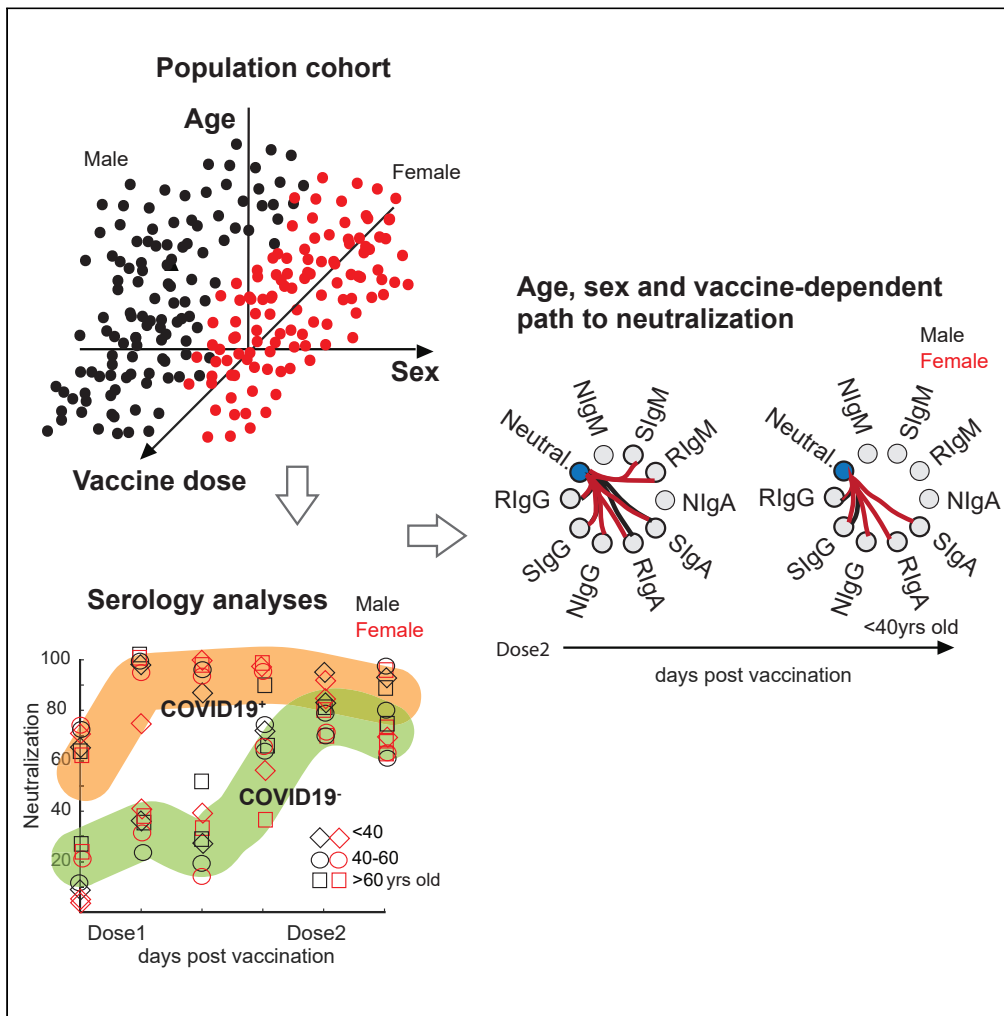**Article**

Multivariate analyses and machine learning link sex and age with antibody responses to SARS-CoV-2 and vaccination

Miroslava
Cuperlovic-Culf,
Steffany A.L.
Bennett, Yannick
Galipeau, ..., Jörg
H. Fritz, Ciriaco A.
Piccirillo, Angela
M. Crawley

acrawley@ohri.ca

Highlights

Network analysis of the interplay of complex antibody characteristics and sex and age

Age and sex distinguish temporal dynamics of infection- and vaccine-induced immunity

Less isotype switching and neutralization in older individuals with vaccination

SARS-CoV-2 neutralization relation with antibodies is dependent on COVID-19 status

Article

Multivariate analyses and machine learning link sex and age with antibody responses to SARS-CoV-2 and vaccination

Miroslava Cuperlovic-Culf,[1,2,13] Steffany A.L. Bennett,[2] Yannick Galipeau,[2] Pauline S. McCluskie,[2] Corey Arnold,[2] Salman Bagheri,[2,3,4,5] Curtis L. Cooper,[3,4,5,6,7,8] Marc-André Langlois,[2,3] Jörg H. Fritz,[3,9,10,11,12,13] Ciriaco A. Piccirillo,[3,9,10,11,12,13] and Angela M. Crawley[2,3,4,5,13,14,*]

SUMMARY

Prevention of negative COVID-19 infection outcomes is associated with the quality of antibody responses, whose variance by age and sex is poorly understood. Network approaches identified sex and age effects in antibody responses and neutralization potential of *de novo* infection and vaccination throughout the COVID-19 pandemic. Neutralization values followed severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)-specific receptor binding immunoglobulin G (RIgG), spike immunoglobulin G (SIgG) and spike and receptor immunoglobulin G (S, and RIgA) levels based on COVID-19 status. Serum immunoglobulin A (IgA) antibody titers correlated with neutralization only in females 40–60 years old (y.o.). Network analysis found males could improve IgA responses after vaccination dose 2. Complex correlation analyses found vaccination induced less antibody isotype switching and neutralization in older persons, especially in females. Sex-dependent antibody and neutralization decayed the fastest in older males. Shown sex and age characterization can direct studies integrating cell-mediated responses to define yet elusive correlates of protection and inform age and sex precision-focused vaccine design.

INTRODUCTION

Sex and age differences in antiviral responses, vaccination, and long-term recovery are increasingly recognized in COVID-19[1–4] as key determinants to disease susceptibility and severity. During viral infections, females have greater inflammatory, antiviral, and humoral immune responses compared to males. Early in the COVID-19 pandemic, it was observed that infection prevalence is sex independent or possibly slightly higher in females. In contrast, males are more likely to experience severe disease, combined with increased hospitalization rates and higher mortality.[5,6] Increased disease severity was particularly significant in men over 60 years of age.[2] In a large, hospitalized patient cohort study of 308,010 patients, the mortality rate was higher in males regardless of age, race, or comorbidities.[7] This is also observed with other coronaviruses including the Middle East respiratory syndrome epidemic, where mortality rate is also higher in males.[8] Differences in sex-related mortality risk have been observed in animal models of acute respiratory viral infections,[9] further suggesting significant sex effects influencing the outcomes. The molecular basis for these divergent immunological outcomes in males and females is thought to be related at least in part to differences in sex steroid synthesis, which can have a crucial role on the function and regulation of the immune response.[10,11] Indeed, sex steroids have been reported to exert a suppressive role in immune functions suppressing immune cell activity by promoting anti-inflammatory mediators. However, the nature and dynamics of antibody production throughout infection and vaccination in men and women are not well understood.[12]

[1]Digital Technologies Research Centre, National Research Council of Canada, 1200 Montreal Road, Ottawa, ON K1A 0R6, Canada
[2]Department of Biochemistry, Microbiology, and Immunology, University of Ottawa, Ottawa, ON K1H 8M5, Canada
[3]Coronavirus Variants Rapid Response Network (CoVaRR-Net), Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada
[4]Chronic Disease Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada
[5]Centre for Infection, Immunity, and Inflammation (CI3), University of Ottawa, Ottawa, ON, Canada
[6]Clinical Epidemiology, Ottawa Hospital Research Institute, Ottawa, ON, Canada
[7]Division of Infectious Diseases, Department of Medicine, University of Ottawa and the Ottawa Hospital Research Institute, Ottawa, ON, Canada
[8]Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada
[9]Department of Microbiology and Immunology, McGill University, Montréal, QC, Canada
[10]Program in Infectious Diseases and Immunology in Global Health, The Research Institute of the McGill University Health Centre (RI-MUHC), Montréal, QC, Canada
[11]Centre of Excellence in Translational Immunology (CETI), Montréal, QC, Canada
[12]McGill University Research Centre on Complex Traits (MRCCT), Montréal, QC, Canada
[13]Senior author
[14]Lead contact
*Correspondence: acrawley@ohri.ca
https://doi.org/10.1016/j.isci.2024.110484

Table 1. Overview of study groups and time points assembled for this data analysis

	Unvaccinated (male/female)	Dose 1 (male/female)	Dose 2 (male/female)
Unique study participants (n=970)	198/394	122/244	140/290
Samples	393/643	205/422	440/1,012
Positive	177/163	84/90	119/168
Negative	216/480	121/332	321/844
<40 y.o.	99 (30.3 ± 4.5)/249 (30.6 ± 4.9)	44 (30.4 ± 5.2)/118 (31.2 ± 4.8)	111 (31.5 ± 4.6)/368 (31.4 ± 4.5)
Age (40–60) y.o.	173 (48.0 ± 6.3)/275 (49.1 ± 6.2)	53 (49.6 ± 5.7)/134 (50.6 ± 5.60)	172 (46.8 ± 5.4)/411 (50.3 ± 6.0)
≥60 y.o.	121 (66.6 ± 4.8)/119 (66.0 ± 4.3)	55 (68.0 ± 5.0)/67 (65.7 ± 4.7)	104 (65.7 ± 5.2)/135 (65.8 ± 4.7)

Dose #: 0 unvaccinated; dose 1 and 2: minimum day 14 post-vaccination and before next dose.
Positive: seropositive for IgG-N (nucleoprotein); negative: seronegative for IgG-N; y.o.: years old.

Adaptive immunity to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection depends on the concerted action of both humoral and cell-mediated immune components. Immunoglobulin (Ig)G, IgM, and IgA against spike and receptor-binding domain (RBD) are detected in plasma 7–8 days following the onset of symptoms in infection with IgA persisting for over 400 days after symptom onset.[13] Detectable IgM and IgG Abs occur concurrently with increased plasma B cells. While IgG levels reportedly decline modestly ~8 weeks after symptom onset, recovered patients maintain high spike-specific IgG and IgA titers. Neutralizing antibodies (nAbs) as well as binding antibodies (bAbs) can control SARS-CoV-2 infection; convalescent plasma containing nAb and bAb administered as a therapeutic can improve clinical symptoms.[14] In SARS-CoV-2 convalescent subjects, IgG1, IgA1, and IgM antibodies against spike and RBD all have capacity for virus neutralization.[13] Based on regression analysis, IgM and IgG1 show the strongest contribution to neutralization; however, direct analysis shows that IgA was also able to neutralize.[13]

Severe COVID-19 patients display an acute respiratory distress syndrome, which can be mediated directly by the virus and/or a hyperactive immune response. SARS-CoV-2 antigen-specific CD4+ and CD8+ T cells have been identified in recovered and active disease patients and contribute to protection from re-infection.[15–18] The immune profiles of COVID-19 patients with moderate disease indicate a protective T cell response, while patients with severe disease exhibit exaggerated systemic inflammation, signs of T cell exhaustion and lymphopenia, and decreased polyfunctionality or cytotoxicity, all features of dysregulated adaptive immunity. Different immunotypes predict the course of infection, whereby coordinated action between CD4+, CD8+ T cells, and antibodies contributes to optimal viral control, efficient immune protection, and better disease prognosis.

Viral-induced inflammation increases with age and is associated with a hyper-inflammatory responses leading to significant lung damage as well as increased morbidity and mortality rates in older adults. Elevated levels of inflammatory cytokines (e.g., interleukin [IL]-6), increased inflamm-aging, and T cell senescence, as well as higher titers of anti-S/-N IgG and IgM, correlate with worse clinical readouts with older age, suggesting adverse effects of the immune response on disease severity.[16,19,20]

Immune responses are known to differ in males and females leading to sex-specific responses to vaccination and infections.[21] Female mice have a more robust humoral and cellular immunity than their male counterparts following a primary challenge with influenza virus whereas vaccination yields equivalent protection in both sexes.[22] Transfer of immune sera from vaccinated females to naive mice provided better protection compared to transfer of sera from vaccinated males. This superior sex-associated protection is thought to be a consequence of an increase in antibody titers, IgG antibodies of greater neutralization capacity, and effective Ig isotype switching in females contributing to more effective antibody function and neutralization capacity. A meta-analysis of 117 peer-reviewed articles showing serum levels of IgA, IgG, or IgM in adult human beings[23] suggests that older age and male sex are associated with higher vaccine-specific IgA levels and that lower IgM and IgG levels were not influenced by age and sex. A population-based study[24] supports these findings. Increased levels of serum IgA and decreased IgM with age and higher IgA titers in males than females were noted. Interestingly, IgG level decreases with age (up to 60).[24]

The use of computed network analyses can enable the extraction of important host response patterns from complex data and has been applied in the COVID-19 era.[4] In an effort to better understand the observed sex and age differences in SARS-CoV-2 infection risk, disease outcome, and response to COVID-19 vaccination, we applied a novel network analysis strategy[25–27] to track the evolution of antibody generation and functionality following natural infection and/or vaccination in a large longitudinal cohort referred to as *Stop the Spread Ottawa* (SSO),[28] pinpointing changes in the response to infection and/or vaccination in immune-competent or immune-compromised individuals.

RESULTS

Age and sex weakly affect overall response to vaccination and natural infection

In this analysis, serially collected serum samples from 970 participants (3,666 samples in total) were separated into groups by the vaccination period, age, and sex of the participant as well as by SARS-CoV-2 infection-acquired immunity status (Table 1). The primary goal of the study was to explore the relationship between serology measurements and viral neutralization in males and females separately in three age groups. Thus, we aimed to first ensure sufficient statistical power for each group and obtain correlative relationships (Figure S1). The specific
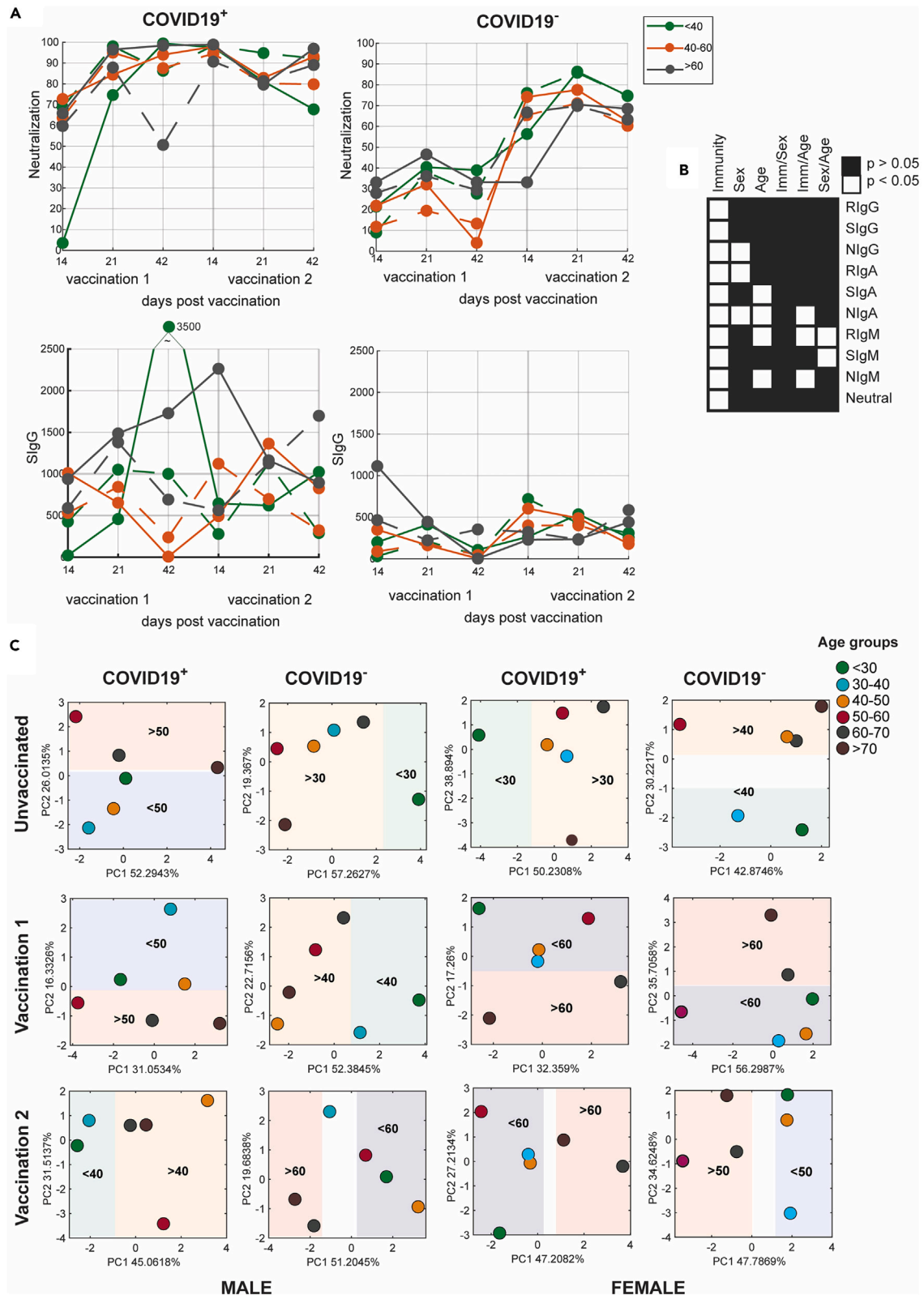
Figure 1. Small differences between sex and age groups in response to vaccination and infection are observed

Samples are divided by COVID-19 status (COVID-19⁻ = no COVID-19 infection history, confirmed by negative NIgG serology; COVID-19⁺ = history of COVID-19 infection, confirmed by positive NIgG serology).

(A and B) Graphical representation of the mean values for neutralization and SIgG at three time points after vaccination 1 and vaccination 2 (days 14, 21, and 42 post-vaccination) in cohorts grouped by sex and age (in three age groups <40, 40–60, and over 60 years of age). (B) Three-way ANOVA evaluation of the mean differences between measurements grouped by acquired immunity (COVID-19⁻ or COVID-19⁺), sex (male/female), or age (under 40, 40–60, or over 60 years of age). Indicated as white boxes are statistically significant mean differences for direct and combined effect.

(C) gPCA for log transformed and Z score scaled dataset of 9 different antibody responses and neutralization (10 features) with included PC1 for guided PCA (gPCA) for age groups for each subset divided by vaccination dose and sex; age groups are under 40, 40–60, and over 60 years of age. Shown are loadings for each group.

SARS-CoV-2 strain that infected the study participants with known infection is not specified. All samples were collected between October 21, 2020, and August 20, 2021, before the emergence of Omicron COVID-19 variant. Nine different antibody responses were measured from each serum sample corresponding to IgG, IgA, and IgM titers specific for SARS-CoV-2 RBD (R), spike (S), and nucleoprotein (N) as well as neutralization measurements.

An initial data overview was performed for the complete set using unsupervised visualization and clustering methods to determine dominant patterns in the data. Principal-component analysis (PCA) (Figure S2A) indicates a dominant effect of vaccination on serology profiles in all age groups with the most apparent separation between the unvaccinated and vaccination 1 groups. Change over time with vaccination is once again significantly different between COVID-19⁺ and COVID-19⁻ groups, showing major value difference in SIgG between these two groups as well as lower level of neutralization in COVID-19⁻ cohorts in all age groups and both sexes even after vaccination (Figure 1A). However, from observing individual serology factors, differences between age and sex groups is not apparent. However, effect of age and sex on the serology measure starts to be visible in the ANOVA analysis (Figure 1B) as well as guided PCA (gPCA) analysis, a semi-supervised method that shows group loadings in the PCA (Figure 1C) (Reese, Archer et al. 2013). Furthermore, t-distributed stochastic network embedding (t-SNE) analysis (Figure S2B), providing projection of sample points while preserving relationship between sample points, indicates some grouping of sample points based on sex as is shown with p values of the significance of the separation by sex in different subgroups (Figure S2C). The association of the effect of sex on antibody responses was particularly significant between male and female participants <40 years old (y.o.) who were either SARS-CoV-2 antibody negative or positive across all three first doses of vaccination. Meanwhile, the 40–60 y.o. group showed significance only in the SARS-CoV-2-positive subjects after vaccination 1 and 3, and there were limited/transient sex effects in the >60 y.o. group, reserved for SARS-CoV-2-seropositive individuals after the second vaccination dose. Thus, t-SNE suggests that age is a primary determinant of antibody response with complex dependence between the response and sex.

Further, hierarchical cluster analysis (HCL, Figure S3) indicates comparable separation between samples collected pre-and post-vaccination in all groups, independent of sex and infection status. Interestingly, vaccination-driven separation of samples is independent, in this type of analysis, of the time since vaccination. In all four groups of samples (including male and female groups distinguished by SARS-CoV-2 infection-acquired immunity status), HCL shows that neutralization values cluster most closely with receptor binding immunoglobulin G (RIgG), spike binding immunoglobulin G (SIgG), and spike binding immunoglobulin A (SIgA) levels. In COVID-19+ samples, neutralization is also clustered with receptor binding immunoglobulin A (RIgA). Uniquely in HCL of COVID-19- males, neutralizing immunoglobulin G (NIgG) clusters with neutralizing immunoglobulin M (NIgM), spike and receptor binding immunoglobulin M (SIgM and RIgM) while it is most strongly grouped with NIgA, RIgA, and SIgA in the other groups (Figure 1B). Additionally, Ig factors primarily cluster by size with exceptions of NIgG which in all four groups shows closer proximity to non-IgG features. Once again, these clustering results suggest a level of sex-dependent behavior in the relationships between measured antibody responses (Figure S3).

Co-behavior between antibody measures and neutralization is sex, age, and immunity dependent

Detailed analysis of feature co-clustering (i.e., similarities in behavior between neutralization and antibodies) was performed using the fuzzy C-means (FCM) clustering method. This approach provides information about the level of co-clustering for each feature to each group, allowing for possible multiple relationships. Data in Figure 2 show a weight for each antibody response feature for belonging to the cluster relative to their respective neutralization value. Cluster analysis is performed separately for subject groups divided by age, sex, and COVID-19 infection-acquired immunity status. Unlike HCL analyses of all age groups combined (Figure 1), this approach is applied to three age groups separately and shows differences between males and females, particularly in older and COVID-19-positive groups. Interestingly, in females >40 y.o., RIgA and SigA co-cluster with neutralization, and males co-cluster with RigM and SigM responses. In the <40y.o. groups, SigG and RigG levels correlate with neutralization in both males and females.

Differences in antibody responses between males and females are dependent on age and infection-acquired immunity

We used three different feature selection methods (see STAR Methods) to determine the most significant feature differences (Figure 3). Features that are selected by all three methods are indicated. The data for all time points were combined in the analysis, but a separate analysis was performed for COVID-19-negative and COVID-19-positive samples across the age groups and is reported here (Figures S2 and S3 show feature selection in each vaccination group). Differences in IgG and IgA factors are identified in participants <40y.o. COVID-19-negative group. Differences in IgM levels are noted in individuals 40–60 y.o. and >60 y.o. The features selected as statistically different using these methods are largely in agreement with features showing major clustering differences in Figure 2.
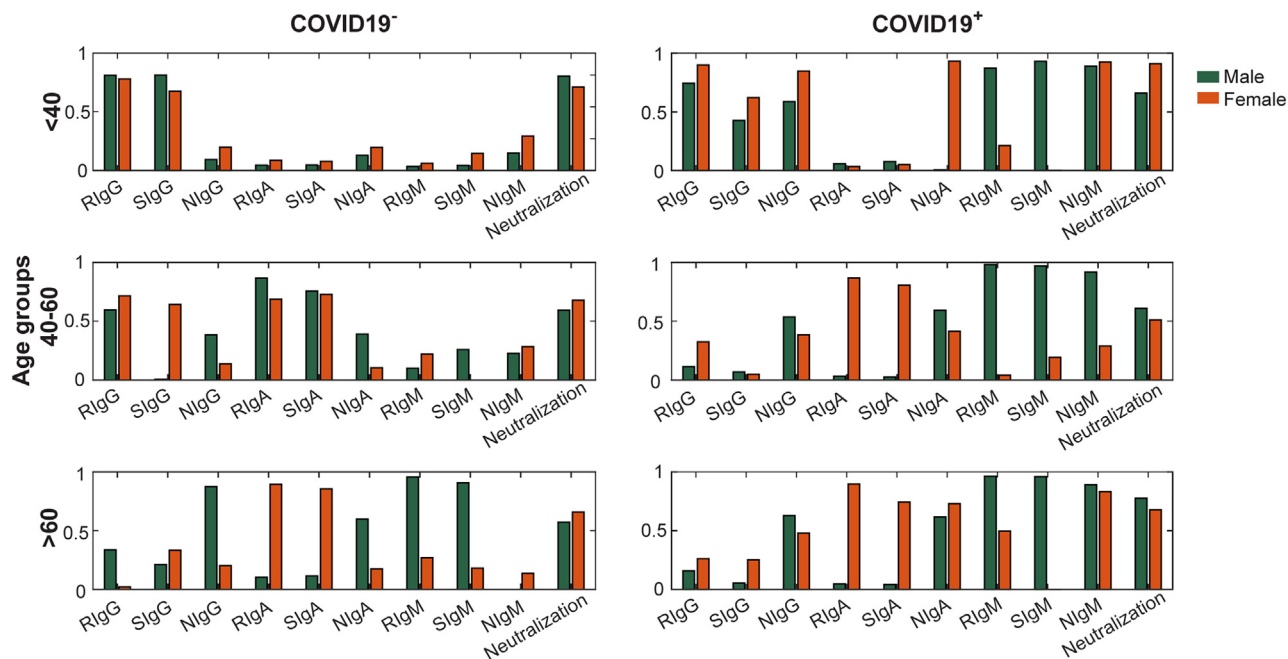
Figure 2. Cluster analysis of antibody responses relative to neutralization shows differences in co-behavior in males and females by age

Fuzzy C-means clustering of antibody response features across data in subgroups divided by age, sex (male in black, female in orange), and COVID-19 infection status (negative or positive SARS-CoV-2 history, confirmed by NIgG level). Shown are membership values, i.e., cluster association of features in different subject cohorts with clusters that have the strongest neutralization membership. Fuzzy C-means here clusters feature separately for each of the groups of data divided by age, sex, and COVID-19 infection status. Shown are weights (0.0–1.0), i.e., cluster belongings for each antibody response feature relative to neutralization values (weights approaching 1.0 indicate strong clustering with neutralization).

It is important to observe that SIgM and NIgA come up as significant in 4 age and immunity subgroups and both antibodies show different clustering patterns in males and females (Figure 2). In addition to overall differences in these antibody features established by the three methods of feature selection analysis (Figure 3), there are significant changes in antibody responses over this period. In males, there is a rise in the RIgA, NIgA, and SIgA serum levels generally following dose 2 vaccination. In contrast, higher IgM serum levels are noted in females over 60 y.o. in both vaccination 1 and vaccination 2 periods.

Correlation analysis provides quantitative links between sex and antibody responses

Further investigation of the relationship between serological profile and sex was performed using point-biserial correlation specifically developed for analysis of correlations between binary and continuous variables. Figure 5 shows antibody responses correlate with subjects' sex. Once again, SIgM and RIgM are positively correlated with sex in individuals over 60 y.o. following both dose 1 and dose 2 vaccination. As is typical of antibody response evolution through Ig isotype switching with repeated vaccination (or even repeated infection), IgM levels show a negative correlation with sex in the younger age groups with values higher in males than in females.

In the point-biserial correlation analysis, antibody neutralization, an important indicator of protection against serious COVID-19 disease outcomes,[29,30] is only statistically significantly correlated with sex in the >60 group (Figure 4). However, time course analysis (Figure 5) as well as feature selection (Figure 3) indicate some differences in neutralization trends between males and females. To determine whether there are specificities in the contributions of antibodies to neutralization in age and sex groups, a distance correlation analysis was performed between antibody neutralization measurements and other antibody response features as a measure of co-behavior. In addition, relative square distance correlation was tested as a measure of contribution of each antibody response factor to the neutralization value. This analysis showed that, after vaccination 2, neutralization potential decays the fastest among older persons who are male than any other group. This co-behavior also suggests that females retain antibody-neutralization correlations longer than their male counterparts in all age groups.

Correlation between antibodies and neutralization varies over time differently in males and females

Square of the correlation between two features represents a measure of the contribution of one feature to the variance of the other. Analysis of distance correlation between neutralization and each antibody in male and female groups over time following vaccination dose 1 and dose 2 shows major differences in the pattern of antibody contribution to neutralization (Figure 6). As an example, the contribution of RIgG and NIgA to neutralization in males and females under 40 y.o. is positively correlated over time and highly comparable by
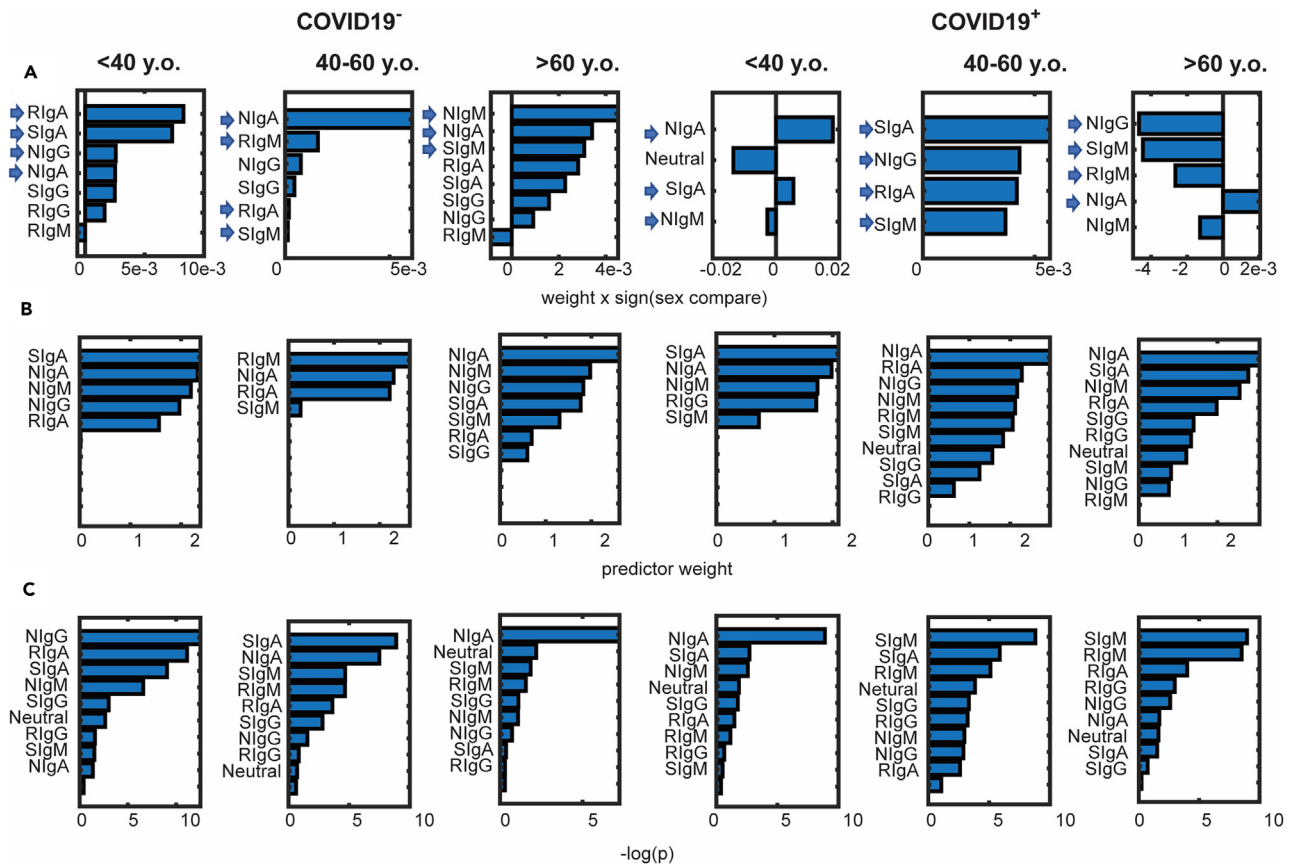
Figure 3. Selection of antibodies showing the largest difference between male and female cohorts in different age and acquired immunity subgroups

Three different feature selection methods were used to identify sex differences in antibody responses in each of the six sample subgroups separated based on the COVID-19 status and participants' age (<40, 40–60, and >60 y.o.).

(A–C) Results from the (A) Relieff machine learning feature selection method, (B) feature selection method using neighborhood component analysis for regression (*fsrnca*), and (C) univariate feature ranking for regression using F-tests (*fsrftest*) method show major antibody measures that distinguish males from females within specified subgroups. Solid arrows indicate features selected by all three methods. Relieff plots show only antibody features with positive weight (i.e., features that contribute to significant sex separation), and the sign (i.e., +/−) in this representation is related to the male to female mean group ratios with positive sign indicating larger mean value in a male group.

sex. However, neutralization and SIgM and RIgM are negatively correlated over time in these same groups. In contrast, following dose 2 vaccination in those over 60 y.o., neutralization is similarly related to RIgG in males and females but inversely correlated with SIgG in the two groups. Thus, although the overall pattern in neutralization measurements is not correlated with sex, the observed levels of neutralization show different and complex relationships with antibody responses. Furthermore, the analysis shown in Figure 6 suggests changes in the correlation of individual antibody responses over time, differing between males and females and by age group. In the under 40-y.o. male group following vaccination 1, neutralization is initially strongly correlated to RIgG and NIgA concentrations and then more strongly to SIgG, NIgG, IgA, and IgM antibodies. In females of this age group assessed at earlier time points, the relationship with neutralization is the strongest with SIgG, NIgG, and SIgA and all three IgM specificities. Following administration of vaccination 2 in this group, the relative contribution of RIgG and SIgG to neutralization initially, at early time points post-vaccination, drops.

Changes in the contribution of antibody response features to neutralization over time in different groups can be further understood by exploring correlations in different time periods. For the ease of presentation and increased number of samples in this case, time was binned into 21-day periods. Distance correlation was calculated using samples for each period, and the square of individual values is divided with the sum of values for each time bin to provide a relative contribution. Thus, Figure 7 shows relative square distance correlation contributions to neutralization of each of the nine antibody response factors at 4 time points for each period. This representation is useful in that it highlights sex and age differences that were otherwise less apparent in an unsupervised PCA (Figure 7).

This relative contribution of different antibody isotypes to neutralization over time depends on age and sex with differences observed in Ig isotype responses to all specific targeted proteins measured and is an important finding of this work. With this novel approach to data analysis, we show that the relationship between antibody responses with neutralization depends on the subject's age and sex, even in cases where
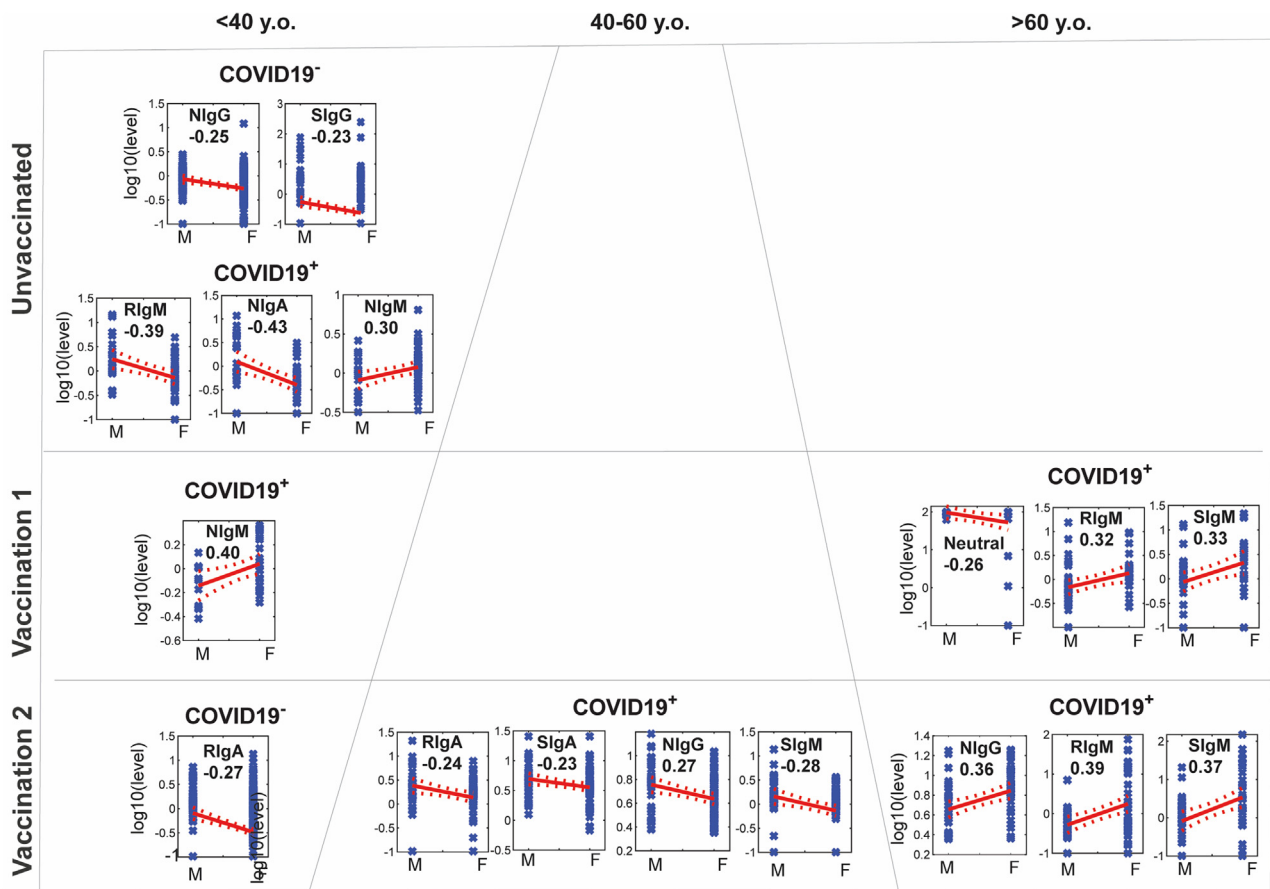
Figure 4. Direct correlation between antibody level and subjects' sex
Point-biserial correlation between sex and serological factors. Shown are correlations with $p < 0.05$ and absolute correlation value >0.20 within subgroups. A correlation value between sexes and log10 transformed concentration levels is indicated.

the actual value of neutralization is not significantly different between these subgroups. For example, following vaccination 2 in the oldest group, compared to all studied Ig, the relative contribution of IgG is smaller than that in younger groups particularly in females, while in under 40-y.o. males, the relative contribution of IgM is more significant than in females.

DISCUSSION

We applied novel network analyses to a complex set of antibody response data collected longitudinally from a large cohort of individuals over an important period of the COVID-19 pandemic, bridging naive immunity, infection-acquired immunity, and immunity afforded across serial vaccination. The size and breadth of the cohort age range offered robust analytical outcomes by sex and age across multiple antibody parameters and repeated measures. Sex differences emerged with t-SNE analysis demonstrating that complex, non-linear analyses of these data would be required to understand factors driving antibody responses to COVID-19. Whether pre-immune status influences vaccine/infection/hybrid-mediated immunity remains to be determined, but sex effects identified by HCL analysis suggest that the SARS-CoV-2 serology patterns of unvaccinated COVID-19-negative males were categorically different than those of any other group pre- and post-vaccination (Figure S3). Correlations revealed important sex-associated co-behaviors of antibody responses with neutralization, some of which suggest the important finding that young females may have stronger linkages in mucosal immunity and protection as their serum IgA responses are associated well with neutralization levels. Feature selection methods continued to confirm clustering analysis findings by age, with particularly prominent sex effects in IgA and NIgG responses in individuals <60 y.o. The capacity for significant Ig isotype switching with repeated vaccination was found to be weaker in older females whereas IgA responses in males could improve after vaccination 2. This phenomenon was extended to yet-additional new observations of this study that highlighted an appreciation for the rate of antibody decay, which is quite rapid in males >60 y.o. and where females exhibited a superior retention of Ig isotype switching correlating with neutralization over time. These two latter results emerge as important findings of this study, alongside the expected observation that antibody response profiles are inherently complex patterns in relation to neutralization that differ by sex and age, as shown by distance correlations, and these changed over time.
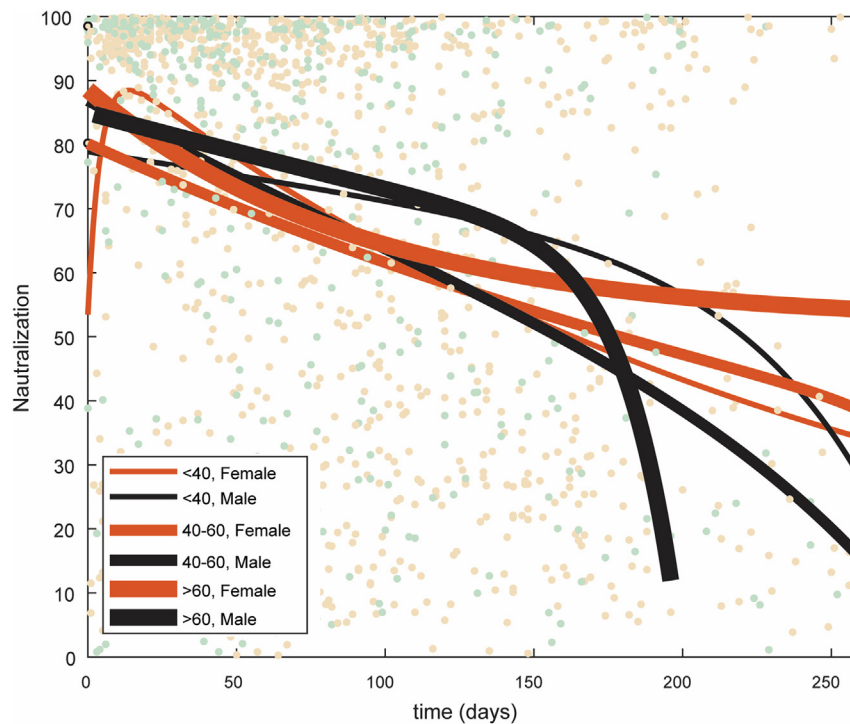
Figure 5. Antibody neutralization levels for over 60-year-old males decay faster than those for other sex and age groups

Values present neutralization measurements in males and females in 3 different age groups following the second dose of COVID-19 vaccination. Exponential fit is optimized separately for each represented group. Neutralization following vaccination 2 for males and females. Shown are exponential best fit lines (as a*exp(b*time)+c*exp(d*time)). Time is shown in days from vaccine 2. Y axis shows the original value. Lines represent dark green, male, and red, female. Also included are actual sample measures showing major individual differences.

These analytical approaches afforded several different angles by which to evaluate the complex relationships between antibody responses and neutralization showing that considerations for sex and age were highly important.

Demographic factors including age and sex are known to be associated with responses to both viral infection and immunization.[1,22,31] Our data show differences in antibody responses across vaccination periods, in convalescent individuals with SARS-CoV-2 acquired immunity compared to individuals that did not show naturally acquired SARS-CoV2 immunity. Understanding the extent and type of this association and its change over time in relation to the resilience to an infection is crucial in attempts toward achieving more protective immunization and effective antiviral approaches. Extensive work on COVID-19 immune responses led to an appreciation that antigen specificity and Ig isotype can contribute to ACE2 neutralization.[13,32,33] While typical primary immune responses dominated by IgM evolve into recall immune responses composed of IgG, IgA, and IgE antibodies, there is variation in which factors contribute to protection against SARS-CoV-2, predictable in part based on antibody neutralization levels. Serum IgA is a dominant antibody soon after the onset of symptoms which peaks during the third week of the disease and provides the greatest contribution to neutralization. Assessment of bronchial alveolar fluid suggests this contributes predominantly to antibody-mediated virus neutralization activity in the lung compared to IgG.[34] Additionally, Ruggiero et al. showed that a coordinated response implicating both SIgM and SIgG is associated with protective immunity following vaccination.[33] While sex differences have been reported in disease outcomes, whether complex features of antibody responses to infection and vaccination play a role in determining immune protection to SARS-CoV-2 remains to be determined.

To identify potential differences in the coordination of the response over time in males and females of different ages, we have developed a novel methodology for analysis of correlation over time and applied it to a large longitudinal dataset spanning the COVID-19 pandemic period, prior to and through vaccination dose series, with robust representation of subjects across different age groups and in both sexes. Our regression analysis of the age dependence of the measured serology profile is consistent with earlier studies showing age of ~40 and ~60 as significant points in the relationship between serology and response to the virus. Analyses of major features separating measurements in males and females in different age groups (<40 y.o., 40–60 y.o., and >60 y.o.) show changes at different points in the vaccination series, thus indicating differences in progression of antibody responses over time. In males, RIgA and SIgA show significantly faster increases following vaccination while the opposite is observed in SIgM and RIgM, particularly in the <40 y.o. and >60 y.o. groups. Although many selected antibodies have low weight in feature selection analysis, some are significantly correlated with subject's sex. This is demonstrated in point-biserial correlation for different age and vaccination dose groups. RIgM is significantly correlated with sex in both young adult and senior groups but with opposite ratios, where RIgM is higher in males <40 y.o. yet lower in females >60 y.o. Similarly, SIgM and NIgG are significantly
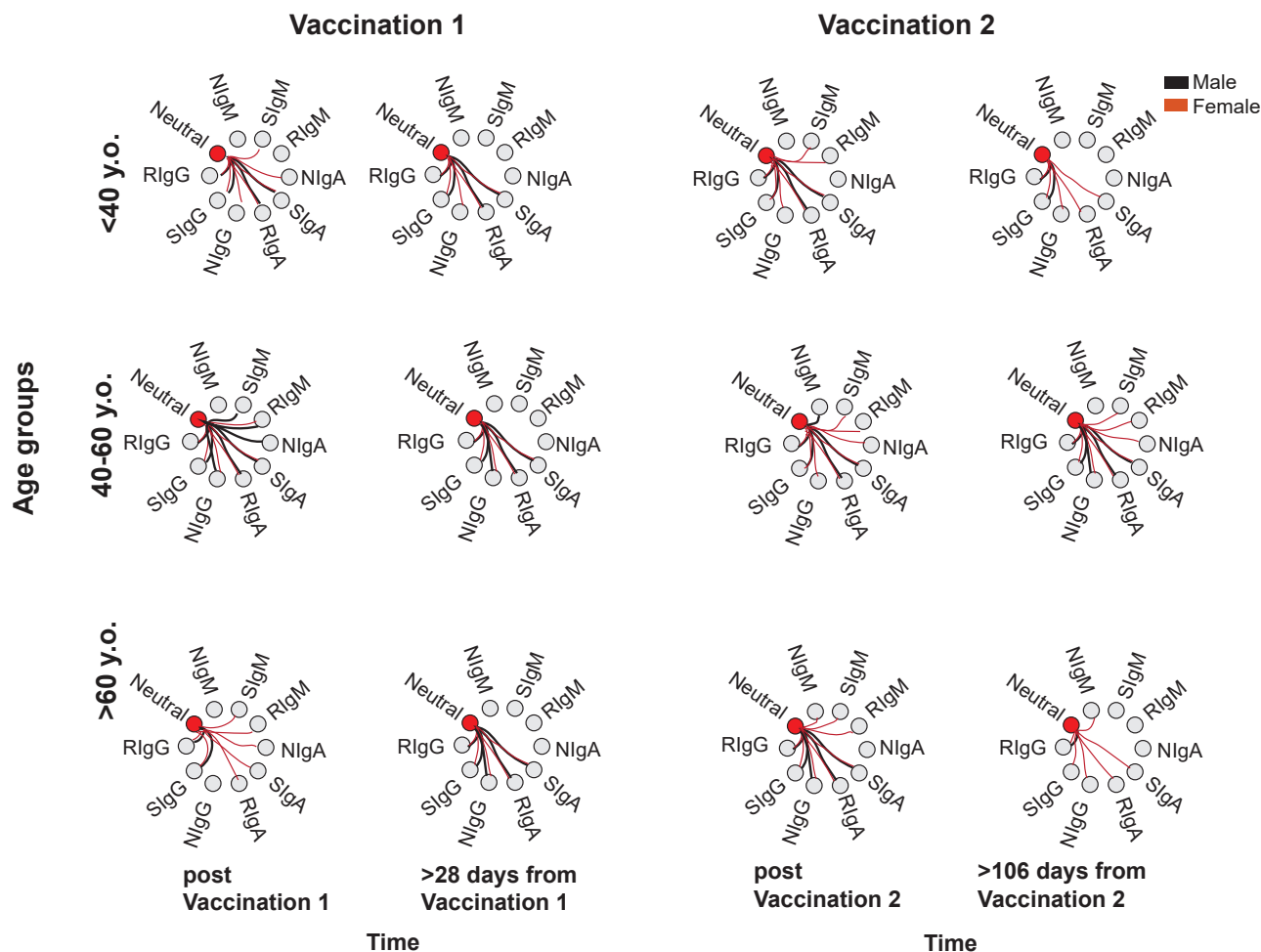
Figure 6. Correlation between neutralization and antibody isotypes is different in males and females in distinct age groups

Distance correlation between neutralization and all other features (pairwise) is calculated, and thresholds were set using sample size appropriate correlation thresholds and p values below 0.003 for measurements grouped by sex and age and time from vaccination. Vaccination groups include samples collected prior to half-time between dose 1 and 2 (vaccination 1) and dose 2 and 3 (vaccination 2) showing correlation for all time points following one dose and prior to the next and correlation after the half-time post-vaccination. Shown are correlation edges that pass the hard threshold separately for males (black line) and females (dark red line).

correlated with sex in 40–60 y.o. and over 60 y.o. but are higher in males in the 40–60 y.o. and in females in the over >60 y.o. group. However, neutralization levels are not correlated with sex in any of the age or vaccination groups suggesting different contributors to neutralization. To determine the relationship between neutralization and antibody production in these groups, we first performed a cluster analysis which revealed sex differences in the grouping of antibodies and neutralization in different groups. Furthermore, distance correlation and squared distance correlation analyses across 21-day period bins were employed to determine the contribution of different antibody isotypes to neutralization. The level of correlation contribution differed for different antibody responses over time and by sex and age. Squared distance correlation, representing the relative contribution of a variable to the level of another variable, was also calculated for each antibody measurement pairwise to normalization. Representation of this analysis shows change of contributing factors levels over time. Importantly, contributions in males and females in different age groups are sufficiently distinct allowing PCA separation of male and female samples based on this information. Unlike in the direct analysis of antibody serum where the data based on sex were not clearly separated by PCA, analysis of contributions of these factors to neutralization is clearly very different between sexes.

In sum, our study exploits novel, multivariate, machine learning, and network analysis tools integrating complex immunological parameters over time and across a diverse population to delve into the temporal dynamics of antibody responses through the COVID-19 pandemic. Specifically, unsupervised and supervised machine learning strategies, coupled with square distance correlation, have enabled us to describe changes in the role of individual antibody responses in neutralization of ACE2. Contribution program changes over time, and vaccination dose clearly differs between sex and age groups with major differences in infected vs. uninfected cohorts. In addition to the anticipated
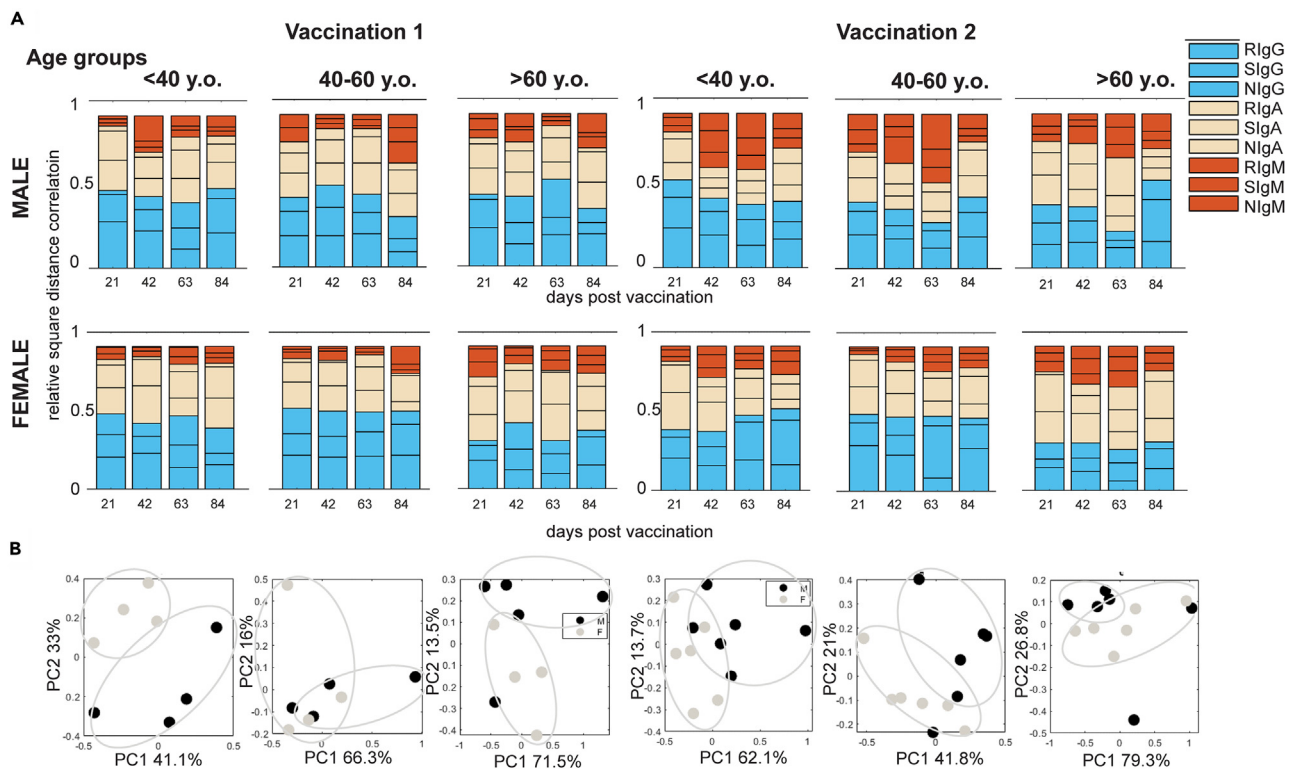
Figure 7. All measured antibody responses contribute to neutralization values at a level that is dependent on sex and age and changes over time
Relative square distance correlation (Y axis) between each antibody response was measured, and neutralization was calculated for 21-day periods post-vaccination 1 and post-vaccination 2 in each subgroup of patients. COVID-19+ and COVID-19- subjects are combined in this analysis. (A) Data are shown in bar graphs with relative contributions by immunoglobulin isotype responses (color-coded) over time (days 21, 42, 63, and 84 post-vaccination 2), and sex effects are depicted in plot groupings. (B) Unsupervised PCAs show separation of males and females square correlation profiles at 4 or 5, 21-day periods in vaccination 1 and 2 (as indicated) for different age groups. Ellipsoids indicate visual groupings of male and female data points.

effects of age on the retention antibody responses, our analyses suggest that female antibody responses to SARS-CoV-2 vaccination may be of superior quality in their co-behavior with protection (i.e., neutralization). Our work demonstrates that understanding antibody response profiles and their role in predicting neutralization goes beyond merely measuring absolute titers. Thus, rather than aiming for antibody titer achievements and changing only individual factors therein, our analyses support the pursuit of optimized vaccination strategies that aim to alter the complete network of antibody responses. Furthermore, differences in antibody-mediated responses highlighted in this work suggest the possibility of underlying cell-mediated differences between groups. Immune response differences associated with aging are often associated with immunosenescence and inflamm-aging, which impact the T cell repertoire, innate cell populations, and more.[35] In addition, biological mechanisms underlying sexual dimorphism in immunity have been focused on sex hormone signaling, and variations in Toll-like receptor expression and downstream cytokine production have been[36] observed. It remains unclear how these directly impact SARS-CoV-2 antibody-mediated responses. This sophisticated analysis of antibody responses hereby provides guidance for the parallel study of cell-mediated immunity evaluating sex and age considerations. There is an urgent need for such highly integrated indices of antibody and cell-mediated immunity as the search for reliable correlates of protection against COVID-19 and other emerging pathogens continues, to best inform the design of highly effective vaccine approaches and aid in clinical care.

Limitations of the study

Although the presented study includes nearly 3,700 longitudinal samples from nearly 1,000 individuals, and permitted analysis by sex, age, infection status (i.e., NIgG+), statistical power decreased in subgroups based on vaccination dose and brand or comorbidities; hence these factors were not considered here.

Conclusion

Using a variety of novel tools for knowledge discovery in datasets, we show that age and sex are important variables which impact the nature and temporal dynamics of infection-induced, vaccination-induced, and hybrid immunity. Overall, this work describes a novel analytical pipeline with the potential to pinpoint changes in the response to infection and/or vaccination in the general or immune-compromised individuals.

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
- METHOD DETAILS
  - Serology assay
  - Antibody neutralization assay
  - Sample collection and processing
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Knowledge discovery in data

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.110484.

AUTHOR CONTRIBUTIONS

M.C.-C. accessed and verified the data and contributed to the conceptualization, data curation, formal analysis, software, validation, visualization, writing – original draft, and writing – review and editing of the revised manuscript. S.B. accessed and verified the data and contributed to the data curation, formal analysis, software, visualization, and writing – original draft. Y.G., P.S.M., and C.A. contributed to the data curation, formal analysis, investigation, and writing – original draft. S.B. contributed to the conceptualization, methodology, validation, writing – review and editing, and editing of the revised manuscript. C.L.C. contributed to the conceptualization, methodology, resources, validation, and writing – original draft. M.-A.L. accessed and verified the data and contributed to the conceptualization, funding acquisition, methodology, project administration, resources, and supervision. J.F. and C.A.P. contributed to the conceptualization, methodology, validation, and writing – review and editing. A.M.C. accessed and verified the data and contributed to the conceptualization, funding acquisition, methodology, project administration, resources, supervision, and writing – review and editing. All authors have seen and approved of the final text.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Scully, E.P., Haverfield, J., Ursin, R.L., Tannenbaum, C., and Klein, S.L. (2020). Considering how biological sex impacts immune responses and COVID-19 outcomes. Nat. Rev. Immunol. 20, 442–447.

2. Arnold, C.G., Libby, A., Vest, A., Hopkinson, A., and Monte, A.A. (2022). Immune mechanisms associated with sex-based differences in severe COVID-19 clinical outcomes. Biol. Sex Differ. 13, 7.

3. Chen, L., Liao, H., Huang, G., Ding, S., Guo, W., Huang, T., and Cai, Y. (2022). Identification of DNA Methylation Signature and Rules for SARS-CoV-2 Associated with Age. Front. Biosci. 27, 204–211.

4. Huang, F., Chen, L., Guo, W., Zhou, X., Feng, K., Huang, T., and Cai, Y. (2022). Identifying COVID-19 Severity-Related SARS-CoV-2 Mutation Using a Machine Learning Method. Life 12, 806.

5. Takahashi, T., Ellingson, M.K., Wong, P., Israelow, B., Lucas, C., Klein, J., Silva, J., Mao, T., Oh, J.E., Tokuyama, M., et al. (2020). Sex differences in immune responses that underlie COVID-19 disease outcomes. Nature 588, 315–320.

6. Dehingia, N., and Raj, A. (2021). Sex differences in COVID-19 case fatality: do we know enough? Lancet Global Health 9, e14–e15.

7. Nguyen, N.T., Chinn, J., de Ferrante, M., Kirby, K.A., Hohmann, S.F., and Amin, A. (2021). Male gender is a predictor of higher mortality in hospitalized adults with COVID-19. PLoS One 16, e0254066.

8. Alghamdi, I.G., Hussain, I.I., Almalki, S.S., Alghamdi, M.S., Alghamdi, M.M., and El-Sheemy, M.A. (2014). The pattern of Middle east respiratory syndrome coronavirus in Saudi Arabia: A descriptive epidemiological analysis of data from the Saudi Ministry of Health. Int. J. Gen. Med. 7, 417–423.

9. Channappanavar, R., Fett, C., Mack, M., Ten, E.P.P., Meyerholz, D.K., and Perlman, S. (2017). Sex-Based Differences in

Susceptibility to Severe Acute Respiratory Syndrome Coronavirus Infection. J. Immunol. *198*, 4046–4053.

10. Peckham, H., de Gruijter, N.M., Raine, C., Radziszewska, A., Ciurtin, C., Wedderburn, L.R., Rosser, E.C., Webb, K., and Deakin, C.T. (2020). Male sex identified by global COVID-19 meta-analysis as a risk factor for death and ITU admission. Nat. Commun. *11*, 6317.

11. Lott, N., Gebhard, C.E., Bengs, S., Haider, A., Kuster, G.M., Regitz-Zagrosek, V., and Gebhard, C. (2023). Sex hormones in SARS-CoV-2 susceptibility: key players or confounders? Nat. Rev. Endocrinol. *19*, 217–231.

12. Taneja, V. (2018). Sex hormones determine immune response. Front. Immunol. *9*, 1931.

13. Klingler, J., Weiss, S., Itri, V., Liu, X., Oguntuyo, K.Y., Stevens, C., Ikegame, S., Hung, C.T., Enyindah-Asonye, G., Amanat, F., et al. (2021). Role of Immunoglobulin M and A Antibodies in the Neutralization of Severe Acute Respiratory Syndrome Coronavirus 2. J. Infect. Dis. *223*, 957–970.

14. Meyer, E.G., Simmons, G., Grebe, E., Gannett, M., Franz, S., Darst, O., Di Germanio, C., Stone, M., Contestable, P., Prichard, A., et al. (2021). Selecting COVID-19 convalescent plasma for neutralizing antibody potency using a high-capacity SARS-CoV-2 antibody assay. Transfusion (Paris) *61*, 1160–1170.

15. Diao, B., Wang, C., Tan, Y., Chen, X., Liu, Y., Ning, L., Chen, L., Li, M., Liu, Y., Wang, G., et al. (2020). Reduction and Functional Exhaustion of T Cells in Patients With Coronavirus Disease 2019 (COVID-19). Front. Immunol. *11*, 827.

16. Giamarellos-Bourboulis, E.J., Netea, M.G., Rovina, N., Akinosoglou, K., Antoniadou, A., Antonakos, N., Damoraki, G., Gkavogianni, T., Adami, M.E., Katsaounou, P., et al. (2020). Complex Immune Dysregulation in COVID-19 Patients with Severe Respiratory Failure. Cell Host Microbe *27*, 992–1000.e3.

17. Grifoni, A., Weiskopf, D., Ramirez, S.I., Mateus, J., Dan, J.M., Moderbacher, C.R., Rawlings, S.A., Sutherland, A., Premkumar, L., Jadi, R.S., et al. (2020). Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals. Cell *181*, 1489–1501.e15.

18. Peng, Y., Mentzer, A.J., Liu, G., Yao, X., Yin, Z., Dong, D., Dejnirattisai, W., Rostron, T., Supasa, P., Liu, C., et al. (2020). Broad and strong memory CD4+ and CD8+ T cells induced by SARS-CoV-2 in UK convalescent individuals following COVID-19. Nat. Immunol. *21*, 1336–1345.

19. Del Valle, D.M., Kim-Schulze, S., Huang, H.H., Beckmann, N.D., Nirenberg, S., Wang, B., Lavin, Y., Swartz, T.H., Madduri, D., Stock, A., et al. (2020). An inflammatory cytokine signature predicts COVID-19 severity and survival. Nat. Med. *26*, 1636–1643.

20. Lucas, C., Wong, P., Klein, J., Castro, T.B.R., Silva, J., Sundaram, M., Ellingson, M.K., Mao, T., Oh, J.E., Israelow, B., et al. (2020). Longitudinal analyses reveal immunological misfiring in severe COVID-19. Nature *584*, 463–469.

21. Leung, G.M., Medley, A.J., Ho, L.M., Chau, P., Wong, I.O.L., Thach, T.Q., Ghani, A.C., Donnelly, C.A., Fraser, C., Riley, S., et al. (2004). The epidemiology of severe acute respiratory syndrome in the 2003 Hong Kong epidemic: An analysis of all 1755 patients. Ann. Intern. Med. *141*, 662–673.

22. Fink, A.L., Engle, K., Ursin, R.L., Tang, W.Y., and Klein, S.L. (2018). Biological sex affects vaccine efficacy and protection against influenza in mice. Proc. Natl. Acad. Sci. USA *115*, 12477–12482.

23. Khan, S.R., van der Burgh, A.C., Peeters, R.P., van Hagen, P.M., Dalm, V.A.S.H., and Chaker, L. (2021). Determinants of Serum Immunoglobulin Levels: A Systematic Review and Meta-Analysis. Front. Immunol. *12*, 664526.

24. Khan, S.R., Chaker, L., Ikram, M.A., Peeters, R.P., van Hagen, P.M., and Dalm, V.A.S.H. (2021). Determinants and Reference Ranges of Serum Immunoglobulins in Middle-Aged and Elderly Individuals: a Population-Based Study. J. Clin. Immunol. *41*, 1902–1914.

25. Sha, C., Cuperlovic-Culf, M., and Hu, T. (2021). SMILE: systems metabolomics using interpretable learning and evolution. BMC Bioinf. *22*, 284.

26. Cuperlovic-Culf, M., Nguyen-Tran, T., and Bennett, S.A.L. (2023). Machine Learning and Hybrid Methods for Metabolic Pathway Modeling. Methods Mol. Biol. *2553*, 417–439.

27. Monti, F., Stewart, D., Surendra, A., Alecu, I., Nguyen-Tran, T., Bennett, S.A.L., and Čuperlović-Culf, M. (2023). Signed Distance Correlation (SiDCo): an online implementation of distance correlation and partial distance correlation for data-driven network analysis. Bioinformatics *39*, btad210.

28. Collins, E., Galipeau, Y., Arnold, C., Bosveld, C., Heiskanen, A., Keeshan, A., Nakka, K., Shir-Mohammadi, K., St-Denis-Bissonnette, F., Tamblyn, L., et al. (2022). Cohort profile: S top the Spread Ottawa (SSO) - a community-based prospective cohort study on antibody responses, antibody neutralisation efficiency and cellular immunity to SARS-CoV-2 infection and vaccination. BMJ Open *12*, e062187.

29. Ju, B., Zhang, Q., Ge, J., Wang, R., Sun, J., Ge, X., Yu, J., Shan, S., Zhou, B., Song, S., et al. (2020). Human neutralizing antibodies elicited by SARS-CoV-2 infection. Nature *584*, 115–119.

30. Piccoli, L., Park, Y.J., Tortorici, M.A., Czudnochowski, N., Walls, A.C., Beltramello, M., Silacci-Fregni, C., Pinto, D., Rosen, L.E., Bowen, J.E., et al. (2020). Mapping Neutralizing and Immunodominant Sites on the SARS-CoV-2 Spike Receptor-Binding Domain by Structure-Guided High-Resolution Serology. Cell *183*, 1024–1042.e21.

31. Klein, S.L., and Flanagan, K.L. (2016). Sex differences in immune responses. Nat. Rev. Immunol. *16*, 626–638.

32. Ma, H., Zeng, W., He, H., Zhao, D., Jiang, D., Zhou, P., Cheng, L., Li, Y., Ma, X., and Jin, T. (2020). Serum IgA, IgM, and IgG responses in COVID-19. Cell. Mol. Immunol. *17*, 773–775.

33. Ruggiero, A., Piubelli, C., Calciano, L., Accordini, S., Valenti, M.T., Carbonare, L.D., Siracusano, G., Temperton, N., Tiberti, N., Longoni, S.S., et al. (2022). SARS-CoV-2 vaccination elicits unconventional IgM specific responses in naïve and previously COVID-19-infected individuals. EBioMedicine *77*, 103888.

34. Sterlin, D., Mathian, A., Miyara, M., Mohr, A., Anna, F., Claër, L., Quentric, P., Fadlallah, J., Devilliers, H., Ghillani, P., et al. (2021). IgA dominates the early neutralizing antibody response to SARS-CoV-2. Sci. Transl. Med. *13*, eabd2223.

35. Fagnoni, F.F., Vescovini, R., Passeri, G., Bologna, G., Pedrazzoni, M., Franceschi, C., Passeri, M., Sansoni, P.G., Lavagetto, A., and Casti, C. (2000). Shortage of circulating naive CD8(+) T cells provides new insights on immunodeficiency in aging. Blood *95*, 2860–2868.

36. Torcia, M.G., Nencioni, L., Clemente, A.M., Civitelli, L., Celestino, I., Limongi, D., Fadigati, G., Perissi, E., Cozzolino, F., Garaci, E., and Palamara, A.T.E. (2012). Sex differences in the response to viral infections: TLR8 and TLR9 ligand stimulation induce higher IL10 production in males. PLoS One *7*, e39853.

37. Colwill, K., Galipeau, Y., Stuible, M., Gervais, C., Arnold, C., Rathod, B., Abe, K.T., Wang, J.H., Pasculescu, A., Maltseva, M., et al. (2022). A scalable serology solution for profiling humoral immune responses to SARS-CoV-2 infection and vaccination. Clin. Transl. Immunol. *11*, e1380.

38. Belacel, N., Čuperlović-Culf, M., Laflamme, M., and Ouellette, R. (2004). Fuzzy J-Means and VNS methods for clustering genes from microarray data. Bioinformatics *20*, 1690–1701.

39. Reese, S.E., Archer, K.J., Therneau, T.M., Atkinson, E.J., Vachon, C.M., De Andrade, M., Kocher, J.P.A., and Eckel-Passow, J.E. (2013). A new statistic for identifying batch effects in high-throughput genomic data that uses guided principal component analysis. Bioinformatics *29*, 2877–2883.

40. Székely, G.J., Rizzo, M.L., and Bakirov, N.K. (2007). Measuring and testing dependence by correlation of distances. Ann. Stat. *35*, 6.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
CR3022 anti-spike human IgG	Absolute Antibody	Ab01680–10.0; RRID: AB_2848080; *Not associated with Absolute Antibody, just with the CR3022 clone
CR3022 anti-spike human IgM	Absolute Antibody	Ab01680–15.0; RRID: AB_2848080; *Not associated with Absolute Antibody, just with the CR3022 clone
CR3022 anti-spike human IgA	Absolute Antibody	Ab01680–16.0; RRID: AB_2848080; *Not associated with Absolute Antibody, just with the CR3022 clone
HC2003 anti-Nucleocapsid human IgG	Genscript	A02039
CR3018 (03–018) anti-Nucleocapsid human IgM	Absolute Antibody	Ab01690–15.0; RRID: AB_2833185; *Not associated with Absolute Antibody, just with the CR3018 clone
CR3018 (03–018) anti-Nucleocapsid human IgA	Absolute Antibody	Ab01690–16.0; RRID: AB_2833185; *Not associated with Absolute Antibody, just with the CR3018 clone
human IgG	Sigma	I4506; RRID: AB_1163606
human IgM	Sigma	I8260; RRID: AB_1163621
human IgA	Sigma	I4036
Anti-human IgG-HRP fusion	Dr Yves Durocher (NRC, Montréal)	PRO1146
Goat anti human IgM -HRP	Jackson Immuno Research	109-035-011
Goat anti human IgA - HRP	Jackson Immuno Research	109-035-129; RRID: AB_2337580
SARS-CoV-2 Spike	NRC metrology	SmT1
SARS-CoV-2 RBD	Dr Yves Durocher (NRC, Montréal)	PRO6961
SARS-CoV-2 Nucleocapsid (NP)	NRC metrology	NCAP
Angiotensin-converting enzyme 2	NRC metrology	ACE2
Streptavidin-Peroxidase Polymer	Sigma	S2438
384-well high-binding polystyrene Nunc plate	Thermo Fisher Scientific	460372
SuperSignal ELISA Pico Chemiluminescent	Thermo Fisher Scientific	37069
Deposited data		
Serology measurements	This paper	Shared upon request by the lead contact
Software and algorithms		
MATLAB 2023b	Matworks Inc.	https://www.mathworks.com/
MATLAB functions	This paper	https://github.com/complimet/Analysis_SARS-CoV-2-infection-and-vaccination

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Dr. Angela Crawley (acrawley@ohri.ca).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Data reported in this paper will be shared by the lead contact upon reasonable request.
- Original code is provided at: https://github.com/complimet/Analysis_SARS-CoV-2-infection-and-vaccination. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.
- The relevant data are available in the manuscript and the appendix. Data that are not presented in the Article or appendix are available upon reasonable request from the corresponding author.

METHOD DETAILS

Serology assay

Antibodies against SARS-CoV-2 were measured using a high-throughput direct chemiluminescent ELISA performed on MicroLab STAR robotic liquid handlers (Hamilton) fitted with a 405TS/LS LHC2 plate washer (Biotek/Agilent) (full methods described previously).[37] Briefly, SARS-CoV-2 antigens (full trimeric Spike (Wuhan), receptor binding domain (RBD) (Wuhan) and the Nucleocapsid protein) were generously provided by Dr. Yves Durocher (National Research Council of Canada (NRC), Montréal) or purchased from the Metrology division of NRC and coated on assay plates. Human serum samples or controls were either diluted 1:100 for measurement of all antibody-isotype combinations, or 1:10,000 to resolve saturating signals for Spike and RBD due to vaccination. Controls included an isotype-antigen specific calibration curve pooled human sera from SARS-CoV-2 naive, convalescent, vaccinated individuals as well as a general Ig control. Isotype-specific secondary antibodies detected bound Igs. Antibody titers were determined as lab-specific ug/mL concentrations in relation to the isotype-antigen specific calibration curve or, for IgG, converted to international units (binding antibody units (BAU)/mL) via modelized WHO Standard (NIBSC 20/136).

Antibody neutralization assay

A surrogate protein-based neutralization assay was used to determine the neutralization efficiency in sera. The detailed methods, validation and calibration of the assay is available here.[37] Briefly, SARS-CoV-2 full-trimeric Spike (Wuhan)-coated plates were blocked, and then human serum samples and controls were assayed. Controls included a calibration curve (NRCoV2-20-Fc, NRC) as well as pooled human sera from naive, convalescent, and vaccinated individuals. Subsequent incubation with biotinylated ACE2 (NRC) followed by streptavidin-HRP linkage and chemiluminescent substrate addition permitted the determination of a luminescent signal. Raw luminescence values were blank (serum-free, no ACE2) adjusted using ACE2-only wells (representing unimpeded ACE2-Spike interaction). Percentage inhibition was established by measuring the reduction in ACE2-Spike interaction (0% inhibition = maximal ACE2-Spike interaction).

Sample collection and processing

All biological specimens from SSO are stored in the Coronavirus Variants Rapid Response Network (CoVaRR-Net) Biobank, following the highest standards of biobanking practices in sample collection, processing, storage, access, and distribution (upon application). Sample collection followed approved ethics protocols (REB # H-09-20-6135).

QUANTIFICATION AND STATISTICAL ANALYSIS

Knowledge discovery in data

Data preprocessing

Data points generated from serology and neutralization assays that were below the level of detection were imputed using 1/5 of the lowest measurement for the given feature. Values above the upper level of quantification were imputed with the highest measured value for the assay. In some analyses, data for each feature across samples is normalized or log transformed, and this transformation is indicated. No other data manipulation was performed. Antibody and neutralization responses were evaluated in specimens collected at baseline and days 14, 21 and 42 post-vaccination (wherever possible).

Data overview

General overview of the serology measurements was performed using Principal Component Analysis (PCA), guided PCA (gPCA), t-distributed stochastic network embedding (t-SNE) and hierarchical cluster (HCL) analysis. Cluster analysis was done using Fuzzy C-means clustering (function *fcm* in MATLAB). Briefly this approach determines the level of belonging of each feature to each cluster in this way allowing for possible multiple groupings. Fuzzy c-means (FCM) and related methods fuzzy j-means[38] allows each feature to belong to more than one group by providing a degree of membership, "belonging", to each cluster by maximizing proximity between similar features and distance between dissimilar features. FCM is based on the minimization of the objective function: $J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^m \left\| x_i - c_j \right\|^2$ where $m \in (1, \infty)$ is the "fuzzyfication" factor, u_{ij}^m is the membership degree for feature x_i to the cluster j with c_j defining the cluster center. FCM clustering assigns objects to groups with features belonging to the same clusters showing more similarity to each other than to features in other clusters. Higher membership value indicates stronger belonging to the cluster with membership value of 1 ultimately indicating that feature is only associated with the single cluster. An optimal number of clusters in FCM analysis was determined using within-cluster sum of squares and silhouette methods applied to k-means clustering as a directly related crisp method. Clustering with this approach was performed to determine grouping of features in different subject cohorts. Function used to determine number of clusters in FCM is provided on GitHub.

Guided PCA (gPCA) was performed following method previously presented for identification and batch analysis.[39] MATLAB function for gPCA application presented here is provided on GitHub. Briefly, this approach is focusing on the investigation of the importance of each group corresponding to the corresponding principal component (PC). In order to determine group loading contribution "batch" in our case group indicator matrix G, defined as $n \times b$ matrix where n is number of samples and b number of groups as assembled such as: $g_{ik} = \begin{cases} 1 \text{ if sampe } i \text{ is in group } k \\ 0 \text{ otherwise} \end{cases}$. Then single value decomposition of the guided matrix calculated as $G'X$ gives group loadings in PCA and can be used to determine separation of groups in a guided sense.

Correlation analysis

Correlations were calculated using Pearson, distance and point bi-serial correlation. Pearson correlation is a standard method used for linear correlation assessment and was used for the determination of correlation between distance correlation level values for different subgroups of the cohort.

Distance correlation was performed using previously published applications[27] and MATLAB function provided on GitHub. Briefly, distance correlation[40] was calculated using distance covariance, rather than direct covariance used in Pearson method. In this way, distance correlation provides information about both linear and non-linear correlations. Formally, distance correlation between features X and Y is calculated as:

$$dCor(X, Y) = \frac{dCov(X, Y)}{\sqrt{dVar(X)dVar(Y)}}$$

distance covariance is determined as:

$$dCov(X, Y)^2 = \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} A_{j,k} B_{j,k}$$

where A and B are calculated as simple linear functions of the pairwise distances between elements in samples X and Y. A and B are doubly centered distance matrices for variables X and Y, respectively, calculated from the pairwise distance between elements in each sample set calculated as.

$$A_{j,k} = a_{j,k} - \underline{a}_{j.} - \underline{a}_{k.} + \overline{\overline{a}}_{..}$$

where $a_{j,k} = \sqrt{(x_j - x_k)(x_j - x_k)'}$ and $\underline{a}_{j.}$ and $\underline{a}_{k.}$ are respectively the j-row and k-column mean values and $\overline{\overline{a}}_{..}$ is the overall mean of matrix A. The distance between x_j and x_k is calculated using Euclidian distance. Matrix B is populated using equivalent measures for variable Y.

Square distance correlation, i.e., the coefficient of determination, is used to represent the fraction of the variation in a variable that can be explained by the other variable. Relative square distance correlation is calculated as a square distance correlation divided by the sum of square distance correlation values for all features.

Correlation between binary values (e.g., sex) and continuous values (e.g., serological measurements) was determined using **point-biserial correlation** method as:

$$\rho_{pb} = \frac{M_1 - M_0}{s_n} \sqrt{\frac{n_1 n_o}{n^2}}$$

where correlation is calculated between binary and continuous variable by dividing continuous variable into two groups, 0 and 1, based on values of the binary feature and where M_0 and M_1 are mean values, n_0 and n_1 are numbers of samples of two subgroups and s_n is the standard deviation for the complete continuous variable set.

Polynomial function regression analysis was performed using second order polynomial and exponential functions and bisquare robustness analysis. Briefly, the bisquare robustness method for polynomial fit determination minimizes a weight sum of squares with points near the line getting full weight and points further from the line getting a reduced weight. Points that are farther from the line are expected by random chance and hence are given a weight of zero. In this way, the bisquare method aims to find a function of best fit using a least squares approach while at the same time minimizing the effect of outliers.

Feature selection

Selection of the most significantly different features between different groups of samples was performed using three different methods to determine method-independent major features: 1) statistical feature selection using F-test (function *fsrftest* running under MATLAB); 2) machine learning methods Relieff (function relief in MATLAB); and 3) neighborhood component analysis for regression (*fsrnca* function in MATLAB). Features were considered significantly different if selected by all three methods. Data was Z score normalized for each feature prior to analysis. In F-test features with $p < 0.05$, $-\log(p) > 3$ is selected as significant. In Relieff, all features with positive weight are determined as possibly contributing to group selection. Similarly, in fsrnca, all features with positive weights are kept as possibly significant. *Fsrnca* aims to find, i.e., learn, a distance metric to maximize leave-one-out classification performance. Thus, the feature selection analysis used here combines two different ML methods including filter (Relieff) and wrapper (fsrnca) with a statistical method (F-test) to determine consistently significant features. MATLAB Notebook for Relieff feature selection is provided on GitHub.