

# BMJ Open Implementation of serological and molecular tools to inform COVID-19 patient management: protocol for the GENCOV prospective cohort study

Jennifer Taher,<sup>1,2</sup> Chloe Mighton,<sup>2,3,4,5</sup> Sunakshi Chowdhary,<sup>2,3</sup> Selina Casalino,<sup>2,3</sup> Erika Frangione,<sup>2,3</sup> Saranya Arnoldo,<sup>1,6</sup> Erin Bearss,<sup>7,8</sup> Alexandra Binnie,<sup>6</sup> Yvonne Bombard ,<sup>4,5</sup> Bjug Borgundvaag,<sup>8,9</sup> Howard Chertkow,<sup>10</sup> Marc Clausen,<sup>4</sup> Luke Devine,<sup>11,12</sup> Hanna Faghfoury,<sup>13</sup> Steven Marc Friedman ,<sup>8,14</sup> Anne-Claude Gingras,<sup>2,3,15</sup> Zeeshan Khan,<sup>16</sup> Tony Mazzulli,<sup>1,17</sup> Allison McGeer,<sup>1,3,5,17</sup> Shelley L McLeod,<sup>8,18</sup> Trevor J Pugh,<sup>19,20,21</sup> David Richardson,<sup>6</sup> Jared Simpson,<sup>21,22</sup> Seth Stern,<sup>16</sup> Lisa Strug,<sup>23,24</sup> Ahmed Taher,<sup>14,16,25</sup> Jordan Lerner-Ellis  <sup>1,2,3</sup>

**To cite:** Taher J, Mighton C, Chowdhary S, *et al*. Implementation of serological and molecular tools to inform COVID-19 patient management: protocol for the GENCOV prospective cohort study. *BMJ Open* 2021;**11**:e052842. doi:10.1136/bmjopen-2021-052842

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2021-052842>).

Received 26 April 2021  
Accepted 24 August 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Dr Jennifer Taher;  
Jennifer.Taher@sinahealth.ca and  
Dr Jordan Lerner-Ellis;  
jordan.lerner-ellis@sinahealth.ca

## ABSTRACT

**Introduction** There is considerable variability in symptoms and severity of COVID-19 among patients infected by the SARS-CoV-2 virus. Linking host and virus genome sequence information to antibody response and biological information may identify patient or viral characteristics associated with poor and favourable outcomes. This study aims to (1) identify characteristics of the antibody response that result in maintained immune response and better outcomes, (2) determine the impact of genetic differences on infection severity and immune response, (3) determine the impact of viral lineage on antibody response and patient outcomes and (4) evaluate patient-reported outcomes of receiving host genome, antibody and viral lineage results.

**Methods and analysis** A prospective, observational cohort study is being conducted among adult patients with COVID-19 in the Greater Toronto Area. Blood samples are collected at baseline (during infection) and 1, 6 and 12 months after diagnosis. Serial antibody titres, isotype, antigen target and viral neutralisation will be assessed. Clinical data will be collected from chart reviews and patient surveys. Host genomes and T-cell and B-cell receptors will be sequenced. Viral genomes will be sequenced to identify viral lineage. Regression models will be used to test associations between antibody response, physiological response, genetic markers and patient outcomes. Pathogenic genomic variants related to disease severity, or negative outcomes will be identified and genome wide association will be conducted. Immune repertoire diversity during infection will be correlated with severity of COVID-19 symptoms and human leucocyte antigen-type associated with SARS-CoV-2 infection. Participants can learn their genome sequencing, antibody and viral sequencing results; patient-reported outcomes of receiving this information will be assessed through surveys and qualitative interviews.

**Ethics and dissemination** This study was approved by Clinical Trials Ontario Streamlined Ethics Review System (CTO Project ID: 3302) and the research ethics boards at

## Strengths and limitations of this study

- This study will link serological, genomic and patient characteristics to provide a comprehensive understanding of factors that contribute to variability in clinical symptoms and outcomes among patients with COVID-19.
- Data will be generated using multiple methodologies, including multiple serological assays, host genome sequencing, T-cell and B-cell receptor sequencing and viral genome sequencing in order to provide real-time genetic and immunological risk factor information needed for the prevention, treatment and management of patients with COVID-19 disease.
- We will broadly share study data to enhance international collaborative efforts aimed at mitigating the spread of COVID-19.
- Use of a surrogate neutralisation ELISA will allow us to identify which antibodies have neutralising ability, which could aid in the selection, development and implementation of appropriate serology immunoassays for detection of patients that have and maintain viral neutralising ability.
- A limitation is that patients who do not return to provide convalescent samples could limit our ability to evaluate trends in immune response over time, and may introduce attrition bias; follow-up calls will be made to patients to help increase rate of return for convalescent samples.

participating hospitals. Study findings will be disseminated through peer-reviewed publications, conference presentations and end-users.

## INTRODUCTION

SARS-CoV-2 causes COVID-19, which spread rapidly to become a global pandemic.<sup>1</sup>

There is considerable variability in symptom severity and outcomes among patients infected by SARS-CoV-2.<sup>2</sup> While some infected individuals are asymptomatic or experience only mild symptoms, others have severe symptoms requiring hospitalisation.<sup>3</sup> Known risk factors include age and pre-existing comorbidities,<sup>4</sup> however, there are likely additional risk factors that have yet to be characterised, including immunity, host genetics or viral lineage.

Serological antibody testing can identify individuals with active COVID-19 and those who have previously been infected.<sup>5</sup> However, the presence of antibodies does not necessarily indicate immunity<sup>6</sup> as some patients produce antibodies that do not neutralise the virus. However, it is also possible to have protection without neutralisation, such as through Fc-mediated complement activation, antibody-dependent cell-mediated cytotoxicity and antibody-dependent cellular phagocytosis. Serology assays primarily target two of the virus's four main structural proteins; the spike protein (containing the receptor-binding domain which may be targeted directly) and the nucleocapsid protein. Patients show variable immune response to COVID-19,<sup>6</sup> which has been shown correlate with disease severity. The variable immune response observed in patients can include differences in antibody titres, isotype, antigen target and viral neutralisation. The adaptive immune system responds to infection through a process of gene translocation or gene-shuffling to produce antibodies against antigens. Molecular profiling of T-and B-cell receptor (TCR/BCR) dynamics over time can provide a comprehensive examination of immune response that cannot be determined from serological findings (antigenic epitopes) alone.<sup>7</sup> Cataloguing TCR/BCR repertoire among patients with COVID-19 could inform diagnostics and vaccine development,<sup>8</sup> and monitoring T-cell response is important as a correlate of immunity.<sup>7</sup>

The host genome may also affect susceptibility to COVID-19 and severity of infection,<sup>9 10</sup> however, the relationship between common and rare genetic variation, COVID-19 antibody production, developed immunity and patient outcomes (eg, respiratory failure, kidney failure, death) has yet to be determined. Genetic factors are known to contribute to differences in response to other viral pathogens.<sup>11-13</sup> For example, human leucocyte antigen (HLA) haplotypes have been associated with susceptibility and severity of infectious diseases including HIV, tuberculosis, hepatitis B, influenza, dengue and SARS-CoV-2.<sup>14-17</sup> Several studies have identified familial clustering of severe COVID-19, further suggesting that there may be a hereditary basis to severe outcomes.<sup>9 18 19</sup> Additionally, there are genetic disorders which may affect COVID-19 outcomes<sup>20</sup> including conditions that predispose to thrombotic crises (eg, hereditary thrombophilia) or cardiopulmonary complications (eg, cystic fibrosis) that may be induced by severe illness.<sup>20</sup>

Over time, new variants of the SARS-CoV-2 virus have emerged, including lineage B.1.1.7 first identified in England,<sup>21</sup> B.1.351 first identified in South Africa and P.1 first identified in Brazil. New lineages are associated with

increased transmission, disease severity and mortality.<sup>22</sup> Correlating viral genome data to infection and severity has important implications for the development of vaccines and therapies and managing the response to SARS-CoV-2.<sup>23</sup>

Variable physiological responses in SARS-CoV-2-infected patients can be identified clinically through symptom monitoring, or through biochemistry and haematology laboratory testing. There are now recommendations to measure specific laboratory testing profiles and efforts to identify symptoms and other related clinical conditions<sup>2 24-26</sup> to better characterise physiological responses that lead to poor or favourable outcomes.

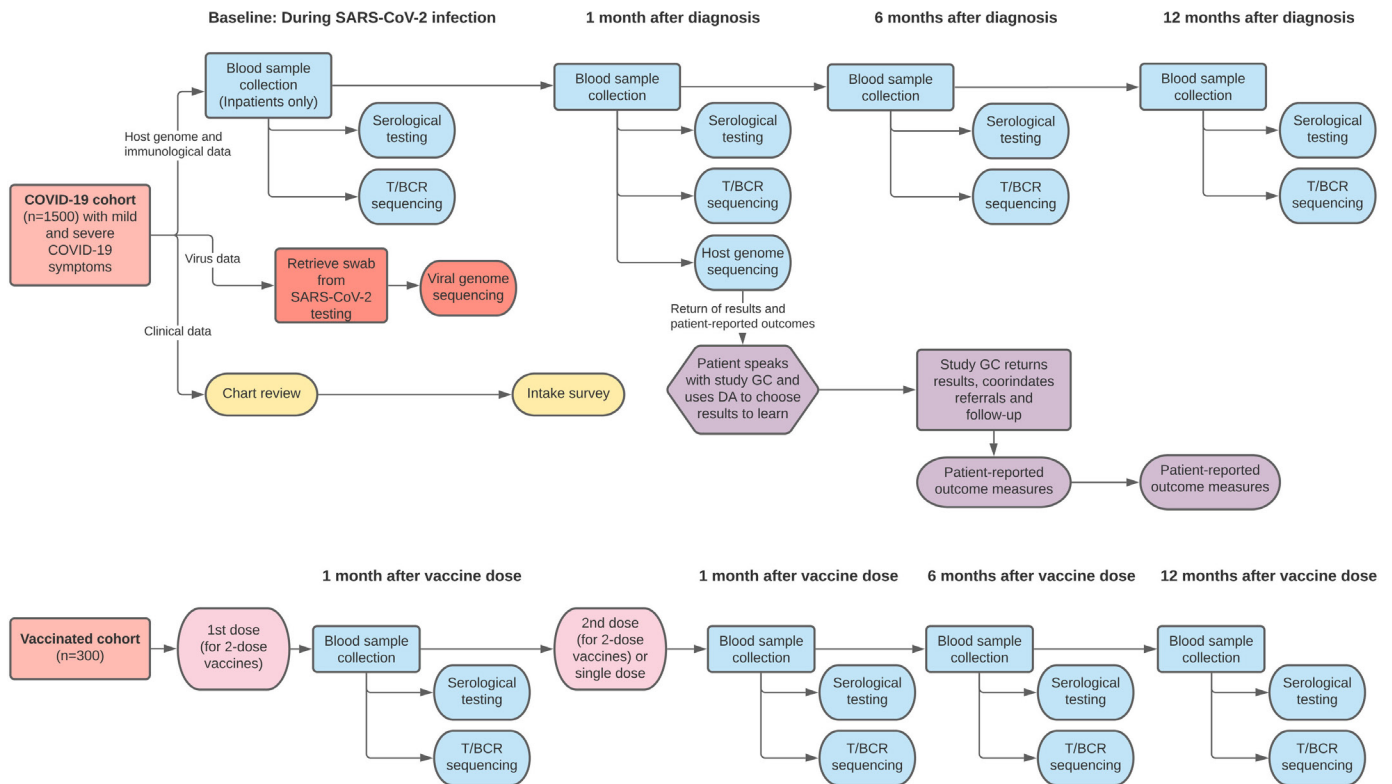
In the context of COVID-19 research and clinical care, biomarkers that are used to assess COVID-19 may be returned to patients (eg, antibody results, genome sequencing (GS) results, viral lineage). For example, host GS to identify genetic markers associated with SARS-CoV-2 susceptibility or severity may reveal information about inherited predispositions to multiple other diseases. It is recommended that medically actionable genomic results (results associated with established treatments or preventive strategies) should be offered to patients undergoing clinical GS<sup>27</sup> and additional results be offered based on patients' preferences. Recent policy supports the return of clinically actionable genomic results to research participants.<sup>28</sup> Previous research has found that participants value learning a broad range of genomic results, beyond results that are clinically actionable.<sup>29-32</sup> It is unknown how returning this information to patients with COVID-19 will impact patients' well-being and behaviour, and ultimately how these factors may impact the healthcare system (eg, health service utilisation). While previous studies have assessed return of results among patient populations affected by hereditary conditions,<sup>33 34</sup> less work has addressed return of results in the general population.

In summary, outcomes among patients with COVID-19 may be influenced by differences in short-term and long-term immune response, acute physiological response to infection, host genetic variation and viral lineage. Returning information on biomarkers associated with COVID-19 to patients may impact patients' well-being, behaviour and healthcare service use. This prospective cohort study aims to (1) identify the characteristics of the antibody and TCR/BCR response that result in maintained immune response and better patient outcomes, (2) identify host genetic differences that impact COVID-19 infection severity and immune response, (3) assess the impact of viral lineage on antibody response and patient outcomes and (4) evaluate patient-reported outcomes of receiving host genome, antibody and viral lineage results.

## METHODS

### Study design overview

A prospective cohort study is being conducted at six hospitals in the Greater Toronto Area in Ontario, Canada, with a target recruitment of 1500 positive patients with



**Figure 1** Study flow diagram with data collection time points. DA, decision aid; GC, genetic counsellor; T/BCR, T/B cell receptor.

COVID-19. Enrolment began in October 2020, and is anticipated to be completed by November 2021; data collection is expected to be completed by November 2022. Participants are consented for blood draws at baseline (for inpatients), and at 1, 6 and 12 months since PCR positive date (figure 1). Antibody isotype, titre, antigen targets and viral neutralisation will be assessed at all time points alongside TCR/BCR sequencing. Host and virus genomes will be sequenced. Chart review and surveys will be performed to obtain patient characteristics (age, sex, ancestry, symptoms, outcome, comorbidities, treatment) and biological response via laboratory test results. Immune response, genetic variation, viral variation, biochemical response and patient characteristics will be correlated with clinical outcomes. Participants will have the option to learn their own results from host GS, antibody testing and/or viral lineage results; we will assess patient-reported outcomes of receiving this information.

### Study setting

The primary study site is Mount Sinai Hospital, part of Sinai Health (SH) in Toronto, Ontario, Canada. Patients with COVID-19 seen at SH, Mackenzie Health, University Health Network and William Osler Health System will be recruited.

### Patient population and recruitment

Over 1 year, 1500 patients will be enrolled in the study across four health systems, comprising six individual hospitals. We will recruit patients admitted to hospital (~300

patients) as well as those with mild or no symptoms that test positive for COVID-19 seen in the emergency department or COVID-19 assessment centres (~1200 patients). The inpatient cohort will also include deceased patients (~100–150), from whom samples will be obtained retrospectively or prospectively from participating GENCOV study sites. Inclusion criteria are as follows: age 18 years or older, and have a positive result from a COVID-19 nasopharyngeal, nasal or oral swab taken at one of the participating sites. Patients who have received COVID-19 vaccination or are vaccinated during the course of the study are still eligible to participate. A population that has not had COVID-19 but has received a Health Canada approved SARS-CoV-2 vaccine will be enrolled as a control group to compare immunological responses (antibody response and TCR/BCR repertoire) to patients who have had COVID-19. As there are currently only four Health Canada approved vaccines, and three in widespread use, we will recruit ~300 patients in this cohort with ~100 from each vaccine.

### Sample collection

For participants in the COVID-19 cohort, blood samples will be collected at 1, 6 and 12 months post-COVID-19 diagnosis. For COVID-19 inpatients, blood from routine in-hospital testing will also be collected, if available, or a new blood sample will be drawn if the patient is in-hospital and within 14 days of a COVID-19 diagnosis. At each time point, 10 mL of blood per tube (two tubes: one EDTA and



**Table 1** Sample collection overview for COVID-19 cohort

Activity	Sample type	Sample requirements	Sample collection time points
Serology	Blood (Li-plasma or serum)	500 µL	Baseline, 1 month, 6 months, 1 year
Host genome sequencing	Blood or DNA	Blood: 2–5 mL blood needed at minimum for DNA extraction . DNA: 1 µg DNA minimum. 1–2 ug of DNA. Minimum volume of 20 µL. Minimum concentration of 20 ng/µL	1 month
T/B cell receptor sequencing	Buffy coat, blood or DNA, genomic DNA library	Blood: 10 mL preferred (2 mL acceptable) DNA: 500 ng library accepted; Buffy coat: derived from 2 to 10 mL of blood	Baseline, 1 month, 6 months, 1 year
Viral genome sequencing	Nasopharyngeal, nasal or oral swab		Baseline

one Li-Heparin) will be collected (table 1). COVID-19 viral swabs will be retrieved to isolate and sequence the virus. For the vaccine cohort, a sample will be taken at 1 month following the first dose and at 1, 6 and 12 months following the second dose. For vaccines requiring only one dose, samples will be taken at 1, 6 and 12 months.

### **Aim 1: identify the characteristics of the antibody response that result in maintained immune response and better patient outcomes**

#### **Assessing antibody levels**

Total antibody levels will be assessed on two Health Canada approved Roche immunoassays: (1) the Elecsys Anti-SARS-CoV-2 qualitative assay targeting the nucleocapsid protein and (2) the Elecsys Anti-SARS-CoV-2-S quantitative assay targeting the spike protein. Both have undergone a method evaluation (assessment of the assay and approval for clinical use) in the core biochemistry laboratory at SH. Serology reports from the assays accompanied by a summary letter will be released back to the patient by registered mail or email, depending on participant preference (online supplemental appendix 1). In the COVID-19 cohort, when the two results do not match, the results will be considered inconclusive. This can occur due to borderline samples, potential false positive or false negative results on one platform, or the possibility of having antibodies with stronger affinity to spike vs nucleocapsid antigens. In the vaccinated cohort, the presence of spike antibodies will be reported as a positive result, as all Health Canada approved vaccines currently generate antibodies to spike but not nucleocapsid proteins. Therefore, discordance between the two assays would be expected and not considered inconclusive in this cohort.

Further antibody characterisation (isotype, relative levels and antigen target) will be performed at the Lunenfeld-Tanenbaum Research Institute on a high throughput, research developed, automated ELISA. The assay has been validated on samples from convalescent and active patients, including panels obtained through the National Microbiology Laboratory, the Canadian Blood Services and the Toronto Invasive Bacterial

Diseases Network<sup>35</sup>; assays were recently standardised with the National Research Council of Canada. In parallel, a surrogate neutralisation ELISA has been developed that evaluates the inhibition of the spike-ACE2 interaction,<sup>36</sup> providing a scalable assay for neutralising antibodies. The results of this assay correlate well with those of virus-based assays, including plaque reduction neutralisation titre assays and pseudotyped lentiviral assays, on samples from the Canadian Blood Services.

#### **Chart extraction and intake questionnaires**

Clinical data including COVID-19 symptoms and comorbidities will be obtained for the COVID-19 cohort through chart review and questionnaires. Participants' medical chart information will be accessed from the recruiting hospital and any clinics to which referrals were made. At each site, study personnel will extract data from charts using a standardised data extraction sheet (online supplemental file 2). Data will be collected on the date of PCR positivity, date of symptom onset, symptom severity (based on triage data, vital signs, chief complaint and ward admitted into), whether supportive care was required (eg, intensive care unit (ICU) admission, ventilation, oxygen therapy), treatment (eg, ACE inhibitors, interleukin 6 inhibitors, antivirals), comorbidities and outcome (discharge, death), as recommended by the International Severe Acute Respiratory and emerging Infections Consortium.<sup>37 38</sup> Data on viral load as indicated by Ct values from the PCR instrument will be obtained from the microbiology laboratory where PCR testing was conducted. For inpatients, we also will assess a panel of 10 laboratory tests at baseline to help define patient physiological response as recommended by the International Federation of Clinical Chemistry<sup>26</sup> and others (table 2).<sup>24 25</sup> For the deceased cohort, no personally identifying health information will be provided to the study team, and data will only be collected on age, sex and ancestry.

Intake questionnaires will be administered to participants online, and will assess patients' characteristics including age, sex assigned at birth, self-reported

**Table 2** Clinical data points and collection methods

	Chart review	Baseline (Intake questionnaires)	Questionnaires 6 months after return of results
Underlying conditions Hypertension, chronic cardiac disease, coronary artery disease, congestive heart failure, obesity, chronic metabolic disease, chronic lung disease, chronic neurologic disease, dementia, renal disease, chronic immunosuppressive disease, liver disease, blood disorder, autoimmune/rheumatological disorder, chronic haematological/blood disorder, liver disease, pregnancy, endocrine disorder, cardiovascular disorder, cancer/malignant neoplasm, gastrointestinal disorder, lipid disorders, viral infection, tuberculosis.	x	x	
COVID-19 symptoms and duration Cough, fever, sore throat, runny nose wheezing, shortness of breath, lower chest wall in-drawing, chest pain, conjunctivitis, lymphadenopathy, headache, loss of smell (anosmia), loss of taste (ageusia), seizures, fatigue/malaise, anorexia, altered consciousness/confusion, muscle aches (myalgia), joint pain (arthralgia), inability to walk, abdominal pain, diarrhoea, vomiting/nausea, skin rash, bleeding, other symptoms.	x	x	
COVID-19 complications Pneumonia, acute respiratory distress syndrome, pneumothorax, pleural effusion, bronchiolitis, seizure, stroke, congestive heart failure, heart infection, cardiac arrhythmia, cardiac ischaemia, cardiac arrest, bacteraemia, coagulation clotting disorder, anaemia, rhabdomyolysis/myositis, kidney failure/injury, gastrointestinal haemorrhage, pancreas infection, liver injury/failure, hyperglycaemic, hypoglycaemic, other.	x	x	
Procedures performed during hospitalisation Non-invasive ventilation (eg, BIPAP, CPAP), invasive ventilation, extra corporeal life support, high-flow nasal canula oxygen therapy, dialysis/haemofiltration, medications (eg, dopamine, epinephrine, vasopressin, neuromuscular blocking agents, antibiotics), tracheostomy, nitric oxide inhalation, prone positioning, other intervention or procedure.	x	x	
Tests performed during hospitalisation Chest CT, chest X-ray, ECG, ultrasound, other tests	x	x	
Laboratory tests to assess physiological response Haemoglobin, white cell count, lymphocyte count, neutrophil count, haematocrit, platelets, blood gases, activated partial thromboplastin time ratio, prothrombin time, ALT/SGPT, total bilirubin, AST/SGOT, procalcitonin, BNP, NT-pro-BNP, C reactive protein, ferritin, cytokine IL-6, glucose, lactate, sodium, potassium, albumin, LDH, creatinine, urea, cardiac troponin I, D-dimer, fibrinogen	x		
Risk factors Smoking/vaping, diet, BMI	x	x	
COVID-19 vaccination Vaccination status, vaccine type, number of doses, side effects.		x	x

ALT, alanine aminotransferase; AST, aspartate aminotransferase; BIPAP, bilevel positive airway pressure; BMI, body mass index; BNP, B-type natriuretic peptide; CPAP, continuous positive airway pressure; LDH, lactate dehydrogenase; NT-pro-BNP, N-terminal pro b-type natriuretic peptide; SGOT, serum glutamic-oxaloacetic transaminase; SGPT, serum glutamic-pyruvic transaminase.

ethnicity/ancestry, clinical conditions, risk factors (eg, smoking, body mass index), as well as COVID-19 symptoms and complications. Clinical data points are summarised in [table 2](#), and the full questionnaire is available in the supplemental materials (online supplemental file 3).<sup>2</sup>

### Statistical analysis

Analyses will be hypothesis-generating and exploratory; a sample size of 1500 is sufficient for hypothesis-generation. Appropriate regression models (eg, linear regression for continuous outcomes and logistic regression for dichotomous outcomes) will be used to test associations between antibody response, physiological response (laboratory and clinical characteristics) and patient outcome

(eg, severity of COVID-19 disease), adjusting for patient characteristics (eg, age, ancestry, comorbidities) and stratifying by sex. COVID-19 severity will be defined as recommended by the Host Genetics Initiative.<sup>39</sup> Severe disease will be defined as laboratory confirmed SARS-CoV-2 infection and hospitalisation for COVID-19.<sup>40</sup> Non-severe disease will be defined as laboratory-confirmed SARS-CoV-2 infection and not hospitalised 21 days after the test.<sup>40</sup> Mixed effect models will be used to account for within-patient measurements that change over time.

We expect that variations in severity of COVID-19 infection will correlate with differences in antibody response (eg, antibody titre, duration of antibodies). We further

expect that the antibody response will correlate with change in biochemical, haematological and/or clinical characteristics when the patient is acutely infected. Furthermore, we expect antigen targets (nucleocapsid and spike) to result in differential ability to neutralise virus. Patients who have non-neutralising antibodies may correlate with different biochemical/haematological responses or patient outcomes.

## Aim 2: determine impact of host genetic differences on COVID-19 infection severity and immune response

### Host GS

DNA will be extracted from blood lymphocytes for sequencing; RNA may also be extracted for confirmatory or sequencing validation purposes. Genomes will be sequenced through the Canadian Genomics COVID-19 Network (CanCOGeN) at the Centre for Applied Genomics at the Hospital for Sick Children. GS data will be generated from DNA libraries according to the manufacturer's protocol (Illumina) and sequenced on the NovaSeq6000 S4 flow cell to an average depth of at least 30×. Illumina provided software, bcl2fastq, will be used to convert the per-cycle binary base call files generated by the Illumina Sequencing systems to standard primary sequencing output in FASTQ format. During the conversion step, demultiplexing of samples will also be performed. Quality control (QC) metrics will be computed to assess the quality of the experiments. Reads will be aligned to the reference human genome using Burrows-Wheeler Aligner. Germline variant detection using Genome Analysis Toolkit (GATK) and copy number variant (CNV) and structural variant (SV) detection using read-depth and paired-end/split-read based methods will be performed. Variants will be annotated using a custom pipeline developed in-house. Files in standard output file formats will be generated including gvcf/vcf files for SNV and indel variants, vcf files for SV and CNV calls, tsv for annotated variants, and Binary Alignment Map (BAM) and index files for visualisation of the data. Per sample alignment and variant summaries will also be generated. Genome data will be stored in the CanCOGeN infrastructure with access control and linkable to administrative and other national databases and transferred to SH for clinical analysis by JL-E's laboratory.

### Host GS interpretation

Participants' whole genome data will be analysed for pathogenic variation according to current clinical standards<sup>41</sup> using a suite of software tools, disease and control databases, including both public sources and those housed locally at SH and The Hospital for Sick Children. Data will be analysed using custom in-house bioinformatics pipelines that follow GATK best practices and a third party software platform. The genome will be examined for clinically significant variation (eg, antibody deficiency, complement system, immune dysregulation, innate immunity, phagocyte defects, combined immunodeficiencies; autoinflammation, haematological, lung

or cardiovascular function and metabolism). We will also analyse HLA status, ABO blood group and genetic ancestry.

Reanalysis of genomic data will be feasible as new information is learnt from larger association studies. The reanalysis of any genomic results is patient-centric and will be initiated if there are any additional symptoms being reported by the patient or through their recruiting physician at the recruiting hospital to the study team, provided this occurs within the study time frame and sufficient resources are available.

### Data analysis

We will assess if host genomic variations contribute to differences in antibody response and disease severity. Logistic mixed models will be used for binary traits, and linear mixed models will be used for quantitative traits, according to the Host Genetics Initiative protocols.<sup>39 40 42</sup> Logistic mixed models will be used to test associations between molecular markers (eg, HLA subtypes; blood group genotype) and severe or non-severe illness (as described under aim 1). We will also test variants previously found to be associated with COVID-19 severity (eg, rs10735079, rs74956615, rs2236757)<sup>43</sup> to determine if these findings can be replicated in our sample. To correlate genomic and serological data, mixed linear regression models will be used to test associations between molecular markers (eg, HLA subtypes; blood group genotype) and differences in antibody titre, adjusting for covariates (eg, age, comorbidities) and stratifying by sex. Per recommendations from the Host Genetics Initiative, analyses will also be run separately for males and females, participants over and under 60 years of age at time of SARS-CoV-2 infection, and for each major ancestry group.<sup>42</sup>

To identify novel genetic loci associated with COVID-19 severity, genome-wide association studies (GWAS) will be conducted iteratively as new data becomes available through CanCOGeN using a variety of packages depending on outcome measures and custom methodology (Variant Integration Kit for NGS [ViKNGS]) for rare and common variants.<sup>44</sup> GENCOV data will be combined with other data collected through HostSeq to increase statistical power for GWAS.<sup>45</sup> GWAS will aim to identify genetic variants associated with severe or mild disease, using phenotypes defined by the Host Genetics Initiative. Per the HostSeq protocol, all results will be shared in an outward-facing permission-based control access portal and summary statistics will be shared with the international community.

### TCR/BCR sequencing and analysis

TCR/BCR sequencing will be performed on inpatients (n=300) and vaccinated patients at three different time points (baseline, 1 month, 6 or 12 months). To enable a high-resolution map of T/B-cell clonality and dynamics over time, we will profile T/B-cell repertoire in the serial

blood samples from patients during and after resolution of COVID-19 infection or postvaccination.

Samples from the first time point undergoing GS will follow the genome library preparation described above. At subsequent time points, Illumina-compatible next-generation sequencing libraries will be constructed from 100 to 1550 ng of fragmented DNA using the KAPA Hyper-Prep Kit (Sigma).<sup>8</sup> Hybrid capture will be performed according to the CapTCR-seq sequencing protocol. Hybrid capture probes will be directed against all V and J regions from all four TCR/BCR loci (alpha, beta, delta and gamma) annotated by the international ImMunoGeneTics database (<http://imgt.cines.fr>), following Roche SeqCap (Roche) conditions with xGen blocking oligos (IDT) and human Cot-1 blocking DNA (Invitrogen).<sup>8</sup> Following hybridisation, libraries will be amplified by PCR and sequenced on an Illumina NextSeq 500 instrument.<sup>8</sup> Clonotypes will be called using the MiXCR algorithm and clonal diversity calculated for every sample and compared within each patient over time.<sup>8</sup> Deidentified TCR/BCR sequences will be analysed in part using a third party software vendor.

Patients with intact (more diverse) immune systems may have a more robust immune response compared with those with compromised immune systems. We will determine if infection with SARS-CoV2 has a consistent diversifying or bottlenecking effect on the immune repertoire relative to controls. The immune repertoire diversity will be scored at the beginning of infection with duration and severity of COVID-19 symptoms. We will compare specific complementarity determining region three sequences across patients of the same HLA-type to nominate TCR sequences and potential immunogenic peptides associated with SARS-CoV-2 infection that may be leads for early diagnostics and second-generation vaccine development.

### **Aim 3: determine impact of viral lineage on antibody response and patient outcomes**

#### **Viral GS**

Viral samples will be taken from participants' primary nasopharyngeal, nasal and oral swabs, which are frozen and banked by the testing laboratory. Samples testing positive for SARS-CoV-2 RNA with qPCR cycle threshold <30 will be whole GS at the Ontario Institute for Cancer Research (OICR) as part of CanCOGeN Virus-Seq, the national SARS-CoV-2 sequencing initiative. Library construction and sequencing will follow the amplicon-based ARTIC V.3 protocol.<sup>46</sup> ARTIC V.3 primers will be used to amplify the viral genome, which will then be sequenced using Illumina instruments. The sequencing data will be processed with pipelines developed by CanCOGeN (<https://github.com/oicr-gsi/ncov2019-artic-nf>). QC will be performed using *ncov-tools* (<https://github.com/jts/ncov-tools/>). The use of this standardised protocol will allow the genomes sequenced in this project to be integrated with similar QC to the large collection of viral genomes produced by CanCOGeN and international projects, and deposited in the National Center for Biotechnology

Information (NCBI) and GISAID (<https://www.gisaid.org/>). Viral sequences will be assigned to viral lineages using pangolin.

#### **Data analysis**

Viral lineage will be associated with viral load (low/medium/high or as a continuous trait) and with other interactions. Recurrent viral mutations will be associated with serological response (eg, antibody titre, duration of antibody response, neutralisation) and clinical outcomes (eg, COVID-19 severity). We will evaluate how viral lineage interacts with the host genome findings and if it is associated with serological response. For example, we will assess whether novel variants in the spike protein are correlated with differences in viral neutralisation ability or patient symptom severity. Among patients in our cohort who have been reinfected by SARS-CoV-2, we will assess which viral variants are present in each lineage of SARS-CoV-2 by which they were infected, to identify variants or lineages associated with reinfection. Viral lineage will be reported to the clinician and participant in the genome report.

### **Aim 4: evaluate patient-reported outcomes of receiving host genome, antibody and viral lineage results**

#### **GS results reporting**

GS is performed on all participants in the COVID-19 cohort, however, participants have a choice as to whether they would like to learn their individual GS results, and the types of results that they would be willing to learn. Protocols are in place at SH to return genomic data to patients and to share variant classification data in public databases.<sup>47–49</sup> Participants will speak with a genetic counsellor and use an online decision aid<sup>50 51</sup> to decide which types of GS results they would like to learn. Reports will be issued to the participants, and results requiring clinical follow-up may be shared with their family doctor, with participant consent. Participants with clinically actionable or rare disease results will have their results communicated over the phone or by videoconference by the study genetic counsellor. All other results (eg, HLA status, ABO blood group, ancestry, viral lineage) will be communicated through the genomic report and summary letter only, and participants will be able to contact the genetic counsellor if they have questions. The study genetic counsellor and medical geneticist will determine the recommendation for each result and coordinate any necessary follow-up care.

Genome reports will include reason for referral, elected gene panels, genomic findings by disease group (eg, cardiology, neurology, metabolic, immunology, HLA status, blood group genotype, ancestry information, polygenic risk scores, viral lineage type), variant information, disease and inheritance information, treatment options, management recommendations and testing methods and limitations. Benign or common sequence variants of unlikely clinical significance will not be reported. Variants of uncertain



**Table 3** Patient-reported outcomes of returning genome, serology and viral lineage results and data collection time points

	Baseline (Intake survey)	Immediately after pretest counselling	Immediately after return of results	6 months after return of results
Decisional Conflict Scale <sup>54</sup>	x	x	x	x
Knowledge of genome sequencing (GS) benefits and limitations <sup>59</sup>	x	x	x	x
Knowledge of antibody testing	x		x	x
Knowledge of viral lineage	x		x	x
Expectations for antibody results	x		x	x
Hospital Anxiety and Depression Scale (HADS) <sup>52</sup>	x	x		x
Feelings About genomC Testing Results (FACToR) <sup>53</sup>			x	x
Quality of Life (12-Item Short Form Survey [SF-12]) <sup>55</sup>	x			x
Perceived personal and clinical utility of GS results <sup>57 58</sup>	x		x	x
Clinical actions taken based on GS, antibody and viral lineage results				x
Health behaviour change based on GS, antibody and viral lineage results				x
Genetic discrimination <sup>56</sup>				x
Brief Health Literacy Screening Tool (BRIEF) <sup>60</sup>	x			
Attitudes toward genetics <sup>61</sup>	x			
Attitudes toward healthcare <sup>62</sup>	x			
COVID-19 vaccination questions (vaccination status, vaccine type, number of doses, side effects)	x			x

significance may be reported if they occur in genes that match or are related to a specific clinical phenotype. Pathogenic and likely pathogenic variants will be reported. Secondary findings or variants in genes unrelated to the clinical phenotype may be included on the report based on the participant's preferences. A research consult letter will be drafted by the study genetic counsellor and made available to the participant, to summarise their findings and any recommendations based on their results.

#### Patient-reported outcome measures

In the COVID-19 cohort, we will assess patient-reported outcomes of learning GS results, antibody results and viral lineage through quantitative measures administered at multiple time points. We hypothesise that patient-reported outcomes will differ between patients who receive results requiring clinical follow-up and those who do not. Surveys will be administered at baseline (before pretest counselling), immediately after pretest counselling, immediately after the return of results, and 6 months following return of results (table 3). Novi Survey will be used for data collection and management. Each participant will be sent a unique link to complete the survey online; surveys will be self-administered. Surveys include validated questionnaires and measures as well as items developed by the study team. The outcomes that we will assess are distress (Hospital Anxiety and Depression Scale),<sup>52</sup> genetic-test related emotions (Feelings About genomC Testing Results),<sup>53</sup> decisional conflict

(decisional conflict scale),<sup>54</sup> quality of life (12-Item Short Form Survey [SF-12]),<sup>55</sup> genetic discrimination,<sup>56</sup> perceived utility of GS results,<sup>57 58</sup> clinical actions and health behaviour changes attributable to GS results/antibody results/viral lineage, knowledge of GS,<sup>59</sup> knowledge of antibody testing and knowledge of viral lineage. We will also assess health literacy,<sup>60</sup> attitudes towards genetics<sup>61</sup> and attitudes towards healthcare,<sup>62</sup> to be included in analyses as covariates as these characteristics may influence how participants respond to and act on their results. A limitation is that patients are not randomised, which increases the risk of bias. All measures and collection time points are listed in table 3, and the full survey can be found in online supplemental file 3.

#### Qualitative interviews

We will conduct qualitative interviews with a subset of participants (up to n=50) approximately 6 months after the return of GS results. Through semistructured qualitative interviews (online supplemental file 4), we will explore participants' experiences related to learning GS, antibody and viral lineage results, and the clinical, behavioural and psychosocial impacts of their results. Participants will be purposively sampled to participate in interviews based on the type of results they received (eg, results that required clinical follow-up).



## Data analysis

The yield of reported results from GS and the proportion of cases for which GS results required clinical follow-up will be reported descriptively. Participants' responses for each outcome measure will be summarised using descriptive statistics. To assess the impact of GS results, antibody results and viral lineage results on each outcome described above, we will compare each outcome between participants who receive different types of results (eg, for GS medically actionable results vs other categories of results; for antibody results, positive vs inconclusive vs negative results) using appropriate regression models to adjust for covariates that may also influence the outcomes (eg, attitudes toward genetics, health literacy). For all measures administered at multiple time points (table 3), we will use mixed-effects models to examine whether trajectories over time differ between groups.

Qualitative data analysis will draw on interpretive description methodology; interviews will be analysed thematically using constant comparison.<sup>63</sup> Interviews will be audiorecorded and transcribed. Two or more coders will review transcripts to generate a coding framework, which will be applied to transcripts and updated iteratively as new themes are identified in the data.<sup>63</sup> We will integrate quantitative and qualitative data using a mixed-methods matrix to better understand the impact of receiving GS, serology and viral results.<sup>64</sup>

## Patient and public involvement

Patients and the public were not involved in the design of the study. One aim of the study is to assess patient-reported outcomes related to receiving genome (including COVID-19-related results, and secondary findings), serology and viral lineage results.

## Data storage and sharing

Deidentified data will be shared through the HostSeq Databank (table 4). Sequence data can be deposited in the research domain such as dbGAP in the future.

Data will be shared with other researchers that request it and that are using it for research, for example, gene-disease linkages. Linkages will be made between laboratory data (eg, diagnostic/biochemical, sequencing data, serology; viral and host genome data) genomic sequences and HostSeq database. Viral GS performed by OICR through CanCOGeN funding requires mandatory data sharing. Viral genomes with deidentified metadata (including sample collection date, originating lab, host age and host sex), but without any supporting raw data, will be submitted to GISAID (<https://www.gisaid.org/>), the standard repository for SARS-CoV-2 genomes. Viral genomes and raw sequencing reads, excluding any host reads, will be uploaded to the NCBI and archived at the National Microbiology Laboratory.

Chart extraction data and laboratory data generated outside of the SH core lab will be stored securely in a centralised excel database and entered in a local instance of REDCap software on the Mount Sinai Hospital server. Roche data will be securely stored in the laboratory information system with the unique study ID of each patient. All samples will be deidentified on collection or receipt in the core lab at SH.

## ETHICS AND DISSEMINATION

### Ethics

This study has been approved at participating sites by the Clinical Trials Ontario Streamlined Ethics Review System (CTO-3302), with Mount Sinai Hospital as the board of record and which conducts ethical review and provides oversight for studies involving multiple sites in the province of Ontario. All participants provide informed consent to participate. A member of the research team reviews the consent form with participants over the phone or in person and answers participants' questions prior to obtaining consent. All participants consent to blood draws, GS, viral sequencing, serology testing, access to their banked samples and medical records, and data sharing.

**Table 4** Databases and privacy

Data type	Site data generated	Storage location	Where data is sent	How data is securely transferred
Host genome data	TCAG/The Hospital for Sick Children	The Hospital for Sick Children and Mount Sinai Hospital	Lunenfeld Tanenbaum Research Institute, Mount Sinai Hospital	Secure download link
Third party genome analysis software	N/A	US Server	Uploaded via secure portal	Secure upload link
Viral genome data	OICR-Ontario Institute for Health Research	OICR/Mount Sinai Hospital	Lunenfeld Tanenbaum Research Institute, Mount Sinai Hospital/NCBI/GISAID	Secure download or upload link
Deidentified clinical data	Mount Sinai Hospital and recruiting hospitals	REDCap - Mount Sinai Hospital Server	Mount Sinai Hospital	N/A
Survey data (Novi Survey)	Mount Sinai Hospital/Novi Survey	Novi Survey-Mount Sinai Hospital server	St. Michael's Hospital	Secure download link

N/A, not available; OICR, Ontario Institute for Cancer Research; TCAG, The Centre for Applied Genomics.



## Dissemination

Results from this study will be disseminated through peer-reviewed publications and presented at national and international conferences. Our knowledge translation strategy is to include the clinicians and laboratory professionals directly involved in patient care as members of the research team to inform the design, implementation and evaluation of the research findings, and to aid with prompt dissemination and application of results to clinical care. Members of the team are involved and can coordinate with other initiatives with established funding, including CanCOGeN, through which viral and host genome data will be shared, and the Canadian Open Genetics Repository<sup>47 48 65</sup> through which host genome variant data will be shared. Study data will be shared with the scientific community through open access and controlled access databases.

## Significance

This study will link serological, genomic, virology and patient characteristics to provide a comprehensive understanding of factors that contribute to variability in clinical symptoms and outcomes among patients with COVID-19. Healthcare systems and public health will be able to: (1) determine clinical utility of serology testing and aid in development and implementation of appropriate serology assays (2) select treatments based on understanding the antibody response and (3) identify priority populations for vaccinations if there is limited supply. Genomic approaches will (1) amplify the CanCOGeN research platform making data directly available to clinicians and patients, (2) enable harmonisation of genomic analyses methodologies and enable data sharing through existing clinical research platforms and (3) allow for strategic patient treatment and management plans and provide a resource for future analysis of secondary disease complications and long-term economic evaluation. Correlating viral genome data to infection and severity has important implications for the development of vaccines and therapies and managing the response to SARS-CoV-2. This research will also generate evidence on the patient-reported impact of learning results from GS as well as antibody results, and viral lineage, which could inform the adoption of GS and COVID-19 serological testing in clinical practice.

## Author affiliations

<sup>1</sup>Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada

<sup>2</sup>Pathology and Laboratory Medicine, Mount Sinai Hospital, Sinai Health, Toronto, Ontario, Canada

<sup>3</sup>Lunenfeld-Tanenbaum Research Institute, Sinai Health, Toronto, Ontario, Canada

<sup>4</sup>Li Ka Shing Knowledge Institute, St. Michael's Hospital, Unity Health Toronto, Toronto, Ontario, Canada

<sup>5</sup>Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada

<sup>6</sup>William Osler Health System, Brampton, Ontario, Canada

<sup>7</sup>Mount Sinai Academic Family Health Team, Mount Sinai Hospital, Sinai Health, Toronto, Ontario, Canada

<sup>8</sup>Department of Family and Community Medicine, University of Toronto, Toronto, Ontario, Canada

<sup>9</sup>Department of Emergency Medicine, Mount Sinai Hospital, Sinai Health, Toronto, Ontario, Canada

<sup>10</sup>Baycrest Health Sciences, Toronto, Ontario, Canada

<sup>11</sup>Division of General Internal Medicine, Mount Sinai Hospital, Sinai Health, Toronto, Ontario, Canada

<sup>12</sup>Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada

<sup>13</sup>Fred A Litwin and Family Centre in Genetic Medicine, University Health Network & Mount Sinai Hospital, Toronto, Ontario, Canada

<sup>14</sup>Emergency Medicine, University Health Network, Toronto, Ontario, Canada

<sup>15</sup>Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada

<sup>16</sup>Mackenzie Health, Richmond Hill, Ontario, Canada

<sup>17</sup>Department of Microbiology, Mount Sinai Hospital, Sinai Health, Toronto, Ontario, Canada

<sup>18</sup>Schwartz/Reisman Emergency Medicine Institute, Sinai Health System, Toronto, Ontario, Canada

<sup>19</sup>Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada

<sup>20</sup>Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada

<sup>21</sup>Ontario Institute for Cancer Research, Toronto, Ontario, Canada

<sup>22</sup>Department of Computer Science, University of Toronto, Toronto, Ontario, Canada

<sup>23</sup>The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, Canada

<sup>24</sup>Department of Statistical Sciences, University of Toronto, Toronto, Ontario, Canada

<sup>25</sup>Division of Emergency Medicine, University of Toronto, Toronto, Ontario, Canada

**Acknowledgements** We thank Jo-Anne Hebrick and Miranda Lorenti of The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada for assistance with DNA extraction. We thank Karan Singh at William Osler Health System for recruitment support. We thank Pirammiya Shanmugathas, Michael Puopolo, and Jordan Fung at Sinai Health for their support with recruitment and study coordination. We thank Andrew Wong at Sinai Health for IT support. We thank Paul Krzyzanowski at the Ontario Institute for Cancer Research (OICR), and Andrew McArthur at McMaster University for their contributions to generating viral sequence data.

**Contributors** Conceptualisation: JT, CM, TJP, YB, MC and JL-E. Methodology: JT, JL-E, YB, MC, A-CG, TJP, JS and TM. Software: EF, TJP, JS and JL-E. Validation: JT, SA, TJP, JS, TM and JL-E. Formal analysis: JT, CM, Sch, Sca, EF, SA, EB, AB, YB, BB, HC, MC, LD, HF, SMF, A-CG, ZK, TM, AM, SLM, TJP, DR, JS, SS, LS, AT and JL-E. Investigation: JT, CM, Sch, Sca, EF, SA, EB, AB, YB, BB, HC, MC, LD, HF, SMF, A-CG, ZK, TM, AM, SLM, TJP, DR, JS, SS, LS, AT and JL-E. Resources: JT, SA, EB, AB, YB, BB, HC, MC, LD, HF, SMF, A-CG, ZK, TM, AM, SLM, TJP, DR, JS, SS, LS, AT and JL-E. Data curation: EF. Writing—original draft: JT, CM, JL-E. Writing—review and editing: JT, CM, Sch, Sca, EF, SA, EB, AB, YB, BB, HC, MC, LD, HF, SMF, A-CG, ZK, TM, AM, SLM, TJP, DR, JS, SS, LS, AT and JL-E. Supervision: JT and JL-E. Project administration: JT, Sch and JL-E. Funding acquisition: JT, JL-E, SA, BB, HC, LD, HF, SMF, A-CG, TM, AM, SLM, TJP, DR, JS and LS.

**Funding** This work was supported by the Canadian Institutes of Health Research (Funding Reference Number VR4-172753).

**Competing interests** None declared.

**Patient consent for publication** Not applicable.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

## ORCID iDs

 Yvonne Bombard <http://orcid.org/0000-0002-9516-4539>

 Steven Marc Friedman <http://orcid.org/0000-0002-7871-3583>

 Jordan Lerner-Ellis <http://orcid.org/0000-0003-3685-5679>

## REFERENCES

- 1 Stadlbauer D, Amanat F, Chromikova V, *et al.* SARS-CoV-2 seroconversion in humans: a detailed protocol for a serological assay, antigen production, and test setup. *Curr Protoc Microbiol* 2020;57:e100.
- 2 Garg S, Kim L, Whitaker M, *et al.* Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-Confirmed Coronavirus Disease 2019 - COVID-NET, 14 States, March 1-30, 2020. *MMWR Morb Mortal Wkly Rep* 2020;69:458-64.
- 3 Zhao J, Yuan Q, Wang H, *et al.* Antibody responses to SARS-CoV-2 in patients with novel coronavirus disease 2019. *Clin Infect Dis* 2020;71:2027-34.
- 4 To KK-W, Tsang OT-Y, Leung W-S, *et al.* Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. *Lancet Infect Dis* 2020;20:565-74.
- 5 Winter AK, Hegde ST. The important role of serology for COVID-19 control. *Lancet Infect Dis* 2020;20:758-9.
- 6 Wu F, Wang A, Liu M. Neutralizing antibody responses to SARS-CoV-2 in a COVID-19 recovered patient cohort and their implications, 2020. Available: <https://www.medrxiv.org/content/10.1101/2020033020047365v2>
- 7 Sette A, Crotty S. Adaptive immunity to SARS-CoV-2 and COVID-19. *Cell* 2021;184:861-80.
- 8 Mulder DT, Mahé ER, Dowar M, *et al.* CapTCR-seq: hybrid capture for T-cell receptor repertoire profiling. *Blood Adv* 2018;2:3506-14.
- 9 LoPresti M, Beck DB, Duggal P, *et al.* The role of host genetic factors in coronavirus susceptibility: review of animal and systematic review of human literature. *Am J Hum Genet* 2020;107:381-402.
- 10 Ellinghaus D, Degenhardt F, *et al.* Severe Covid-19 GWAS Group. Genomewide association study of severe Covid-19 with respiratory failure. *N Engl J Med* 2020;383:1522-34.
- 11 Allers K, Schneider T. CCR5Δ32 mutation and HIV infection: basis for curative HIV therapy. *Curr Opin Virol* 2015;14:24-9.
- 12 Duggal P, Thio CL, Wojcik GL, *et al.* Genome-wide association study of spontaneous resolution of hepatitis C virus infection: data from multiple cohorts. *Ann Intern Med* 2013;158:235-45.
- 13 Schulert GS, Zhang M, Fall N, *et al.* Whole-exome sequencing reveals mutations in genes linked to hemophagocytic lymphohistiocytosis and macrophage activation syndrome in fatal cases of H1N1 influenza. *J Infect Dis* 2016;213:1180-8.
- 14 Shi Y, Wang Y, Shao C, *et al.* COVID-19 infection: the perspectives on immune responses. *Cell Death Differ* 2020;27:1451-4.
- 15 Lin M, Tseng H-K, Trejaut JA, *et al.* Association of HLA class I with severe acute respiratory syndrome coronavirus infection. *BMC Med Genet* 2003;4:9.
- 16 MacDonald KS, Fowke KR, Kimani J, *et al.* Influence of HLA supertypes on susceptibility and resistance to human immunodeficiency virus type 1 infection. *J Infect Dis* 2000;181:1581-9.
- 17 Stephens HAF, Klaythong R, Sirikong M, *et al.* HLA-A and -B allele associations with secondary dengue virus infections correlate with disease severity and the infecting viral serotype in ethnic Thais. *Tissue Antigens* 2002;60:309-18.
- 18 Yousefzadegan S, Rezaei N. Case report: death due to COVID-19 in three brothers. *Am J Trop Med Hyg* 2020;102:1203-4.
- 19 Ikitimur H, Borku Uysal B, Cengiz M. Determining host factors contributing to disease severity in a family cluster of 29 hospitalized SARS-CoV-2 patients: could genetic factors be relevant in the clinical course of COVID-19? *J Med Virol* 2020.
- 20 Stergachis AB, Weiss ST, Green RC. Biobanks could identify medically actionable findings relevant for COVID-19 clinical care. *Nat Med* 2020;26:991.
- 21 Volz E, Mishra S, Chand M. Transmission of SARS-CoV-2 lineage B.1.1.7 in England: insights from linking epidemiological and genetic data. *medRxiv* 2021.
- 22 Dumonteil E, Fusco D, Drouin A, *et al.* Genomic signatures of SARS-CoV-2 associated with patient mortality. *Viruses* 2021;13:227.
- 23 van Dorp L, Acman M, Richard D, *et al.* Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infect Genet Evol* 2020;83:104351.
- 24 Lippi G, Plebani M. Laboratory abnormalities in patients with COVID-2019 infection. *Clin Chem Lab Med* 2020;58:1131-4.
- 25 Henry BM, Lippi G, Plebani M. Laboratory abnormalities in children with novel coronavirus disease 2019. *Clin Chem Lab Med* 2020.
- 26 International Federation of Clinical Chemistry and Laboratory Medicine. IFCC Information Guide on COVID-19 Tuesday 14 April updates - IFCC, 2020. Available: <https://www.ifcc.org/ifcc-news/2020-03-26-ifcc-information-guide-on-covid-19>
- 27 Miller DT, Lee K, Gordon AS. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2021 update: a policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genet Med* 2021.
- 28 Global Alliance for Genomics and Health. *Policy on clinically actionable genomic research results*, 2021.
- 29 Mackley MP, Fletcher B, Parker M, *et al.* Stakeholder views on secondary findings in whole-genome and whole-exome sequencing: a systematic review of quantitative and qualitative studies. *Genet Med* 2017;19:283-93.
- 30 Sanderson SC, Linderman MD, Suckiel SA, *et al.* Motivations, concerns and preferences of personal genome sequencing research participants: baseline findings from the HealthSeq project. *Eur J Hum Genet* 2016;24:14-20.
- 31 Middleton A, Morley KI, Bragin E, *et al.* Attitudes of nearly 7000 health professionals, genomic researchers and publics toward the return of incidental results from sequencing research. *Eur J Hum Genet* 2016;24:21-9.
- 32 Mighton C, Carlsson L, Clausen M, *et al.* Quality of life drives patients' preferences for secondary findings from genomic sequencing. *Eur J Hum Genet* 2020;28:1178-86.
- 33 Hart MR, Biesecker BB, Blout CL, *et al.* Secondary findings from clinical genomic sequencing: prevalence, patient perspectives, family history assessment, and health-care costs from a multisite study. *Genet Med* 2019;21:1100-10.
- 34 Robinson JO, Wynn J, Biesecker B, *et al.* Psychological outcomes related to exome and genome sequencing result disclosure: a meta-analysis of seven clinical sequencing exploratory research (CSER) Consortium studies. *Genet Med* 2019;21:2781-90.
- 35 Isho B, Abe KT, Zuo M, *et al.* Persistence of serum and saliva antibody responses to SARS-CoV-2 spike antigens in COVID-19 patients. *Sci Immunol* 2020;5:eabe5511. doi:10.1126/sciimmunol.abe5511
- 36 Abe KT, Li Z, Samson R, *et al.* A simple protein-based surrogate neutralization assay for SARS-CoV-2. *JCI Insight* 2020;5 doi:10.1172/jci.insight.142362
- 37 ISARIC4C. *ISARIC coronavirus clinical characterisation consortium*, 2020.
- 38 Docherty AB, Harrison EM, Green CA, *et al.* Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. *BMJ* 2020;369:m1985.
- 39 COVID-19 Host Genetics Initiative. The COVID-19 host genetics initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *Eur J Hum Genet* 2020;28:715-8.
- 40 COVID-19 Host Genetics Initiative. *Phenotype definitions for analysis: 27 April 2020 version 2.0*, 2020.
- 41 Richards S, Aziz N, Bale S, *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of medical genetics and genomics and the association for molecular pathology. *Genet Med* 2015;17:405-24.
- 42 Initiative C-HG. *Analysis Plan, August 2020 - version 1.1*, 2020.
- 43 Pairo-Castineira E, Clohisey S, Klaric L, *et al.* Genetic mechanisms of critical illness in COVID-19. *Nature* 2021;591:92-8.
- 44 Baskurt Z, Mastromatteo S, Gong J, *et al.* ViNGS: a C++ variant integration kit for next generation sequencing association analysis. *Bioinformatics* 2020;36:1283-5.
- 45 Lin Y-C, Brooks JD, Bull SB, *et al.* Statistical power in COVID-19 case-control host genomic study design. *Genome Med* 2020;12:115.
- 46 Farr B, Rajan D. DNA Pipelines R&D. *COVID-19 ARTIC V3 Illumina library construction and sequencing protocol*, 2020.
- 47 Lebo MS, Zakoor K-R, Chun K, *et al.* Data sharing as a national quality improvement program: reporting on BRCA1 and BRCA2 variant-interpretation comparisons through the Canadian open genetics Repository (COGR). *Genet Med* 2018;20:294-302.
- 48 Lerner-Ellis J, Wang M, White S, *et al.* Canadian open genetics Repository (COGR): a unified clinical genomics database as a community resource for standardising and sharing genetic interpretations. *J Med Genet* 2015;52:438-45.
- 49 Shickh S, Gutierrez Salazar M, Zakoor K-R, *et al.* Exome and genome sequencing in adults with undiagnosed disease: a prospective cohort study. *J Med Genet* 2021;58:275-83.





- 50 Bombard Y, Clausen M, Mighton C, *et al.* The genomics ADVISER: development and usability testing of a decision aid for the selection of incidental sequencing results. *Eur J Hum Genet* 2018;26:984–95.
- 51 Bombard Y, Clausen M, Shickh S, *et al.* Effectiveness of the genomics ADVISER decision aid for the selection of secondary findings from genomic sequencing: a randomized clinical trial. *Genet Med* 2020;22:727–35.
- 52 Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983;67:361–70.
- 53 Li M, Bennette CS, Amendola LM, *et al.* The feelings about genomic testing results (factor) questionnaire: development and preliminary validation. *J Genet Couns* 2019;28:477–90.
- 54 O'Connor AM. Validation of a decisional conflict scale. *Med Decis Making* 1995;15:25–30.
- 55 Johnson JA, Coons SJ. Comparison of the EQ-5D and SF-12 in an adult US sample. *Qual Life Res* 1998;7:155–66.
- 56 Bombard Y, Veenstra G, Friedman JM, *et al.* Perceptions of genetic discrimination among people at risk for Huntington's disease: a cross sectional survey. *BMJ* 2009;338:b2175.
- 57 Zoltick ES, Linderman MD, McGinniss MA, *et al.* Predispositional genome sequencing in healthy adults: design, participant characteristics, and early outcomes of the PeopleSeq Consortium. *Genome Med* 2019;11:10.
- 58 Linderman MD, Nielsen DE, Green RC. Personal genome sequencing in ostensibly healthy individuals and the PeopleSeq Consortium. *J Pers Med* 2016;6 doi:10.3390/jpm6020014
- 59 Kaphingst KA, Facio FM, Cheng M-R, *et al.* Effects of informed consent for individual genome sequencing on relevant knowledge. *Clin Genet* 2012;82:408–15.
- 60 Haun J, Luther S, Dodd V, *et al.* Measurement variation across health literacy assessments: implications for assessment selection in research and practice. *J Health Commun* 2012;17:141–59.
- 61 Sanderson SC, Wardle J, Michie S. The effects of a genetic information leaflet on public attitudes towards genetic testing. *Public Underst Sci* 2005;14:213–24.
- 62 Bombard Y, Miller FA, Hayeems RZ, *et al.* Public views on participating in newborn screening using genome sequencing. *Eur J Hum Genet* 2014;22:1248–54.
- 63 Thorne S. *Interpretive description. developing qualitative inquiry.* Left Coast Press, 2008.
- 64 O'Cathain A, Murphy E, Nicholl J. Three techniques for integrating data in mixed methods studies. *BMJ* 2010;341:c4587.
- 65 Mighton C, Smith AC, Mayers J, *et al.* Data sharing to improve concordance in variant interpretation across laboratories: results from the Canadian Open Genetics Repository. *J Med Genet* 2021. doi:10.1136/jmedgenet-2021-107738. [Epub ahead of print: 19 Apr 2021].