Image-based and machine learning-guided multiplexed serology test for SARS-CoV-2

Pietiäinen, Vilja ; Polso, Minttu ; Migh, Ede ; Guckelsberger, Christian ; Harmati, Maria ; Diosdi, Akos ; Turunen, Laura ; Hassinen, Antti ; Potdar, Swapnil ; Koponen, Annika ; Sebestyen, Edina Gyukity ; Kovacs, Ferenc ; Kriston, Andras ; Hollandi, Reka ; Burian, Katalin ; Terhes, Gabriella ; Visnyovszki, Adam ; Fodor, Eszter ; Lacza, Zsombor ; Kantele, Anu ; Kolehmainen, Pekka ; Kakkola, Laura ; Strandin, Tomas ; Levanov, Lev ; Kallioniemi, Olli ; Kemeny, Lajos ; Julkunen, Ilkka ; Vapalahti, Olli ; Buzas, Krisztina ; Paavolainen, Lassi ; et al ; Hepojoki, Jussi

DOI: https://doi.org/10.1016/j.crmeth.2023.100565

Originally published at:
Pietiäinen, Vilja; Polso, Minttu; Migh, Ede; Guckelsberger, Christian; Harmati, Maria; Diosdi, Akos; Turunen, Laura; Hassinen, Antti; Potdar, Swapnil; Koponen, Annika; Sebestyen, Edina Gyukity; Kovacs, Ferenc; Kriston, Andras; Hollandi, Reka; Burian, Katalin; Terhes, Gabriella; Visnyovszki, Adam; Fodor, Eszter; Lacza, Zsombor; Kantele, Anu; Kolehmainen, Pekka; Kakkola, Laura; Strandin, Tomas; Levanov, Lev; Kallioniemi, Olli; Kemeny, Lajos; Julkunen, Ilkka; Vapalahti, Olli; Buzas, Krisztina; Paavolainen, Lassi; et al; Hepojoki, Jussi (2023). Image-based and machine learning-guided multiplexed serology test for SARS-CoV-2. Cell Reports Methods, 3(8):100565.
DOI: https://doi.org/10.1016/j.crmeth.2023.100565

Cell Reports Methods

CellPress
OPEN ACCESS

Article

Image-based and machine learning-guided multiplexed serology test for SARS-CoV-2

Vilja Pietiäinen,[1,24,*] Minttu Polso,[1,23] Ede Migh,[2,23] Christian Guckelsberger,[1,3,4,5,23] Maria Harmati,[2] Akos Diosdi,[2] Laura Turunen,[1] Antti Hassinen,[1] Swapnil Potdar,[1] Annika Koponen,[6,7] Edina Gyukity Sebestyen,[2] Ferenc Kovacs,[2,8] Andras Kriston,[2,8] Reka Hollandi,[2] Katalin Burian,[9] Gabriella Terhes,[9] Adam Visnyovszki,[10] Eszter Fodor,[11]

(Author list continued on next page)

[1]Institute for Molecular Medicine Finland (FIMM), Helsinki Institute of Life Sciences (HiLIFE), University of Helsinki, Helsinki, Finland
[2]Laboratory of Microscopic Image Analysis and Machine Learning, Institute of Biochemistry, Biological Research Centre, Szeged, Hungary
[3]Department of Computer Science, Aalto University, Espoo, Finland
[4]Finnish Center for Artificial Intelligence, Espoo, Finland
[5]School of Electronic Engineering and Computer Science, Queen Mary University of London, London, UK
[6]Minerva Foundation Institute for Medical Research, Helsinki, Finland
[7]Department of Anatomy, Faculty of Medicine, University of Helsinki, Helsinki, Finland
[8]Single-Cell Technologies Ltd., Szeged, Hungary

(Affiliations continued on next page)

MOTIVATION Serological assays are essential for studying and controlling infectious disease outbreaks, such as the current coronavirus disease 2019 (COVID-19) pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Immunofluorescence assays relying on cells infected with a given virus provide rapidly available means for demonstrating antibody response against emerging pathogens. However, setting up such assays may require a high-biosafety-level facility for handling the virus. In addition, such assays are often low throughput, and interpretation of the results can be labor intensive and subjective. Here, we describe a mini-immunofluorescence assay (mini-IFA), an automated, high-throughput, microscopy-based serology assay requiring a low-biosafety-level environment. We demonstrate its efficacy using SARS-CoV-2 as a model.

SUMMARY

We present a miniaturized immunofluorescence assay (mini-IFA) for measuring antibody response in patient blood samples. The method utilizes machine learning-guided image analysis and enables simultaneous measurement of immunoglobulin M (IgM), IgA, and IgG responses against different viral antigens in an automated and high-throughput manner. The assay relies on antigens expressed through transfection, enabling use at a low biosafety level and fast adaptation to emerging pathogens. Using severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) as the model pathogen, we demonstrate that this method allows differentiation between vaccine-induced and infection-induced antibody responses. Additionally, we established a dedicated web page for quantitative visualization of sample-specific results and their distribution, comparing them with controls and other samples. Our results provide a proof of concept for the approach, demonstrating fast and accurate measurement of antibody responses in a research setup with prospects for clinical diagnostics.

INTRODUCTION

Here, we describe a scalable and automated high-content, microscopy-based mini-immunofluorescence assay (mini-IFA) for serological testing; i.e., detection of antibodies. Unlike conventional IFA, which often relies on use of cells infected with the target pathogen, our assay employs transfected cells expressing individual viral antigens. The assay builds on a custom neural network-based image analysis pipeline for the automated and multiplexed detection of immunoglobulins (immunoglobulin G [IgG], IgA, and IgM) in patient samples. As a proof of concept, we employed high-throughput equipment to set up the assay

Zsombor Lacza,[11] Anu Kantele,[12,13] Pekka Kolehmainen,[14] Laura Kakkola,[14] Tomas Strandin,[15] Lev Levanov,[15] Olli Kallioniemi,[1,16] Lajos Kemeny,[17] Ilkka Julkunen,[14,18] Olli Vapalahti,[15,19,20] Krisztina Buzas,[2,21] Lassi Paavolainen,[1] Peter Horvath,[1,2,8,23,*] and Jussi Hepojoki[15,22,23,*]

[9]Department of Medical Microbiology, Faculty of Medicine, University of Szeged, Szeged, Hungary
[10]1st Department of Internal Medicine, Faculty of Medicine, University of Szeged, Szeged, Hungary
[11]Department of Sports Physiology, Institute of Sports and Health Sciences, University of Physical Education, Budapest, Hungary
[12]Meilahti Infectious Diseases and Vaccine Research Center (MeiVac), University of Helsinki and Helsinki University Hospital, Helsinki, Finland
[13]Human Microbiome Research Program, University of Helsinki, Helsinki, Finland
[14]Institute of Biomedicine, University of Turku, Turku, Finland
[15]Department of Virology, Medicum, University of Helsinki, Helsinki, Finland
[16]Department of Oncology-Pathology, Karolinska Institutet, Science for Life Laboratory, Stockholm, Sweden
[17]HCEMM-USZ Skin Research Group, Department of Dermatology and Allergology, University of Szeged, Szeged, Hungary
[18]Turku University Hospital, Turku, Finland
[19]Department of Veterinary Biosciences, University of Helsinki, Helsinki, Finland
[20]HUS Diagnostic Center, HUSLAB, Clinical Microbiology, Helsinki University Hospital, Helsinki, Finland
[21]Department of Immunology, Faculty of Medicine, Faculty of Science and Informatics, University of Szeged, Szeged, Hungary
[22]University of Zurich, Vetsuisse Faculty, Institute of Veterinary Pathology, Zürich, Switzerland
[23]These authors contributed equally
[24]Lead contact
*Correspondence: vilja.pietiainen@helsinki.fi (V.P.), horvath.peter@brc.hu (P.H.), jussi.hepojoki@helsinki.fi (J.H.)
https://doi.org/10.1016/j.crmeth.2023.100565

for measuring antibody response against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection with spike (S), membrane (M), and nucleo (N) proteins and the receptor-binding domain (RBD) as the antigens. We compared the automated mini-IFA results from hundreds of patient samples to the visual observations by human experts and to the results obtained with conventional ELISA as well as the conventional virus IFA. The comparisons demonstrated a high correlation with both, suggesting high sensitivity and specificity of the mini-IFA. By testing pre-pandemic samples and those collected from patients with RT-PCR-confirmed SARS-CoV-2 infection, we found the mini-IFA to be most suitable for IgG and IgA detection. The results demonstrated N and S proteins as the ideal antigens, and the use of these antigens can serve to distinguish between vaccinated and infected individuals. The assay principle described enables detection of antibodies against practically any pathogen, and none of the assay steps require a high-biosafety-level (BSL) environment, unlike many IFA assays[1]. The simultaneous detection of multiple Ig classes allows for distinguishing between recent and past infection. The mini-IFA pipeline is highly flexible because it does not necessarily require complex automation or a specific microscope. The basic requirement is only a basic fluorescence microscope capable of imaging multiwell plates. The unique key element is the analysis pipeline, which can be performed regardless of the location, sample preparation equipment, or microscope used.

RESULTS

Mini-IFA offers a robust and scalable method for analysis of serum antibodies against SARS-CoV-2 antigens

Laboratory method

See Figure 1, steps 1–4, and STAR Methods for the standard operating procedure [SOP]. The cell line of interest (here, African green monkey kidney cells, Vero E6) is separately transfected with plasmids encoding four SARS-CoV-2 antigens (S with a His tag [S-pCAGGS];[2] M with a hemagglutinin [HA] tag [M-pEBB]; N with a His tag [NP-pCAGGS];[3] and the RBD of S with a His tag [RBD-pCAGGS][2]) (see STAR Methods for materials and methods and the Key resources table). The transfected cells in suspension are transferred to 384-well assay plates with high-throughput automation. Transient transfection, as opposed to stable transfection, allows simultaneous control of the background signal and detection of autoantibodies based on non-transfected cells included in the same well. Patient sera and assay controls (Figures 2A and S1), as well as a DNA-binding dye (Hoechst 33342) are added to the wells with an acoustic dispenser to simultaneously monitor the accuracy and success of the sample transfer. Each of the assays (and plates) includes several positive (and negative) controls, which allows for comparing the results between assays and the measurement of transfection percentage, enabling cross-assay standardization. Ig classes present in the samples are detected by multiplexed immunostaining with fluorescently labeled antibodies for IgM (Alexa Fluor 488 [AF488]), IgG (DyLight 55 [DL550]), and IgA (AF647). Transfection efficiency (Table S1A) is controlled by staining with antibodies (Key resources table) targeting the respective (His/HA -tagged) antigens and/or antigen-specific antisera. We image four different fluorescent channels with high-content microscopy to detect (1) IgG, (2) IgA, (3) IgM, and (4) cell nuclei (Figures 2A, 2B, S1). The method can be modified to include any antibodies/markers; however, any new marker will require a new training set for machine learning (ML) models (STAR Methods). Here, we excluded the cytoplasmic marker to include the transfection controls/additional Ig classes in the basic four-channel microscopy set-up.

Computational analysis

Computational analysis is detailed in Figure 1, steps 5–8. The analysis pipeline includes the processing of the microscopy images, per-cell and per-well predictions, and the visualization of the results. The images were processed with the BIAS software,[4,5] facilitating pre-processing,[6] deep-learning-based cell segmentation[7] (Figure 2C), and feature extraction (STAR

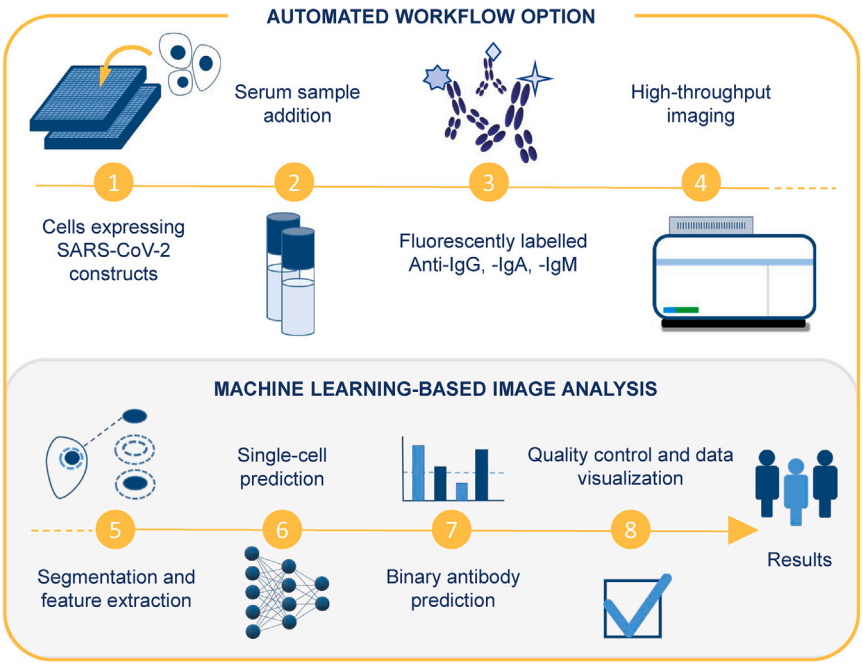


Figure 1. The miniaturized immunofluorescence assay (mini-IFA) pipeline

Vero E6 cells, transfected to express SARS-CoV-2 Ags (N, M, R, and S), are fixed in 384-well plates for incubation with serum samples. The immunoglobulins (IgG, IgA, and IgM) are detected simultaneously with fluorophore-labeled secondary Abs using automated high-content fluorescence microscopy. Machine learning is employed for (1) nuclei and cell segmentation and (2) phenotypic cell classification. The results are presented via a decision support system, which also allows interactive visual observation of the raw images. For a full explanation with numbers, see the main text.

Methods). We developed dedicated Python scripts to use these features for prediction of antibody response through supervised ML (GitHub: https://github.com/fimm-covid-19-hca/mini-IFA_paper). More specifically, we trained models to classify each segmented immunostaining phenotype, represented by the extracted image features, into five classes: ''positive,'' ''negative,'' ''atypical,'' ''small bright,'' and ''trash'' (Figures 2D and 2E; see class design rationale in STAR Methods). As training data, we labeled 55,496 cells across 16 plates (four plates for each antigen) with roughly equal proportions per antigen (Table S2). Labels were assigned in BIAS software[4,5] as previously described,[8] using an active learning feature[9] to increase training data quality. We avoid model bias through class imbalances by including class weights in the training. To account for potential signal variations across plates and batches, image features were normalized based on per-plate controls. We employed leave-one-plate-out cross-validation to (1) determine the best of three normalization approaches and (2) select the best classification model and hyperparameters for each Ig class-antigen pair. Performance was scored on prediction sensitivity × specificity, and one plate was left out as a validation set in each fold to assess the method's generalization potential as prediction quality on unseen plates—a typical scenario for the assay's clinical application. We found that the model can classify individual cells accurately with a specificity of 0.96–0.97, 0.95–0.96, and 0.96–0.97 and a sensitivity of 0.84–0.89, 0.79–0.84, and 0.82–0.86 for IgG, IgA, and IgM, respectively (STAR Methods; Figures S2A and S2B; Table S3). Following this selection, we re-trained the best classification model/hyperparameter configuration on the entire training dataset, processed with the best normalization scheme. We applied the trained model to a separate experimental test dataset, including samples from four plates, separately for each antigen to produce per-cell predictions. Based on these predictions, we calculated a per-well positivity score as the ratio of predicted positive cells over all segmented cells. The user can define the cutoff of this value based on positive and negative controls (see STAR Methods and webpage: https://fimm-covid-19-hca.github.io/ for data visualization graphs showing the sample distribution compared with the positive and negative controls), resulting in binary per-well antibody response predictions.

Visualization

Quality control (QC) assessment with an automated script is available to control the assay's quality, including cell seeding and sample transfer accuracy as well as signal distribution for positive and negative controls (Figure 3; Table S1B). The data are visualized with (interactive) plots and heatmaps (Figure 3; Data S1).

The test can be adapted and scaled to be used in other laboratories

To (1) evaluate the assay's technical scalability and adaptability to a laboratory without advanced automation and (2) test a variation that measures the IgG/IgA signal only in antigen-transfected cells (recognized with the His tag), shipped ready-to-go assay plates (cells transfected with the virus antigen and fixed) or manually prepared assay plates were utilized in another laboratory (Biological Research Centre, Szeged, Hungary; see STAR Methods and Figure S3 for the semi-automated method). Single-cell phenotypic analysis was performed with BIAS software as described previously.[5] The results are displayed in detail in the STAR Methods and Figures S5–S7.

Altogether, we processed 948 (n = 583 at the Institute for Molecular Medicine Finland [FIMM]) serum or plasma specimens collected in 2017 and 2020, with ethics approval and informed consent (STAR Methods; Table S4) from 890 (n = 542 at FIMM) donors, of which 181 (n = 42 at FIMM) were individuals with a positive SARS-CoV-2 RT-PCR and/or ELISA test result.

The interactive visualization of mini-IFA assay results provides a further decision support tool

Because the results include thousands of data points, we developed automated scripts for the visualization of the quality

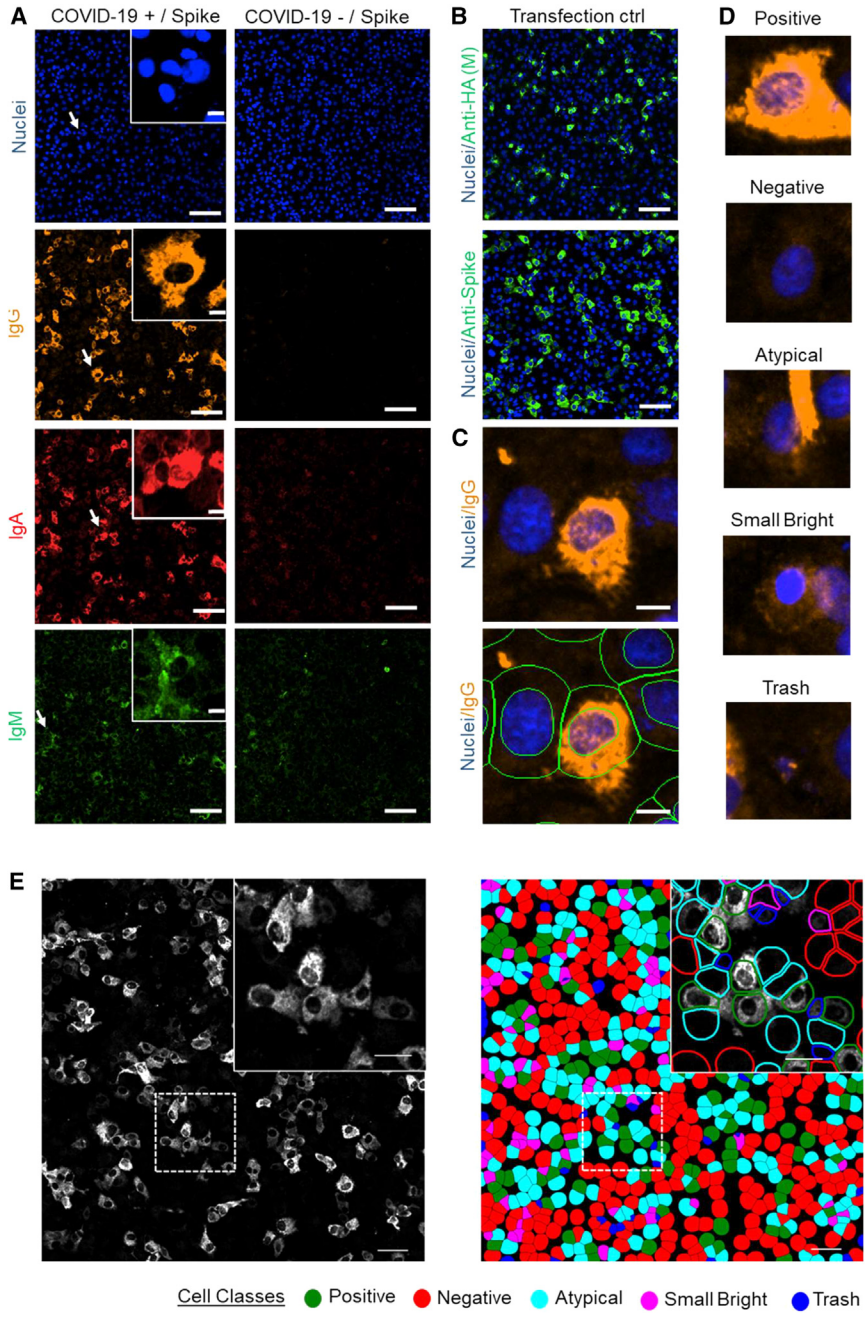


Figure 2. Examples of microscopy images, classifications, and predictions

(A) Examples of microscopy images of specific IgG (DL550, orange), IgA (AF647, red), and IgM (AF488, green) responses against SARS-CoV-2 spike (S) protein in assay control serum samples obtained from positive (COVID-19+) and neg (COVID-19–) patients. Scale bars, 100 μm (overview images) and 10 μm (zoomed-in images).

(B) Transfection efficiency of viral Ags is determined by immunostaining. Here, AF488-conjugated anti-HA (green) Ab for HA-tagged M protein and in-house rabbit anti-S Ab (AF488, green) for S protein are shown. Scale bars, 100 μm.

(C) Visualization of non-segmented and segmented cells. After segmentation of cell nuclei, an additional mask is created for the "whole cell" by dilating the nucleus area to a maximum of 7 μm, and the "cytoplasm" is determined as an area without a segmented nucleus. Scale bars, 10 μm.

(D) Examples of manual labeling of cell classes for the training set. Cells were divided into five categories: positive, negative, atypical, small bright, and trash.

(E) Example of single-cell predictions. Shown is an example image (IgG response against S-protein), scaled linearly from range (quantile [0.5], quantile [0.995]) to 0255 8-bit values for visualization purposes (left). Single-cell class predictions of the same image were made by the model trained for S-protein IgG Ab images (right). The close up shows the contours of cell boundaries colored with the predicted class color. Scale bars, 50 μm (whole-field-of-view images) and 25 μm (close-up images). Classification of cells to five different categories as in (D) is highlighted with different colors.

Unless otherwise stated, the microscopy images presented in the figure show cells expressing the S protein of SARS-CoV-2, and they have been linearly adjusted from the original 16-bit image format to improve visual appearance.

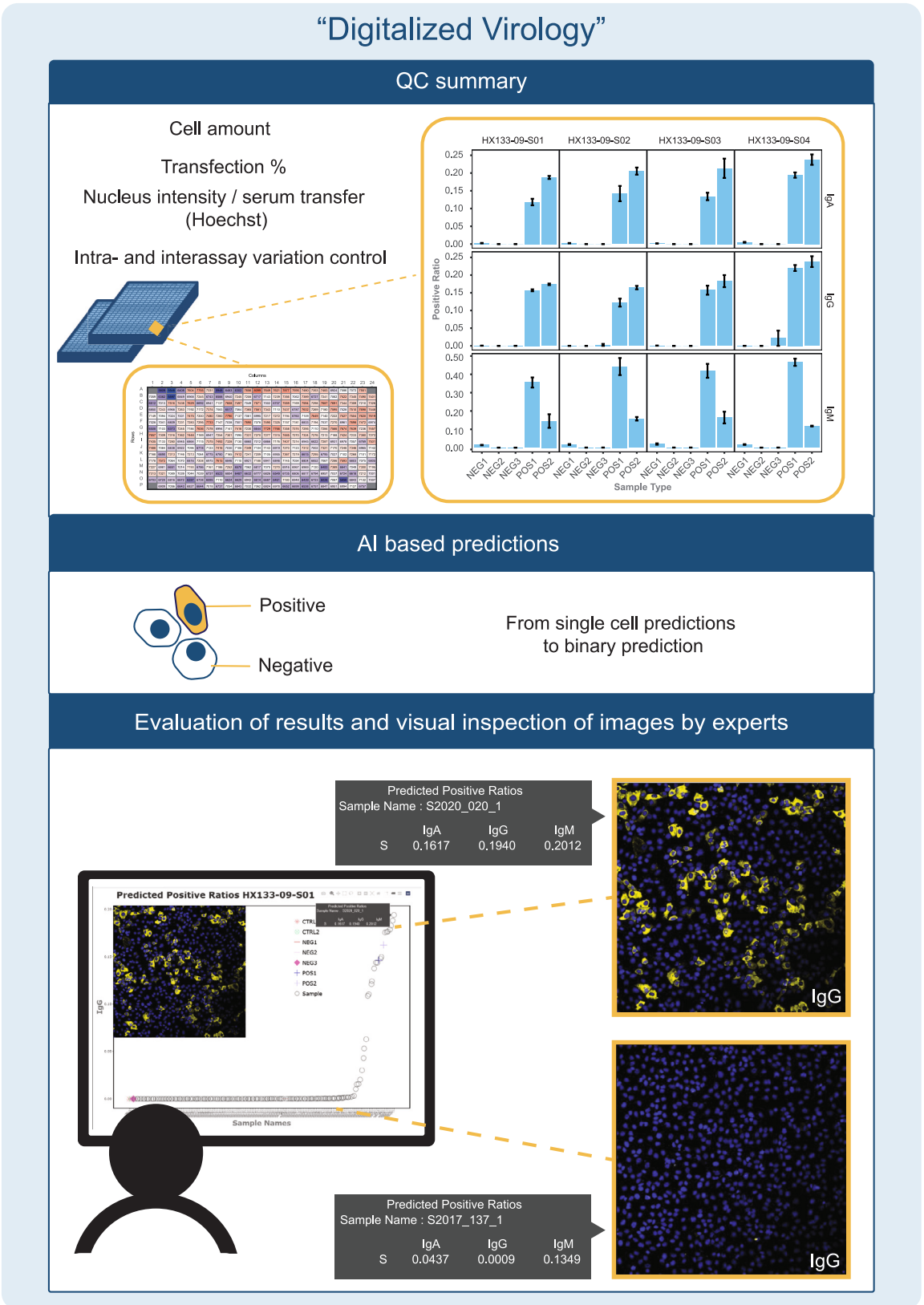
assessment as well as for the assay results. The visualization includes assay control data and plate-based heatmaps (Figure 3). The results, allowing the observer to view images of each (patient) sample separately for all antibody (Ab)/antigen (Ag) scores, are visualized as interactive plots (plate-based; STAR Methods). The plots allow the clicking of the data point and opening of the image of the selected sample in a separate window for additional visual evaluation and as heatmaps (Figure 3; Data S1).

This enables a fast digital (re)view of the results, independent of place and time, with quantitatively scored findings, and fast prioritization of the serum samples for further tests; e.g., for a neutralizing assay or for the re-evaluation of technically challenging samples or borderline cases.

Different Igs responses to multiple Ags can be simultaneously distinguished

As shown in Figure 4A, the assay could distinguish between serum samples from donors that provided SARS-CoV-2-positive (coronavirus disease 2019 [COVID-19]) and negative samples in RT-PCR with a high level of significance, as indicated by low p values (Table S5). The p values were lowest for the N and S proteins for both IgA and IgG. IgM produced the least-specific responses (Figure S4A; Table S5), mostly explained by the appearance of unspecific staining in some samples. For IgA against the SARS-CoV-2 M protein, rarely employed in ELISA tests, our assay gave a high signal (i.e., positive ratio) for samples taken within less than 2 weeks of the positive RT-PCR test result (Figure 4A). Based on these findings, the assay enables

CellPress
OPEN ACCESS



(legend on next page)

simultaneous determination of different Ig class responses against multiple Ags. This is beneficial (1) for separation of SARS-CoV-2-positive and -negative cases, (2) for retrospective timing of development of the infection and immunity against it, and (3) for distinguishing between vaccine- and infection-induced Ig responses.

The assay performance, compared with an expert opinion, results in high AUC values

An indirect IFA is commonly carried out on virus-infected cells fixed on glass slides, for which the findings rely on an expert's visual inspection under a fluorescence microscope. To evaluate the quality of our per-well predictions, we first performed a comparison against the consensus finding of six experts' visual inspections. We randomly selected a balanced number of images (IgG/IgA × S/N/R plates) to represent the Ab reactivity for SARS-CoV-2 Ags of expected positive (COVID-19 patient sera) and expected negative (samples collected in 2017; Table S4) cases. Images from IgM and M plates were not included because of unspecific or unclear staining patterns. Six experienced virologists/cell biologists annotated a set of 576 images as positive, negative, or unclear, one by one (STAR Methods). 26 images for which the experts did not reach a consensus (majority vote: all IgA, with 23 opinions for IgA R Ag) were removed from the evaluation (Figure 5A). Model quality was assessed by measuring the area under the receiver operating characteristic curve (AUC) (Figure 4B), which relates to the prediction's sensitivity and specificity and is considered optimal at a value of 1. For IgG, the model performed at an AUC of 0.98, 0.97, and 0.98, and for IgA it performed at an AUC of 0.96, 0.96, and 0.89 for S, N, and R Ags, respectively.

Validation of mini-IFA provides the highest correlation with IgG against SARS-CoV-2 Ags

Next, we compared our per-well predictions with the results of SARS-CoV-2 ELISA tests (IgG, IgA, and IgM against N, R, and S; see STAR Methods for more details), performed as in Amanat et al., [2] Rusanen et al.,[3] and Stadlbauer et al.[10] The comparison used 83 samples from 45 patients (#F1a; #F1b) with confirmed SARS-CoV-2 infection detected by RT-PCR (Table S4) and 500 pre-COVID-19 samples collected in 2017 (#F2). Spearman correlations between the predicted positive ratio of our assay and the N, R, and S ELISA results were between 0.81 and 0.82 for IgG and 0.56 and 0.80 for IgA, while IgM demonstrated the lowest correlations, characterized by values ranging between 0.17 and 0.44 (Figures 4C and S4B; Table S6; STAR Methods).

To evaluate whether severely ill patients would demonstrate a specific Ab pattern, we plotted the predictive positive ratios as a function of the severity of COVID-19 disease (treated at home, hospitalized/non-intensive care unit [ICU], and hospitalized/ICU) (Figures 4D and S4C). Interestingly, ICU patients had higher anti-N protein IgG and IgA levels compared with the other groups; however, no definite conclusions can be drawn from these findings due to the low number of samples.

Further evaluation of the assay quality with conventional IFA and ELISA shows similarity to the conventional IFA assay

To demonstrate the quantitative possibilities of the assay, we compared the positive ratios generated from the mini-IFA assay with the titer values of two conventional IFA methods using either (1) cells infected with SARS-CoV-2 or (2) cells transiently transfected with SARS-CoV-2 S Ag (Figures 5B and 5C). To compare the methods, we selected 38 SARS-CoV-2-positive samples (confirmed with PCR), of which 33 were ELISA positive and 5 ELISA negative for anti-SARS-CoV-2 IgG Abs. We observed that the mini-IFA assay resulted in an estimative quantitative value similar to the titer values in conventional IFAs, although higher amounts of samples would need to be tested for more definite conclusions. Altogether, our findings highlight the flexibility and the robustness of the mini-IFA method as well as the utility of the produced data for serological analysis.

DISCUSSION

The automated mini-IFA method enables high-throughput screening of Abs from a small volume of serum with good correlation to ELISA and traditional IFAs, as exemplified here by analysis of serum specimens from SARS-CoV-2-infected patients. The ability of the mini-IFA to simultaneously detect three Ig classes and immune responses against multiple Ags can improve diagnostic accuracy compared with single-Ag tests. Ag presentation by transient expression in cells allows (1) execution of the assay in a laboratory with a low BSL and (2) use of complex Ags, including those of any new virus variants, without the need for protein purification; thus, it has the potential to overcome the limitations of conventional IFAs (for 1) and ELISA (for 2). In mini-IFA, the cell segmentation is based on a well-established algorithm, and the pre-defined features ensure the specificity of the deep convolutional neural network. All codes are shared; however, it is essential to note that a new MLmodel needs to be trained for each Ag. Here, we labeled, on average, 4,500 cells in each Ag-Ab combination, which took an expert approximately a workday.

The cell-based format used here was considered the most physiological way to express the Ags, which is not possible with ELISA or bead-based immunoassays. The benefit of the assay format over ELISA or similar methods is that the assay can be set up without purifying the expressed Ag because cells expressing the Ag are enough for mini-IFA. The additional benefits could include higher local Ag concentration, better folding (although potentially distorted by fixation), and detection of auto-Abs as well as all of the feature-based information the images can provide for scientific studies.

Figure 3. Digitalized virology with quality control and data visualization

The mini-IFA pipeline offers a comprehensive quality control (QC) report, assessing the technical performance of each assay and displaying inter- and intra-assay variation. The QC report, prediction results, and microscopy images are digitally visualized and can be reviewed by an expert for an additional QC step if needed, as well as for reporting of the results.

CellPress
OPEN ACCESS

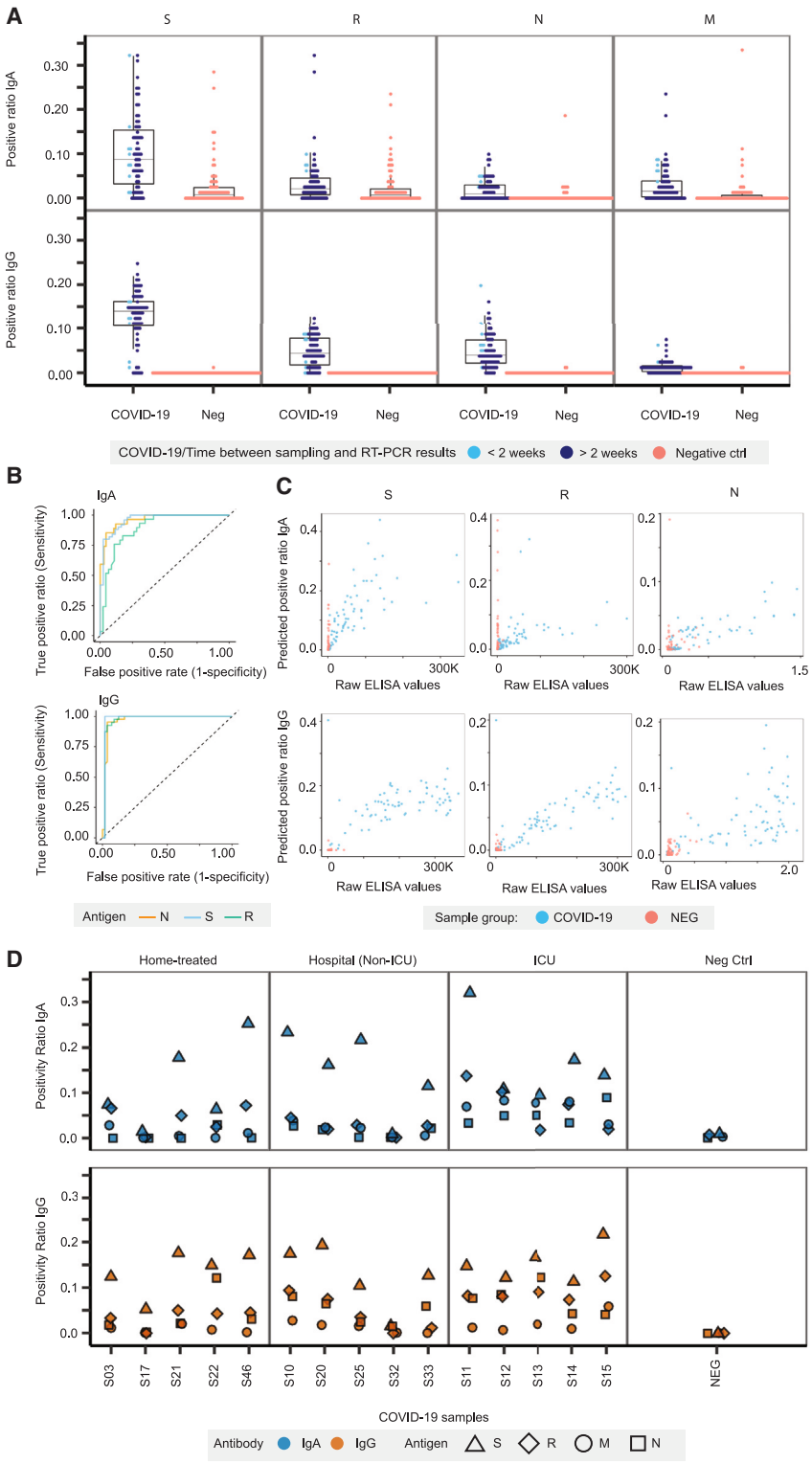


Figure 4. Assay performance

(A) The dot plots show the distribution of sample prediction values of IgA and IgG Ab responses for S, R, N, and M Ags present in the serum samples obtained from COVID-19-positive (COVID-19; collected in 2020) and negative (neg; collected in 2017) patients. The sera obtained from COVID-19+ patients are marked to show the time from a SARS-CoV-2 positive RT-PCR result. See Table S5 for the statistics.

(B) Comparison of the predicted values (IgG and IgA for N, S, and R Ags) with expert data, displayed as ROC curves.

(C) Correlation of the assay-predicted values (positivity ratios) with those of ELISA for IgA and IgG against the S and R Ags (high-throughput ELISA: n = 46 COVID-19 patients, 80 samples; neg: n = 80 patients, 80 samples) and against the N Ag (traditional ELISA: n = 46 COVID-19 patients, 79 samples; neg: n = 80 patients, 80 samples). ELISA values for S/R vs. N Ag differ in scale because they were obtained via chemiluminescence vs. colorimetric detection, respectively. See Table S6 for the statistics.

(D) Patient-specific Ab responses at different time points from onset of symptoms or from obtaining a positive SARS-CoV-2 RT-PCR result, plotted according to COVID-19 severity (treated at home, hospitalized/non-ICU, and hospitalized/ICU).

The mini-IFA pipeline provides several QC steps lacking from other serological methods. For example, we can follow up on the nuclear intensity (providing information on the accuracy of the serum transfer to the assay plates) and cell amount, as well as inter- and intra-assay variability (with positive, negative, and transfection controls).

A similar approach has been suggested[1] and tested[11] previously for SARS-CoV-2 using virus-infected cells grown under biosafety level 3 (BSL-3) conditions. There, the analysis focuses on scoring the ratio of the median Ab signal between infected and non-infected cells and using this ratio in receiver operating characteristic (ROC) analysis to set an optimal threshold to deem sample either positive or negative for IgG, IgA, and IgM Abs separately. However, in the setup of Pape et al.,[1] the response against the virus as such does not reveal any specific responses to different virus Ags, whereas in our assay, the measurements of different virus Ags can be used to differentiate vaccinated (showing Igs only toward S /R Ag used in the vaccine) individuals from those that suffered the disease (also showing Igs to other virus proteins).

Another benefit of our methodology is the simplified visualization of the results, which would enable a straightforward clinical application. End users of the assay can easily compare the visualizations of the semi-quantitative results expressed as positivity ratios with sample images. This facilitates digital archiving, reduces bias caused by intra-/interobserver variability, reduces microscopy workload/time, and

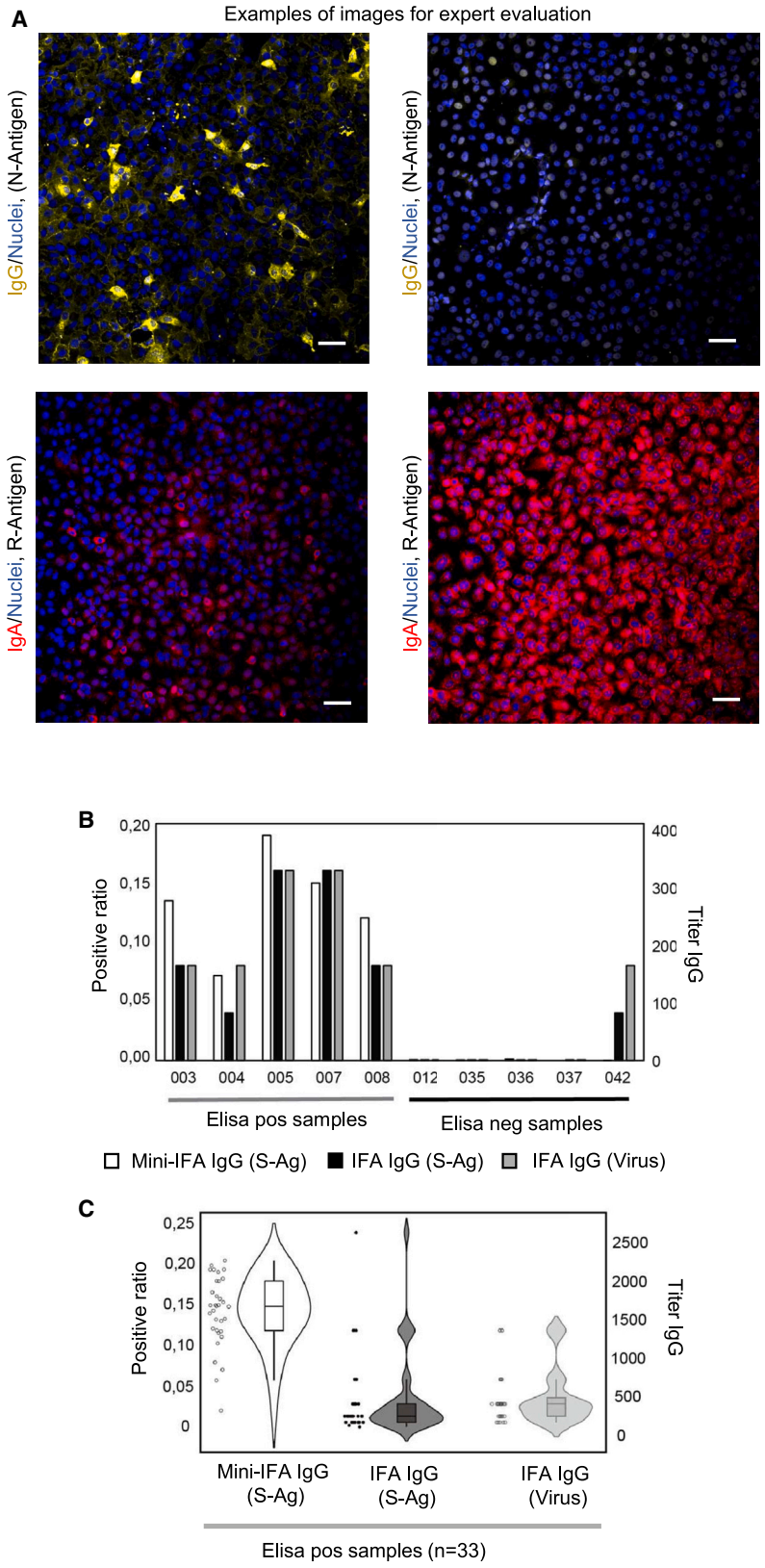


Figure 5. Assay validation and comparison with other methods

(A) Examples of expert annotation task images. Experts were asked to annotate a total of 576 images as Ab positive (pos) or negative (neg). In case they were uncertain, the image was marked as unclear. Top left: image of IgG Ab/N, annotated as pos with full consensus by the experts. Top right: image of IgG Ab/N, annotated as neg with full consensus by the experts. Bottom left: image of IgA Ab/R-Ag with mixed expert opinions. Three experts annotated the image as neg, two as pos, and one as unclear. Bottom right: image of IgA Ab/R Ag annotated as unclear by all experts and, thus, removed from the evaluation. Scale bar, 50 μm.

(B) A comparison between mini-IFA (using cells transiently transfected with S Ag) and two conventional IFA methods (utilizing cells infected with SARS-CoV-2 and cells transfected with S Ag) was performed. Ten patient samples were selected for the analysis based on their pos SARS-CoV-2 RT-PCR results. Among these samples, five also tested pos for IgG Abs against S Ag in ELISA, and five samples tested as neg in ELISA. The barplots represent IgG values either as the positive ratio obtained from the mini-IFA assay or as titer values obtained from the conventional IFAs.

(C) The distribution of data points obtained from three different immunoassays—the mini-IFA (using cells transiently transfected with S Ag) and two conventional IFA methods (either assay with SARS-CoV-2-infected cells or assay with S Ag-transfected cells)—was assessed. The results are presented as violin plots depicting the distribution patterns. The analysis included a total of 33 samples that tested pos for SARS-CoV-2 in RT-PCR and exhibited IgG reactivity in ELISA against the S Ag. The data points in the violin plots represent IgG values either as the positive ratio obtained from the mini-IFA assay or as titer values obtained from the conventional IFAs.

increases the reproducibility of the results. Thus, the assay offers unique benefits compared with similar methods utilized for virus diagnostics, which use infectious viruses, manual immunostaining of slides, and visual inspection by an expert. The assay can be used, for example, for kinetics and longevity studies of Ab responses induced by infection, and it could serve as a useful post-vaccination tool for SARS-CoV-2 serosurveillance studies. Moreover, the assay could work with virus-infected cells in the early stages of any pandemic before the relevant viral Ags are cloned. However, here we optimized the method for Ags produced via transfection of plasmids carrying SARS-CoV-2 structural proteins, allowing detection of Ab response against individual Ags and working in a low BSL, something not achievable using, e.g., traditional IFAs.

The combination of high-throughput imaging and ML makes the proposed assay an appealing alternative to existing serological assays by providing robustness and rapid setup combined with reasonable sensitivity. The richness of the data derived from the images of Ab-Ag interactions gives additional information to other immunoassays; e.g., for studies of the immune responses to specific virus proteins, some of which cannot be expressed in any other system. This is beneficial for any vaccine development and understanding the immunogenic responses to infection.

As exemplified by SARS-CoV-2, readily available, low-cost serology tests are key to control epidemics. Our proof-of-concept study for automated mini-IFA serodiagnostics, presented here, suggests that the assay translates to various pathogens, with high potential for wide-scale research and clinical applications.

Limitations of the study

We set up the assay at an early phase of the SARS-CoV-2 pandemic; because of this, the patient samples used for estimating assay performance represent the first wave of SARS-CoV-2 in Finland (early 2020). Because of the number of samples analyzed, we cannot draw firm conclusions on the sensitivity of the assay against ELISA, even though we found the mini-IFA results to be in good agreement with ELISA results. Furthermore, we only applied a couple of serum dilutions for the mini-IFA, because of which we were unable to estimate patient titers in the mini-IFA vs. the respective ELISA. Testing patient samples collected in later waves of the pandemic could allow estimating the assay's competence against newer SARS-CoV-2 variants as well as completing a through comparison of the sensitivity and specificity of the mini-IFA with traditional methods such as ELISA.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Cell lines and plasmids
  - Patient serum samples and ethical permissions
- METHOD DETAILS
  - Automated mini-IFA assay
  - Transfection
  - Immunostaining
  - High-content imaging
  - Semi-automated mini-IFA assay
  - Image processing and feature extraction
  - Random Forest (RF)
  - Cross-validation
  - Visual inspection data
  - Comparison to visual inspection data
  - Comparison to ELISA
  - Comparison to conventional IFA
  - QC and visualization of the results
  - Evaluation of the semi-automated assay results
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistical analysis was performed with Python scripts
  - Enzyme-linked immunosorbent assay (ELISA)
  - Conventional immunofluorescence assay (IFA)
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.crmeth.2023.100565.

ACKNOWLEDGMENTS

The authors thank the Minerva Institute (Helsinki, Finland) for providing utilities for the project, Prof. Perttu Hämäläinen (Aalto University, Finland) for providing the expertise of his group for the project, the FIMM High Throughput Biomedicine Unit for providing access to high-throughput robotics, the FIMM High Content Imaging and Analysis Unit for HC imaging and analysis (HiLIFE, University of Helsinki and Biocenter Finland; EuroBioImaging, ISIDORe partner), and the CSC – IT Center for Science, Finland, for computational resources. We acknowledge support from the LENDULET-BIOMAG grant (2018-342), from the European Regional Development Funds (GINOP-2.3.2-15-2016-00006, GINOP-2.3.2-15-2016-00026, and GINOP-2.3.2-15-2016-00037), from the H2020-discovAIR (874656), from the H2020 ATTRACT-Spheroid-Picker, and from the Chan Zuckerberg Initiative, Seed Networks for the HCA-DVP. The Finnish TEKES/BusinessFinland FiDiPro Fellow Grant 40294/13 (to V.P., O.K., L.P., and P.H.), grants awarded by the Academy of Finland (iCOIN-336496 to O.K., V.P., and O.V.; 308613 to J.H.; 321809 to T.S.; 310552 to L.P.; 337530 to I.J.; and FIRI2020-337036 to FIMM-HCA, A.H., L.P., V.P., and P.H.), the EU H2020 VEO project (O.V.), and a Minerva Foundation for COVID-19 Research project grant (to V.P.) are also acknowledged. C.G. is funded by the Academy of Finland Flagship program, Finnish Center for Artificial Intelligence. OrthoSera Ltd. was funded by NKFIH grants (2020-1.1.6-JÖVŐ-2021-00010 and TKP2020-NKA-17). The authors thank Dora Bokor, PharmD, for proofreading the manuscript.

AUTHOR CONTRIBUTIONS

J.H., P.H., and V.P. originally designed and led the project, and K. Buzas, J.H., P.H., and V.P. supervised the project. J.H., E.M., V.P., M.P., and L.T. designed the laboratory experiments, and J.H., M.H., A. Koponen, M.P., E.G.S., and L.T. executed the experiments. L.L., I.J. and L. Kakkola provided the constructs. K. Burian, A. Kantele, L. Kemeny, G.T., A.V., and O.V. provided serum samples. E.F. and Z.L. collected 165 positive sera with patient data and performed

ELISAs in Hungary in co-operation with Orthosera Ltd. A.H., E.M., L.P., and M.P. performed high-content imaging, and J.H., I.J., P.K., L.L., E.G.S., and T.S. analyzed IFA samples. C.G., F.K., and L.P. developed the analysis pipeline. C.G. and L.P. facilitated the prediction model selection and its comparison with ELISA and expert annotations. P.H., A. Kriston, and F.K. developed the image analysis software (BIAS). R.H. prepared the model for nuclei detection and segmentation. S.P. prepared the QC and visualization scripts. A.D., E.M., and M.P. performed expert training. C.G., J.H., M.H., P.H., L.P., M.P., and V.P. wrote the main parts of the manuscript. A.D., C.G., M.H., L.P., M.P., S.P., V.P., and L.T. prepared the figures, tables, and data files included in the paper. All authors read, reviewed, and approved the final version of the manuscript.

DECLARATION OF INTERESTS

P.H. is the founder and shareholder and A. Kriston and F.K. are employees of Single-Cell Technologies Ltd. This study has been protected with invention disclosures (ID965/2020 and ID115/2021 University of Helsinki, Finland), and the patent application has been filed (Hungary, no. P2100295).

Received: December 16, 2022
Revised: July 25, 2023
Accepted: July 31, 2023
Published: August 22, 2023

REFERENCES

1. Pape, C., Remme, R., Wolny, A., Olberg, S., Wolf, S., Cerrone, L., Cortese, M., Klaus, S., Lucic, B., Ullrich, S., et al. (2021). Microscopy-based assay for semi-quantitative detection of SARS-CoV-2 specific antibodies in human sera: A semi-quantitative, high throughput, microscopy-based assay expands existing approaches to measure SARS-CoV-2 specific antibody levels in human sera. Bioessays *43*, e2000257. https://doi.org/10.1002/bies.202000257.
2. Amanat, F., Stadlbauer, D., Nguyen, T.H.O., Strohmeier, S., Chromikova, V., McMahon, M., Jiang, K., Arunkumar, G.A., Jurczyszak, D., Polanco, J., et al. (2020). A serological assay to detect SARS-CoV-2 seroconversion in humans. Nat. Med. *26*, 1033–1036. https://doi.org/10.1038/s41591-020-0913-5.
3. Rusanen, J., Kareinen, L., Levanov, L., Mero, S., Pakkanen, S.H., Kantele, A., Amanat, F., Krammer, F., Hedman, K., Vapalahti, O., and Hepojoki, J. (2021). A 10-Minute "Mix and Read" Antibody Assay for SARS-CoV-2. Viruses *13*, 143. https://doi.org/10.3390/v13020143.
4. Daly, J.L., Simonetti, B., Klein, K., Chen, K.E., Williamson, M.K., Antón-Plágaro, C., Shoemark, D.K., Simón-Gracia, L., Bauer, M., Hollandi, R., et al. (2020). Neuropilin-1 is a host factor for SARS-CoV-2 infection. Science *370*, 861–865. https://doi.org/10.1126/science.abd3072.
5. Mund, A., Coscia, F., Hollandi, R., Kovács, F., Kriston, A., Brunner, A.-D., Bzorek, M., Naimy, S., Rahbek Gjerdrum, L.M., Dyring-Andersen, B., et al. (2021). AI-driven Deep Visual Proteomics defines cell identity and heterogeneity. Preprint at bioRxiv, accepted for publication in Nature Biotechnology. https://doi.org/10.1101/2021.01.25.427969.
6. Smith, K., Li, Y., Piccinini, F., Csucs, G., Balazs, C., Bevilacqua, A., and Horvath, P. (2015). CIDRE: An illumination-correction method for optical microscopy. Nat. Methods *12*, 404–406. https://doi.org/10.1038/nmeth.3323.
7. Hollandi, R., Szkalisity, A., Toth, T., Tasnadi, E., Molnar, C., Mathe, B., Grexa, I., Molnar, J., Balind, A., Gorbe, M., et al. (2020). nucleAIzer: A Parameter-free Deep Learning Framework for Nucleus Segmentation Using Image Style Transfer. Cell Syst. *10*, 453–458.e6. https://doi.org/10.1016/j.cels.2020.04.003.
8. Piccinini, F., Balassa, T., Szkalisity, A., Molnar, C., Paavolainen, L., Kujala, K., Buzas, K., Sarazova, M., Pietiainen, V., Kutay, U., et al. (2017). Advanced Cell Classifier: User-Friendly Machine-Learning-Based Software for Discovering Phenotypes in High-Content Imaging Data. Cell Syst. *4*, 651–655.e5. https://doi.org/10.1016/j.cels.2017.05.012.
9. Smith, K., and Horvath, P. (2014). Active learning strategies for phenotypic profiling of high-content screens. J. Biomol. Screen *19*, 685–695. https://doi.org/10.1177/1087057114527313.
10. Stadlbauer, D., Amanat, F., Chromikova, V., Jiang, K., Strohmeier, S., Arunkumar, G.A., Tan, J., Bhavsar, D., Capuano, C., Kirkpatrick, E., et al. (2020). SARS-CoV-2 Seroconversion in Humans: A Detailed Protocol for a Serological Assay, Antigen Production, and Test Setup. Curr. Protoc. Microbiol. *57*, e100. https://doi.org/10.1002/cpmc.100.
11. Tönshoff, B., Müller, B., Elling, R., Renk, H., Meissner, P., Hengel, H., Garbade, S.F., Kieser, M., Jeltsch, K., Grulich-Henn, J., et al. (2021). Prevalence of SARS-CoV-2 Infection in Children and Their Parents in Southwest Germany. JAMA Pediatr. *175*, 586–593. https://doi.org/10.1001/jamapediatrics.2021.0001.
12. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: machine learning in python. J. Mach. Learn. Res. *12*, 2825.
13. Cohen, J. (1960). Coefficient of Agreement for Nominal Scales. Educ. Psychol. Meas. *20*, 37–46. https://doi.org/10.1177/001316446002000010.
14. Fleiss, J.L. (1971). Measuring nominal scale agreement among many raters. Psychol. Bull. *76*, 378–382. https://doi.org/10.1037/h0031619.
15. Regier, D.A., Narrow, W.E., Clarke, D.E., Kraemer, H.C., Kuramoto, S.J., Kuhl, E.A., and Kupfer, D.J. (2013). DSM-5 Field Trials in the United States and Canada, Part II: Test-Retest Reliability of Selected Categorical Diagnoses. Am. J. Psychiatr. *170*, 59–70. https://doi.org/10.1176/appi.ajp.2012.12070999.
16. Potdar, S., Ianevski, A., Mpindi, J.P., Bychkov, D., Fiere, C., Ianevski, P., Yadav, B., Wennerberg, K., Aittokallio, T., Kallioniemi, O., et al. (2020). Breeze: an integrated quality control and data analysis application for high-throughput drug screening. Bioinformatics *36*, 3602–3604. https://doi.org/10.1093/bioinformatics/btaa138.
17. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag)978-3-319-24277-4.
18. Sievert, C. (2020). Interactive Web-Based Data Visualization with R, Plotly, and Shiny (Chapman and Hall/CRC)9781138331457.
19. Yang, H.S., Costa, V., Racine-Brzostek, S.E., Acker, K.P., Yee, J., Chen, Z., Karbaschi, M., Zuk, R., Rand, S., Sukhu, A., et al. (2021). Association of Age With SARS-CoV-2 Antibody Response. JAMA Netw. Open *4*, e214302. https://doi.org/10.1001/jamanetworkopen.2021.4302.
20. Hjelholt, A., Christiansen, G., Sørensen, U.S., and Birkelund, S. (2013). IgG subclass profiles in normal human sera of antibodies specific to five kinds of microbial antigens. Pathog. Dis. *67*, 206–213. https://doi.org/10.1111/2049-632X.12034.
21. Sui, Y., Bekele, Y., and Berzofsky, J.A. (2021). Potential SARS-CoV-2 Immune Correlates of Protection in Infection and Vaccine Immunization. Pathogens *10*, 138. https://doi.org/10.3390/pathogens10020138.
22. Long, Q.X., Tang, X.J., Shi, Q.L., Li, Q., Deng, H.J., Yuan, J., Hu, J.L., Xu, W., Zhang, Y., Lv, F.J., et al. (2020). Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. Nat. Med. *26*, 1200–1204. https://doi.org/10.1038/s41591-020-0965-6.
23. Grossberg, A.N., Koza, L.A., Ledreux, A., Prusmack, C., Krishnamurthy, H.K., Jayaraman, V., Granholm, A.C., and Linseman, D.A. (2021). A multiplex chemiluminescent immunoassay for serological profiling of COVID-19-positive symptomatic and asymptomatic patients. Nat. Commun. *12*, 740.
24. Haveri, A., Smura, T., Kuivanen, S., Österlund, P., Hepojoki, J., Ikonen, N., Pitkäpaasi, M., Blomqvist, S., Rönkkö, E., Kantele, A., et al. (2020). Serological and molecular findings during SARS-CoV-2 infection: the first case study in Finland, January to February 2020. Euro Surveill. *25*, 2000266. https://doi.org/10.2807/1560-7917.ES.2020.25.11.2000266.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit anti-SARS-CoV-2 RBD-His (antiserum)	Rusanen et al. 2021[3]	N/A
Rabbit anti-SARS-CoV-2 NP-His (antiserum)	Rusanen et al. 2021[3]	N/A
Goat anti-Human IgG Fc Cross-Adsorbed Secondary Antibody, DyLight 550	Invitrogen	CAT#SA5-10135, RRID:AB_2556715
Goat anti-Human IgM (Heavy chain) Alexa Fluor 488	Invitrogen	CAT#A-21215, RRID:AB_2535800
Alexa Fluor 647 AffiniPure Goat Anti-Human Serum IgA, α-chain specific	Jackson ImmunoResearch	CAT#109-605-011, RRID: AB_2337883
Mouse monoclonal anti-His-tag antibody, Alexa Fluor 647	BioLegend	CAT#652513, RRID:AB_2716153
Mouse monoclonal anti-HA-tag antibody 16B12, Alexa Fluor 488	Invitrogen	CAT#A-21287, RRID:AB_2535829
Goat anti-rabbit IgG(H + L) cross-adsorbed secondary antibody, Alexa Fluor 488	Invitrogen	CAT#A-11008, RRID:AB_143165
Goat anti-Human IgG (H + L) Cross-adsorbed secondary antibody, Alexa Fluor 546	*Invitrogen*	CAT#*A-21089*, RRID:AB_2535745
Mouse monoclonal anti-6x-His Tag antibody (HIS.H8)	*Invitrogen*	CAT#*MA1-21315*, RRID:AB_557403
Goat anti-Mouse IgG (H + L) Highly cross-adsorbed secondary antibody, Alexa Fluor Plus 488	*Invitrogen*	CAT#*A32723*, RRID:AB_2633275
Goat anti-Mouse IgG (H + L) highly cross-adsorbed secondary antibody, Alexa Fluor Plus 647	*Invitrogen*	CAT#*A32728, RRID:*AB_2633277
Biological samples		
Human Serum samples	Described in the STAR Methods text with ethical permissions, see also Rusanen et al. 2021.[3]	N/A
Chemicals, peptides, and recombinant proteins		
MEM: Minimal Essential Medium Eagle	Sigma	CAT#M2279-500ML
L-glutamine 200 mM	Gibco® Life Technologies, USA	CAT#25030-024
Fetal Bovine Serum; FBS	Gibco® Life Technologies, USA	CAT#10270-106
Penicillin-Streptomycin (5,000 U/mL penicillin, 5000 μg/mL streptomycin)	Gibco® Life Technologies, USA	CAT#15070-063
Penicillin-Streptomycin (10,000 U/mL) (10,000 U/mL penicillin, 10,000 μg/mL streptomycin)	Gibco® Life Technologies, USA	CAT#15140-122
0.25% Trypsin-EDTA	Gibco® Life Technologies, USA	CAT#25200-072
S-pCAGGS, SARS-CoV-2 spike protein/plasmid	Amanat et al. 2020[2]	N/A
RBD-pCAGGS, SARS-CoV-2 receptor-binding domain/plasmid	Amanat et al. 2020[2]	N/A
NP-pCAGGS, SARS-CoV-2 nucleoprotein/plasmid	Rusanen et al. 2021[3]	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
M-pEBB, SARS-CoV-2 membrane protein/plasmid	hCoV-19/Finland/1/2020 (GenBank accession MT020781) with pEBB-N-HA mammalian expression plasmid. Synthetized and cloned to the plasmid by GeneArt/Thermo Fisher Scientific	N/A
FuGENE HD	Promega	CAT#E2312
OptiMEM I Reduced Serum Medium	Gibco® Life Technologies, USA	CAT#31985070
Hoechst 33342	Invitrogen	CAT#H1399
PBS tablets	Medicago	CAT#09-9400-100
TBS tablets	Medicago	CAT#09-7500-100
Tween 20	Sigma	CAT#P9416-100ML
BSA (Bovine Serum Albumin)	Biowest	CAT#P6154-100G
Triton X-100	Sigma	CAT#T8787-100ML
DAPI	*PanReac AppliChem*	CAT#*A1001*
DPBS (10X), calcium, magnesium	*Gibco® Life Technologies, USA*	CAT#*14080048*
Pierce™ 20X TBS Buffer	*Thermo Scientific*	CAT#*28358*
Bovine Serum Albumin	*Sigma-Aldrich*	CAT#*A4503-100G*
OmniPur Triton X-100 Surfactant - Calbiochem	*Millipore*	CAT#*9410-1L*
Deposited data		
Extracted single-cell features from microscopic images	This paper	https://doi.org/10.5281/zenodo.6352550
Experimental models: Cell lines		
Vero E6, African green monkey kidney cell line	ATCC	N/A
Software and algorithms		
BIAS	Single Cell Technologies	https://single-cell-technologies.com/bias-2/
Source code	This paper	https://doi.org/10.5281/zenodo.8158791
Other		
T-75 cell culture flask	CELLSTAR®, Greiner Bio-one	CAT#658175
Cellcarrier ultra (PE) 384 imaging plates	PerkinElmer	CAT#6057302
T-75 cell culture flask	*Corning*	CAT#*430641U*
384 Well Microplate, PS, μClear®	*Greiner Bio-One*	CAT#*781090*
96 Well Microplate, PS, μClear®	*Greiner Bio-One*	CAT#*655090*
Labcyte 384PP (Echo® Qualified 384-Well Polypropylene Source Microplate)	Labcyte	CAT#PP-0200
Costar® 50 mL Reagent Reservoirs	Corning	CAT#4871
5 μL cassette MultiFlo FX	Biotek	N/A
8 channel 5 μL dispensing EL406 cassette (1260016)	Biotek	N/A
"Brad" BioTek MultiFlo FX dispenser	Biotek	N/A
"Elmeri" EL406 -plate washer/dispenser	Biotek	N/A
"Echo" Echo 525 -acoustic dispenser	Labcyte	N/A
Certus FLEX 0.3 (air pressure)	Gyger	N/A
Opera Phenix confocal microscope	PerkinElmer	https://www2.helsinki.fi/en/infrastructures/bioimaging/fimm-hca
ThermoScientific SL40R	Thermo Fisher	ThermoScientific SL40R
Biomek FXp	BeckmanCoulter	Biomek FXp
Operetta CLS High-Content Analysis System	*PerkinElmer*	N/A
TIPOR-M+ 8 Channel micropipettor	*Orange Scientific*	N/A
Research plus 12 channel pipette	*Eppendorf*	N/A

CellPress
OPEN ACCESS

RESOURCE AVAILABILITY

Lead contact

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request (vilja.pietiainen@helsinki.fi).

Materials availability

This study did not generate new unique reagents. The detailed info of all materials and reagents used is given in the key resources table. The automated assay pipeline protocol is described in Methods S1.

Data and code availability

- All of the data reported in this paper will be shared by the lead contact upon request.
- Source code is available at GitHub repository: https://github.com/fimm-covid-19-hca/mini-IFA_paper.git. Extracted single-cell features from microscopic images to train and test models are available at Zenodo: 10.5281/zenodo.6352549. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell lines and plasmids

The study made use of the following plasmids for recombinant expression of three SARS-CoV-2 proteins and receptor binding domain (RBD; or R) of the spike (S) protein: S protein with His-tag (S-pCAGGS; described in[2]), RBD with His-tag (RBD-pCAGGS; described in Amanat et al.[2]), membrane (M) protein with HA-tag (M-pEBB; HA-Tag) and nucleoprotein (NP; or N) with His-tag (NP-pCAGGS; described in Rusanen et al.[3]) (see also key resources table). For M-pEBB, M gene, according to hCoV-19/Finland/1/2020 (GenBank accession MT020781) sequence, was synthesized and cloned into pEBB-N-HA mammalian expression vector by GeneArt (Thermo Fisher Scientific).

Vero E6, African green monkey kidney cell line (ATCC; mycoplasma tested), were grown in Minimal Essential Medium Eagle (SIGMA, US) supplemented with 10% fetal bovine serum (Gibco, US), 2 mM L-glutamine (Gibco), 100 IU/mL penicillin and 100 μg/mL streptomycin (Gibco/Sigma). Vero E6 cells were cultured to a 90–95% confluence in T-75 cell culture flasks and detached using 0.25% Trypsin-EDTA (Gibco).

Patient serum samples and ethical permissions

Information of patient samples is provided in Table S4. For automated IFA-assay (Helsinki, Finland), the SARS-CoV-2 patient sample panels 1&2 comprised of 83 blood samples (serum, plasma) drawn from 45 individuals, of whom 42 had been tested positive in clinically validated SARS-CoV-2 RT-PCR test (HUSLAB, Helsinki University Hospital Laboratory Diagnostics); the panel also included three samples from COVID-19 suspects with negative/no result in SARS-CoV-2 RT-PCR. The RT-PCR testing was performed from nasopharyngeal swab samples. The serum samples from COVID-19 patients were drawn 8 to 81 days after onset of symptoms. In their research article, Rusanen et al.[3] established rapid LFRET immunoassays using partially overlapping samples. We are unable to provide information on the consideration of age, gender or ethnicities in recruitment due to the anonymization of the relevant data. The data and samples for this panel were collected under research permit HUS/211/2020 and ethics committee approval HUS/853/2020 and HUS/1238/2020 (Helsinki University Hospital, Finland). A panel of 500 serum samples collected during 2017 from patients with suspected Puumala virus infection served as SARS-CoV-2 antibody negative controls (Research permits HUS/167/2016, HUS/38/2018 and HUS/244/2021). Written informed consent was obtained from all participants. The study was conducted in accordance with the Declaration of Helsinki.

For semi-automated mini-IFA assay, a total of 200 negative control sera and 165 positive samples derived from 145 patients were provided by the Hungarian National Blood Transfusion Service, the Department of Dermatology and Allergology of the University of Szeged and the Orthosera Ltd. We are unable to provide information on the consideration of gender or ethnicities in recruitment due to the anonymization of the relevant data. The age range of the participants was 16–96, with the median being at age 43 (see Figure S7A). Samples predicted positive were collected from 139 RT-PCR-and/or ELISA-confirmed patients and from 6 further patients having symptoms or tested by another method under general informed consent. The sera were stored at −80°C and heat inactivated at 56°C for 60 min before the assay.

Sample collection and this study was approved by the Scientific and Research Ethics Committee of the Hungarian Medical Research Council (clearance numbers IV/3457-2/2020/EKU and IV/3757-4/2020/EKU). The study was conducted in accordance with the Declaration of Helsinki.

METHOD DETAILS

Automated mini-IFA assay

The assay pipeline protocol is provided in Methods S1

Transfection

Vero E6 cells were transfected using FuGENE HD (Promega, US) at 3.5:1 reagent to plasmid ratio. The FuGENE HD-plasmid mixes were prepared in Opti-MEM (Gibco) following the manufacturer's protocol, after 10–15 min complex formation, a suspension of trypsinized Vero E6 cells in fully supplemented media was added, the resulting suspension was incubated at RT for 15–30 min in rotation, and the cells were seeded onto plates either manually or with automation.

In the automated assay, the transfected cells were seeded to CellCarrier Ultra 384-well microplates (PerkinElmer, US) with a seeding density of approximately 3–4 x 10^3 cells/well in 25 μL per well using MultiFlo FX RAD dispenser (BioTek, US) and incubated at 37°C humidified atmosphere for 48 h in the growth media. Cells were fixed in 4% PFA for 15 min at RT, washed twice with PBS, and filled with fresh PBS. The plates were stored at 4°C prior to use. Liquid dispensing and aspiration were performed using EL406 -plate washer/dispenser (Biotek).

Immunostaining

Washing, liquid dispensing and aspiration were performed with EL406, buffer dispensing with Certus FLEX 0.3 (Gyger) and serum dispensing with Echo 525 (LabCyte). Cell permeabilization and blocking of the unspecific binding were performed using 0.25% Triton X-100 in Tris-buffered saline (TBS) (0.05 M Tris-HCl, 0.15M NaCl, pH 7.6), containing 3% bovine serum albumin (Biowest, US) for 15 min at RT. Cells were washed three times with TBS before serum samples were applied on wells. Two different dilutions of each serum sample were used; 1:25 (for detection of IgA; IgM) and 1:100 (for detection of IgG) in 0.5% BSA in TBS with 1:5000 dilution of Hoechst 33342 (Lifetech) and incubated at RT for 1.5 h. After washing steps, the fluorochrome-conjugated secondary antibodies, goat anti-human IgG DyLight 550 (1:500; Invitrogen, US), goat anti-human IgM AlexaFluor (AF) 488 (1:1000, Invitrogen) and goat anti-human Serum IgA AF647 (1:1000; Jackson ImmunoResearch, US), were used for detecting different immunoglobulins from patient sera by incubating them in the wells for 45 min at RT. To determine the transfection percentage, mouse anti-His-tag antibody AF647 (1:1000, BioLegend, US; for S, N, and R-protein transfections), mouse anti-HA tag AF488 (1:1000, Invitrogen; for M-protein transfection) and in-house rabbit anti-N and anti-S/R stained with AF488 secondary antibody (1:1000, Invitrogen) were used. All assay plates had five wells reserved for the transfection controls. For more accurate calculations of transfection efficacy, a parallel set of plates with transfected Vero E6 cells were stained with antibodies for each antigen as described above, and a transfection control and mean transfection percentages were calculated from images captured as described below. The transfection percentage was determined by selecting a mean intensity cut-off for immuno-stained transfected cells and calculating the percentage of transfected cells of all cells. The transfection rates for N, R, S, and M are shown in Table S1A.

High-content imaging

Opera Phenix confocal spinning-disk high content screening microscope (PerkinElmer, Inc., Waltham, MA, USA) was used for imaging of 384-well plates for automated IFA-assay (High Content Imaging and Analysis Unit, FIMM, HiLIFE, University of Helsinki, Finland). Screening was conducted with a 20x water immersion objective (NA 1.0, working distance 1.7 mm, depth of focus 1.8 μm, effective xy resolution 0.66 μm) and four excitation lasers (405 nm with emission band-pass filter 435/480; ex 488, em 500/550; ex 561, em 570/630; ex 647, em 650/760). Nine fields-of-view with 5% overlap were imaged per well using two predetermined Z focus planes with laser-based autofocusing. The images were captured with two Andor Zyla sCMOS cameras (16-bit, 2160 x 2160 px, 6.5 μm pixel size).

Semi-automated mini-IFA assay

First, we tested the protocol in 96-well plate set-up for the first 40 samples, with manually transfected plates (for transfection efficiency, see Table S1A). In the manual protocol, after 15 min incubation with the transfection mix, the cells were seeded to black 96-well tissue culture plates (Greiner, Austria) with a density of 12,000 cells/well. After 48 h incubation, culture medium was removed and the cells were washed with DPBS, fixed with 4% PFA for 10 min at RT, washed again with DPBS twice and stored in DPBS at 4°C. For the manual protocol, 8- and 12-channel pipettes and 50 mL reagent reservoirs were used.

Additional 325 samples were investigated on ready-to-go plates produced by the automated method. Permeabilization was performed in 3% BSA with 0.2% Triton X-100 in TBS for 5 min at RT and blocking was done in 5% BSA in TBS for 60 min at RT. Serum samples diluted in 1:25 ratio in 0.5% BSA in TBS were incubated on the plates for 90 min at RT. AF546-conjugated anti-human IgG (1:500; Invitrogen) along with AF488-conjugated anti-human IgM (1:1000; Invitrogen) or AF647-conjugated anti-human IgA (1:1000; Jackson ImmunoResearch) were applied for 120 min at RT. The 6x-His Tag mouse antibody (1:1000; Invitrogen) was used for overnight at 4°C and AFPlus488-conjugated anti-mouse IgG (1:500; Invitrogen) or AFPlus647-conjugated anti-mouse IgG (1:500; Invitrogen) was added for 60 min at RT to detect the His-tagged virus antigens (S, N, R) in the transfected cells in each well. Each antibody was diluted in 1.2% BSA in TBS. Nuclei were counterstained with 200 ng/mL DAPI for 15 min at RT. Wells were washed with TBS 3 times for 5 min between each step and covered with TBS for imaging.

Images were obtained using the Operetta high-content imaging system (PerkinElmer) with a 20x air objective (20× long WD; NA 0.45, working distance: 7.8 mm; depth of focus: 4.6 μm). Images were acquired with a Peltier cooled CCD camera (14-bit, field of view: 675 × 509 μm2; optical xy resolution: 0.67 μm). Six fields of view were acquired per well using one Z focus plane determined by the laser-based autofocus system. For detecting fluorophores, the following excitation and emission filters were used: DAPI (380/40; 445/35); Alexa Fluor 488 (475/30; 525/50); Alexa Fluor 546 (535/30; 595/70); Alexa Fluor 647 (630/20; 675/50).

Image processing and feature extraction

Images obtained by both methods described were processed using BIAS software.[5,8] Pipeline created for the analysis consisted of three main steps: 1) pre-processing of the images, 2) segmentation and 3) feature extraction. In the pre-processing, a maximum intensity projection was created from each stack of images in different focus depths. Non-uniform illumination was corrected separately for each channel using the CIDRE method.[6] Deep learning segmentation method[7] was applied to segment nuclei in images. From these nuclei regions, two additional segmented regions were defined: 1) the cells were defined by dilating nuclei regions with maximum 7 μm radius so that adjacent cells did not overlap, and 2) cytoplasm regions were defined by subtracting nuclei segmentation from the cell segmentation. Finally, morphological properties of these three different regions as well as intensity and texture features from all channels were extracted (in total 255 features) for cell classification.

Single-cell phenotyping annotations (classes, numbers, distribution in cells)

For the automated method, we employed supervised machine learning to predict five different cell types: Positive (P), Negative (N), Atypical (A) and Small Bright (S) cells as well as other artifacts that can be considered Trash (T). These classes were chosen based on clinical diagnostics classification of microscopic findings in virological IFA assay (P, N, A, T) as well as to exclude e.g., rounded or dying cells (S). Cells with evenly distributed high specific intensity across the whole cells were labeled as Positive, whilst cells with low intensity were labeled as Negative. Cells with abnormally small, bright nuclei and a strong intensity of unspecific antibody staining were considered as Small Bright, to differentiate the dying or dividing cells from the true positive cells. Cells with atypical staining patterns, for example strong nucleus intensity, were considered as Atypical. We include an ''atypical'' class to capture cells that phenotypically resemble positives, but present unusual staining patterns compared to the positive controls, e.g., strong nucleus intensity. Moreover, atypical cells show higher non-specific fluorescence, and their classification thus serves as additional quality control to highlight unspecific antibody binding. Small, segmented areas, caused by the background of the serum samples were labeled as Trash. To enable the classifiers to capture the differences in these five classes having different staining patterns, we used all extracted features, including intensity, morphological and texture-based features, and let the classifier to optimize the weighting between the features instead of running any feature selection methods.

To gather training data for our predictive model, the cells were annotated using an active learning feature within the BIAS software. We thus employed a simple machine learning setup to generate data for training a more sophisticated predictive model later on. For active learning, a Support Vector Machine (SVM) classifier within BIAS was trained on a set of initial annotations to classify samples from outside the training pool. The annotator then iteratively chose candidates from these new samples to extend the training data. The annotator then retrained the active learning model, allowing for a gradual increase in classification quality. Additional samples were added to the training set until a cross-validation accuracy of at least 0.85 per antigen and antibody was achieved. The annotations were initially performed by one expert to ensure consistency, and then checked against by other experts. We instructed the annotators to aim for a similar number of annotations in each of the five categories. Yet, the negative class includes approximately double the number of annotations than other classes, as it is the most typical class in the data. The annotators tried to focus on increasing the number of rare cases (atypical, small bright, trash) to make the training dataset more balanced.

In total, we annotated 55 496 cells across four plates from two experimental batches, covering four viral antigens and three antibody classes (Table S2). Roughly the same number of cells were annotated per antigen. The number of annotations vary slightly between plates, as sometimes more annotations were required to surpass the 0.85 cross-validation accuracy threshold. They vary more strongly between cell classes due to the frequency of specific examples in the respective samples.

Feature normalization. We included an additional quality control for feature normalization based on control wells as not all control samples were transferred correctly into plates. As the serum sample and nuclear stain (Hoechst) were transferred simultaneously, we based the quality control on the mean Hoechst signal in nuclei. We measured the mean and the standard deviation of Hoechst intensity from nuclei mean intensities in each well of a plate and ruled out all control wells that were not inside [mean - 1.0*std, mean +1.0*std] range. The range was defined by optimizing the multiplier of std using wells that were known not to have correctly transferred samples. We found that mean +- 1.3*std was the most optimal range but to make the rule stricter we decided to use 1.0 as a multiplier.

Based on the controls, we normalized the image features per-plate to allow for model training and prediction across plates with potentially varying signal strength. As a common requirement for some of our machine learning models, we moreover standardized the features to approximately zero mean and unit variance. We compared three techniques, each drawing on image features from different sets of control samples.

PosNeg: All transferred positive and negative controls

PosNegBal: The same amount of positive and negative transferred controls, based on sampling n controls from the set of transferred positive and negative controls respectively, with n being the highest number of transferred controls in both sets.

NegOnly: Only negative transferred controls

(PosNeg) allows us to obtain a data sample which is as representative as possible for the signal in positive and negative samples, respectively. However, the number of successfully transferred wells in either set might vary, which can lead to a shift of the mean toward a lower or higher signal. We hence evaluated (PosNegBal) where both types of controls contribute equally to normalization and standardization. As the signal in negative controls present mostly the background variation of the assay whereas the positive controls have more variation due to biologically meaningful differences, and because obtaining positive control samples in the first place can be more difficult, we also evaluated (NegOnly) which only leverages the set of negative controls. The best approach was determined based on its performance in per-cell predictions, assessed through cross-validation.

Training, models, and used hyper-parameters (single-cell predictions). We trained and evaluated three different types of machine learning models commonly used for the multi-class cell phenotype classification, drawing on the Python implementation in Scikit-learn.[12] For each classifier, we performed a comprehensive parameter grid search to determine the best model. Following an initial, coarse search, we fine-tuned the following model-hyperparameter configurations.

Random Forest (RF)

We search the number of decision trees, $N \in \{50,100,200,300,500\}$, to avoid an overfitting of the estimator while keeping the estimator complexity low. We moreover consider different numbers of features for each tree node $M \in \{\#features, \sqrt{(\#features)}\}$to trade-off the reduction of variance with the increase of bias.

Support Vector Machine (SVM) with a radial basis function kernel. We evaluated the regularization parameter $C \in \{0.1,1,10,50,100\}$ to achieve a good trade-off between a smooth decision surface, hence avoiding overfitting, and the correct classification of all training samples. We moreover evaluated different values of the kernel parameter $\gamma \in \{1.0,1\times [\![10]\!] \hat{}n,5\times [\![10]\!] \hat{}n \,|n = -1, \ldots,-4\}$ which controls the influence of individual training examples and thus the complexity of the decision boundary.

Multilayer Perceptron (ANN) with logistic activation functions. We compared neural network architectures with one to three hidden layers of varying size{(256,128,64),(128,64,32),(256,128),(128,64),(64,32),(32,32),(128),(64),(32)} controlling the model's complexity. We moreover evaluated regularization parameters $\alpha \in \{1.0,0.1,0.05,0.01,0.005,0.001\}$to avoid overfitting.

We avoid model bias through class imbalances by including class weights in the training. For each antigen-antibody combination, we selected the best model based on a cross-validation of cell type predictions (Table S3).

Cross-validation

For the practical use of this assay, it is essential that the trained model can produce predictions on new, previously unseen plates for which no training data is available. To evaluate the prediction quality of different model candidates for use in per-well predictions, we consequently performed a 4-fold leave-one-out cross-validation on the level of plates. It has also been used to identify the best normalization technique for cross-plate prediction.

Each cross-validation fold comprised training data of three plates and used the remaining plate as validation data. The ratio of training to validation data per fold was thus 3:1 (e.g., for S/IgG 3920:1301; Table S2). We calculated the sensitivity, specificity, and accuracy of predictions, and selected the best performing model (and normalization approach) based on the mean sensitivity×specificity across all folds. This implements a maximum ignorance assumption on the later clinical use. If an application would benefit to trade-off one factor against the other, the model can be re-trained based on a corresponding weighting of both measures. The cross-validation was performed individually for each antigen-antibody combination, and the results for all tested classifiers and normalization approaches with optimized hyperparameters (Table S7) are presented in Table S3. We also show the confusion matrices for the ANN classifier for NegOnly and PosNegBal normalization scheme in Figures S2A and S2B.

Visual inspection data

Following the single cell cross-validation analysis we proceeded to evaluate how well the selected optimal model works in the sample positivity prediction task. In this task we wanted to evaluate 1) how well the predicted positive cell ratio correlates with the ground-truth positivity measurement of the image data prepared with the presented workflow, and 2) the performance of the model on the positivity prediction task using the same images as ground-truth.

We hence performed a separate labeling study to obtain a second type of ground-truth data, reflecting expert virologists' judgements of sample positivity and negativity based on visual inspection of the very same images that are used for predictions in the proposed assay. The study was designed to closely mimic standard clinical practice. Six experts in immunofluorescence microscopy were asked to rate each in a selection of samples, shown next to a positive and negative control, as either 'positive', 'negative' or 'unclear'.

The sample images were procedurally selected and processed to remove experimenter bias (Figure 5A). For each combination of the IgG and IgA antibodies and the N, S, R antigens, we randomly picked 96 sample images from the test dataset. We selected the same amount of 24 samples from each of the four plates in the set. We moreover enforced a distribution of 50% assumed (samples taken before the virus outbreak vs. RT-PCR confirmed COVID-19 patients) positive and negative samples in each category to allow for well-differentiated sensitivity/specificity curves in the later evaluation. Each image was transformed using the same color transfer function, and only included the Hoechst 33342 channel for nuclear stain and one antibody channel. The positive and negative controls were specific to a plate and antibody/antigen combination.

The study materials were also procedurally composed, which allowed for straight-forward randomization of samples per participant to account for ordering effects. A screen calibration image was included in the beginning of the study material pack to ensure that all participants experience the sample and control images similarly. Each participant rated the same set of 576 samples, and the results were parsed automatically later to avoid any errors in translating the findings. Participation in the study took approximately 2 h per participant. No sample was overlooked, resulting in a total of 1884 'negative', 1412 'positive', and 160 'unclear' ratings (Figure 5A).

We assessed the inter-rater reliability, i.e., the degree to which our participants agreed with each other about their ratings, based on Fleiss' Kappa, an extension of Cohen's Kappa to more than two raters.[13,14] Calculated on ratings for the three classes 'negative', 'positive' and 'unclear', we found kappa = 0.71, i.e., very good[15] reliability. Following this analysis, we removed samples that were rated 'unclear' at least as often as they were rated 'positive' or 'negative'. This only applied to 26 samples, i.e., the final dataset contains 550 samples. We calculated the mean score for each sample (between 0 = negative, 1 = positive), ignoring 'unclear' ratings. Only 16 samples have a mean rating between 0.4 and 0.6, indicating a good consensus of sample positivity and negativity throughout. We finally binarized the results by considering any sample with mean score >0.5 as positive.

Per-well predictions. We used the best cell type classifier and normalization scheme determined in the cross-validation to generate binary predictions of antibody reactivity for a specific antigen and antibody. To this end, we applied the selected model to all segmented cells identified in the respective sample well. Our per-well predictions are given as the positive ratio as the number of positively classified cells divided by the total segmented cell count in the well.

Comparison to visual inspection data

We compared our predictions to the visual inspection data obtained from human experts as ground truth for this assay. We expect different assay applications to favor different trade-offs between assay sensitivity and specificity and hence do not report a fixed value. Instead, we highlight all possible trade-offs between these metrics by plotting the receiver operating characteristic (ROC) curve. Each point on the curve reflects a different threshold to binarize our predicted positive ratio for comparison with the binarized visual inspection ground truth. This threshold is a model hyperparameter and has an intuitive interpretation: it defines how many cells in the well image must be predicted positive to deem the entire well positive. For application of the assay, the threshold can be chosen arbitrarily or by evaluation on a set of validation ground truth data to obtain a specific sensitivity-specificity trade-off. We also report the area under this curve (AUC) as a scalar, aggregate measure of classification performance across all possible thresholds (see Figure 4B).

Comparison to ELISA

We compared our per-well predictions with ELISA results for the N, R, and S antigens of a sample subset across several test plates. ELISA represents an alternative to our assay, but, in contrast to our visual inspection data, cannot be considered ground truth. We consequently chose to facilitate a comparison based on Spearman's correlation. We selected Spearman as a non-parametric, rank-based correlation coefficient because no linear relationship between ELISA titers and our positivity ratio can be expected, and the normality assumptions required for parametric tests might be violated. To not bias our correlations by the substantially higher number of negative samples, we draw a random subset of negative samples to exactly match the number of positive samples for the given antibody and antigen combination (see Figures 4C, and S4B). In Table S6, we report the correlation mean and standard deviation for 1000 repetitions of this sampling strategy.

Comparison to conventional IFA

Two separate comparisons were performed to evaluate the quantitative performance of mini-IFA assay in comparison to conventional IFA methods. The mini-IFA described here utilized cells transiently transfected with S antigen, while the conventional IFA methods used cells infected with SARS-CoV-2 and cells transfected with S antigen (described in STAR methods). In the first comparison, ten patient samples were selected for analysis based on their positive SARS-CoV-2 PCR results. Among these samples, five also tested positive for IgG antibodies against S antigen in ELISA, and five samples tested as negative in ELISA. The results were assessed in a barplot illustrating the IgG values either as the positive ratio obtained from the mini-IFA assay, or the titer values obtained from the conventional IFAs (Figure 5B).

To further evaluate the performance of the mini-IFA assay, a separate comparison was conducted showing the distribution of results. A total of 33 samples tested positive for SARS-CoV-2 (in PCR) and exhibiting IgG reactivity in ELISA against the S antigen was selected. The results were presented as violin plots, providing a visualization of the distribution patterns and allowing for more comprehensive assessment of the mini-IFAs quantitative performance (Figure 5C)

QC and visualization of the results

Quality control (QC) statistics give valuable information about the technical issues within the plate quality, variability, and striping due to the dispensing errors as well as edge effects. It also helps in understanding the performance and behavior of positive and negative controls as well as samples.[16] Here, we also monitored the sample transfer by adding nuclear stain to the samples being transferred with an acoustic dispenser. Quality control data analysis pipeline was established to follow the performance of the assay and how different variable parameters behave (Figure 3). This included cell amount/well, and the intensity of DAPI staining, automatically

represented as plate layout heatmaps (false color images). The positive and negative controls were plotted for each assay plate (Table S1B), and plate-specific scatterplots were created for each sample/control data point to follow any variation and data distribution. The controls can be visualized in graphs to help follow the inter- and intra-plate variation as presented in Figure 3. As an example, inter assay variability across the S-plates (using the POS1 control on IgG values) was CV% 12.2 when normalized to the transfection rate. The intra assay variability ranged from CV% 3.3–14.2 on S-plates (using POS1 and POS2 controls on IgG values) when normalized to transfection rate. In the automatic method the transfection was controlled by adding transfection controls (antibodies against HA- or HIS-tag and rabbit antibodies against N or S/R) to pre-determined wells (Figure 2B; Table S1A). As an example, the transfection percentage for the four assay S-plates, calculated using the rabbit antibodies against S antigen (5 wells per plate), were 14.5% (STD 1.6), 10.1% (STD 0.8), 13.6% (STD 2.1) and 17.9% (STD 1.3). In the semi-automated method, the transfection control (HIS-tag) was included in all the wells (Figure S3B; Table S1A). Interactive graphics like scatterplots in html format help in exploratory data analysis by showing additional information like well annotations and raw values on hovering over the points. For the data visualization per sample, the heatmaps were created for Igs/proteins (Data S1 for IgA and IgG). R programming based ggplot2, plotly packages and custom functions were used for data processing and visualization.[17,18]

Evaluation of the semi-automated assay results

Similarly, to the automated version of IFA assay, the images were processed (illumination correction, cell segmentation and feature extraction) with BIAS software. In addition, they were further analyzed in BIAS by supervised machine-learning to predict either positive or negative status of the samples. To determine the positive ratio for the antibodies against SARS-CoV-2 proteins, the number of cells classified as ''positive'' were divided by the cell-count of transfected (His-tag expressing) cells; parallel values were averaged, excluding the outliers caused by technical errors. To compare the mini-IFA assay results with those from a commercially available antibody test, we performed a recomWell SARS-CoV-2 IgG ELISA, which uses the N antigen (Mikrogen, Germany).

GraphPad Prism Version 5.03 was used to illustrate the results and perform statistical analyses including correlations with patient data. An unpaired t test with Welch's correction was used to compare two groups showing significantly different variances. Due to the deviation from normality, demonstrated by the Shapiro–Wilk test, correlation analyses were performed by calculating the nonparametric Spearman correlation coefficient and the related p value. In general, p values less than 0.05 were considered statistically significant.

Using the semi-automated method (Figure S3), the human IgG and IgA responses and the transfection efficacy (for S, N or R antigens) were investigated side by side in each well on the patient sample set #H1 containing 165 samples from 145 COVID-19 patients (median age 44, IQR 34–53) and #H2 containing 200 negative control sera (Table S4). In set #H1, the SARS-CoV-2 infection was confirmed with PCR and/or nucleocapsid-specific IgG ELISA test. Results of IgM analysis are not discussed here due to the lower specificity and sensitivity in the assay performance.

Assay performance. Single-cell level phenotypic analysis was done in this case using all analyzed plates and there was no separate test dataset. Confusion matrices show 10-fold cross-validation results with high accuracy for predictions (all >92%) (Figure S5). Positive and negative sample groups were separated from each other using a dot blot distribution diagram (Figure S6A). SARS-CoV-2 -infected patients did not show IgG and/or IgA positivity against all the three virus proteins, leading to high distribution ranges of the positive samples (Figure S6A). The N-protein specific IgG data from our assay correlated strongly with the same data analyzed by ELISA ($n = 117$, $r = 0.8789$, $p < 0.0001$) (Figure S6B).

Clinical observations

Using the transfection-controlled mini-IFA protocol, both IgG and IgA responses showed positive correlations with age – as previously observed also by Yang et al.[19]– potentially related with trained immunity[20,21] (Figure S7A). IgG positivity, but not IgA, also showed weak correlations with length of symptomatic period, which may be a consequence of a prolonged virus replication and stronger immune responses (Figure S7B). Furthermore, the ratio of IgG positive cells (S protein) was significantly different in symptomatic and asymptomatic patients, suggesting disease severity to correlate with higher antibody levels as reported earlier[22,23] (Figure S7B).

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis was performed with Python scripts

Our code is shared in the project GitHub repository. We evaluated our models' accuracies by measuring sensitivity and specificity using expert annotation as the ground-truth. The single-cell classification prediction was evaluated using single-cell annotations made by an expert with leave-one-plate-out cross-validation strategy. The presented values in Results section show mean results from four iterations of cross-validation. We evaluated the positivity ratio prediction of our models to the consensus of six experts (describes in the *Visual Inspection Data* section) and measured the area under the ROC curve. The expert annotated 576 images as positive, negative, or unclear. The expert evaluation did not reach consensus in 26 images which were removed from the evaluation.

The sample number in the study represent patient serum samples collected in 2020 and negative control samples from a study collected in 2017. The number of these samples are presented in *Patient serum samples* section. When evaluating our model accuracy on single-cell classification, the *n* represents the number of expert annotated cells described in *Image processing and feature extraction* section. In the sample positivity evaluation, the *n* represents the number of images where the expert reached consensus. The coefficient of variation (CV%) was assessed to evaluate intra- and inter assay variability using assay plates, with a focus on the S-antigen plates. To determine the interassay variability (CV%), the positivity ratios of POS1 control (present in four wells on each

plate) were used. First, the positive ratios from each plate were normalized to the transfection controls and the mean of the four control wells were calculated for each plate. The overall mean and standard deviation (SD) of the means was determined. The CV% was calculated by dividing the SD of the plate means by mean of plate means and multiplying the results by 100. For the assessment of intra-assay variability (CV%) both POS1 and POS2 controls on the S-antigen plates were utilized. The CV% was calculated for each sample duplicate by determining the SD of the duplicate, dividing it with the duplicate mean, and multiplying the result with 100. Subsequently, the average of the CV% values was calculated.

Enzyme-linked immunosorbent assay (ELISA)

An in-house ELISA for SARS-CoV-2 with antigens (R, N, and S) was used for generating the reference data for sample sets of all sample donors in Finland (Table S4) as described 2,3,10. In addition to traditional ELISA 2,3,10, used here for SARS-CoV-2 N protein with colorimetric detection, we set up an automated high-throughput ELISA on 384-well plate format as follows: Plates (Nunc 384-well MaxiSorp, non-sterile, #460372 for S and R) were coated with purified antigens3 in 20 μL 0.05 M carbonate buffer (pH 9.6, Medicago, Sweden) overnight at 4°C, S protein 1 μg/mL, and R antigen 1.5 μg/mL. Dispensing was performed with FritzGyger CertusFlex (0.3 mm nozzle, 0.2 bar for all dispenses). For each 384-well plate with serum samples, six plates were coated with each antigen to give two replicate plates for each detection antibody conjugate (anti-IgG/A/M). Plates were covered with MicroClime lids (Labcyte/BeckmanCoulter) filled with ultra-pure water during all incubations through the assay. After antigen coating, the wells were emptied with BioTek EL406 96 pin washer manifold, and 30 μL blocking buffer (3% non-fat milk in PBS with 0.1% Tween 20) added to all wells with EL406 syringe pump. Blocking solution was incubated for 1h RT, and the wells were washed once with 60 μL of PBST (PBS with 0.05% Tween 20) using EL406. Washing solution (PBST) was dispensed with a syringe pump. For serum sample dilution, 15 μL of assay buffer (1% non-fat milk in PBS with 0.05% Tween 20) was added with CertusFlex. Serum samples were pre-diluted in concentration of 1/25 to assay buffer on 384 intermediate plate (Axygen 384-well Clear V-Bottom 240 μL Polypropylene Deep Well Not Treated Plate, #P-384-240SQ-C) pipetting with Biomek FXp 96 head (BeckmanCoulter). Biomek FXp 384 head was used to pipette 5 μL of 1/25 serum dilution to 15 μL of buffer on assay plates to end up with 20 μL of 1/100 dilution of serum. Serum dilution was incubated for 2 h RT with orbital shaking 600 rpm (Heidolph Titramax 1000). Plates were washed with EL406 washer manifold and syringe pump dispensing three times with 80 μL of PBST and aspirated empty. Horseradish peroxidase (HRP) conjugated secondary antibodies were diluted in an assay buffer as follows: goat anti human IgG-HRP 1/6000, goat anti human IgA-HRP 1/5000, and goat anti human IgM-HRP 1/1500. Antibody solutions were dispensed to 20 μL by CertusFlex. Wells without samples and without antigen coating were filled with an assay buffer only. After incubation of 1 h RT plates were washed three times with 80 μL PBST as previously, and then wells filled with Pierce ECL Western Blotting Substrate (#32106) using CertusFlex. After 10 min of incubation, the plates were read for luminescence BMG Pherastar FS, with measurement interval of 1 s, and focal height 12.5 mm for Nunc plate.

For Hungarian SARS-CoV-2 positive serum samples (Orthosera Ltd., see Table S4), the commercially available recomWell SARS-CoV-2 IgG ELISA (Mikrogen, Germany) with purified recombinant SARS-CoV-2 N protein was performed following the manufacturer's instructions.

Conventional immunofluorescence assay (IFA)

"Conventional" IFA based on Vero E6 cells infected with SARS-CoV-24 served for estimating the performance of the mini-IFA. Shortly, Vero E6 cells were inoculated with SARS-CoV-2 (first Finnish isolate described in Haveri et al.[24]), at two days post inoculation the cells were trypsinized, washed with PBS and mixed at approximately 1:2 ratio with trypsinized non-infected Vero E6 cells washed with PBS. The mixture of SARS-CoV-2 infected and non-infected cells in PBS was dispensed on 10-well microscope slides with reaction wells (Marienfeld, No. 1216521), and after air-drying fixed with ice-cold acetone. The diluted serum samples were applied to the wells, and after 30–45 min incubation at 37°C, the slides were washed three times with PBS and once with milliQ water, and air-dried. Fluorescein isothiocyanite (FITC)-conjugated anti-human IgM or IgG antibodies diluted in PBS were applied to the wells, followed by 30 min incubation at 37°C. Before mounting, the slides were washed three times with PBS, once with milliQ water, and air-dried. The results were visually inspected using a fluorescence microscope. In addition, slides prepared as above but from Vero E6 cells transfected with S-antigen expression plasmid were utilized.

ADDITIONAL RESOURCES

A webpage was created for the visualization of quantitative sample-specific IgG results and their distribution compared to controls and other samples, together with microscopic images. It is located at a webpage: https://fimm-covid-19-hca.github.io/. The webpage contains separate files (.html) for interactive detection of samples' IgG responses toward SARS-CoV-2 antigens (S, R, N, M) on 384-well plates (S01-S04) with corresponding microscopy images. The graphs show the positive ratio (Y axis on the left; IgG) for IgG with S/R/N/M antigen for each patient sample (X axis; Sample names). Regions of the graph can be zoomed in, and the individual data points can be hovered to view 1) the microscopic image for an IgG/Ag pair for each sample and 2) the positive ratios of the antigen with different Igs (IgA, IgG, IgM) for each sample. The results shown are created with the automated mini-IFA assay, and include donors' serum sample sets #F1a, #F1b, and #F2 (see Table S4). Images of controls used in the analysis pipeline are also shown. Webpage: https://fimm-covid-19-hca.github.io/