R. L., CARRILERO, L., GREEN, L. R., GROVES, D. C., JOHNSON, K. J., KEELEY, A. J., LINDSEY, B. B., PARSONS, P. J., RAZA, M., ROWLAND-JONES, S., SMITH, N., TUCKER, R. M., WANG, D., WYLES, M. D., MCDANAL, C., PEREZ, L. G., TANG, H., MOON-WALKER, A., WHELAN, S. P., LABRANCHE, C. C., SAPHIRE, E. O. & MONTEFIORI, D. C. 2020. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell,* 182**,** 812-827.e19.

LEE, J., KIM, S. Y., SUNG, H., LEE, S. W., LEE, H., ROH, K. H., YOO, C. K. & HONG, K. H. 2020. Challenges and issues of SARS-CoV-2 pool testing. *Lancet Infect Dis,* 20**,** 1232-1233.

LI, Q., GUAN, X., WU, P., WANG, X., ZHOU, L., TONG, Y., REN, R., LEUNG, K. S. M., LAU, E. H. Y., WONG, J. Y., XING, X., XIANG, N., WU, Y., LI, C., CHEN, Q., LI, D., LIU, T., ZHAO, J., LIU, M., TU, W., CHEN, C., JIN, L., YANG, R., WANG, Q., ZHOU, S., WANG, R., LIU, H., LUO, Y., LIU, Y., SHAO, G., LI, H., TAO, Z., YANG, Y., DENG, Z., LIU, B., MA, Z., ZHANG, Y., SHI, G., LAM, T. T. Y., WU, J. T., GAO, G. F., COWLING, B. J., YANG, B., LEUNG, G. M. & FENG, Z. 2020a. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. 382**,** 1199-1207.

LI, R., PEI, S., CHEN, B., SONG, Y., ZHANG, T., YANG, W. & SHAMAN, J. 2020b. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science (New York, N.Y.),* 368**,** 489-493.

LI, W., MOORE, M. J., VASILIEVA, N., SUI, J., WONG, S. K., BERNE, M. A., SOMASUNDARAN, M., SULLIVAN, J. L., LUZURIAGA, K., GREENOUGH, T. C., CHOE, H. & FARZAN, M. 2003. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature,* 426**,** 450-454.

LONG, S. W., OLSEN, R. J., CHRISTENSEN, P. A., BERNARD, D. W., DAVIS, J. J., SHUKLA, M., NGUYEN, M., SAAVEDRA, M. O., YERRAMILLI, P. & PRUITT, L. 2020. Molecular architecture of early dissemination and massive second wave of the SARS-CoV-2 virus in a major metropolitan area. *MBio,* 11.

LU, R., ZHAO, X., LI, J., NIU, P., YANG, B., WU, H., WANG, W., SONG, H., HUANG, B., ZHU, N., BI, Y., MA, X., ZHAN, F., WANG, L., HU, T., ZHOU, H., HU, Z., ZHOU, W., ZHAO, L., CHEN, J., MENG, Y.,

WANG, J., LIN, Y., YUAN, J., XIE, Z., MA, J., LIU, W. J., WANG, D., XU, W., HOLMES, E. C., GAO, G. F., WU, G., CHEN, W., SHI, W. & TAN, W. 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet,* 395**,** 565-574.

MAJID, F., OMER, S. B. & KHWAJA, A. I. 2020. Optimising SARS-CoV-2 pooled testing for low-resource settings. *The Lancet. Microbe,* 1**,** e101-e102.

MAJUMDER, J. & MINKO, T. 2021a. Recent Developments on Therapeutic and Diagnostic Approaches for COVID-19. *The AAPS journal,* 23**,** 14-14.

MAJUMDER, J. & MINKO, T. 2021b. Recent Developments on Therapeutic and Diagnostic Approaches for COVID-19. *Aaps j,* 23**,** 14.

MALLAPATY, S. 2020. The mathematical strategy that could transform coronavirus testing. *Nature,* 583**,** 504-505.

MARTINES, R. B., RITTER, J. M., MATKOVIC, E., GARY, J., BOLLWEG, B. C., BULLOCK, H., GOLDSMITH, C. S., SILVA-FLANNERY, L., SEIXAS, J. N., REAGAN-STEINER, S., UYEKI, T., DENISON, A., BHATNAGAR, J., SHIEH, W. J. & ZAKI, S. R. 2020. Pathology and Pathogenesis of SARS-CoV-2 Associated with Fatal Coronavirus Disease, United States. *Emerg Infect Dis,* 26**,** 2005-2015.

MEHTA, P., MCAULEY, D. F., BROWN, M., SANCHEZ, E., TATTERSALL, R. S. & MANSON, J. J. 2020. COVID-19: consider cytokine storm syndromes and immunosuppression. *Lancet,* 395**,** 1033-1034.

MORAN, K. R. & DEL VALLE, S. Y. 2016. A Meta-Analysis of the Association between Gender and Protective Behaviors in Response to Respiratory Epidemics and Pandemics. *PLoS One,* 11**,** e0164541.

MUJAWAR, M. A., GOHEL, H., BHARDWAJ, S. K., SRINIVASAN, S., HICKMAN, N. & KAUSHIK, A. 2020. Nano-enabled biosensing systems for intelligent healthcare: towards COVID-19 management. *Materials Today Chemistry,* 17**,** 100306.

MUTESA, L., NDISHIMYE, P., BUTERA, Y., SOUOPGUI, J., UWINEZA, A., RUTAYISIRE, R., MUSONI, E., RUJENI, N., NYATANYI, T., NTAGWABIRA, E., SEMAKULA, M., MUSANABAGANWA, C., NYAMWASA, D., NDASHIMYE, M., UJENEZA, E., MWIKARAGO, I. E., MUVUNYI, C. M., MAZARATI, J. B., NSANZIMANA, S., TUROK, N. & NDIFON, W. 2020a. A strategy for finding people infected with SARS-CoV-2: optimizing pooled testing at low prevalence. 2020.05.02.20087924.

MUTESA, L., NDISHIMYE, P., BUTERA, Y., SOUOPGUI, J., UWINEZA, A., RUTAYISIRE, R., NDORICIMPAYE, E. L., MUSONI, E., RUJENI, N. & NYATANYI, T. 2020b. A pooled testing strategy for identifying SARS-CoV-2 at low prevalence. *Nature***,** 1-8.

NICD. 2020a. *FIRST CASE OF COVID-19 CORONAVIRUS REPORTED IN SA* [Online]. Available: https://www.nicd.ac.za/first-case-of-covid-19-coronavirus-reported-in-sa/ [Accessed 16 June 2021].

NICD. 2020b. *LATEST CONFIRMED CASES OF COVID-19 IN SOUTH AFRICA* [Online]. National Institute for Communicable Diseases. Available: https://www.nicd.ac.za/latest-confirmed-cases-of-covid-19-in-south-africa/ [Accessed 12 Nov. 2020].

NICD. 2021. *LATEST CONFIRMED CASES OF COVID-19 IN SOUTH AFRICA (29 MAY 2021)* [Online]. Available: https://www.nicd.ac.za/latest-confirmed-cases-of-covid-19-in-south-africa-29-may-2021/ [Accessed 02 June 2021].

NIKOLAI, L. A., MEYER, C. G., KREMSNER, P. G. & VELAVAN, T. P. 2020. Asymptomatic SARS Coronavirus 2 infection: Invisible yet invincible. *International Journal of Infectious Diseases*.

ONG, D. S. Y., DE MAN, S. J., LINDEBOOM, F. A. & KOELEMAN, J. G. M. 2020. Comparison of diagnostic accuracies of rapid serological tests and ELISA to molecular diagnostics in patients with suspected coronavirus disease 2019 presenting to the hospital. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases,* 26**,** 1094.e7-1094.e10.

PALIWAL, P., SARGOLZAEI, S., BHARDWAJ, S. K., BHARDWAJ, V., DIXIT, C. & KAUSHIK, A. 2020. Grand challenges in bio-nanotechnology to manage the COVID-19 pandemic. *Frontiers in Nanotechnology,* 2**,** 5.

PARASKEVIS, D., KOSTAKI, E. G., MAGIORKINIS, G., PANAYIOTAKOPOULOS, G., SOURVINOS, G. & TSIODRAS, S. 2020. Full-genome evolutionary analysis of the novel corona virus (2019-nCoV) rejects the hypothesis of emergence as a result of a recent recombination event. *Infection, Genetics and Evolution,* 79**,** 104212.

PARK, M., COOK, A. R., LIM, J. T., SUN, Y. & DICKENS, B. L. 2020. A Systematic Review of COVID-19 Epidemiology Based on Current Evidence. *Journal of clinical medicine,* 9**,** 967.

PUSHPAKOM, S., IORIO, F., EYERS, P. A., ESCOTT, K. J., HOPPER, S., WELLS, A., DOIG, A., GUILLIAMS, T., LATIMER, J., MCNAMEE, C., NORRIS, A., SANSEAU, P., CAVALLA, D. & PIRMOHAMED, M. 2019. Drug repurposing: progress, challenges and recommendations. *Nature Reviews Drug Discovery,* 18**,** 41-58.

RAO, K., VERMA, P., KUMAR, K., VERMA, M. K., SIDDIQUI, A. H., SINGH, S., DHAKAR, J. S., KURU, C. I., SHARMA, P. K. & SONI, N. O. 2020. Review on Newly Identified Coronavirus and its Genomic Organization.

REDDY, K. P., SHEBL, F. M., FOOTE, J. H. A., HARLING, G., SCOTT, J. A., PANELLA, C., FITZMAURICE, K. P., FLANAGAN, C., HYLE, E. P., NEILAN, A. M., MOHAREB, A. M., BEKKER, L.-G., LESSELLS, R. J., CIARANELLO, A. L., WOOD, R., LOSINA, E., FREEDBERG, K. A., KAZEMIAN, P. & SIEDNER, M. J. 2020. Cost-effectiveness of public health strategies for COVID-19 epidemic control in South Africa: a microsimulation modelling study. 2020.06.29.20140111.

REIJNS, M. A., THOMPSON, L., ACOSTA, J. C., BLACK, H. A., SANCHEZ-LUQUE, F. J., DIAMOND, A., PARRY, D. A., DANIELS, A., O'SHEA, M. & UGGENTI, C. 2020a. A sensitive and affordable multiplex RT-qPCR assay for SARS-CoV-2 detection. *PLoS biology,* 18**,** e3001030.

REIJNS, M. A. M., THOMPSON, L., ACOSTA, J. C., BLACK, H. A., SANCHEZ-LUQUE, F. J., DIAMOND, A., PARRY, D. A., DANIELS, A., O'SHEA, M., UGGENTI, C., SANCHEZ, M. C., O'CALLAGHAN, A., MCNAB, M. L. L., ADAMOWICZ, M., FRIMAN, E. T., HURD, T., JARMAN, E. J., CHEE, F. L. M., RAINGER, J. K., WALKER, M., DRAKE, C., LONGMAN, D., MORDSTEIN, C., WARLOW, S. J., MCKAY, S., SLATER, L., ANSARI, M., TOMLINSON, I. P. M., MOORE, D., WILKINSON, N., SHEPHERD, J., TEMPLETON, K., JOHANNESSEN, I., TAIT-BURKARD, C., HAAS, J. G., GILBERT, N., ADAMS, I. R. & JACKSON, A. P. 2020b. A sensitive and affordable multiplex RT-qPCR assay for SARS-CoV-2 detection. *PLoS Biol,* 18**,** e3001030.

ROBYN, R., JOCELYNE, L., TIANSHENG, Z., MAGIE, F., BEI, X., MELISSA, R., ALI TOLOUE, O., SALVATORE, R., NICHOLAS, J. D., MOHAMMED, N. A.-A., DAVID, J. K., CHRISTOPHER, D. R., JASON, K., DARRYL, F. & ALYSON ANNE, K. 2020. 2019-nCoV (Wuhan virus), a novel Coronavirus: human-to-human transmission, travel-related cases, and vaccine readiness. *The Journal of Infection in Developing Countries,* 14.

ROGERS, T. F., ZHAO, F., HUANG, D., BEUTLER, N., BURNS, A., HE, W. T., LIMBO, O., SMITH, C., SONG, G., WOEHL, J., YANG, L., ABBOTT, R. K., CALLAGHAN, S., GARCIA, E., HURTADO, J., PARREN, M., PENG, L., RAMIREZ, S., RICKETTS, J., RICCIARDI, M. J., RAWLINGS, S. A., WU, N. C., YUAN, M., SMITH, D. M., NEMAZEE, D., TEIJARO, J. R., VOSS, J. E., WILSON, I. A., ANDRABI, R., BRINEY, B., LANDAIS, E., SOK, D., JARDINE, J. G. & BURTON, D. R. 2020. Isolation of potent SARS-CoV-2 neutralizing antibodies and protection from disease in a small animal model. *Science,* 369**,** 956-963.

SHENTAL, N., LEVY, S., SKORNIAKOV, S., WUVSHET, V., SHEMER-AVNI, Y., PORGADOR, A. & HERTZ, T. 2020. Efficient high throughput SARS-CoV-2 testing to detect asymptomatic carriers. 2020.04.14.20064618.

SINGANAYAGAM, A., PATEL, M., CHARLETT, A., BERNAL, J. L., SALIBA, V., ELLIS, J., LADHANI, S., ZAMBON, M. & GOPAL, R. 2020. Duration of infectiousness and correlation with RT-PCR cycle threshold values in cases of COVID-19, England, January to May 2020. *Eurosurveillance,* 25**,** 2001483.

SKYNEWS. 2021. *how COVID spread and storeis of human tragedy* [Online]. Available: https://news.sky.com/story/coronavirus-deaths-pass-one-million-how-covid-19-spread-and-stories-of-human-tragedy-12083278 [Accessed 01 June 2021].

TAO, K., TZOU, P. L., NOUHIN, J., GUPTA, R. K., DE OLIVEIRA, T., KOSAKOVSKY POND, S. L., FERA, D. & SHAFER, R. W. 2021. The biological and clinical significance of emerging SARS-CoV-2 variants. *Nature Reviews Genetics*.

TEGALLY, H., WILKINSON, E., GIOVANETTI, M., IRANZADEH, A., FONSECA, V., GIANDHARI, J., DOOLABH, D., PILLAY, S., SAN, E. J. & MSOMI, N. 2020a. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv*.

TEGALLY, H., WILKINSON, E., GIOVANETTI, M., IRANZADEH, A., FONSECA, V., GIANDHARI, J., DOOLABH, D., PILLAY, S., SAN, E. J., MSOMI, N., MLISANA, K., VON GOTTBERG, A., WALAZA, S., ALLAM, M., ISMAIL, A., MOHALE, T., GLASS, A. J., ENGELBRECHT, S., VAN ZYL, G., PREISER, W., PETRUCCIONE, F., SIGAL, A., HARDIE, D., MARAIS, G., HSIAO, M., KORSMAN, S., DAVIES, M.-A., TYERS, L., MUDAU, I., YORK, D., MASLO, C., GOEDHALS, D., ABRAHAMS, S., LAGUDA-AKINGBA, O., ALISOLTANI-DEHKORDI, A., GODZIK, A., WIBMER, C. K., SEWELL, B. T., LOURENÇO, J., ALCANTARA, L. C. J., POND, S. L. K., WEAVER, S., MARTIN, D., LESSELLS, R. J., BHIMAN, J. N., WILLIAMSON, C. & DE OLIVEIRA, T. 2020b. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. 2020.12.21.20248640.

TIAN, H., LIU, Y., LI, Y., KRAEMER, M. U. G., CHEN, B., WU, C.-H., CAI, J., LI, B., XU, B., YANG, Q., WANG, B., YANG, P., CUI, Y., SONG, Y., ZHENG, P., WANG, Q., BJORNSTAD, O. N., YANG, R., GRENFELL, B. T., PYBUS, O. G. & DYE, C. 2020. Early evaluation of transmission control measures in response to the 2019 novel coronavirus outbreak in China. 2020.01.30.20019844.

VAHIDY, F. S., DREWS, A. L., MASUD, F. N., SCHWARTZ, R. L., BOOM, M. L. & PHILLIPS, R. A. 2020. Characteristics and outcomes of COVID-19 patients during initial peak and resurgence in the Houston metropolitan area. *Jama,* 324**,** 998-1000.

VANGUILDER, H. D., VRANA, K. E. & FREEMAN, W. M. 2008. Twenty-five years of quantitative PCR for gene expression analysis. *Biotechniques,* 44**,** 619-26.

VOGELS, C. B. F., BREBAN, M. I., OTT, I. M., ALPERT, T., PETRONE, M. E., WATKINS, A. E., KALINICH, C. C., EARNEST, R., ROTHMAN, J. E., GOES DE JESUS, J., MORALES CLARO, I., MAGALHÃES FERREIRA, G., CRISPIM, M. A. E., BRAZIL, U. K. C. G. N., SINGH, L., TEGALLY, H., ANYANEJI, U. J., NETWORK FOR GENOMIC SURVEILLANCE IN SOUTH, A., HODCROFT, E. B., MASON, C. E., KHULLAR, G., METTI, J., DUDLEY, J. T., MACKAY, M. J., NASH, M., WANG, J., LIU, C., HUI, P., MURPHY, S., NEAL, C., LASZLO, E., LANDRY, M. L., MUYOMBWE, A., DOWNING, R., RAZEQ, J., DE OLIVEIRA, T., FARIA, N. R., SABINO, E. C., NEHER, R. A., FAUVER, J. R. & GRUBAUGH, N. D. 2021. Multiplex qPCR discriminates variants of concern to enhance global surveillance of SARS-CoV-2. *PLOS Biology,* 19**,** e3001236.

WALLS, A. C., PARK, Y.-J., TORTORICI, M. A., WALL, A., MCGUIRE, A. T. & VEESLER, D. 2020. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell,* 181**,** 281-292.e6.

WANG, M., HAO, H., LEEPER, N. J. & ZHU, L. 2018. Thrombotic Regulation From the Endothelial Cell Perspectives. *Arterioscler Thromb Vasc Biol,* 38**,** e90-e95.

WHO. 2020. Available: https://www.who.int/news/item/13-01-2020-who-statement-on-novel-coronavirus-in-thailand [Accessed 31/05/2021 2021].

WHO. 2021. *WHO Coronavirus (COVID-19) Dashboard* [Online]. world health organisation. Available: https://covid19.who.int/ [Accessed 15/04 2021].

WORLDOMETER. 2021. *COVID19-Coronavirus Pandemic* [Online]. Available: https://www.worldometers.info/coronavirus/ [Accessed 02 June 2021].

WROBEL, A. G., BENTON, D. J., XU, P., ROUSTAN, C., MARTIN, S. R., ROSENTHAL, P. B., SKEHEL, J. J. & GAMBLIN, S. J. 2020. SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. *Nature Structural & Molecular Biology,* 27**,** 763-767.

WU, J. T., LEUNG, K. & LEUNG, G. M. 2020. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet,* 395**,** 689-697.

WU, Z. & MCGOOGAN, J. M. 2020. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *Jama,* 323**,** 1239-1242.

XIONG, H. & YAN, H. 2020. Simulating the infected population and spread trend of 2019-nCov under different policy by EIR model. 2020.02.10.20021519.

YOSHIKAWA, T., HILL, T., LI, K., PETERS, C. J. & TSENG, C. T. 2009. Severe acute respiratory syndrome (SARS) coronavirus-induced lung epithelial cytokines exacerbate SARS pathogenesis by modulating intrinsic functions of monocyte-derived macrophages and dendritic cells. *J Virol,* 83**,** 3039-48.

YOUNG, B. E., ONG, S. W. X., KALIMUDDIN, S., LOW, J. G., TAN, S. Y., LOH, J., NG, O.-T., MARIMUTHU, K., ANG, L. W. & MAK, T. M. 2020. Epidemiologic features and clinical course of patients infected with SARS-CoV-2 in Singapore. *Jama,* 323**,** 1488-1494.

YUKI, K., FUJIOGI, M. & KOUTSOGIANNAKI, S. 2020. COVID-19 pathophysiology: A review. *Clinical immunology (Orlando, Fla.),* 215**,** 108427-108427.

ZHAO, W.-M., SONG, S.-H., CHEN, M.-L., ZOU, D., MA, L.-N., MA, Y.-K., LI, R.-J., HAO, L.-L., LI, C.-P., TIAN, D.-M., TANG, B.-X., WANG, Y.-Q., ZHU, J.-W., CHEN, H.-X., ZHANG, Z., XUE, Y.-B. & BAO, Y.-M. 2020a. The 2019 novel coronavirus resource. *Yi chuan = Hereditas,* 42**,** 212-221.

ZHAO, X., CHEN, D., SZABLA, R., ZHENG, M., LI, G., DU, P., ZHENG, S., LI, X., SONG, C., LI, R., GUO, J. T., JUNOP, M., ZENG, H. & LIN, H. 2020b. Broad and Differential Animal Angiotensin-Converting Enzyme 2 Receptor Usage by SARS-CoV-2. *J Virol,* 94.

ZHOU, P., YANG, X.-L., WANG, X.-G., HU, B., ZHANG, L., ZHANG, W., SI, H.-R., ZHU, Y., LI, B., HUANG, C.-L., CHEN, H.-D., CHEN, J., LUO, Y., GUO, H., JIANG, R.-D., LIU, M.-Q., CHEN, Y., SHEN, X.-R., WANG, X., ZHENG, X.-S., ZHAO, K., CHEN, Q.-J., DENG, F., LIU, L.-L., YAN, B., ZHAN, F.-X., WANG, Y.-Y., XIAO, G.-F. & SHI, Z.-L. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature,* 579**,** 270-273.

ZHU, N., ZHANG, D., WANG, W., LI, X., YANG, B., SONG, J., ZHAO, X., HUANG, B., SHI, W., LU, R., NIU, P., ZHAN, F., MA, X., WANG, D., XU, W., WU, G., GAO, G. F. & TAN, W. 2020. A Novel Coronavirus from Patients with Pneumonia in China, 2019. 382**,** 727-733.

APPENDIX 1: Biomedical Ethics Approval

UNIVERSITY OF KWAZULU-NATAL
INYUVESI YAKWAZULU-NATALI

25 April 2021

Mr Ugochukwu Jacob Anyaneji (217080488)
School of Lab Med & Medical Sc
Medical School

Dear Mr Anyaneji,

Protocol reference number: BREC/00002647/2021
Project title: Implementation of an efficient pooling strategy for high throughput diagnostic testing of Severe Acute Respiratory Syndrome Corona Virus-2.
Degree: MMedSci

EXPEDITED APPLICATION: APPROVAL LETTER

A sub-committee of the Biomedical Research Ethics Committee has considered and noted your application.

The conditions have been met and the study is given full ethics approval and may begin as from 25 April 2021. Please ensure that outstanding site permissions are obtained and forwarded to BREC for approval before commencing research at a site.

This approval is subject to national and UKZN lockdown regulations, see (http://research.ukzn.ac.za/Libraries/BREC/BREC_Lockdown_Level_1_Guidelines.sflb.ashx). Based on feedback from some sites, we urge PIs to show sensitivity and exercise appropriate consideration at sites where personnel and service users appear stressed or overloaded.

This approval is valid for one year from 25 April 2021. To ensure uninterrupted approval of this study beyond the approval expiry date, an application for recertification must be submitted to BREC on the appropriate BREC form 2-3 months before the expiry date.

Any amendments to this study, unless urgently required to ensure safety of participants, must be approved by BREC prior to implementation.

Your acceptance of this approval denotes your compliance with South African National Research Ethics Guidelines (2015), South African National Good Clinical Practice Guidelines (2006) (if applicable) and with UKZN BREC ethics requirements as contained in the UKZN BREC Terms of Reference and Standard Operating Procedures, all available at http://research.ukzn.ac.za/Research-Ethics/Biomedical-Research-Ethics.aspx.

BREC is registered with the South African National Health Research Ethics Council (REC-290408-009). BREC has US Office for Human Research Protections (OHRP) Federal-wide Assurance (FWA 678).

The sub-committee's decision will be noted by a full Committee at its next meeting taking place on 11 May 2021.

Yours sincerely,

Prof D Wassenaar
Chair: Biomedical Research Ethics Committee

Biomedical Research Ethics Committee
Chair: Professor D R Wassenaar
UKZN Research Ethics Office Westville Campus, Govan Mbeki Building
Postal Address: Private Bag X54001, Durban 4000
Email: BREC@ukzn.ac.za
Website: http://research.ukzn.ac.za/Research-Ethics/Biomedical-Research-Ethics.aspx

APPENDIX 2: Supplementary information for Chapter two; published article

Table S1. Slicing pattern for sub-pools using a 27-sample (3 dimensional) hypercube experiment

Slice 1 sub-pool	Slice 2 sub-pool	Slice 3 sub-pool	Slice 4 sub-pool	Slice 5 sub-pool	Slice 6 sub-pool	Slice 7 sub-pool	Slice 8 sub-pool	Slice 9 sub-pool
1	10	19	1	4	7	1	2	3
2	11	20	2	5	8	4	5	6
3	12	21	3	6	9	7	8	9
4	13	22	10	13	16	10	11	12
5	14	23	11	14	17	13	14	15
6	15	24	12	15	18	16	17	18
7	16	25	19	22	25	19	20	21
8	17	26	20	23	26	22	23	24
9	18	27	21	24	27	25	26	27

Table S2. Slicing pattern for sub-pools using an 81-sample (4 dimensional) hypercube experiment

Slice 1 sub-pool	Slice 2 sub-pool	Slice 3 sub-pool	Slice 4 sub-pool	Slice 5 sub-pool	Slice 6 sub-pool	Slice 7 sub-pool	Slice 8 sub-pool	Slice 9 sub-pool	Slice 10 sub-pool	Slice 11 sub-pool	Slice 12 sub-pool
1	28	55	1	10	19	1	4	7	1	2	3
2	29	56	2	11	20	2	5	8	4	5	6
3	30	57	3	12	21	3	6	9	7	8	9
4	31	58	4	13	22	10	13	16	10	11	12
5	32	59	5	14	23	11	14	17	13	14	15
6	33	60	6	15	24	12	15	18	16	17	18
7	34	61	7	16	25	19	22	25	19	20	21
8	35	62	8	17	26	20	23	26	22	23	24
9	36	63	9	18	27	21	24	27	25	26	27
10	37	64	28	37	46	28	31	34	28	29	30
11	38	65	29	38	47	29	32	35	31	32	33
12	39	66	30	39	48	30	33	36	34	35	36
13	40	67	31	40	49	37	40	43	37	38	39
14	41	68	32	41	50	38	41	44	40	41	42
15	42	69	33	42	51	39	42	45	43	44	45
16	43	70	34	43	52	46	49	52	46	47	48
17	44	71	35	44	53	47	50	53	49	50	51
18	45	72	36	45	54	48	51	54	52	53	54
19	46	73	55	64	73	55	58	61	55	56	57
20	47	74	56	65	74	56	59	62	58	59	60
21	48	75	57	66	75	57	60	63	61	62	63
22	49	76	58	67	76	64	67	70	64	65	66
23	50	77	59	68	77	65	68	71	67	68	69
24	51	78	60	69	78	66	69	72	70	71	72
25	52	79	61	70	79	73	76	79	73	74	75

26	53	80	62	71	80	74	77	80	76	77	78
27	54	81	63	72	81	75	78	81	79	80	81

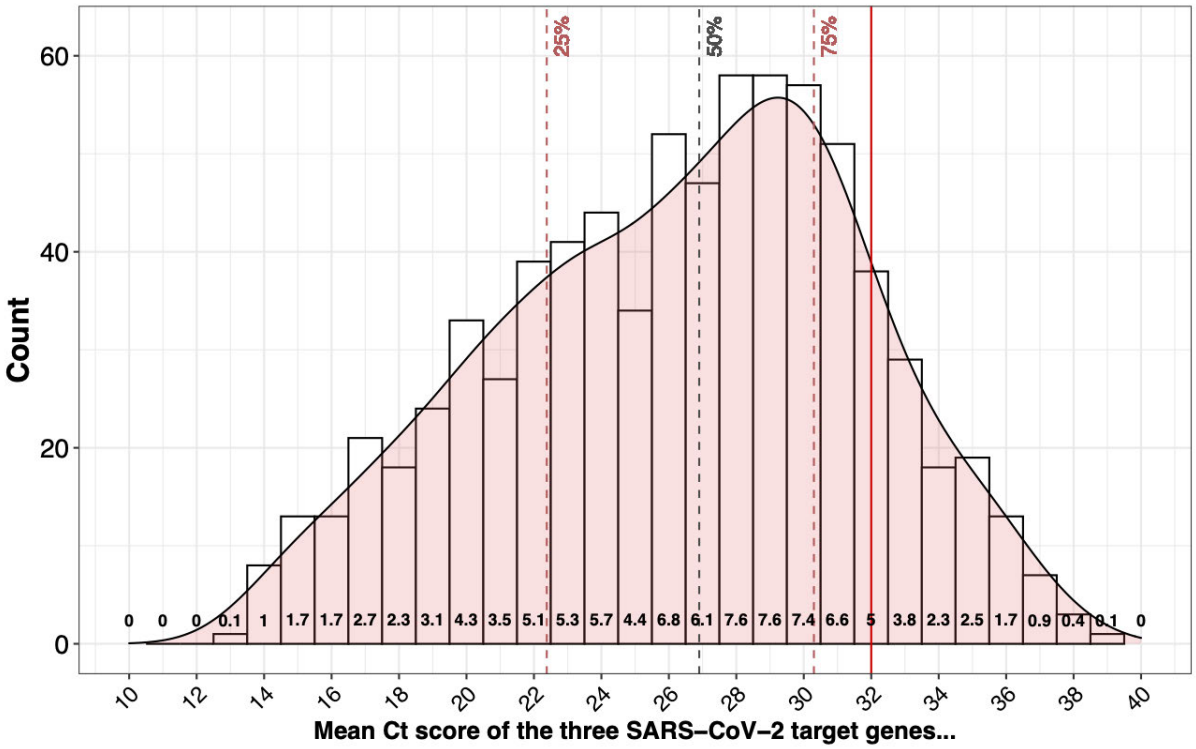


Fig. S1. Distribution of Cycle Threshold (Belouzard et al.) scores from population of samples tested in KwaZulu-Natal; the figures inside the bars represent the percentage of samples that fall within this Ct score range. The solid red vertical line represents the Ct at which sensitivity of the assay is compromised (Ct ≥ 32).

Table S3. 27-sample 3-d hypercube results

Experiment ID	**Description**	**S-gene Ct**	**N-gene Ct**	**ORF1ab Ct**	**Result**
A	27-pool sample	21.85	21.07	21.4	Positive
	9-pool slice 1	Undetermined	Undetermined	Undetermined	Negative
	9-pool slice 2	Undetermined	Undetermined	Undetermined	Negative
	9-pool slice 3	19.99	19.22	19.55	Positive
	9-pool slice 4	Undetermined	Undetermined	Undetermined	Negative
	9-pool slice 5	19.92	19.39	19.63	Positive
	9-pool slice 6	Undetermined	Undetermined	Undetermined	Negative

	9-pool slice 7	Undetermined	Undetermined	Undetermined	Negative
	9-pool slice 8	Undetermined	Undetermined	Undetermined	Negative
	9-pool slice 9	20.53	19.62	20.09	Positive
	Standard deviation	0.33	0.20	0.29	
B	27-pool sample	28.81	28.06	28.71	Positive
	9-pool slice 1	Undetermined	Undetermined	Undetermined	Negative
	9-pool slice 2	Undetermined	Undetermined	Undetermined	Negative
	9-pool slice 3	26.94	26.92	27.51	Positive
	9-pool slice 4	Undetermined	Undetermined	Undetermined	Negative
	9-pool slice 5	26.4	26.14	26.39	Positive
	9-pool slice 6	Undetermined	Undetermined	Undetermined	Negative
	9-pool slice 7	Undetermined	Undetermined	Undetermined	Negative
	9-pool slice 8	Undetermined	Undetermined	Undetermined	Negative
	9-pool slice 9	26.14	26.22	26.51	Positive
	Standard deviation	0.41	0.43	0.61	
C	27-pool sample	Undetermined	38.03	37.64	Positive
	9-pool slice 1	Undetermined	Undetermined	Undetermined	Negative
	9-pool slice 2	Undetermined	Undetermined	Undetermined	Negative
	9-pool slice 3	30.91	30.31	30.9	Positive
	9-pool slice 4	Undetermined	Undetermined	Undetermined	Negative
	9-pool slice 5	29.97	29.77	30.54	Positive
	9-pool slice 6	Undetermined	Undetermined	Undetermined	Negative
	9-pool slice 7	Undetermined	Undetermined	Undetermined	Negative
	9-pool slice 8	Undetermined	Undetermined	Undetermined	Negative
	9-pool slice 9	30.21	29.97	30.06	Positive
	Standard deviation	0.49	0.27	0.42	

Pool = 27-fold dilution; Slice = 9-fold dilution

Based on the consistent amplification of the positive sample across the respective slices, the location of the positive sample can be deduced to be in position 24 within each of the three 3-d hypercubes.

Table S4. 81- sample 4-d hypercube results

Experiment ID	Description	S-gene Ct	N-gene Ct	ORF1ab Ct	Result
A	Undiluted positive	19.9	17.62	19.16	Positive
	81-pool sample	26.33	23.92	25.95	Positive
	27-pool slice 1	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 2	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 3	24.24	22.04	23.92	Positive
	27-pool slice 4	24.52	22.35	24.55	Positive
	27-pool slice 5	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 6	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 7	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 8	24.67	22.37	24.52	Positive
	27-pool slice 9	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 10	24.25	21.94	23.92	Positive
	27-pool slice 11	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 12	Undetermined	Undetermined	Undetermined	Negative
	Standard deviation	0.21	0.22	0.36	
B	Undiluted positive	29.21	29.20	29.16	Positive
	81-pool sample	33.42	34.55	33.85	Positive
	27-pool slice 1	31.23	34.29	31.35	Positive
	27-pool slice 2	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 3	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 4	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 5	31.68	34.92	31.53	Positive
	27-pool slice 6	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 7	Undetermined	Undetermined	Undetermined	Negative

	27-pool slice 8	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 9	32.05	34.04	32.23	Positive
	27-pool slice 10	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 11	31.47	34	31.08	Positive
	27-pool slice 12	Undetermined	Undetermined	Undetermined	Negative
	Standard deviation	0.35	0.42	0.49	
C	Undiluted positive	24.64	24.91	24.41	Positive
	81-pool sample	27.82	31.13	28.86	Positive
	27-pool slice 1	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 2	25.83	29.48	27.22	Positive
	27-pool slice 3	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 4	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 5	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 6	26.86	29.47	27.17	Positive
	27-pool slice 7	26.11	29.44	27.21	Positive
	27-pool slice 8	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 9	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 10	24.64	29.34	27.25	Positive
	27-pool slice 11	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 12	Undetermined	Undetermined	Undetermined	Negative
	Standard deviation	0.92	0.06	0.03	
D	Undiluted positive	32.46	32.60	32.03	Positive
	81-pool sample	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 1	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 2	Undetermined	Undetermined	Undetermined	Negative

	27-pool slice 3	Undetermined	Undetermined	35.2	Inconclusive
	27-pool slice 4	34.82	39.37	Undetermined	Positive
	27-pool slice 5	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 6	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 7	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 8	34.5	39.93	Undetermined	Positive
	27-pool slice 9	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 10	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 11	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 12	Undetermined	Undetermined	Undetermined	Negative
	Standard deviation	0.23	0.40	NA	
E	Undiluted positive	30.7	30.35	30.6	Positive
	81-pool sample	39.59	37.67	Undetermined	Positive
	27-pool slice 1	39.9	36.56	38.46	Positive
	27-pool slice 2	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 3	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 4	38.97	35.88	39.44	Positive
	27-pool slice 5	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 6	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 7	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 8	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 9	36.52	35.7	37.61	Positive
	27-pool slice 10	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 11	35.88	36	39.03	Positive
	27-pool slice 12	Undetermined	Undetermined	Undetermined	Negative

	Standard deviation	1.92	0.37	0.79	
F	Undiluted positive	31.08	29.85	30.8	Positive
	81-pool sample	Undetermined	38.63	Undetermined	Inconclusive
	27-pool slice 1	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 2	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 3	37.33	34.55	35.78	Positive
	27-pool slice 4	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 5	38.79	34.38	37.11	Positive
	27-pool slice 6	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 7	36.86	35	36.75	Positive
	27-pool slice 8	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 9	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 10	35.88	34.88	42.48	Positive
	27-pool slice 11	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 12	Undetermined	Undetermined	Undetermined	Negative
	Standard deviation	1.21	0.29	3.02	
G	Undiluted positive	27.12	27.52	25.66	Positive
	81-pool sample	33.06	33.26	31.71	Positive
	27-pool slice 1	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 2	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 3	30.96	31.44	30.31	Positive
	27-pool slice 4	31.47	31.47	29.91	Positive
	27-pool slice 5	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 6	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 7	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 8	31.65	31.86	30.72	Positive
	27-pool slice 9	Undetermined	Undetermined	Undetermined	Negative

	27-pool slice 10	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 11	31.91	32.18	31.44	Positive
	27-pool slice 12	Undetermined	Undetermined	Undetermined	Negative
	Standard deviation	0.40	0.35	0.65	
H	Undiluted positive	33.01	34.96	32.65	Positive
	81-pool sample	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 1	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 2	34.70	36.34	Undetermined	Positive
	27-pool slice 3	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 4	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 5	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 6	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 7	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 8	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 9	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 10	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 11	Undetermined	Undetermined	Undetermined	Negative
	27-pool slice 12	Undetermined	Undetermined	Undetermined	Negative
	Standard deviation	NA	NA	NA	

Pool = 81-fold dilution; Slice = 27-fold dilution

NA=Not available

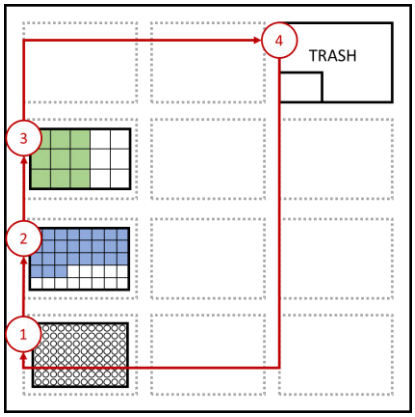


Fig. S2. Top view of the Opentrons2 liquid-handling robot and electronic pipette workflow path: position 1 – pipette tips; position 2 – source plate containing individual samples; position 3 – target plate containing sample pools; position 4 – trash compartment to discard used pipette tips

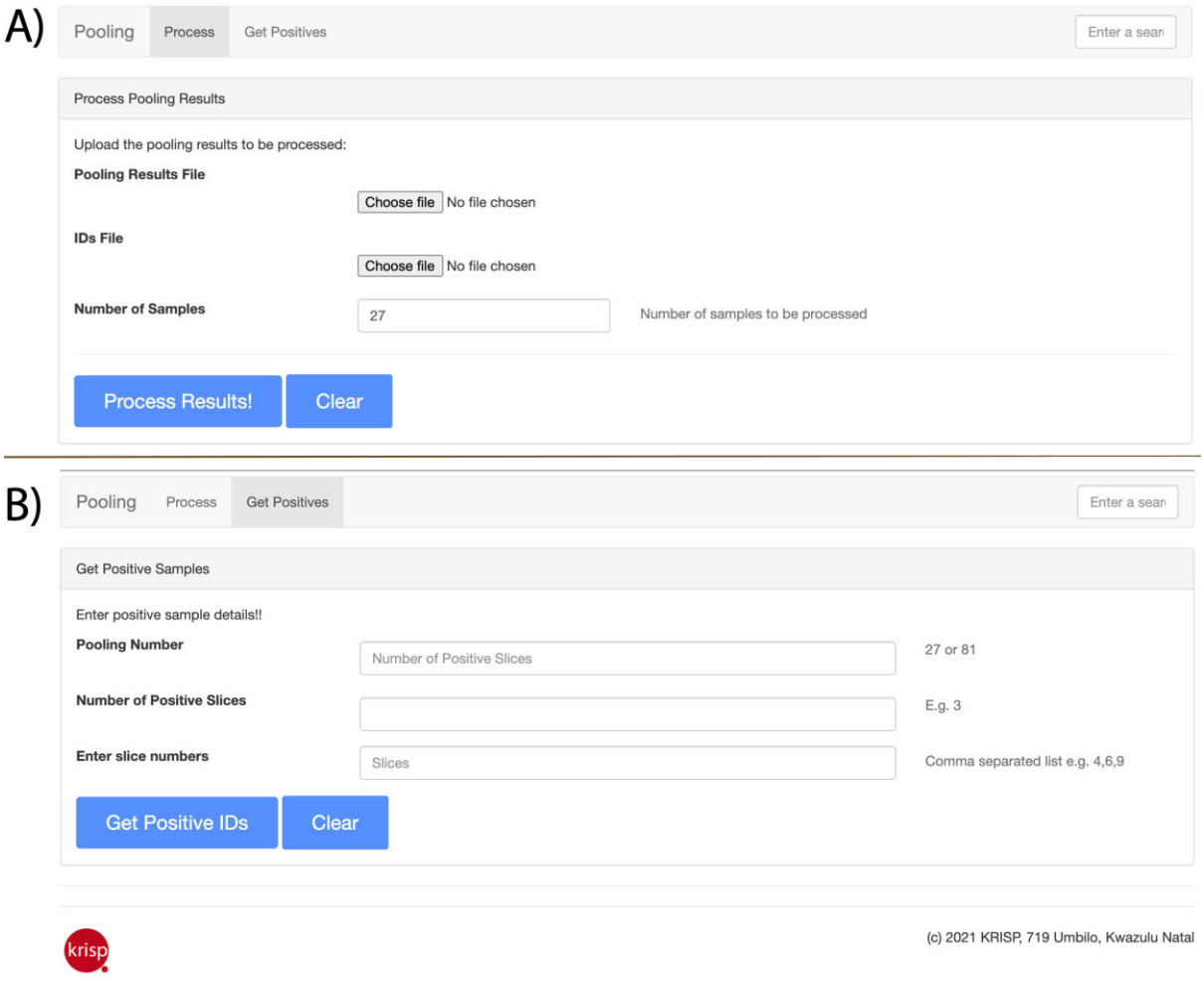


Fig. S3. Automated web application for processing pooling results. (A) Interface for inference of positive sample(s) in the pool. (B) Interface for de-bundling of pooled results into individual results.

Cost comparison of hypercube-based pooled tests vs. lateral flow antigen tests

Here we answer the following question: how much more will it cost to detect infectious individuals with a defined level of accuracy by using a lateral flow antigen (Bendavid et al. 2020) test vs. a hypercube-based (HC) pooled test with ≤81 samples per pool?

To illustrate the analytical framework that we will use, we denote by $C = (0, 40]$ the range of possible Ct values that an infected individual can have. Suppose that an individual is infectious when their Ct value is within an interval $I \subset C$. We assume that individuals are infectious when their Ct value is below 30 *(11, 20)*, so we set $I = (0, 30)$.

Now, let $\{U_i\}$, $U_i = (c_{i-1}, c_i]$, $i = 1,2,\ldots,n$, be a set of sub-intervals induced by a finite partition $\{c_i\}$ of the interval C, with $0 = c_1 < c_2 < \cdots < c_n = 40$. Let the probability that a particular testing method will produce a positive result when the Ct value is in the interval U_i be denoted p_i. Then, given an individual's Ct trajectory the probability that the individual will test positive while infectious is given by

$$p_{\text{pos}} = \sum_{i=1}^{n} \frac{|U_i^t \cap I^t|}{|I^t|} p_i, \qquad (1)$$

where $|I|$ denotes the length of the interval I, $\cap$ denotes the intersection of intervals, and I^t and U_i^t are the time intervals for which the Ct values are in the corresponding range. Equation (Covid-19 National Emergency Response Center, 2020.) follows from the assumption that an individual can be tested at any time point, chosen uniformly at random.

If the length of the infectious period is t_I days, then infectiousness will be detected with probability p_{pos} if tests are conducted every t_I days. If $p_{\text{pos}} < 1$, then it may be necessary to increase the frequency of testing to above $1/t_I$ day^{-1} in order to detect infectiousness with a sufficiently high probability.

Suppose that two testing methods A and B require testing frequencies of f_q^A and f_q^B, respectively, to achieve a detection probability q. In addition, suppose that the testing cost per individual is x_A (respectively x_B) for method A (respectively B). Then, the ratio of the cost required to achieve a detection probability q using method A vs. method B is given by

$$r = \frac{f_q^A \, x_A}{f_q^B \, x_B}. \qquad (2)$$

To compare the cost required to achieve a desired detection probability by using HC vs. LFA, we need the following information: (i) Sample Ct trajectories from infected individuals; (ii) the probability that a given method will produce a positive test as a function of the Ct value; and (iii) the testing cost per individual.

For (i), we use the posterior distribution of Ct trajectories generated by a mathematical model that was previously parametrized using experimental data *(21)*. A representative output from this model is shown in Fig. S4.

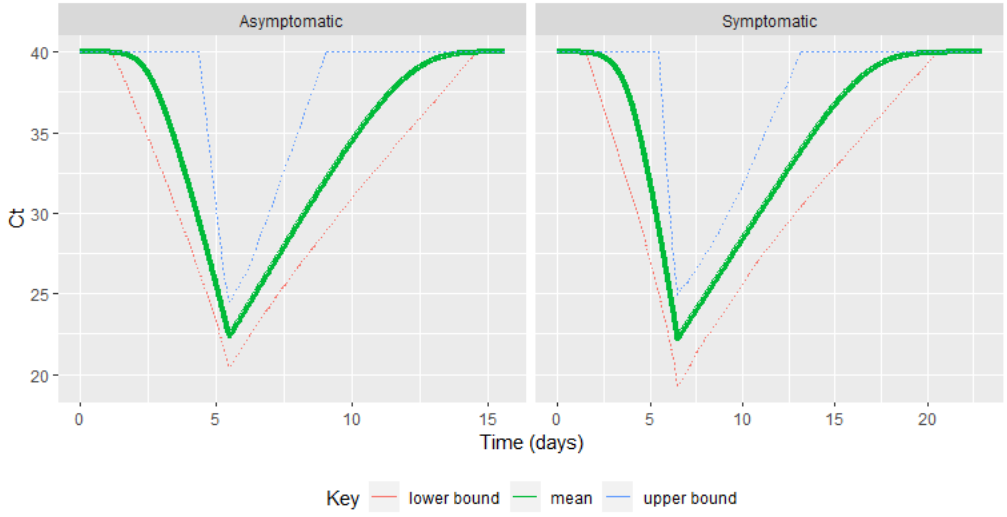


Fig. S4. Ct trajectories for asymptomatic (left panel) and symptomatic (right panel) individuals simulated using a mathematical model based on experimental data *(21)*. This model is conservative in allowing a sharp spike of infectiousness, so that the period of infectiousness is relatively brief. The mean trajectory calculated from 1000 simulations is shown together with envelopes (dotted lines) showing lower and upper 95% credible bounds for these trajectories at each time.

In addition, for (ii) and (iii), we make the following assumptions:

1) Based on data collected in our lab (see main text) and elsewhere *(21)*, for simplicity we assume that an HC test that dilutes positive samples by 81-fold produces a positive result 100% of the time when the positive sample's Ct value is below 30. For LFA, we use the following sensitivity estimates *(20)*, which are optimistic in the light of published data *(22, 23)*: LFA yields a positive result 95% of the time when the Ct value is less than 27, and 65% of the time when the Ct value is between 27 and 30.

2) We assume that LFA costs USD 5 per individual *(24)*. The cost of each PCR test is about USD 56. The cost of HC is calculated as USD 56 times the expected number of tests per individual, which depends on the viral prevalence. At a viral prevalence that is just small enough to allow 81 samples to be pooled (i.e. a prevalence of 0.35/81 = 0.43%) the cost is USD 3.4. The cost per individual is lower at viral prevalences less than 0.43%.

Using the calculated posterior distribution of Ct trajectories (Fig. 5) *(21)*; we find that the mean length of the infectious period is 4.16±0.75 days for asymptomatic individuals and 5.8±1.03 days for symptomatic individuals. In addition, using Equation (Covid-19 National Emergency Response Center et al.), we estimate that HC testing will detect 100% of infectious individuals if applied once during the infectious period, an average of once every 4 days for asymptomatics and once every 6 days for symptomatics. In contrast, LFA will detect only 83% of infectious individuals under the same conditions. We estimate that, in order to detect infection with a probability of 99.9%, an asymptomatic infectious individual will have to be tested with LFA up to 4 times during the infectious period, *i.e.*, daily. This is equivalent to the testing frequency calculated using the following simple formula based on the assumption that test outcomes are Bernoulli random variables:

$$f_q^{\mathrm{LFA}} = \frac{\log(1-q)}{t_I \log(1-p_{\mathrm{pos}})}\ \mathrm{day}^{-1}, \qquad (3)$$

where q = 0.999, p_{pos} = 0.83, and t_1 = 4 days.

Therefore, using Equation (2), we estimate that, at a viral prevalence of 0.43%, it will cost 6 times more to detect infectious individuals with a 99.9% probability using LFA compared to using HC. The cost

advantage of HC relative to LFA increases as the viral prevalence falls. Note that in this comparison we have neglected the possibility of detecting infection prior to the Ct score rising above 30. Since HC is more sensitive than LFA, it is more likely to succeed in detecting infection early, adding to its relative advantage.

Supplementary Information References

1. Singanayagam, A. et al. Duration of infectiousness and correlation with RT-PCR cycle threshold values in cases of COVID-19, England, January to May 2020. *Euro Surveill* **25,** 2001483 (2020).

2. Quilty, B.J. et al. Quarantine and testing strategies in contact tracing for SARS-CoV-2: a modelling study. *Lancet Public Health* **6**, e175-e183 (2021).

3. Kissler, S.M. et al. SARS-CoV-2 viral dynamics in acute infections. *MedRxiv* doi: https://doi.org/10.1101/2020.10.21.20217042 (2020).

4. University of Oxford. Preliminary report from the Joint PHE Porton Down & University of Oxford SARS-CoV-2 test development and validation cell: Rapid evaluation of Lateral Flow Viral Antigen detection devices (LFDs) for mass community testing (8 November 2020). https://www.ox.ac.uk/sites/files/oxford/media_wysiwyg/UK%20evaluation_PHE%20Porton%20Down%20%20University%20of%20Oxford_final.pdf

5. University of Liverpool. Liverpool Community Testing Pilot Interim Evaluation Report (23 December 2020). https://www.liverpool.ac.uk/media/livacuk/coronavirus/Liverpool,Community,Testing,Pilot,Interim,Evaluation.pdf

6. Guglielmi, G. Rapid coronavirus tests: a guide for the perplexed. *Nature* **590**, 203 – 205 (2021).

APPENDIX 3: Cover page for published papers

Publication 1

www.nature.com/scientificreports

scientific reports

Check for updates

OPEN

Implementation of an efficient SARS-CoV-2 specimen poolingstrategy for high throughput diagnostic testing

Lavanya Singh[1,7]✉, Ugochukwu J. Anyaneji[1,7], Wilfred Ndifon[2]✉, Neil Turok[3], Stacey A. Mattison[1], Richard Lessells[1], Ilya Sinayskiy[4,5], Emmanuel J. San[1], Houriiyah Tegally[1], Shaun Barnett[6], Trevor Lorimer[6], Francesco Petruccione[4,5] & Tulio de Oliveira[1]✉

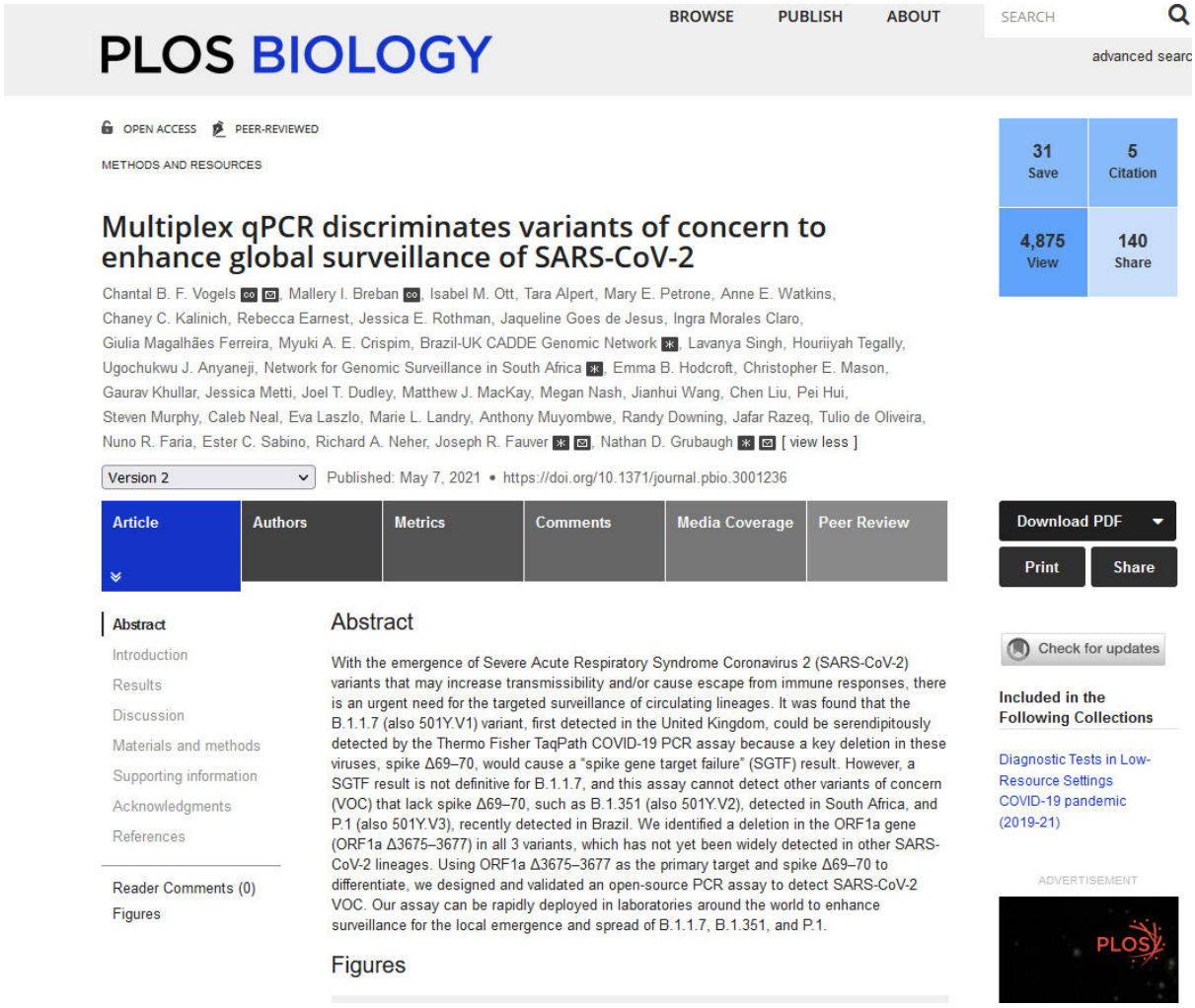
The rapid identification and isolation of infected individuals remains a key strategy for controlling the spread of SARS-CoV-2. Frequent testing of populations to detect infection early in asymptomatic or presymptomatic individuals can be a powerful tool for intercepting transmission, especially when the viral prevalence is low. However, RT-PCR testing—the gold standard of SARS-CoV-2 diagnosis—is expensive, making regular testing of every individual unfeasible. Sample pooling is one approach to lowering costs. By combining samples and testing them in groups the number of tests required is reduced, substantially lowering costs. Here we report on the implementation of pooling strategies using 3-d and 4-d hypercubes to test a professional sports team in South Africa. We have shown that infected samples can be reliably detected in groups of 27 and 81, with minimal loss of assay sensitivity for samples with individual Ct values of up to 32. We report on the automation of sample pooling, using a liquid-handling robot and an automated web interface to identify positive samples. We conclude that hypercube pooling allows for the reliable RT-PCR detection of SARS-CoV-2 infection, at significantly lower costs than lateral flow antigen (LFA) tests.

A novel coronavirus, SARS-CoV-2, emerged at the end of 2019 in the city of Wuhan, China. The highly transmissible nature of SARS-CoV-2[1] has resulted in a pandemic which continues to persist. South Africa (SA) and other African countries are currently facing a resurgence or "third wave" of infections, which in some countries is more severe than previously experienced[2–4]. Although there are approved vaccines against SARS-CoV-2, these are not yet available in amounts sufficient to control the pandemic. To make matters worse, there are new SARS-CoV-2 variants (which are typically identified using genomics surveillance tools[5,6]) that have been identified in SA[7], which may be associated with higher transmissibility and hence a more rapid spread of the virus. There is therefore an urgent need for more efficient population screening and the isolation of infected individuals to reduce the transmission of SARS-CoV-2.

Presymptomatic or asymptomatic individuals, who are infectious viral carriers[8], are the hidden drivers of the pandemic. They represent an estimated proportion ranging from 18 to 81% of infections[9], thereby posing a major challenge to the containment of SARS-CoV-2. If such individuals can be efficiently detected through frequent, repeated population testing at scale and thereby enabled to isolate before they infect others, the spread of the virus can be prevented[8]. Therefore, efficient and affordable, high throughput SARS-CoV-2 testing is highly desirable as a means of controlling the pandemic.

[1]KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), School of Laboratory Medicine & Medical Sciences, University of KwaZulu-Natal, Durban, South Africa. [2]African Institute for Mathematical Sciences, The Next Einstein Initiative, Kigali, Rwanda. [3]Higgs Centre for Theoretical Physics, School of Physics and Astronomy, University of Edinburgh, Edinburgh, UK. [4]School of Chemistry and Physics, University of Kwa-Zulu Natal, Westville, South Africa. [5]National Institute for Theoretical and Computational Sciences (NITheCS), KwaZulu-Natal, South Africa. [6]Discipline of Electrical, Electronic and Computer Engineering, University of KwaZulu-Natal, Durban, South Africa. [7]These authors contributed equally: Lavanya Singh and Ugochukwu J. Anyaneji. ✉email: singhl@ukzn.ac.za; wndifon@nexteinstein.org; tuliodna@gmail.com

Publication 2



RESEARCH ARTICLE

A year of genomic surveillance reveals how the SARS-CoV-2 pandemic unfolded in Africa

EDUAN WILKINSON, MARTA GIOVANETTI, HOURIIYAH TEGALLY, JAMES E. SAN, RICHARD LESSELLS, DIEGO CUADROS, DARREN P. MARTIN, DAVID A. RASMUSSEN, ABDEL-RAHMAN N. ZEKRI, [...] TULIO DE OLIVEIRA +293 authors Authors Info & Affiliations

SCIENCE • 9 Sep 2021 • First Release • DOI: 10.1126/science.abj4336

Abstract

The progression of the SARS-CoV-2 pandemic in Africa has so far been heterogeneous and the full impact is not yet well understood. Here, we describe the genomic epidemiology using a dataset of 8746 genomes from 33 African countries and two overseas territories. We show that the epidemics in most countries were initiated by importations predominantly from Europe, which diminished following the early introduction of international travel restrictions. As the pandemic progressed, ongoing transmission in many countries and increasing mobility led to the emergence and spread within the continent of many variants of concern and interest,

MSc Dissertation-corrected as per examiners comments

Final Audit Report 2022-01-19

Created:	2022-01-19
By:	Cheryl Baxter (cheryl.baxter@caprisa.org)
Status:	Signed
Transaction ID:	CBJCHBCAABAAiRydH6o048ZhpurU7KLHJPtjHFdEGCkb

"MSc Dissertation-corrected as per examiners comments" Histor y

Document digitally presigned by DocuSign\, Inc. (enterprisesupport@docusign.com)
2022-01-19 - 10:01:32 AM GMT- IP address: 146.232.12.180

Document created by Cheryl Baxter (cheryl.baxter@caprisa.org)
2022-01-19 - 12:23:12 PM GMT- IP address: 146.232.12.180

Document emailed to Tulio de Oliveira (cbaxter@sun.ac.za) for signature
2022-01-19 - 12:24:26 PM GMT

Email viewed by Tulio de Oliveira (cbaxter@sun.ac.za)
2022-01-19 - 12:24:41 PM GMT- IP address: 104.47.9.254

Document e-signed by Tulio de Oliveira (cbaxter@sun.ac.za)
Signature Date: 2022-01-19 - 12:25:19 PM GMT - Time Source: server- IP address: 146.232.12.180

Document emailed to Francesco Petruccione (petruccione@ukzn.ac.za) for signature
2022-01-19 - 12:25:21 PM GMT

Email viewed by Francesco Petruccione (petruccione@ukzn.ac.za)
2022-01-19 - 1:12:56 PM GMT- IP address: 172.224.231.128

Document e-signed by Francesco Petruccione (petruccione@ukzn.ac.za)
Signature Date: 2022-01-19 - 1:15:03 PM GMT - Time Source: server- IP address: 105.226.192.64

Agreement completed.
2022-01-19 - 1:15:03 PM GMT