

Article

Not peer-reviewed version

# Bioinformatics-Based Design Strategy and Development of an Antigen Test for SARS-CoV-2 Nucleocapsid Protein

[Partha Ray](#) , Melissa Ledgerwood-Lee , Howard Brickner , Alex E. Clark , Aaron Garretson , Rishi Graham ,  
[Westley Van-Zant](#) , [Aaron F. Carlin](#) , [Elijah S. Aronoff-Spencer](#) \*

Posted Date: 1 April 2024

doi: 10.20944/preprints202404.0108.v1

Keywords: SARS-CoV-2; Nucleocapsid protein; Monoclonal and Polyclonal Antibodies; Enzyme-Linked Immunoassay; Peptide epitope mapping; Viral Quality Assurance; RADx; COVID-19 diagnostics



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Article*

# Bioinformatics-Based Design Strategy and Development of an Antigen Test for SARS-CoV-2 Nucleocapsid Protein

Partha Ray <sup>1</sup>, Melissa Ledgerwood-Lee <sup>1,†</sup>, Howard Brickner <sup>1,†</sup>, Alex E. Clark <sup>1</sup>, Aaron Garretson <sup>1</sup>, Rishi Graham <sup>1</sup>, Westley Van Zant <sup>1</sup>, Aaron F. Carlin <sup>1,2</sup> and Eliah Aronoff-Spencer <sup>1,\*</sup>

<sup>1</sup> Department of Medicine, Division of Infectious Diseases and Global Public Health, University of California, San Diego, CA 92093, USA

<sup>2</sup> Department of Pathology, University of California, San Diego, CA 92093, USA

\* Correspondence: earonoffspencer@health.ucsd.edu

† Equal contribution.

**Abstract:** The continuing mutability of the SARS-CoV-2 virus can result in failures of diagnostic assays. To address this, we describe a generalizable bioinformatics-to-biology pipeline developed for calibration and quality assurance of inactivated SARS-COV-2 variant panels provided to Radical Acceleration of Diagnostics programs (RADx)-radical program awardees. Heuristic genetic analysis based on variant-defining mutations demonstrated the lowest genetic variance in the Nucleocapsid protein (Np)- C-terminal domain (CTD) across all SARS-COV-2 variants. We then employed the Shannon entropy method on (Np) sequences collected from the major variants, verifying the CTD with lower entropy (less prone to mutations) than other Np regions. Polyclonal and monoclonal antibodies were raised against this target CTD antigen and used to develop an Enzyme-linked immunoassay (ELISA) test for SARS-CoV-2. Blinded Viral Quality Assurance (VQA) panels comprising of UV-inactivated SARS CoV-2 variants (XBB.1.5, BF.7, BA.1, B.1.617.2, and WA1) and distractor respiratory viruses (CoV 229E, CoV OC43, RSV A2, RSV B, IAV H1N1, and IBV) were assembled by the RADx-rad Diagnostics core and tested using the ELISA described here. The assay tested positive for all variants with high sensitivity (Limit of Detection: 1.72-8.78 ng/mL) and negative for the distractor virus panel. Epitope mapping for the monoclonal antibodies identified a twenty amino acid antigenic peptide on the Np-CTD that an in-silico program also predicted for the highest antigenicity. This work provides a template for a bioinformatics pipeline to select genetic regions with a low propensity for mutation (low Shannon entropy) to develop robust 'pan-variant' antigen-based assays for viruses prone to high mutational rates.

**Keywords:** SARS-CoV-2; nucleocapsid protein; monoclonal and polyclonal antibodies; enzyme-linked immunoassay; peptide epitope mapping; viral quality assurance; RADx; COVID-19 diagnostics

## Introduction

The emergence of the coronavirus disease 2019 (COVID-19) pandemic sparked global responses to combat the spread of the severe acute respiratory syndrome virus 2 (SARS-CoV-2). Across these efforts, significant emphasis was placed on developing diagnostic reagents and technologies for detecting the virus. These belong primarily to two classes: Nucleic Acid Amplification Tests (NAAT) to detect the viral genome and Affinity ligands (Antibodies and Aptamers) to detect viral antigenic proteins [1]. NAAT is generally more sensitive than antigen testing; however, the assay requirements for these tests are usually a

barrier to their mass deployment. Therefore, antigen-based tests like lateral flow immunoassay (LFIA), which are easy to use and cost-effective, are preferred for point-of-care and home-based testing.

Although some time has passed since the onset of the global pandemic, the development and modification of existing and novel tests have continued to seek improvements in areas including accuracy, cost, and futureproofing against future virus variants [2,3].

In the United States, testing development was supported by the Radical Acceleration of Diagnostics programs (RADx) financed under Operation Warp Speed [4]. These included programs to quickly raise US capacity (RADx-tech), ensure equitable deployment (RADx-UP), and develop the next generation of “radical” diagnostics (RADx-rad) [5].

A primary driver for the further development of diagnostic technologies in the RADx-rad program was to develop technologies to tackle the continuing mutability of SARS-CoV-2. Regions within the structural proteins targeted by diagnostic technologies can mutate to diminish recognition or avoid it altogether. This constant mutation has allowed SARS-CoV-2 to remain a public health threat, forcing constant adaptation in tests and treatments that target evolving regions. This is exemplified by the Spike (S) protein, a common target in therapy, rapid antigen tests, and NAAT due to its outward-facing positioning on the virus's surface. Although its accessibility and direct role in cellular binding and fusion make Spike a sensible target for recognizing functional virions, frequent mutations in several Spike regions allow it to evade detection and neutralization easily [6,7]. While tests can be modified to account for mutations, tests targeting variable regions of the virus remain relevant only to the specific variants they were developed against.

While variant-specific diagnostics are essential in detecting the virus's presence and identifying which variant is causing the infection, a more broadly applicable testing method is still needed. It is crucial to trust that a negative COVID-19 test result is due to a lack of the virus rather than the test's inability to detect an unaccounted-for variant.

An ideal diagnostic test would target an abundant protein's region that is sufficiently general to detect the presence of any variant yet specific enough to avoid detecting other pathogens with similar presentations. Targeting an exposed protein region of the virus far less prone to mutations could allow the test to apply to past and future virus variants.

The SARS-CoV-2 nucleocapsid (N) is the most abundant structural protein [8] with historically high conservation, making it a target for many current diagnostics [9]. The N-protein is essential in the virus's life cycle and is responsible for binding and packaging the viral genome into the ribonucleoprotein complex (RNP) [10]. The protein also plays accessory roles in immune regulation by suppressing viral RNA silencing and assisting in the transcription and replication of viral mRNA [11]. The 419-amino-acid (aa) length N-protein comprises both intrinsically disordered regions (IDRs) and regions with conserved structures, with discrete N-terminal and C-terminal domains present [9,12].

Most at-home COVID-19 test assays target N-protein, utilizing its abundance to offer speed and sensitivity. The protein is also highly immunogenic, allowing an overwhelming proportion of antibodies to be raised against it following infection [9,13]. The N-protein serves as a reasonable target due to its decreased tendency for mutation compared to other structural proteins, though it is still susceptible to mutation. It mutates at fewer amino acid positions than the Spike and Envelope (E) proteins, though it does at slightly more positions than the Membrane (M) protein [13–15]. Despite its improved reliability, it can still mutate and subvert diagnostic recognition [16]. For instance, mutations in its NTD or linker region can reduce the sensitivity or subvert detection from at-home rapid antigen tests [17]. If mutations in the nucleocapsid protein consistently occur in localized areas, other regions may not see frequent mutations. If areas within the N-protein are conserved between viral variants, they could serve as potential antigenic regions for antibody development in the novel, multi-variant COVID-19 assays.

Observing the mutation rate between viral variants within each open reading frame (ORF) allows for identifying specific amino acid positions or regions particularly resistant to mutation. Although many N-protein mutations have been identified within the SARS-CoV-2 variants, most N-

protein mutations do not map to the CTD, indicating resistance [18,19]. This indicates the potential of the CTD region as a target for antibody development. These findings can be corroborated using computational methods to identify regions of lower mutability separately.

An unbiased bioinformatics approach to sequence analysis can be implemented to calculate the entropy of regions within the nucleocapsid's genetic sequence. Regions more prone to mutation will exhibit higher entropy across viral variants, while more conserved regions will exhibit lower calculated entropic values. Low entropy regions can then serve as potential antigenic regions to produce antibodies for diagnostics.

This work describes methods developed by the Diagnostic core of the RADx-rad program to accelerate tests that are sensitive and robust to mutation while remaining specific enough for clinical utility. Two central roadblocks exist to such development. First, there is access to variants for testing that may be difficult to obtain, out of circulation, or poorly standardized; and second, the availability of tools to quickly design reagents targeting constant regions within the viral genome. To achieve this, we developed standardized, inactivated viral quality assurance panels that could be used for test development at lower levels of biohazard containment. At the same time, validating these panels themselves required rapid development and standardization of quantitative tests to both host of and evolving variants. To achieve this, we employed a hybrid bioinformatic approach that could be automated to detect invariant viral regions preserved in SARS-COV-2 in a process that can be readily generalized. We herein describe the computational approach for target selection, reagent, and antigen-based assay development, and we demonstrate its efficacy with validated VQAs employed in the RADx program.

## Materials and Methods

We present a case study of targeting sub-regions of the N-protein. N-protein was chosen due to its abundance and historical use cases for antigen-based testing. The process hinges on analyzing the comparative variability of the CTD and NTD regions of the SARS-COV-2 N-protein. While individual tools allow the assessment of defining mutations [20,21] or amino acid variability [22], there remain no accessible pipelines to conduct flexible variability analysis and comparison at scale quickly. We developed a workflow employing a combination of existing and new tools to facilitate this and similar exploration. This work used genomic sequence data shared via GISAID, the global science initiative [23].

**Community Sourced "Defining Mutations":** When navigating the landscape of SARS-CoV-2 mutations, a primary source of insight emerges from examining consensus SARS-CoV-2 "Defining Mutations," accessible via the CoVariants website [20]. These mutations signify the phylogenetic root of a variant and can be detected by examining a variant's root node on a Nextstrain tree [21]. A machine-readable compendium of defining mutations for each variant is also available on the CoVariants GitHub page. We began our investigation into sequence variability by examining these sources.

**Shannon Entropy Diversity Metric:** Mutation analysis typically incorporates one of three widely accepted metrics for quantifying uncertainty: Wu-Kabat Variability analysis [24], Simpson's Diversity [25], and Shannon entropy [26]. Here, we focus on Shannon entropy. Given a discrete random variable with M distinct realizations and  $P_i$  the probability of realization i, Shannon's Information Entropy (H) is

$$H = -\sum_{i=1}^M P_i \log_2 P_i,$$

Entropy measures the level of 'randomness' or 'disorder' within the random process and is widely used across many disciplines [27–30]. A process with high predictability and low variability has low entropy and vice versa. In biological systems, Shannon's methods have provided a statistically sound measure of system diversity and sequence analyses [31–34]. The entropy information analysis eliminates serious bias and exhibits more stability than the Wu-Kabat and second-generation measures of diversity.



**Single Amino Acid Variability:** In gene sequence alignments, Shannon entropy has been used to quantify conservative locations by comparing the frequency of amino acid [35] or nucleotide [36] realizations at a given position. For amino acid conservation, related sequences are aligned, and entropy is estimated across the alignment for each amino acid position,  $j$ . As the total number of amino acid types is 20, we estimate the Shannon entropy at position  $j$  as

$$\hat{H}_{aj} = -\sum_{i=1}^{20} \left( \frac{k_{ij}}{N} \right) \log_2 \left( \frac{k_{ij}}{N} \right),$$

where  $k_{ij}$  represents the count of amino acid  $i$  at position  $j$ , and  $N$  the total number of aligned sequences. Here, the estimated entropy  $\hat{H}_{aj}$  can range from 0 to 4.322. A value of 0 indicates that only one type of residue is present at that position, signifying complete conservation. On the other hand, a maximum value of 4.322 signifies complete variability, where all 20 types of residues are equally represented at that position [37].

**Epitope Windowed Shannon Entropy:** In our study, we first evaluate the amino acid entropy as a point-based evaluation described above, then consider entropy over an alternate basis to concentrate on short sequences instead of individual residues. That is, we evaluate strings of amino acids that would represent the epitope binding sites. Given that the epitopes typically range in size from 4 to 12 amino acids long [38], we chose to focus our analysis on an epitope window size of 10 amino acids. The entropy of the window centered at position  $j$  is then estimated with:

$$\hat{H}_{wj} = -\sum_{i=1}^{M_w} \left( \frac{s_{ij}}{N} \right) \log_2 \left( \frac{s_{ij}}{N} \right),$$

Where  $M_w = 20^{10}$  is the total number of possible amino acid sequences of length 10, and  $s_{ij}$  is the count of the  $j$ th such sequence at position  $i$  among aligned isolates.

This metric, which applies entropy to a window of 10 amino acids, was designed to more accurately gauge the impact of mutations on the linear epitope structure and subsequent antibody recognition. Our approach more faithfully represents the consequences of comprehensive changes within the linear epitope region than single-point mutation analysis.

**Sample Collection and Preparation:** Sequence analysis in this study used 1800 genomic sequences available on GISAID collected from 5/20/2020 to 4/5/2023. This dataset is accessible at 10.55876/gis8.240302kp, and contains 150 sequences each from the 12 WHO variants (e.g., Alpha, Beta, Gamma, Delta, Iota, Mu, Omicron, Eta, Kappa, Lambda, Epsilon, Zeta). The sequences were screened to be “complete” reads, as many genomic data in GISAID are partial sequences. Any sequences containing “no read” sections were eliminated and replaced with an alternate sample from GISAID. The complete dataset was further processed by aligning and cropping the sequences to limit the ORF portion of the N protein. This dataset was then translated into amino acid form. The entire N protein represents 419 amino acids, the NTD is 131 amino acids, and the CTD is 118 [14]. Only the NTD and CTD amino acids were used for Shannon entropy analysis.

**Statistics:** For our statistical approach, we compared the Shannon entropy values extracted from the N-terminal domain (NTD) and the C-terminal domain (CTD) through a Wilcoxon rank-sum analysis. This non-parametric test is suitable for comparing two independent samples. The Wilcoxon rank-sum analysis enables us to evaluate whether the differences in the entropy values from the two regions are statistically significant, i.e., whether one region exhibits a significantly greater degree of sequence variability than the other. This approach of using a hypothesis test to compare mean entropy values has been used, e.g., to compare mutation frequency across subsequent waves of SARS-CoV-2 in Pakistan [39].

#### **Cloning and Expression of SARS-CoV-2 Nucleocapsid (Np) protein subdomains:**

The Nucleocapsid protein has two major structural and functional units: the N-terminal domain (NTD) and the C-terminal domain (CTD). The NTD (aa 46-176) is a disordered region responsible for RNA binding, achieving this through a positively charged cavity between the domain's core and a basic  $\beta$ -hairpin [9,12]. The CTD (aa 247-364) is responsible for dimerization and can also bind with

RNA [12]. An IDR linker domain (LKD) (aa 177-246) connects the NTD and CTD, and each domain contains an additional N-terminal arm (aa 1-45) and C-terminal arm (aa 365-419) on its outer edges [11]. We cloned and expressed these domains separately for this study.

The Np-CTD was expressed from the pUNO1His-plasmid vector (Invivogen). Briefly, the target portion of the Np open reading frame (ORF) expressing the residues Np-CTD (Lys248-Pro364) was PCR amplified from the pUNO1His-SARS2-N plasmid vector (4.7kb) (Invivogen. Catalog code: p1his-cov2-n) using the primers: F1- Forward primer (Xho1 overhang) and R1- Reverse primer (BamH1 overhang) and ligated to the pUNO1His-plasmid vector, using its Xho1 and BamH1 restriction sites (Supplementary Figure 1). The ligated plasmids were transformed into chemically competent bacterial cells (Subcloning Efficiency™ DH5α Competent Cells Catalog number: 18265017). Individual bacterial colonies were isolated from the Luria Broth (LB) agar plates with a Blasticidin (Invivogen, ant-bl-10p) antibiotic section. Following this, the selected clones were propagated in LB media with Blasticidin, and the plasmid DNA was isolated from the bacterial cells using DNA isolation protocol (Qiagen Miniprep kits, Cat. No. / ID: 27104).

Similarly, the N Terminal portion (Met1-Thr247) of the Np was PCR amplified from the pUNO1His-SARS2-N plasmid vector and subcloned in the pUNO1His-plasmid vector using the primers: F2- Forward primer (Xho1 overhang) and R2-Reverse primers (BamH1 overhang), using the similar method mentioned above (Table 1)

Table 1.

Primers	Sequence	Target
F1	5'- CCGCTCGAGAAGAAATCTGCTGCTGAGGCTTC-3'	Np-CTD (Lys248-Pro364)
R1	5'- CTCGGATCCTTATGGGAATGTTTTGTATGCGTC-3'	Np-CTD (Lys248-Pro364)
F2	5'-  AGGCACTCGAGATGTCCGATAATGGGCCACAGAA-3'	Np-NT (Met1-Thr247)
R2	5'-  CTCTGGATCCGGTAACGGTCTGCCCCTGTTGCTGTTG-  3'	Np-NT (Met1-Thr247)

All the primers were ordered from IDT (Integrated DNA Technologies, Inc., Coralville, Iowa). DNA sequencing was performed on the plasmid DNA vectors from selected clones to ensure the correct ORF sequences.

The clones, pUNO1His-SARS2-N, pUNO1His-SARS2-Np-CTD, and pUNO1His-SARS2- Np-NT, were transfected into HEK293 cells following the standard Lipofectamine 3000 Reagent (Invitrogen, Catalog number: L3000001) protocol for the expression of Np-Full length, Np-CTD, and Np-NT proteins, respectively:

Table 2.

Clones	Protein
pUNO1His-SARS2-N (Invivogen Catalog code: p1his-cov2-n)	Full-length Np (Met1- Ala419)
pUNO1His-SARS2-Np-CTD (This work)	Np-CTD (Lys248-Pro364)
pUNO1His-SARS2- Np-NT (This work)	Np-NT (Met1-Thr247)

The recombinant proteins were expressed in mammalian cells with a Carboxy terminal Histidine tag (6X-HIS) and secreted in the cell media (Conditioned media) (Supplementary Figure 1). Protein expression was additionally verified using the Western blot (data not shown).

**Cell Culture:** HEK293 cells obtained from the (ATCC: CRL-1573) were grown in Dulbecco's Modified Eagle's Medium, DMEM (Gibco, Catalog number: 11965118) media containing 10% Fetal Bovine serum, FBS (Gibco, Catalog number: A5670701) and 1% Penicillin-Streptomycin antibiotics (Gibco, Catalog number: 15140122). VeroE6/TMPRSS2 cells (Sekisui XenoTech) are the VeroE6 cell line modified to express the serine protease TMPRSS2 under Geneticin selection. The cell line is used to produce large stocks of SARS-CoV-2 virus. The cells were grown in DMEM media with 10% Heat inactivated FBS, 2 mM Glutamine, and 1mg/ml of Geneticin (G418). Calu-3 (ATCC: HTB-55) is a lung carcinoma of human epithelial cells used to grow SARS-CoV-2. These cells were grown in DMEM with 1% non-essential amino acids, 2 mM L-glutamine, 1 mM sodium pyruvate, 1.0 g/L glucose, and 20% FBS. All cells were grown at 37°C under standard tissue culture conditions.

**Western Blots:** Whole Cell lysates or the conditioned media were collected from HEK293 cells transiently expressing the Np-FL, Np-CTD, and Np-NT and subjected to the SDS-PAGE technique to resolve the proteins. The proteins were then transferred to a PVDF (Bio-Rad, Catalogue: 1620177) membrane using a gel-transfer apparatus and subjected to Western blotting using mouse monoclonal anti-nucleoprotein primary antibodies at 1:1000 dilution. HRP conjugated Goat anti-mouse (ThermoFisher Scientific Catalogue: 31430) was used as the secondary antibody at 1:1000 dilution, and the antigen-antibody complexes were detected using the ECL system (ThermoFisher Scientific Catalogue: 32209). HRP conjugated beta-tubulin (ThermoFisher Scientific Catalogue: MA5-16308-HRP) at 1:4000 dilution was used as the protein loading control.

**SARS-CoV-2 isolation and culture:** Culture of SARS-CoV-2 variants acquired from BEI (WA1, B.1.351, and B.1.617.2) or isolated from clinical samples (B.1.1.7, BA.1, BA.2.3, BA.2.12.1, BA.5.1, BF.7, BQ.1, and XBB.1.5). Viruses acquired from BEI were propagated on TMPRSS2-VeroE6 (XenoTech) cells as described PMID previously [40]. Viruses from clinical samples were isolated as described previously at UC San Diego under IRBs #200477 (B.1.1.7 [40]), #160524 (BA.1 [41], BA.2.3 [41], and BA.5 [42]) and #200236X (BA.2.12.1 [42]). BF.7, BQ.1, and XBB.1.5 were isolated as described for BA.5 under IRB #160524. Briefly, serial dilutions of the clinical sample were serially diluted in DMEM with 1x Pen/Strep + Amphotericin (Anti/Anti) and 10mM HEPES and applied to monolayers of Calu-3 or TMPRSS2-VeroE6 cells. After one hour, the media above supplemented with 2% FBS was added, and viruses were harvested when the cytopathic effect (CPE) became apparent. Passage 0 stocks were expanded on TMPRSS2-VeroE6 cells and titered by fluorescent focus assay on TMPRSS2-VeroE6 cells. All viral stocks were verified by whole genome sequencing.

All work with infectious SARS-CoV-2 was conducted in BSL3 conditions at UC San Diego following the guidelines approved by the Institutional Biosafety Committee.

**Viral inactivation and VQA panel methods:** Viruses isolated and propagated, as described above, were inactivated at BSL3 by UV<sub>254</sub> irradiation. UV-inactivation was performed with 400mJ/cm<sup>2</sup>

in a UVP Crosslinker CL-3000 6.1 (Analytik Jena) in a thin layer of <4ml in a 10cm dish on a cold block so that the plate bottom is 5.5 inches from irradiation source. Culture media was similarly UV-treated as a control for downstream assays.

Each SARS-CoV-2 sample was confirmed inactive before removal to BSL2. Inactivation confirmation was performed by an extended culture of 10% of the volume on TMPRSS2-VeroE6 cells (or TMPRSS2-VeroE6 and Calu3 cells for variant BA.1 due to reduced infectivity of BA.1 on TMPRSS2-vero) and examination for CPE, followed by passaging of the entire volume of supernatant to new 96-well plates of cells for additional growth and staining with polyclonal nucleocapsid primary (GeneTex, #gtx135357) and AlexaFluor 594 secondary antibody. Images of whole wells were acquired and examined for positive staining. Positive and negative controls were included at each step.

Inactivated viruses were assembled into viral quality assurance panels (VQAs) by dilution in viral transport medium (VTM) (RMBio VTM-CHT-01L) or media (DMEM+ 2% FBS, 1x penicillin/streptomycin, 10 mM HEPES). Dilutions were aliquoted in pre-labeled cryotubes with coded labels and stored at -80°C until assayed.

**Monoclonal and Polyclonal Antibody Production:** ProMab Biotechnologies Inc. performed large-scale Np-CTD protein expression and antibody production. The Histidine tagged-Np-CTD protein was expressed in the Mammalian cell expression system (HEK293) and purified using Ni-NTA affinity column chromatography. The purified protein was resolved in SDS-PAGE and stained with Coomassie Blue (Supplementary Figure 1) to check the yield and purity of the protein. Two Rabbits and five Balb/c Mice were immunized with the protein antigen to generate the Rabbit polyclonal and Mouse monoclonal antibodies.

The mice were injected with the protein antigen five times for monoclonal antibody production, with an interval of three weeks between each injection to generate the mouse monoclonal antibodies. Next, sera collected from the mice were subjected to direct ELISA using the Np-CTD as the antigen. Based on the ELISA result, the mouse with the highest titer was carried on to hybridoma fusion. The top ten hybridoma clones (C1-C10) obtained were next tested by direct ELISA using Full-length Np. Two best-performing clones (C9 and C10) were selected for monoclonal antibody production and purification. For this, the hybridoma clone cells (C9 and C10) were first expanded in DMEM, 10% BSA media using the standard tissue culture technique. The expanded cells were then injected into the peritoneal cavity of five mice through the intraperitoneal (IP)- injection method for antibody generation. Ascites were then collected from the injected mice and purified using an IgG purification column to obtain the monoclonal antibody (mAb 9 and mAb 10).

To generate the rabbit polyclonal antibodies, two rabbits were injected with the Np-CTD protein five times, with an interval of three weeks between injections. After this, the rabbits were sacrificed, their sera were collected, and direct-ELISA was performed using Np-CTD as antigen to test the antibody titer. Then, the sera were precipitated using ammonia persulfate and purified through protein A to get the polyclonal antibody 108.

**Enzyme-Linked Immunosorbent Assay (ELISA):** The sandwich ELISA was developed using the Rabbit polyclonal Antibodies to capture the SARS-CoV-2 Np-antigen and the Mouse monoclonal antibodies for detection. For this, the 96-well microtiter plates (Sarstedt, Catalogue: 82.1581.100) were coated with the capture polyclonal antibodies #108 at dilution (1:1000). The antibodies were first diluted in the carbonate-bicarbonate (pH 9.4) buffer (ThermoFisher Scientific Catalogue: 28382), and 100 µL of this was added to each well. The plates were then sealed with adhesive strips and incubated overnight at 4°C for the antibodies to bind to the microtiter wells through adsorption. The next day, the contents of the plates were discarded, and the wells were washed twice with 200 µL PBS buffer.

Following this, 200 µL of blocking buffer (1x PBS with 1% BSA) was added to each well, covered with adhesive strips, and incubated at room temperature for two hours. The serum albumin proteins in the blocking buffer use the remaining well-surface that is unoccupied by the antibodies and thus improve the assay's sensitivity by reducing the background signal and increasing the signal-to-noise ratio. Next, the blocking solution is discarded, and 100 µL of the Np-antigen (recombinant full-length SARS-CoV-2 Np, AcroBiosystems: NUN-C5227) or the UV-inactivated viruses at specified



concentrations, diluted in the blocking buffer, are added to the antibody-coated wells. The plates are then covered with the adhesive strip and incubated at 37°C for 90 minutes. After this incubation, the solutions in the plates are discarded, and the wells are washed four times with 200 µL PBS buffer. Next, 100 µL of the mouse monoclonal detection antibodies (either mAb 9 or mAb10) at (1: 4400) dilutions in blocking buffer are added to each well, covered with an adhesive strip, and incubated at room temperature for two hours in a rocker. The solutions are discarded, and the wells are washed four times with 200 µL PBS buffer. The Horseradish peroxidase (HRP) conjugated Goat-anti mouse secondary antibodies (ThermoFisher Scientific Catalogue: 31430) at 1:2100 dilution is added to the wells and incubated in a rocker at room temperature for one hour. Following this incubation, the wells are washed with 200 µL PBS buffer four times, and 100 µL of the 3,3',5,5'-Tetramethylbenzidine (TMB) substrate solution (ThermoFisher Scientific Catalogue: N301) is added to each well and incubated for 15 minutes. The chromogenic substrate reaction is terminated by adding 100 µL of stop-solution (ThermoFisher Scientific Catalogue: N600), and the plates are scanned at Abs 450 nm wavelength in a Tecan Multimode microplate reader (Spark®) within 15 minutes. The stock concentration of Rabbit polyclonal antibodies (#108) was 1.3 mg/ml. The mAbs 9 and mAb 10 were at 3 mg/ml and 2.2 mg/ml concentrations, respectively.

The microtiter wells were coated with the Np-antigen at the 250 ng/ml concentration overnight at 4°C using the above-mentioned method to check the antibody titer and select the hybridoma clones (C1-C10) using ELISA. The assay was performed by first incubating the wells with 100 µL of the cell supernatant of each hybridoma clone (1:10 dilution in blocking buffer) for two hours at 37 °C for 90 minutes and washing four times in 200 µL PBS buffer. Goat-anti-mouse secondary antibodies (1:2100 dilution) were added to the wells and incubated for one hour at room temperature. Following this, the wells were washed, the signals were developed, and the plates were scanned at Abs 450 nm in a Tecan using the earlier method. All the ELISA was conducted in triplicates (n=3) at every concentration for statistical significance and Limit of detection (LOD) calculations.

A commercially available Np-ELISA kit (Ray Biotech, COVID-19 / SARS-COV-2 Nucleocapsid Protein ELISA Kit, Catalog Number: ELV-COVID19N) was used to compare our Np-CTD ELISA kit.

**Calculations of Limit of Detection (LOD):** The limit of detection is defined as the lowest concentration of an analyte in a sample that can be consistently detected with a stated probability, typically at 95% certainty. Based on the guidelines provided by the FDA's International Committee on Harmonization, [43] the LOD is expressed as

$$LOD = \frac{3.3\sigma}{S}$$

Where  $\sigma$  is the standard deviation of the blank, and S is the slope of the curve. Concurrently, the limit of quantification (LOQ), which represents the smallest amount or lowest concentration of a substance that can be determined with established accuracy, precision, and uncertainty, is calculated as

$$LOQ = \frac{10\sigma}{S}$$

Current ELISA work on the N protein determines concentrations in ng/ml. To translate these concentrations into nanomolar (nM), the molecular weight of the N protein was calculated based on its size of 419 amino acids (46.09 kDa). Since 1 Da = g/mol, this translates to 46,090 ng/nmol. Therefore, the conversion to nM involves dividing the concentration in ng/ml by 46.09 ng.nmole/ml·L, allowing the analyte's concentration to be expressed in nM.

**Immunofluorescence Assay:** Calu-3 cells were infected at an MOI of 0.05 and fixed with 4% formaldehyde in PBS 30 min at RT 24h later. After PBS washes, cells were permeabilized and blocked with 30 min incubation in 1% BSA and 0.1 % Triton X-100 in PBS. Cells were stained with anti-nucleocapsid primary antibody (GeneTex, gtx135357) or monoclonal and polyclonal anti-N antibodies followed by AlexaFluor 594 or 647-conjugated secondary antibody (Thermo Fisher Scientific) with nuclear counterstain Sytox Green (Thermo Fisher Scientific). Images were acquired on an Incucyte SX5 imager at 20x magnification.

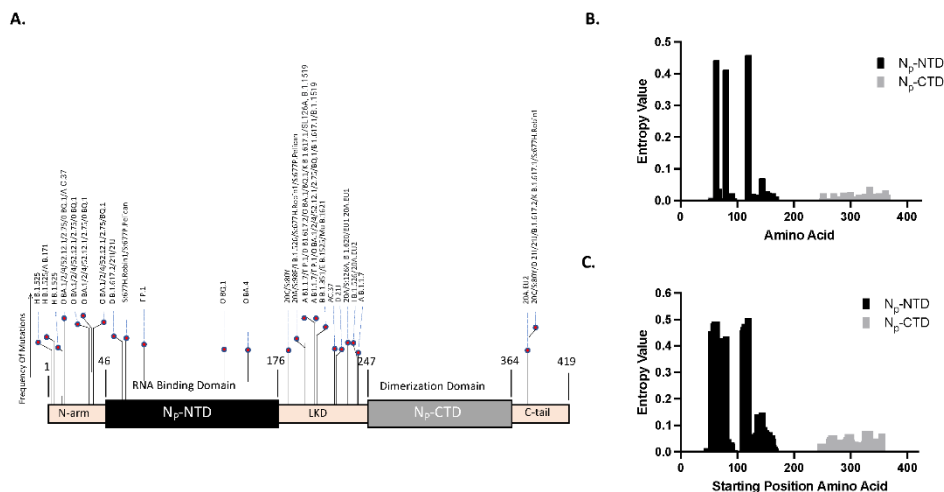
Results

Bioinformatics Approach for Selecting Antigen

a) Examination of the pre-existing “Defining Mutations.”:

The defining mutations for N protein SARS-CoV-2 were collected from the CoVariants website [20] on October 21, 2022. The variants that were included within the defining mutations were: Alpha B.1.1.7, Beta B.1.351, Gamma P.1, Delta B.1.617.2, Delta 21I, Delta 21J, Omicron BA.1, Omicron BA.2, Omicron BA.4, Omicron BA.5, Omicron BA.2.12.1, Omicron BA.2.75, Omicron BQ.1, Kappa B.1.617.1, Eta B.1.525, Iota B.1.526, Lambda C.37, Mu B.1.621, Epsilon 20C/s:452R\*.

These mutation points were plotted with the frequency of mutation representing the number of times this mutation appeared across different variants. In Figure 1A, we illustrate the defining mutations of the N protein through a bar graph schematic, wherein the mutation locations are indicated, and the height of the bars corresponds to the number of variants that possess that specific mutation. Interestingly, while defining mutations appear in various regions of the N protein, including the NTD, no defining mutations were observed in the CTD. This implies that CTD is potentially a highly conserved region, making it a promising target for our antibody development efforts. While these identified mutations provide a helpful starting point, they may only encompass some potential mutations differentiating one variant from another, which requires quantitative mutation analysis such as described here.



**Figure 1. Bioinformatics approach to select the Nucleocapsid protein antigenic region:** **A.** Bar graph illustrating the defining mutations for the N protein. The location of each mutation is plotted on the x-axis, and the height of each bar indicates the number of variants carrying that specific mutation. Notably, no defining mutations are observed in the CTD region. **B.** Single point Shannon entropy of the NTD and CTD regions based on 1800 genomic sequences (150 samples per each of the 12 SARS-CoV-2 variants). Each bar represents the entropy value at a specific amino acid position, indicating the variability at that site. **C.** Windowed Shannon entropy analysis using an epitope-sized window (10 amino acids). Each bar indicates the entropy of the window beginning at the position of the bar on the x-axis. In both (B) and (C), NTD exhibits higher entropy values compared to CTD, indicating more variability in the former, as supported by statistically significant P-values of < 0.005.

**b) Single-Point Shannon Entropy Analysis for Mutation Variability:** Next, we compared the  $\hat{H}_{aj}$ , the point-based amino acid entropy estimates over the regions of interest. Figure 1B presents our analysis of our unique dataset of 1800 genomic samples, which includes 150 samples from each of the 12 SARS-CoV-2 variants. Our study focused solely on the NTD and CTD regions, hence the absence of bars in the N-arm, LKD, and C-tail. The bar graph displays the variability at each

specific amino acid position (x-axis) via entropy values (y-axis). The NTD's three most variable positions were amino acids 119, 63, and 80, with entropy values of 0.458, 0.442, and 0.412, respectively. Conversely, the CTD's top three positions in terms of variability were amino acids 334, 300, and 288, with entropy values of 0.044, 0.028, and 0.025, respectively. Overall, the NTD exhibits higher average entropy values, indicating more variability, than the CTD—a finding further substantiated by a statistically significant p-value of  $< 0.005$ .

- c) **Epitope Windowed Shannon Entropy Analysis:** Next, we considered  $\hat{H}_{wj}$ , the estimated entropy over a multiple amino acid moving window. In Figure 1C, we display the entropy results derived from our analysis performed over a linear epitope-sized window, in this instance, a 10-amino-acid sequence.

Each bar denotes the entropy or variability across the respective window, with the bar positioned at the first amino acid of the window. This analysis further substantiated our previous findings: the NTD shows higher variability, with the most highly variable three window locations starting at positions 119, 63, and 80, presenting entropy values of 0.504, 0.491, and 0.436, respectively. The CTD, in contrast, displayed the highest values at window starting positions 325, 355, and 296, with entropy values of 0.078, 0.069, and 0.067, respectively. A statistically significant p-value of  $< 0.005$  confirms the greater variability in the NTD region. This approach more faithfully represents the consequences of comprehensive changes within the entire linear epitope region than single-point mutation analysis.

**Production of antibodies:** With the Np-CTD identified by our bioinformatics approach as the region with lesser variability than the NTD, it was selected as the target for antibody development. The open reading frame (ORF) corresponding to the CTD region (Lys248-Pro364) was sub-cloned into the pUNO1His-plasmid vector (Invivogen) for protein expression in mammalian HEK293 cells. We used a mammalian protein expression system to introduce the post-translational modifications (PTMs) on the Np-CTD antigen that would mimic similar modifications during infection [44]. The expressed CTD protein with histidine tag was purified using the standard Ni-NTA affinity purification protocol. The purified protein was resolved in SDS-PAGE and Coomassie stained to determine the purity and yield of the protein (Supplementary Figure 1), and it was used to immunize rabbits and mice to raise polyclonal and mouse monoclonal antibodies.

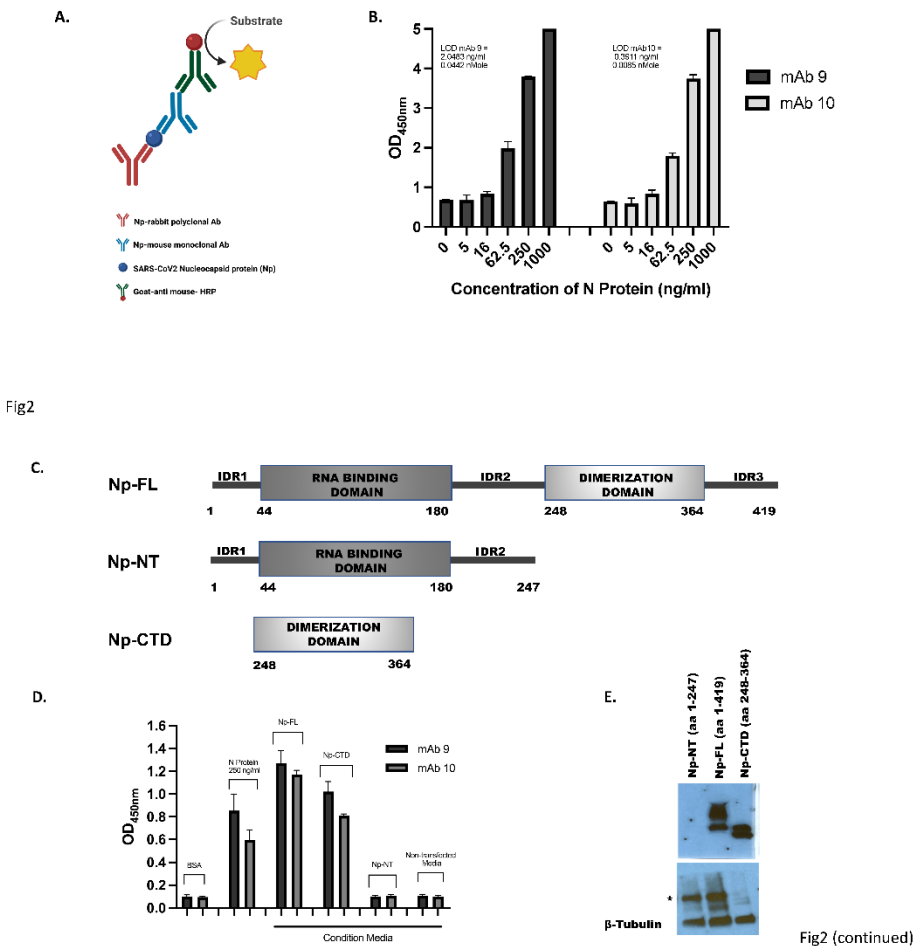
For polyclonal antibody production, the antibody titer from two immunized rabbits (107 and 108) was compared using ELISA to select the polyclonal antibody with stronger CTD binding. The serum samples collected from the immunized animal were tested using purified Np-CTD antigen in ELISA. The pre-immunized serum samples collected from the same animals were used as negative controls. Compared to the pre-immunized serum, strong antibody titers were detected in the post-immunized serum in both animals, and both animals had comparable polyclonal antibody titers. However, since the serum antibody titer in rabbit 108 was stronger than that in rabbit 107, it was selected for polyclonal antibody purification and subsequent assay development (Supplementary Figure 2).

Five mice were injected with the CTD antigen to produce mouse monoclonal antibodies, and the pre- and post-immunization sera from these animals were analyzed. The mice with the highest serum titer were subsequently selected to produce the hybridoma fusion colonies (data not shown). Next, the supernatants from ten hybridoma clones (C1-C10) were analyzed with ELISA to choose the best monoclonal antibody candidates. Of the ten hybridoma clone supernatants analyzed, C9 and C10 had the highest titer and were selected for antibody production, purification, and subsequent development of the diagnostics work (Supplementary Figure 3).

#### *Characterization of the Antibodies*

- a) **Developing sandwich ELISA and determining the assay's sensitivity:** We next developed the ELISA assay using the antibodies produced against the Np-CTD. Figure 2A depicts a schematic representation of the sandwich ELISA developed using rabbit polyclonal and mouse monoclonal

antibodies. The rabbit polyclonal antibody, shown in red (108), was immobilized onto the ELISA micro-titer plates to capture purified recombinant Nucleocapsid protein. The mouse monoclonal antibodies, either mAb 9 or mAb 10, depicted in blue, were used for detection. A secondary goat-anti mouse-HRP antibody was then used for development.



**Figure 2. Developing the RADx sandwich ELISA:** **A:** Schema of the ELISA: The rabbit polyclonal antibodies raised against the antigen, Nucleocapsid protein (Np) CTD, were used to coat the microtiter plates for capturing the SARS CoV-2 Np. The two mouse monoclonal antibodies (either mAb 9 or mAb10) raised against the antigen Np CTD were used as detection antibodies, and the colorimetric assay was developed using Goat-anti-mouse antibodies conjugated to HRP. **B.** Recombinant purified Full-length (FL) Np at indicated concentrations were titrated to determine the ELISA's Limit of Detection (LoD). **C:** Nucleocapsid protein Full length (Np-FL), N-terminal portion (Np-NT), and C-terminal Domain (Np-CTD), as indicated by their size, were cloned and expressed as recombinant proteins in mammalian cells. **D.** The expressed proteins secreted in the conditioned media were used to determine the specificity of the ELISA. **E.** The conditioned media were subjected to the Western blots assay using mAb10. The bands were detected in Np-CTD and Np-FL; no bands were detected in the Np-NT-conditioned media. The higher migrating bands are due to the post-translational modification of the Np. Beta-tubulin was used as the loading marker; the higher migrating bands, indicated by asterisks, were the Beta-tubulin dimer.

To determine the ELISAs' sensitivity, titrations of the N-protein were conducted using the mAb 9 and mAb 10-based sandwich ELISAs (Figure 2B). Two independent ELISA experiments were conducted to estimate the Limit-of-detection (LoD), with each test using the same sandwich ELISA



technique to account for day-to-day variance. The results of Experiments 1 and 2 for each monoclonal antibody are presented in (Table 3). The two experiments for mAb 9 resulted in an average LoD of 1.679 ng/ml (0.036 nMol), while the experiments for mAb 10 resulted in an average LOD of 0.884 ng/ml (0.019 nMol). As the mAb 10-based ELISA exhibited a lower overall LOD, it was identified as the more sensitive assay.

Table 3.

mAb 9	Experiment 1	Experiment 2	Average	SD
Slope	0.125	0.050	0.088	0.053
SD of Blank	0.008	0.002	0.005	0.004
LOD (ng/ml)	2.048	1.310	1.679	0.522
LOD (nMol)	0.044	0.028	0.036	0.011
LOQ (ng/ml)	0.621	0.397	0.509	0.158
LOQ (nMol)	0.013	0.009	0.011	0.003
mAb 10	Experiment 1	Experiment 2	Average	SD
Slope	0.125	0.154	0.140	0.020
SD of Blank	0.001	0.007	0.004	0.004
LOD (ng/ml)	0.391	1.377	0.884	0.697
LOD (nMol)	0.008	0.030	0.019	0.015
LOQ (ng/ml)	1.185	4.589	2.887	2.407
LOQ (nMol)	0.026	0.100	0.063	0.052

- b) Determining the Selectivity of the ELISA:** In addition to determining the sensitivity of the assays, the specificity was also analyzed to ensure no cross-reactivity or false positives would arise while using the assay. To test for cross-reactivity with other regions of the N-protein, three versions of the nucleocapsid were produced: the full-length N-protein (Np-FL), a truncated version containing the NTD and its flanking IDRs (Np-NT), and a version containing only the CTD (Np-CTD). Schematic representations of the three proteins are depicted in Figure

2C. Plasmids expressing these constructs were transfected into mammalian cells, and the conditioned media containing the secreted proteins were collected. Sandwich ELISAs with the conditioned media were conducted in the same manner as before to compare binding with these three proteins (Figure 2D). Both mAb 9 and mAb 10 exhibited CTD-specificity, binding to Np-FL and Np-CTD. Binding was not observed in Np-NT, which showed the same background binding as the BSA control. As a positive control, commercially purchased, purified N-protein (250 ng/ml) was tested, and binding activity was observed to a lesser degree than Np-FL and Np-CTD. In addition to the BSA and Np-NT negative controls, non-transfected media was also tested, and it exhibited the same degree of background binding as the BSA and Np-NT tests. This indicates that both mAb 9 and mAb 10 are specific to the CTD and are resilient to cross-binding against other regions of the N-protein.

A Western Blot was also conducted to corroborate these findings. The conditioned media were probed with mAb 10 antibody. The results of the Western Blot are seen in Figure 2E. No bands were observed in the Np-NT sample, while bands were present in the Np-FL and Np-CTD samples. The additional, higher-migrating bands observed in the Np-FL and Np-CTD result from post-translational modifications in the N-protein [45].  $\beta$ -Tubulin was used as a loading control to verify that each sample was loaded with equal concentrations of the conditioned media. These findings further indicate the CTD-specificity of the assays and the monoclonal antibodies they utilize.

**Immunofluorescent staining of SARS-CoV-2 infected cells using the Antibodies:** The ability of antibodies to detect nucleocapsid by immunofluorescence in cells infected with authentic SARS-CoV-2 was also tested. Calu-3 lung cells were infected for 24h with Omicron sub-variant BA.2.12.1. After fixation, serial dilutions of each antibody were used to stain the infected cells or uninfected controls. Commercial polyclonal anti-nucleocapsid antibody was used as a positive control.

Both monoclonal antibodies (mAb 9 and 10) and polyclonal antibodies specifically detected nucleocapsids in infected cells. Some modest background staining was apparent in uninfected cells stained with the polyclonal antibody, but neither monoclonal antibody exhibited background (Figure 3). Monoclonal antibodies gave equal staining intensities. Interestingly, both monoclonal antibodies detected the punctate localization of nucleocapsid proteins; these are the SARS-CoV-2 genome replication foci on the Endoplasmic Reticulum [46] .

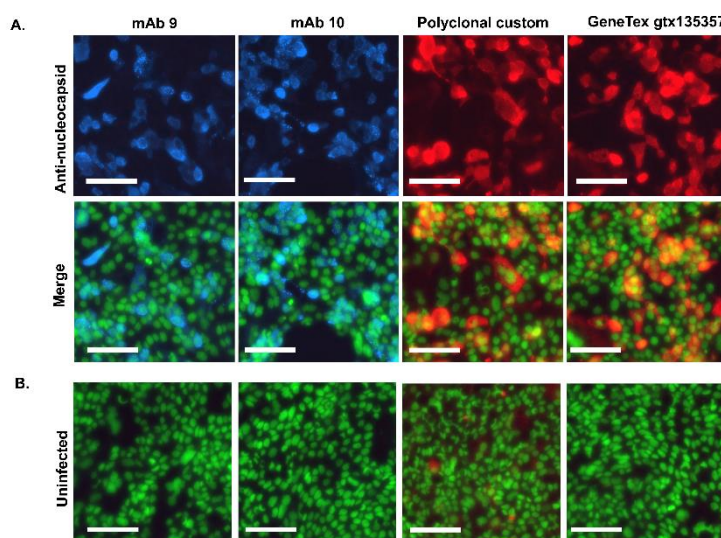


Fig3

**Figure 3. Immunofluorescence assay using the Antibodies to detect nucleocapsid in SARS-CoV-2 infected cells:** Calu-3 cells infected with SARS-CoV-2 variant BA.2.12.1 were fixed at 24 hours post-infection and processed for immunofluorescence. **A.** Nucleocapsid was detected with monoclonal and polyclonal antibodies at a concentration of 1:100 or commercial polyclonal antibody (GeneTex

gtx135357) at 1:1000 (upper panels). Nuclei are counterstained with Sytox Green (lower panels, Merge). **B.** Uninfected controls were treated as above for each antibody (Uninfected, merge). The scale bar is 100µm.

**VQA testing: Quantitative evaluation of variants of SARS-CoV-2:** We next evaluated the ELISA's Limit of Detection (LoD) using authenticated virus samples. The RADx-DCC program prepared panels of blinded dilutions of UV-inactivated SARS-CoV-2 with known concentrations of N protein. Each panel included multiple dilutions of UV-inactivated SARS-CoV-2 in at least triplicate. UV-inactivated viruses serially diluted in viral transport medium (VTM) or VTM-only controls were aliquoted and stored at -80C in blinded tubes before running ELISA. A recombinant N standard curve was also run on the plate. The linear range of the dilution curve (log<sub>10</sub> ng/ml N vs. background-subtracted OD450) was used to produce a best-fit line. The LoD calculated from the recombinant N standard curve was 0.366 ng/ml, and from inactivated XBB.1.5 dilution was 16.7 ng/ml N (Supplementary Figure 4). This result was used to assemble a blinded panel of UV-inactivated SARS-CoV-2 variants of multiple concentrations near the LoD. Measurement by ELISA demonstrated sensitive detection of all variants tested (Figure 4). The LoDs of each variant determined by the dilution series are listed in **Table 4**.

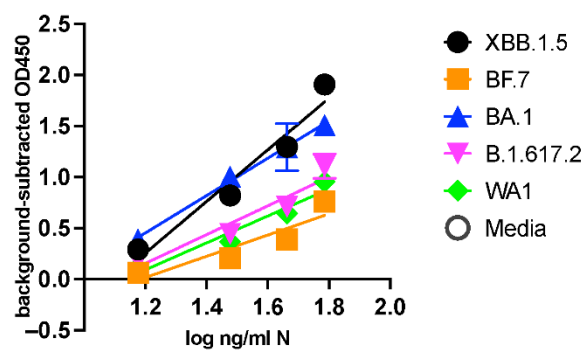


Fig4

**Figure 4. Comparison of ELISA using UV-inactivated SARS-CoV-2 variants:** SARS-CoV-2 variants XBB.1.5, BF.7, BA.1, B.1.617.2, and WA1 were UV-inactivated, and ng/mL N of stocks was determined by commercial ELISA. Stocks were normalized to equal ng/ml N and serially diluted in VTM. Triplicate aliquots of each dilution were measured by ELISA using antibodies by a blinded experimenter. Graphs are mean +/- SD of blank-subtracted OD450 values from triplicate samples. The best-fit line was calculated on log-transformed concentrations in GraphPad Prism 10.

Table 4.

Variant	LoD (ng/ml)
XBB.1.5	2.12
BF.7	8.78

BA.1	1.72
B.1.617.2	3.98
WA1	4.86

Next, potential off-target detection was tested by ELISA on a blinded panel, including SARS-CoV-2 XBB.1.5, Common Cold Coronaviruses (hCoV 229E, hCoV OC43), Respiratory syncytial viruses (RSV A2, RSV B), and Influenza viruses (influenza A H1N1, influenza B). Multiple concentrations of XBB.1.5 were detected with a LoD calculated as 2.75 ng/ml N. All non-SARS-CoV-2 viruses were detected at the background level in the ELISA. It should be noted that using a commercial Np-ELISA KIT (Ray Biotech Catalog Number: ELV-COVID19N) comparator for the same challenge panel provided similar results (XBB.1.5 LoD: 2.08 ng/ml N) (Figure 5).

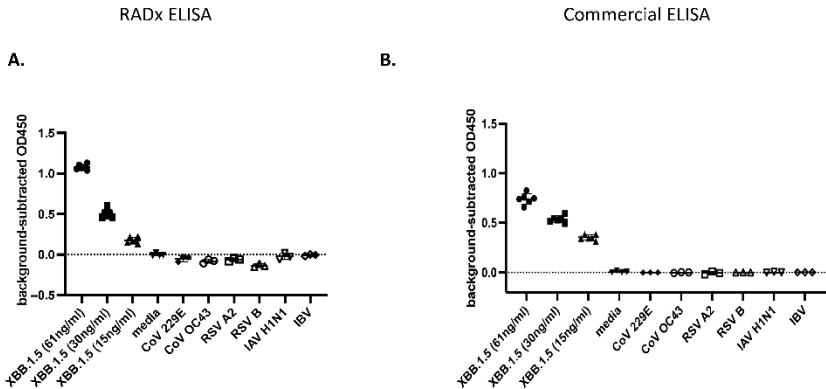


Fig5

**Figure 5. ELISA specificity to SARS-CoV-2:** **A.** SARS-CoV-2 variant XBB.1.5 and additional respiratory viruses were UV-inactivated, diluted in media, and measured by ELISA. **B.** An identical dilution series was prepared at the same time and assayed using the Ray Biotech SARS-CoV-2 nucleocapsid ELISA. Both ELISAs were performed blind on non-SARS-CoV-2 (n=3) and SARS-CoV-2 (n=6) virus samples. Graphs are mean +/- SD of blank-subtracted OD450 values. The best-fit line was calculated on log-transformed concentrations in GraphPad Prism 10.

The results were corroborated by a Western Blot depicted in (Supplementary Figure 5). Here, bands are visible for all the SARS-CoV-2 variants tested. A lack of visible bands indicates no binding occurred for the negative distractor viruses. The ELISA and Western Blot results strongly suggest the antibodies' ability to bind to several SARS-CoV-2 variants while lacking cross-reactivity with other distractor respiratory viruses.

**Epitope mapping for the monoclonal antibodies:** Finally, we wanted to check the Nucleocapsid protein CTD-epitopes to which the monoclonal antibodies (#9 and 10) were binding. We reasoned that since our antibodies could detect the denatured N-protein in Western blots (Figure 2E and Supplementary Figure 5), they were binding to the CTD's linear epitopes. Therefore, we performed peptide mapping experiments using the pepscan method [47]. For this, we designed twelve overlapping peptides, each of 20 amino-acid lengths, spanning the CTD (Figure 6A), and commercially synthesized them using solid-phase peptide chemistry. The HPLC-purified peptides



were then covalently conjugated to the amino group (-NH<sub>2</sub>) of 96-well microtiter plates using the 1-ethyl-3-(3-(dimethylamino)-propyl) carbodiimide (EDC) coupling agent provided in the Peptide coating kit (Takara, MK100). The coated peptides were then subjected to ELISA using the monoclonal antibodies (Figure 6B). The mAb (#9 and #10) bound to only peptide #10; they did not bind to peptides #9 and #11, thus indicating that the epitopes for these antibodies span across the junction of peptides #9 and #11. The data strongly suggests that the two monoclonals (#9 and 10) used in our studies either share the same epitope or their epitopes are close to each other on this 20 amino acid peptide. In this peptide (339-LDDKDPNFKDQVILLNKHID-358), the single point Shannon entropy was estimated to be zero at all amino acid positions except at 342, 345, 349, and 353, where modest entropy values of 0.02, 0.007, 0.023, and 0.007 respectively were estimated. Interestingly, part of this peptide 343-DPNFKD-348 is a major B-cell epitope predicted by BepiPred algorithm [44]. Additionally, our experimentally determined mAbs binding peptide #10 mapped to one of the Np-CTD's major antigenic epitopes predicted by an *in-silico* program (Figure 6C).

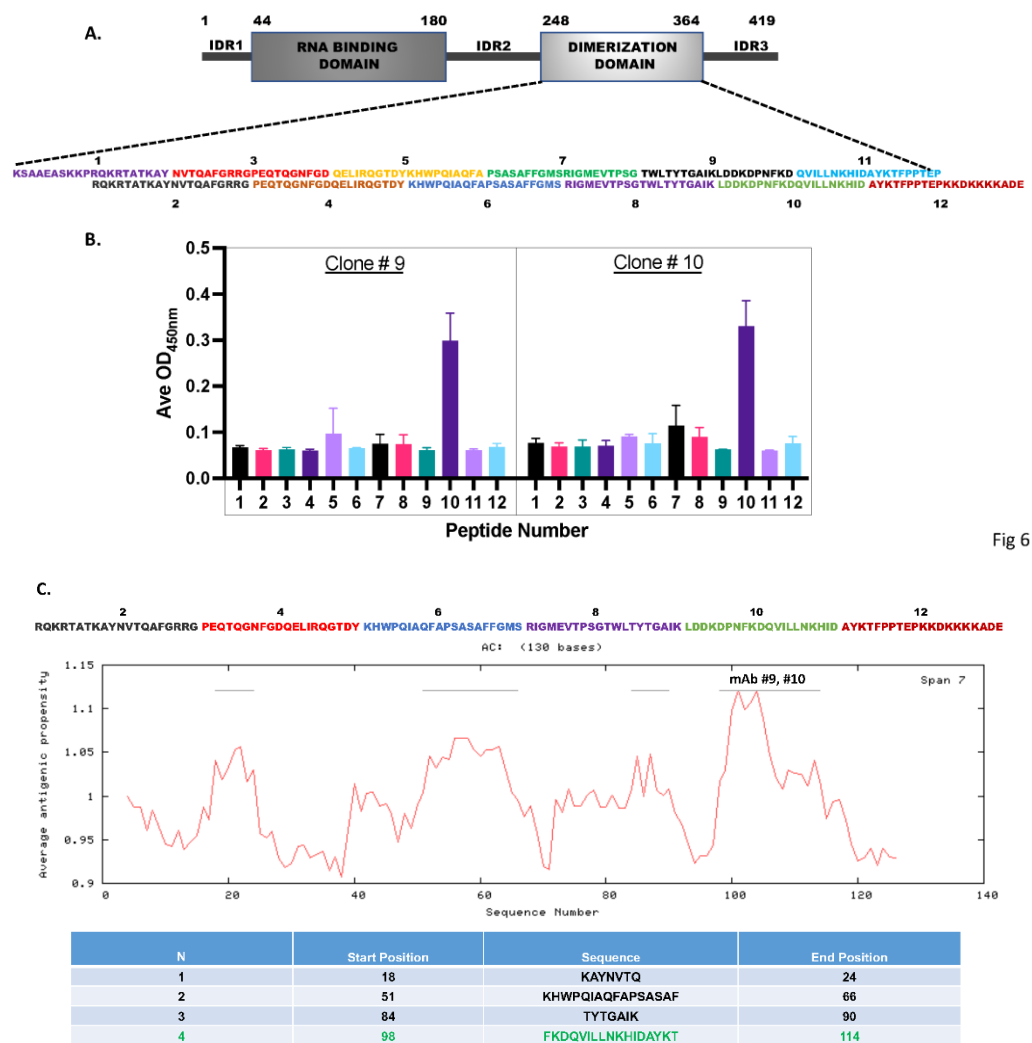


Fig6 (continued) Ref: <http://imed.med.ucm.es/Tools/antigenic.pl>

**Figure 6. Epitope mapping for the monoclonal antibodies:** **A.** Twelve overlapping peptides spanning the Np-CTD were designed to determine the linear epitopes for the mAb 9 and 10 antibodies. **B.** The synthesized peptides were subjected to ELISA using the pepscan method. Peptide #10 demonstrated the highest binding to the mAb 9 and 10, indicating that the major epitope for the monoclonal antibodies is present within these twenty amino acids long linear peptide. **C.** An *in-silico* program predicted the Np-CTD's four major linear epitopes (listed in the table). Notably, the peptide

epitope (highlighted in green) with the predicted highest antigenicity was the same peptide #10 determined by our pepscan assay.

## Discussion

To address the need for continual modifications of SARS-CoV-2 diagnostics due to the mutability of the virus, we aimed to develop an assay that can detect all circulating and historic CoV-2 variants to solve the concomitant problem of providing validated quality assurance panels and support test development. Despite several years since the onset of the COVID-19 pandemic, continual development has still been required for diagnostics as the virus mutates to avoid detection and treatment. Rather than having multiple tests to potentially identify only a handful of variants, having one assay whose detection envelopes several major variants has epidemiological, logistical, and financial benefits in point-of-care settings. However, developing a test capable of detecting many SARS-CoV-2 variants is a fine line; the assay must be specific enough to detect only SARS-CoV-2 but not so broad as to detect other respiratory viruses that present similar symptoms.

The N-protein was first identified as a starting point for sequence analysis due to its functional importance and immunogenicity. The N-protein is the most abundant structural protein, and antibodies produced against it make up a significant amount of the body's immune response [9]. The C-terminal domain of the N-protein is particularly immunogenic. Many anti-nucleocapsid antibodies target CTD, and the antibodies produced against this region are especially specific and selective [48]. Considering the importance of the CTD, we separately identified N-protein regions of low mutability utilizing Python-based biostatistical analysis and Epitope Windowed Shannon entropy Analysis on the N-proteins of major SARS-CoV-2 variants. Because the sequence of the CTD presented significantly lower entropy values than those of the NTD and disordered regions, it was selected as a target for antibody production. We raised several mouse-monoclonal and rabbit polyclonal antibodies against this region, and the strongest-binding antibodies were identified using direct ELISA. We used the highest titer candidates to develop sandwich ELISA assays that successfully identified both full-length N-protein and CTD-containing fragments without cross-binding to fragments lacking the CTD.

We also corroborated our antibodies' ability to detect the N-protein using immunofluorescent (IF) staining of infected cells. Both our monoclonal and polyclonal antibodies successfully identified the nucleocapsid in infected cells. Our monoclonal antibodies exhibited no background binding to the uninfected cells. A very modest background binding was seen with our polyclonal antibodies; for future IF imaging applications, affinity purification of the polyclonal antibodies using the antigen (Np-CTD) can reduce other non-specific IgGs present and eliminate the background binding.

A primary goal of this study was to optimize and validate blinded Viral Quality Assurance (VQA) panels assembled by the RADx-rad diagnostics core as a generalizable method to produce samples that have clinical equivalency without the need for infected patients and high biosafety containment. While this may be of little concern for slowly mutating pathogens, it can be a significant bottleneck in rapidly spreading and mutating outbreaks such as those experienced in COVID-19. The assay was tested using the panel of inactivated CoV-2 variants and distractor pathogens provided to NIH awardees. The assay detected viral loads in the physiological range (LoD 1.72-8.78 ng/mL). With the same VQA panels provided to awardees, it successfully detected every CoV-2 variant while being resilient to cross-binding with the distractor panel of other respiratory viruses tested (analytic sensitivity and specificity = 100). The variability in the limit of detection of variants could be explained by differences in epitope availability or structure in our ELISA compared to the commercial ELISA with which the inactivated viruses were quantified.

## Conclusion

We present a transparent and replicable approach to developing affinity molecules for test validation in a rapidly changing public health emergency. Our combination of heuristics and biostatistical analysis is meant to raise awareness in the scientific community of the need for quality assurance panels as part of the diagnostic development pipeline and the logistical issues of validating

a new virus when there are no existing diagnostics. In our study, we identified the C-terminal domain of the SARS-CoV-2 nucleocapsid protein as a potential antigenic region to produce diagnostic antibodies. We utilized both polyclonal and monoclonal antibodies raised against the CTD to develop an ELISA sandwich. We showed that the ELISA could detect several SARS-CoV-2 variants without cross-binding with other coronaviruses and respiratory viruses. We believe future iterations of our assay that substitute our polyclonal capture antibody for a monoclonal antibody will improve the assay's sensitivity. With the other antibodies produced during the development of this assay, we believe we can create a monoclonal antibody pair with non-overlapping epitopes to optimize capture and detection binding. With these future improvements considered, we believe our assay can serve as a starting point to produce a future multi-SARS-CoV-2 variant test with viability at the point of care testing, such as Lateral Flow Assay.

**Supplementary Materials:** The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

**Acknowledgments:** We gratefully acknowledge all data contributors, i.e., the Authors and their Originating laboratories responsible for obtaining the specimens and their Submitting laboratories for generating the genetic sequence and metadata and sharing via the GISAID Initiative, on which this research is based. The Centers for Disease Control and Prevention deposited the following reagent and obtained it through BEI Resources, NIAID, NIH: SARS-Related Coronavirus 2, Isolate USA-WA1/2020, NR-52281. The following reagent was obtained through BEI Resources, NIAID, NIH: SARS-Related Coronavirus 2, Isolate hCoV-19/South Africa/KRISP-K005325/2020, NR-54009, contributed by Alex Sigal and Tulio de Oliveira. The following reagent was obtained through BEI Resources, NIAID, NIH: SARS-Related Coronavirus 2, Isolate hCoV-19/Japan/TY7-503/2021 (Brazil P.1), NR-54982, contributed by National Institute of Infectious Diseases. The following reagent was obtained through BEI Resources, NIAID, NIH: SARS-Related Coronavirus 2, Isolate hCoV-19/USA/PHC658/2021 (Lineage B.1.617.2; Delta Variant), NR-55611, contributed by Dr. Richard Webby and Dr. Anami Patel. The graphical abstract was made using the BioRender software.

## References

1. Kevadiya, B.D.; Machhi, J.; Herskovitz, J.; Oleynikov, M.D.; Blomberg, W.R.; Bajwa, N.; Soni, D.; Das, S.; Hasan, M.; Patel, M.; et al. Diagnostics for SARS-CoV-2 infections. *Nat Mater* **2021**, *20*, 593-605, doi:10.1038/s41563-020-00906-z.
2. Peeling, R.W.; Heymann, D.L.; Teo, Y.Y.; Garcia, P.J. Diagnostics for COVID-19: moving from pandemic response to control. *Lancet* **2022**, *399*, 757-768, doi:10.1016/S0140-6736(21)02346-1.
3. Dong, T.; Wang, M.; Liu, J.; Ma, P.; Pang, S.; Liu, W.; Liu, A. Diagnostics and analysis of SARS-CoV-2: current status, recent advances, challenges and perspectives. *Chem Sci* **2023**, *14*, 6149-6206, doi:10.1039/d2sc06665c.
4. Lancet Commission on, C.-V.; Therapeutics Task Force, M. Operation Warp Speed: implications for global vaccine security. *Lancet Glob Health* **2021**, *9*, e1017-e1021, doi:10.1016/S2214-109X(21)00140-6.
5. RADx<sup>SM</sup> Radical. Available online: <https://www.radxrad.org/> (accessed on March 8 2024).
6. Weisblum, Y.; Schmidt, F.; Zhang, F.; DaSilva, J.; Poston, D.; Lorenzi, J.C.; Muecksch, F.; Rutkowska, M.; Hoffmann, H.H.; Michailidis, E.; et al. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *Elife* **2020**, *9*, doi:10.7554/eLife.61312.
7. Liu, Z.; VanBlargan, L.A.; Bloyet, L.M.; Rothlauf, P.W.; Chen, R.E.; Stumpf, S.; Zhao, H.; Errico, J.M.; Theel, E.S.; Liebeskind, M.J.; et al. Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. *Cell Host Microbe* **2021**, *29*, 477-488 e474, doi:10.1016/j.chom.2021.01.014.

8. Bojkova, D.; Klann, K.; Koch, B.; Widera, M.; Krause, D.; Ciesek, S.; Cinatl, J.; Munch, C. Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature* **2020**, *583*, 469-472, doi:10.1038/s41586-020-2332-7.
9. Wu, W.; Cheng, Y.; Zhou, H.; Sun, C.; Zhang, S. The SARS-CoV-2 nucleocapsid protein: its role in the viral life cycle, structure and functions, and use as a potential target in the development of vaccines and diagnostics. *Virol J* **2023**, *20*, 6, doi:10.1186/s12985-023-01968-6.
10. Cubuk, J.; Alston, J.J.; Incicco, J.J.; Singh, S.; Stuchell-Brereton, M.D.; Ward, M.D.; Zimmerman, M.I.; Vithani, N.; Griffith, D.; Wagoner, J.A.; et al. The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nat Commun* **2021**, *12*, 1936, doi:10.1038/s41467-021-21953-3.
11. Peng, Y.; Du, N.; Lei, Y.; Dorje, S.; Qi, J.; Luo, T.; Gao, G.F.; Song, H. Structures of the SARS-CoV-2 nucleocapsid and their perspectives for drug design. *EMBO J* **2020**, *39*, e105938, doi:10.15252/embj.2020105938.
12. Bai, Z.; Cao, Y.; Liu, W.; Li, J. The SARS-CoV-2 Nucleocapsid Protein and Its Role in Viral Structure, Biological Functions, and a Potential Target for Drug or Vaccine Mitigation. *Viruses* **2021**, *13*, doi:10.3390/v13061115.
13. Dutta, N.K.; Mazumdar, K.; Gordy, J.T. The Nucleocapsid Protein of SARS-CoV-2: a Target for Vaccine Development. *J Virol* **2020**, *94*, doi:10.1128/JVI.00647-20.
14. Rahman, M.S.; Islam, M.R.; Alam, A.; Islam, I.; Hoque, M.N.; Akter, S.; Rahaman, M.M.; Sultana, M.; Hossain, M.A. Evolutionary dynamics of SARS-CoV-2 nucleocapsid protein and its consequences. *J Med Virol* **2021**, *93*, 2177-2195, doi:10.1002/jmv.26626.
15. Oliveira, S.C.; de Magalhaes, M.T.Q.; Homan, E.J. Immunoinformatic Analysis of SARS-CoV-2 Nucleocapsid Protein and Identification of COVID-19 Vaccine Targets. *Front Immunol* **2020**, *11*, 587615, doi:10.3389/fimmu.2020.587615.
16. Frank, F.; Keen, M.M.; Rao, A.; Bassit, L.; Liu, X.; Bowers, H.B.; Patel, A.B.; Cato, M.L.; Sullivan, J.A.; Greenleaf, M.; et al. Deep mutational scanning identifies SARS-CoV-2 Nucleocapsid escape mutations of currently available rapid antigen tests. *Cell* **2022**, *185*, 3603-3616 e3613, doi:10.1016/j.cell.2022.08.010.
17. Hagag, I.T.; Pyrc, K.; Weber, S.; Balkema-Buschmann, A.; Groschup, M.H.; Keller, M. Mutations in SARS-CoV-2 nucleocapsid in variants of concern impair the sensitivity of SARS-CoV-2 detection by rapid antigen tests. *Frontiers in Virology* **2022**, *2*, doi:10.3389/fviro.2022.971862.
18. Zhao, H.; Nguyen, A.; Wu, D.; Li, Y.; Hassan, S.A.; Chen, J.; Shroff, H.; Piszczek, G.; Schuck, P. Plasticity in structure and assembly of SARS-CoV-2 nucleocapsid protein. *PNAS Nexus* **2022**, *1*, pgac049, doi:10.1093/pnasnexus/pgac049.
19. Cubuk, J.; Alston, J.J.; Incicco, J.J.; Holehouse, A.S.; Hall, K.B.; Stuchell-Brereton, M.D.; Soranno, A. The disordered N-terminal tail of SARS-CoV-2 Nucleocapsid protein forms a dynamic complex with RNA. *Nucleic Acids Res* **2023**, doi:10.1093/nar/gkad1215.
20. Hodcroft, E.B. CoVariants: SARS-CoV-2 Mutations and Variants of Interest. Available online: <https://covariants.org/> (accessed on October 21, 2022).
21. Hadfield, J.; Megill, C.; Bell, S.M.; Huddleston, J.; Potter, B.; Callender, C.; Sagulenko, P.; Bedford, T.; Neher, R.A. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **2018**, *34*, 4121-4123, doi:10.1093/bioinformatics/bty407.
22. Garcia-Boronat, M.; Diez-Rivero, C.M.; Reinherz, E.L.; Reche, P.A. PVS: a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery. *Nucleic Acids Res* **2008**, *36*, W35-41, doi:10.1093/nar/gkn211.



23. Khare, S.; Gurry, C.; Freitas, L.; Schultz, M.B.; Bach, G.; Diallo, A.; Akite, N.; Ho, J.; Lee, R.T.; Yeo, W.; et al. GISAID's Role in Pandemic Response. *China CDC Wkly* **2021**, *3*, 1049-1051, doi:10.46234/ccdcw2021.255.
24. Kabat, E.A.; Wu, T.T.; Bilofsky, H. Unusual distributions of amino acids in complementarity-determining (hypervariable) segments of heavy and light chains of immunoglobulins and their possible roles in specificity of antibody-combining sites. *Journal of Biological Chemistry* **1977**, *252*, 6609-6616, doi:10.1016/s0021-9258(17)39891-5.
25. Simpson, E.H. Measurement of Diversity. *Nature* **1949**, *163*, 688-688, doi:10.1038/163688a0.
26. Shannon, C.E. A mathematical theory of communication. *The Bell System Technical Journal* **1948**, *27*, 379-423, doi:10.1002/j.1538-7305.1948.tb01338.x.
27. Wang, A.; Yao, H.; Estrin, K.; Wang, H.; Yao, K.; Estrin, D. Information-theoretic approaches for sensor selection and placement in sensor networks for target localization and tracking. *Journal of Communications and Networks* **2005**, *7*, doi:10.1109/JCN.2005.6387986.
28. Zhou, R.; Cai, R.; Tong, G. Applications of Entropy in Finance: A Review. *Entropy* **2013**, *15*, 4909-4931, doi:10.3390/e15114909.
29. *Entropy Measures, Maximum Entropy Principle and Emerging Applications*, 1 ed.; Karmeshu, Ed.; Springer Berlin, Heidelberg.
30. Vopson, M.M. A Possible Information Entropic Law of Genetic Mutations. *Applied Sciences* **2022**, *12*, doi:10.3390/app12146912.
31. Baczowski, A.J.; Joanes, D.N.; Shamia, G.M. Range of validity of  $\alpha$  and  $\beta$  for a generalized diversity index  $H(\alpha, \beta)$  due to Good. *Mathematical Biosciences* **1998**, *148*, 115-128, doi:https://doi.org/10.1016/S0025-5564(97)10013-X.
32. Pasechnik, A.; Myllari, A.; Salakoski, T. DYNAMICAL VISUALIZATION OF THE DNA SEQUENCE AND ITS NUCLEOTIDE CONTENT. 2005.
33. Stewart, J.J.; Lee, C.Y.; Ibrahim, S.; Watts, P.; Shlomchik, M.; Weigert, M.; Litwin, S. A Shannon entropy analysis of immunoglobulin and T cell receptor. *Molecular Immunology* **1997**, *34*, 1067-1082, doi:https://doi.org/10.1016/S0161-5890(97)00130-2.
34. Ranjbar, M.M.; Ataei, S.; Nikbakht Brujeni, G.; Golabdar, S. Analysis of variations, structures, and phylogenetic characteristics of bovine leukocyte antigen DRB3 exon2. *Archives of Razi Institute* **2017**, *72*, 147-157, doi:10.22092/ari.2017.111611.
35. Yang, O.O. Candidate vaccine sequences to represent intra- and inter-clade HIV-1 variation. *PLoS One* **2009**, *4*, e7388, doi:10.1371/journal.pone.0007388.
36. Yuan, F.; Wang, L.; Fang, Y.; Wang, L. Global SNP analysis of 11,183 SARS-CoV-2 strains reveals high genetic diversity. *Transbound Emerg Dis* **2021**, *68*, 3288-3304, doi:10.1111/tbed.13931.
37. Litwin, S.; Jores, R. Shannon Information as a Measure of Amino Acid Diversity. In *Proceedings of the Theoretical and Experimental Insights into Immunology*, Berlin, Heidelberg, 1992//, 1992; pp. 279-287.
38. Buus, S.; Rockberg, J.; Forsström, B.; Nilsson, P.; Uhlen, M.; Schafer-Nielsen, C. High-resolution Mapping of Linear Antibody Epitopes Using Ultrahigh-density Peptide Microarrays\*. *Molecular & Cellular Proteomics* **2012**, *11*, 1790-1800, doi:https://doi.org/10.1074/mcp.M112.020800.
39. Ghanchi, N.K.; Nasir, A.; Masood, K.I.; Abidi, S.H.; Mahmood, S.F.; Kanji, A.; Razzak, S.; Khan, W.; Shahid, S.; Yameen, M.; et al. Higher entropy observed in SARS-CoV-2 genomes from the first COVID-19 wave in Pakistan. *PLoS One* **2021**, *16*, e0256451, doi:10.1371/journal.pone.0256451.

40. Li, R.; Mor, M.; Ma, B.; Clark, A.E.; Alter, J.; Werbner, M.; Lee, J.C.; Leibel, S.L.; Carlin, A.F.; Dessau, M.; et al. Conformational flexibility in neutralization of SARS-CoV-2 by naturally elicited anti-SARS-CoV-2 antibodies. *Communications Biology* **2022**, *5*, 789, doi:10.1038/s42003-022-03739-5.
41. Carlin, A.F.; Clark, A.E.; Chaillon, A.; Garretson, A.F.; Bray, W.; Porrachia, M.; Santos, A.T.; Rana, T.M.; Smith, D.M. Virologic and Immunologic Characterization of Coronavirus Disease 2019 Recrudescence After Nirmatrelvir/Ritonavir Treatment. *Clin Infect Dis* **2023**, *76*, e530-e532, doi:10.1093/cid/ciac496.
42. Carlin, A.F.; Clark, A.E.; Garretson, A.F.; Bray, W.; Porrachia, M.; Santos, A.T.; Rana, T.M.; Chaillon, A.; Smith, D.M. Neutralizing Antibody Responses After Severe Acute Respiratory Syndrome Coronavirus 2 BA.2 and BA.2.12.1 Infection Do Not Neutralize BA.4 and BA.5 and Can Be Blunted by Nirmatrelvir/Ritonavir Treatment. *Open Forum Infect Dis* **2023**, *10*, ofad154, doi:10.1093/ofid/ofad154.
43. Q2(R2) Validation of Analytical Procedures. **2024**, 12.
44. Sun, Z.; Zheng, X.; Ji, F.; Zhou, M.; Su, X.; Ren, K.; Li, L. Mass Spectrometry Analysis of SARS-CoV-2 Nucleocapsid Protein Reveals Camouflaging Glycans and Unique Post-Translational Modifications. *Infectious Microbes and Diseases* **2021**, *3*, 149-157, doi:10.1097/im9.0000000000000071.
45. Supekar, N.T.; Shajahan, A.; Gleinich, A.S.; Rouhani, D.S.; Heiss, C.; Chapla, D.G.; Moremen, K.W.; Azadi, P. Variable posttranslational modifications of severe acute respiratory syndrome coronavirus 2 nucleocapsid protein. *Glycobiology* **2021**, *31*, 1080-1092, doi:10.1093/glycob/cwab044.
46. Scherer, K.M.; Mascheroni, L.; Carnell, G.W.; Wunderlich, L.C.S.; Makarchuk, S.; Brockhoff, M.; Mela, I.; Fernandez-Villegas, A.; Barysevich, M.; Stewart, H.; et al. SARS-CoV-2 nucleocapsid protein adheres to replication organelles before viral assembly at the Golgi/ERGIC and lysosome-mediated egress. *Science Advances* **8**, eabl4895, doi:10.1126/sciadv.abl4895.
47. Geysen, H.M.; Meloen, R.H.; Barteling, S.J. Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid. *Proceedings of the National Academy of Sciences* **1984**, *81*, 3998-4002, doi:10.1073/pnas.81.13.3998.
48. Wu, C.; Qavi, A.J.; Hachim, A.; Kavian, N.; Cole, A.R.; Moyle, A.B.; Wagner, N.D.; Sweeney-Gibbons, J.; Rohrs, H.W.; Gross, M.L.; et al. Characterization of SARS-CoV-2 nucleocapsid protein reveals multiple functional consequences of the C-terminal domain. *iScience* **2021**, *24*, 102681, doi:10.1016/j.isci.2021.102681.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.