Journal Pre-proof

Characterization of SARS-CoV-2 N protein reveals multiple functional consequences of the C-terminal domain

Chao Wu, Abraham J. Qavi, Asmaa Hachim, Niloufar Kavian, Aidan R. Cole, Austin B. Moyle, Nicole D. Wagner, Joyce Sweeney-Gibbons, Henry W. Rohrs, Michael L. Gross, J. S. Malik Peiris, Christopher F. Basler, Christopher W. Farnsworth, Sophie A. Valkenburg, Gaya K. Amarasinghe, Daisy W. Leung

PII: S2589-0042(21)00649-0

DOI: https://doi.org/10.1016/j.isci.2021.102681

Reference: ISCI 102681

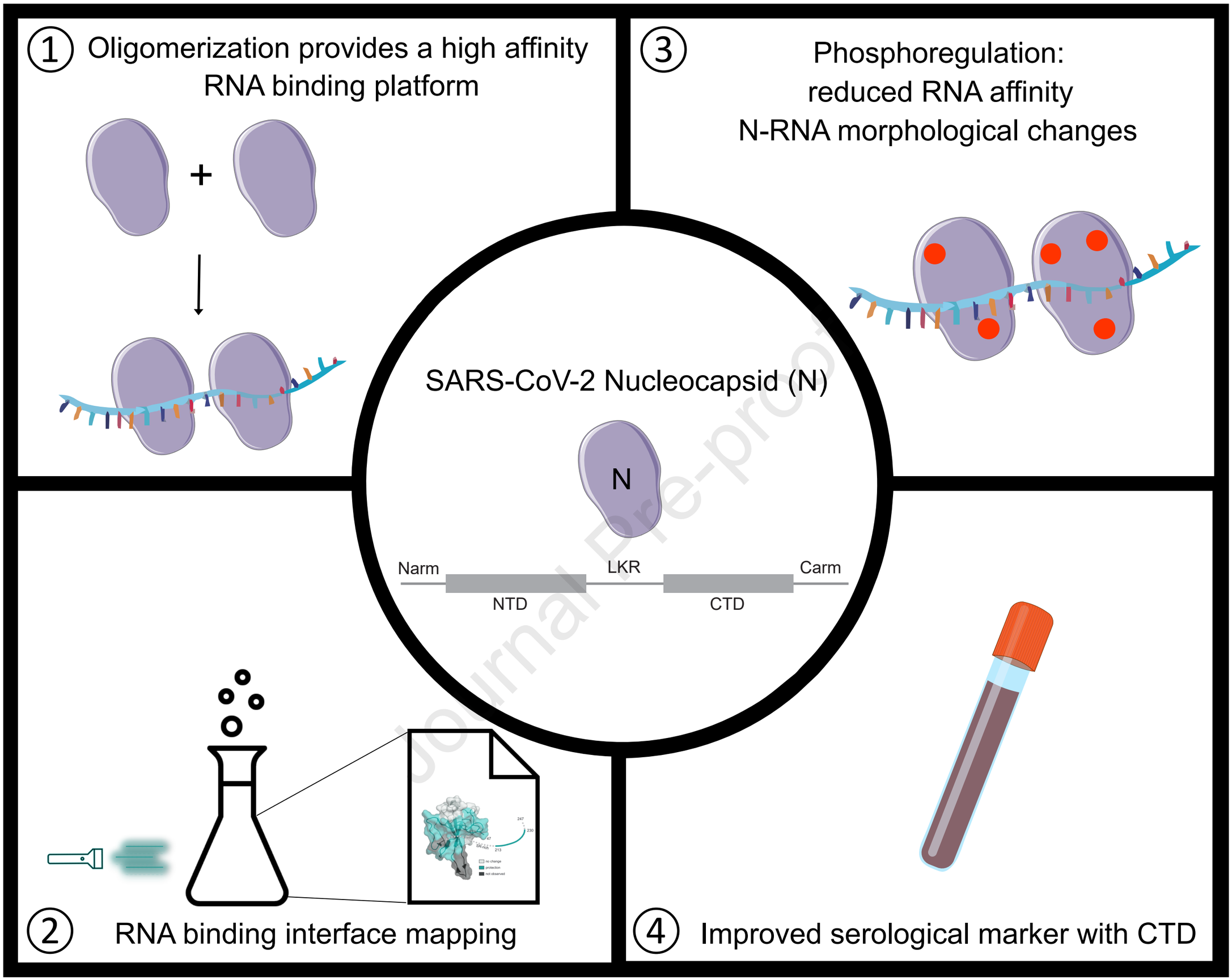
To appear in: *ISCIENCE*

Received Date: 16 December 2020

Revised Date: 13 May 2021

Accepted Date: 28 May 2021

Please cite this article as: 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., Peiris, J.S.M., Basler, C.F., Farnsworth, C.W., Valkenburg, S.A., Amarasinghe, G.K., Leung, D.W., Characterization of SARS-CoV-2 N protein reveals multiple functional consequences of the C-terminal domain, *ISCIENCE* (2021), doi: https://doi.org/10.1016/j.isci.2021.102681.



Characterization of SARS-CoV-2 N protein reveals multiple functional consequences of the C-terminal domain

Chao Wu[1], Abraham J. Qavi[1*], Asmaa Hachim[2*], Niloufar Kavian[2,3,4*], Aidan R. Cole[1*], Austin B. Moyle[6*], Nicole D. Wagner[6*], Joyce Sweeney-Gibbons[7*], Henry W. Rohrs[6], Michael L. Gross[6], J. S. Malik Peiris[2,8], Christopher F. Basler[7], Christopher W. Farnsworth[1#], Sophie A. Valkenburg[2#], Gaya K. Amarasinghe[#1Δ], and Daisy W. Leung[1,5#],

[1]Department of Pathology and Immunology, Washington University School of Medicine in St. Louis, St. Louis, MO, USA.

[2]HKU-Pasteur Research Pole, School of Public Health, The University of Hong Kong, Hong Kong, China

[3]Université Paris Descartes, Sorbonne Paris Cité, Faculté de Médecine, Assistance Publique–Hôpitaux de Paris, Hôpital Universitaire Paris Centre, Centre Hospitalier Universitaire Cochin, Service d'Immunologie Biologique, Paris, France.

[4]Institut Cochin, INSERM U1016, Université Paris Descartes, Sorbonne Paris Cité, Paris, France

[5]Department of Internal Medicine, Washington University School of Medicine in St. Louis, St. Louis, MO, USA.

[6]Department of Chemistry, Washington University in St. Louis, St. Louis, MO, USA.

[7]Center for Microbial Pathogenesis, Institute for Biomedical Sciences, Georgia State University, Atlanta, GA, USA

[8]Division of Public Health Laboratory Sciences, School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China

*equal contributions

[#]corresponding authors. Email addresses: gamarasinghe@wustl.edu, dwleung@wustl.edu, sophiev@hku.hk, cwfarnsworth@wustl.edu.

Δ lead author

Key words: SARS-CoV-2, COVID-19, phase separation, nucleoprotein, serology

Summary

Nucleocapsid (N) encoded by SARS-CoV-2 plays key roles in the replication cycle and is a critical serological marker. Here we characterize essential biochemical properties of N and describe the utility of these insights in serological studies. We define N domains important for oligomerization and RNA binding and show that N oligomerization provides a high affinity RNA binding platform. We also map the RNA binding interface, showing protection in the N-terminal domain and linker region. In addition, phosphorylation causes reduction of RNA binding and redistribution of N from liquid droplets to loose-coils, showing how N/RNA accessibility and assembly may be regulated by phosphorylation. Finally, we find that the C-terminal domain of N is the most immunogenic, based upon antibody binding to patient samples. Together, we provide a biochemical description of SARS-CoV-2 N and highlight the value of using N domains as highly specific and sensitive diagnostic markers.

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a novel coronavirus and the causative agent of COVID-19. Coronavirus has a single-stranded, positive-sense RNA genome encoding for four major structural proteins: spike (S), envelope (E), membrane (M), and nucleocapsid (N). The N protein is the second most proximal to the 3' end of the genome and is one of the most abundantly expressed viral proteins given the multifunctional roles of N during viral replication and assembly (Fung and Liu, 2019; Kim et al., 2020; McBride et al., 2014; Perlman and Netland, 2009). It is estimated that 1,000 copies of N are incorporated into each virion compared to only 100 copies of S (Bar-On et al., 2020). N exists mostly in a phosphorylated state in the cytoplasm, whereas it is predominantly dephosphorylated in mature virions, suggesting that N function is regulated by phosphorylation (Wu et al., 2014; Wu et al., 2009).

A major function of N is to encapsidate the ssRNA viral genome to evade immune detection and to protect the viral RNA from degradation by host factors (Chang et al., 2014; McBride et al., 2014). N has two structural domains (**Figure 1A**): an N-terminal domain (NTD; amino acid residues 44-176) and a C-terminal domain (CTD; amino acid residues 248-369). NTD is generally referred to as the RNA-binding domain (RBD), although regions outside of NTD are implicated in RNA binding as well (Chang et al., 2014; Grossoehme et al., 2009; Gui et al., 2017; Kang et al., 2020; McBride et al., 2014). N_{CTD} exists as a dimer in solution and can also bind RNA (Bouhaddou et al., 2020; Gui et al., 2017; Takeda et al., 2008). A conserved serine/arginine rich-linker region (LKR) connects the NTD and CTD (**Figure 1A** and **Supp. Figure 1A**) (Chang et al., 2014). Phosphorylation of residues in the serine-arginine of LKR is believed to regulate

discontinuous transcription, particularly for shorter subgenomic mRNA closer to the 3' end during early stages of replication (Wu et al., 2014; Wu et al., 2009). The LKR along with residues at the extreme N and C termini (Narm: amino acid residues 1-43 and Carm: amino acid residues 370-419) are intrinsically disordered (Chang et al., 2014; Cubuk et al., 2021). However, relative to the Narm and Carm, the LKR is more conserved (**Figure 1B** and **Supp. Figure 1B**).

Given its abundant expression and conservation within the genome, N has been used as an antigen for serology tests (Chew et al., 2020; Tang et al., 2020a, b). Previous studies showed that N-specific antibodies dominate the overall antibody response (Hachim et al., 2020; Lu et al., 2021). Furthermore, the T cell responses directed towards N are highly immunodominant in SARS-CoV and SARS-CoV-2 infection, with N-specific memory T cell responses evident 17 years after the initial SARS-CoV infection (Le Bert et al., 2020). Thus, N protein stability, RNA binding characteristics, abundance, and conservation are likely to impact T and B cell immunity.

Previous studies, including our own revealed shared and unique functions among viral nucleocapsid proteins (Arragain et al., 2019; Ding et al., 2016; Lu et al., 2020; Luo et al., 2020; Raymond et al., 2010; Su et al., 2018; Wan et al., 2017). These insights include oligomerization, RNA binding, and N-dependent functions such as RNA synthesis and immune evasion. Here we use a series of biochemical and biophysical assays to dissect different regions in N in order to determine how oligomerization, RNA binding, and phosphorylation are coupled to functions of N. Our results reveal that oligomeric N provides a continuous platform for binding RNA with high affinity. Our HDX-MS data identified a novel RNA binding region within the LKR in addition to a positively charged

patch within the NTD. We also show that phosphorylation modulates N-RNA interactions and solution properties, including phase-separated droplets. Finally, we find that the CTD contributes to oligomerization and RNA binding and is efficient at suppressing immune signaling. In SARS-CoV-2 infected patient plasma samples, the truncated CTD of N provides a highly sensitive serological marker. Our studies add to the wealth of information on SARS-CoV-2 N protein and expands upon our knowledge into N domain-specific functions.

Results

Multiple regions within N protein contribute to oligomerization. While several studies on SARS coronavirus N proteins report that N proteins oligomerize and that N oligomers is important for the assembly of viral particles (Chang et al., 2009; Luo et al., 2005; Ye et al., 2020), the contribution of each N domain to oligomerization remains to be defined. To better understand the regions within SARS-CoV-2 N important for promoting N-N interactions, we first used dynamic light scattering (DLS) to determine the hydrodynamic properties of isolated, RNA-free N domains and combinations of domains. Analysis of DLS results show that there are two major oligomeric species for full-length N (N_{WT}; 46 kDa), with hydrodynamic radii (R_h) of 8.9 nm and 450 nm (**Figure 1C and 1D**), suggesting the presence of different oligomeric forms of the protein. For comparison, the R_h values for maltose binding protein (44 kDa) and bovine serum albumin (66 kDa) are 2.9 nm and 3.7 nm, respectively. Removal of the Narm and Carm ($N_{NTD\text{-}LKR\text{-}CTD}$) results in two major species that are similar to N_{WT}. However, both $N_{NTD\text{-}LKR\text{-}CTD}$ populations display reduced polydispersity (narrower peak width, **Supp. Figure**

2A), suggesting that both Narm and Carm contribute to N oligomerization. Further removal of the CTD ($N_{NTD-LKR}$) results in a single peak representing a dimeric species (R_h = 3.9 nm), but with considerable polydispersity. N_{NTD} and N_{CTD} alone form stable domains; N_{NTD} is a monomer (R_h = 2.3 nm) whereas N_{CTD} is a dimer in solution (R_h = 3.5 nm), consistent with previous studies (Takeda et al., 2008). Exact mass measurement by denaturing mass spectrometry yields values corresponding to the mass expected from the amino acid sequence (± 1 Da) and supports the identity of the constructs used here (**Supp. Figure 2B-E**).

Oligomeric N provides a platform for high affinity RNA binding. A critical function for N is to bind and encapsidate viral genomic ssRNA (Chang et al., 2014). This interaction is sequence independent and as a result of N-RNA interactions, newly synthesized viral RNA often avoid immune detection by cellular pattern recognition receptors (Leung and Amarasinghe, 2016). Previous results suggest that regions beyond the NTD are involved in RNA binding, but were limited by experimental detection and only select regions were characterized (Chang et al., 2009; Grossoehme et al., 2009; Keane et al., 2012). To address this limitation and to gain insight into how each domain of SARS-CoV-2 N contributes to RNA binding, we developed a sensitive fluorescence polarization (FP) assay to measure binding of a FITC-labeled 20-nt ssRNA (sequence: UUUCACCUCCCUUUCAGUUU) (**Figure 2A**). From this assay, we find that N_{WT} binds the 20-nt ssRNA with high affinity (K_D = 0.007 ± 0.001 μM). Removal of the Narm and Carm do not impact ssRNA binding (K_D = 0.006 ± 0.002 and 0.006 ± 0.002 μM for $N_{NTD-LKR-CTD-Carm}$ and $N_{NTD-LKR-CTD}$, respectively) (**Figure 2B**). In contrast, the

isolated N_{NTD} and N_{CTD} have low affinity binding (K_D = 20 ± 10 and 13 ± 5 μM, respectively). However, inclusion of the LKR region increased RNA binding affinity significantly (0.50 ± 0.08 and 0.35 ± 0.04 μM for $N_{NTD\text{-}LKR}$ and $N_{LKR\text{-}CTD}$) (**Figure 2B-2D**). Addition of CTD onto NTD-LKR *in cis* increases the binding affinity to the single digit nM range (0.006 ± 0.002 μM) (**Figure 2C**), but not *in trans* (compare $N_{NTD\text{-}LKR}$ + N_{CTD} with $N_{NTD\text{-}LKR\text{-}CTD}$). The increase in binding affinity also occurs when NTD is added to LKR-CTD *in cis* (**Figure 2D**, compare $N_{LKR\text{-}CTD}$ + N_{NTD} with $N_{NTD\text{-}LKR\text{-}CTD}$). Similar binding curves and K_D values were obtained when fluorescence anisotropy values were converted from polarization (**Supp. Figure 3A**). Collectively, our data quantitatively show that the NTD, CTD, and LKR each contribute to ssRNA binding and that the presence of these three domains in tandem confers N with high affinity RNA binding, suggesting a more extensive and coupled RNA binding interface on N than is encompassed within a single domain.

N is proposed to disrupt dsRNA structures formed by transcription regulatory sequences during discontinuous transcription and impact viral RNA synthesis (Grossoehme et al., 2009; Keane et al., 2012; Sola et al., 2015). In order to evaluate the impact of N binding to dsRNA structures, we used an RNA that forms a stable stem-loop structure (slRNA) (sequence: GGAAGAUUAAUAAUUUUCC) (**Figure 2E**). We find that N_{WT} binds this slRNA with relative high affinity (K_D = 0.051 ± 0.004 μM) whereas both N_{NTD} and N_{CTD} alone have significantly weaker binding affinities (K_D = 120 ± 80 and 60 ± 40 μM, respectively). Addition of LKR significantly improves binding, consistent with results observed for ssRNA. Overall, slRNA binding to N appears to be at an order of

magnitude lower than to ssRNA binding (**Supp. Figure 3B**). This is potentially due, in part, to the energetic penalty of unfolding the stem-loop structure. Furthermore, the Narm and Carm may contribute more to slRNA binding than ssRNA because the impact on N binding is more pronounced after removal of the Narm or Carm (**Supp. Figure 3C**).

Hydrogen-deuterium exchange mass spectrometry (HDX-MS) defines an extended RNA binding interface in N. X-ray crystal and NMR structures are available for the SARS-CoV-2 N_{NTD} and N_{CTD} as well as hybrid models for RNA-bound N_{NTD} (Dinesh et al., 2020b; Kang et al., 2020; Ye et al., 2020). While these static structures provide a wealth of information, a critical gap remains in our knowledge of N-RNA interactions. Our studies here show that other N domains *in cis* contribute significantly to RNA binding. To further define the RNA binding interface on N, we performed HDX-MS to locate regions that become protected upon RNA binding. We first tested the $N_{NTD\text{-}LKR\text{-}CTD}$ protein, but observed aggregation upon RNA binding that caused a 100-fold loss in peptide abundances for the bound state, possibly attributable to aggregation-induced poor digestion. To circumvent RNA-binding induced protein aggregation, we reasoned that shorter constructs, such as $N_{NTD\text{-}LKR,}$ could be used based on our binding studies. However, even for this shorter construct, we noticed that the sample became cloudy after mixing with RNA. Next, given phosphorylation of N is a crucial regulatory mechanism in the coronavirus life cycle (Wu et al., 2009), we hypothesized that N phosphorylation likely impacts N-RNA solution properties and stability. Therefore, we examined phosphomimics of N for improved sample stability and solubility that are compatible with HDX-MS studies. For SARS-CoV, glycogen synthase kinase (GSK)-3

was shown to phosphorylate N at Ser177 (corresponding to Ser176 in SARS-CoV-2 N) (Wu et al., 2009). Phosphorylation of Ser177 is preceded by phosphorylation of Ser189 and Ser207 (Ser188 and 206 in SARS-CoV-2 N) by priming kinases (Wu et al., 2009). While more phosphorylation events may occur, we focused on these three better-known positions to evaluate if introduction of S176D/S188D/S206D mutations into $N_{NTD\text{-}LKR}$ ($N_{NTD\text{-}LKR\ S176D/S188D/S206D}$) will resolve N protein aggregation at the concentrations of interest.

With the introduction of the mutations on a shorter construct, $N_{NTD\text{-}LKR\ S176D/S188D/S206D}$, we were able to use sequential FXIII and pepsin digestion to recover 152 peptides, resulting in 93.3% sequence coverage (**Supp. Figure 4**), which enabled us to further examine the system by HDX-MS. HDX-MS analysis of $N_{NTD\text{-}LKR\ S176D/S188D/S206D}$ shows clear protection in four distinct regions upon RNA binding (amino acid residues 41-63, 105-108, 146-171, and 213-230) (**Figure 3A-B** and **Mendeley dataset** doi:10.17632/sv8r6phkzt.1). Residues 133-143 are not perturbed by RNA binding, but peptides covering residues 146-171 show clear protection. The largest differences in HDX are observed where 50-80% of the residues of unbound peptides undergo a burst phase of HDX in the first 10 s (amino acid residues 146-156, 163-171, and 213-230), where the peptides cover regions of little hydrogen bonding in the unbound state. When bound to RNA, the fraction of residues participating in the burst phase decreases, resulting in observed protection. Then, HDX either converges over time (amino acid residues 146-156, 163-171, and 219-223) consistent with protein conformation or RNA binding dynamics, or the HDX never converges in the timescale of the experiment (amino acid residues 222-230), consistent with relatively static binding. Interestingly,

peptides covering amino acid residues 103-108 and 156-159 undergo very little HDX throughout the experiment, consistent with either a hydrophobic pocket or secondary or tertiary structure hydrogen bonding. Of note, HDX decreases for the bound state in these peptides only after 1 h. The low initial HDX limits the dynamic range of binding-induced protection from HDX, but statistically significant protection is still observed.

Overall, HDX analysis revealed that the protected regions (**Figure 3C**) overlap with a basic patch groove in the N_{NTD} structure (**Figure 3D**) that is also observed in recently published NMR titration studies (Dinesh et al., 2020a); however, a region (amino acid residues 213-230) within the LKR domain, C-terminal to the SR-motif, also shows statistically significant HDX protection. This observation is consistent with the C-terminal end of the LKR domain, a region truncated in other RNA-binding studies of similar N proteins, being perturbed by RNA binding; furthermore, the parallel HDX kinetics plots indicate relatively static binding interactions within this region. Interestingly, we did not detect HDX protection in the SR-motif, which was proposed to bind RNA. This may be due to the Ser-to-Asp mutations introduced into this region, changing the RNA binding patterns. Altogether, HDX results along with our biochemical data define an RNA-binding interface within the NTD and LKR regions of N.

N-RNA liquid droplet formation is impacted by N domains. The above oligomerization and RNA binding results were obtained using RNA-free protein. However, oligomerization is an intrinsic property of N and is complicated by RNA binding during copurification due to the high affinity of N proteins for RNA. We found that bacterial RNA copurified even with increased ionic strength in purification buffer.

Size exclusion chromatography revealed three populations of N, including two RNA-bound states (p1 and p2) and an RNA-free state (p3) (**Figure 4A**). Truncation of the Narm results in an increase of the RNA-free peak (p3), suggesting that N truncations can alter the structure of N and correspondingly impact RNA binding and oligomerization. When both Narm and Carm were removed, we observed an even greater shift to p3, suggesting that both arms contribute to RNA binding interactions. To gain a better understanding of the two RNA-bound populations p1 and p2, we visualized these samples by using negative-stain electron microscopy (EM) in near-physiological salt concentrations (150 mM). We observed that N_{WT} p1 contains N-RNA with a loose-coil appearance (**Figure 4B, top left**), similar to that observed for other RNA-bound nucleocapsids (Bharat et al., 2012; Mavrakis et al., 2002). In contrast, we mostly observe spheres in N_{WT} p2 that correspond to liquid droplets separated from the surrounding buffer (**Figure 4B, top right**), consistent with many recent studies that describe liquid-liquid phase separation (LLPS) of N in the presence of RNA (Carlson et al., 2020; Cubuk et al., 2021; Iserman et al., 2020; Jack et al., 2020; Savastano et al., 2020).

To determine if N regions that impact RNA binding also impact liquid droplet formation, we next examined the role of Narm and Carm. $N_{NTD\text{-}LKR\text{-}CTD\text{-}Carm}$ behaves similarly to N_{WT}, having loose-coils in p1 (**Figure 4B, middle left**) and forming spherical liquid droplets in p2 (**Figure 4B, middle right**). However, examination of p2 from $N_{NTD\text{-}LKR\text{-}CTD}$ (**Figure 4B, bottom right**) revealed a much smaller population of liquid droplets (red arrow) and mostly crystal-like needle aggregates, suggesting that the Carm is important for droplet

formation. A transition from spherical liquids to needle-like solids is consistent with the liquid-to-solid transitions observed for other proteins that undergo phase separation (Patel et al., 2015).

Phosphorylation of LKR modulates RNA binding and higher-order assembly. Recently N protein phosphorylation has been qualitatively shown to modulate both RNA binding and phase separation (Carlson et al., 2020; Lu et al., 2020; Savastano et al., 2020), and we successfully utilized the improved solubility of phosphomimics to map the RNA binding interface within the NTD-LKR region. To further test how phosphorylation quantitatively impacts RNA binding and solution properties of SARS-CoV-2 N-RNA, we generated additional N phosphomimics in the context of N_{WT} and $N_{NTD\text{-}LKR}$. Size exclusion chromatography showed that, compared to N_{WT}, $N_{S188D/S206D}$ (priming serine residues for S176 by GSK-3) produced a reduced RNA-free peak (p3) and an increased RNA-bound peak (p1) (**Figure 4C**). Introduction of S176D to generate $N_{S176D/S188D/S206D}$ resulted in an even greater shift in p1 and p3 distributions, showing how phosphorylation can affect N interactions with RNA. The height of p2 remains relatively the same for all preparations. Examination of these protein peaks using electron microscopy revealed that $N_{S188D/S206D}$ displays similar loose-coils in p1 (**Figure 4D, top left**) and spherical droplets in p2 (**Figure 4D, top right**) for the RNA-bound species. Similar observations were made for $N_{S176D/S188D/S206D}$ (**Figure 4D**). To describe this interaction further, we measured ssRNA binding to the N phosphomimics (**Figure 4E** and **Supp. Figure 5A**). $N_{S176D/S188D/S206D}$ displays ~5-fold lower binding affinity to ssRNA compared to N_{WT} binding, a result in trend with previous work examining the impact of

LKR phosphorylation on RNA binding through simulations (Savastano et al., 2020). We observed a similar trend of lower binding to phosphomimics for the $N_{NTD\text{-}LKR}$ construct. Furthermore, binding to slRNA is also affected by these mutations (**Supp. Figure 5B**). Collectively, our data suggest that phosphorylation of the LKR region can impact N interactions with RNA, causing changes in properties, including solvent accessibility and RNA binding. Interestingly, there are 14 serine residues in the SR motif of LKR in SARS-CoV-2 N, of which 13 are found in SARS-CoV N, and an increase in phosphorylation in this region may further enhance these changes for RNA interaction and N associated functions.

N_{CTD} is a sensitive serological marker. Given that the different domains of N impact the various biochemical and physicochemical properties of N, we next assessed if particular domains contribute more significantly to N immunodominance that has been observed worldwide (Hachim et al., 2020). We also asked if there is a correlation between the biophysical properties, including RNA binding by N, and domain-specific responses in COVID-19 patient responses. Plasma samples collected from two cohorts of RT-PCR confirmed COVID-19 patients, one in St. Louis, USA (n = 45) and one in Hong Kong (n = 23), at different time points of infection. Using these samples, we performed enzyme-linked immunosorbent assays (ELISAs) to detect IgG present in COVID-19 patient plasma using different N domains that we have purified and characterized. First, we confirmed that purified N_{WT} is a sensitive serological marker to differentiate between COVID-19 positive and negative individuals (**Figure 5A**). As shown in **Figure 5B**, antibodies against all five N constructs were detected in the

COVID-19 cohort ($p < 0.0001$ versus negative controls for all). A cut-off based on the mean of the negatives plus three standard deviations allowed us to assess the performance of each N construct at detecting IgG antibodies in COVID-19 positive individuals (**Figure 5C**). We find that $N_{NTD\text{-}LKR\text{-}CTD\text{-}Carm}$ shows the lowest sensitivity (41.2%), whereas the truncated $N_{NTD\text{-}LKR\text{-}CTD}$ can detect more COVID-19 positive individuals (70.6%). Furthermore, N_{CTD} shows the highest combination of sensitivity (75%) and specificity (96.4%) over the other N constructs tested. This is demonstrated by the lowest cut-off score for the N_{CTD} for negative control samples, despite a comparable level of amino acid sequence conservation of the N_{CTD} (29-41%) to the N_{NTD} (32-48%) and N_{LKR} (28-42%) domains with common cold corona viruses (**Figure 1B**).

We next compared the immunogenicity of $N_{NTD\text{-}LKR\text{-}CTD\text{-}Carm}$ to N_{CTD} and N_{NTD} on an IgG heatmap during natural infection to an independent panel of 67 COVID-19 samples from Hong Kong. The magnitude of the IgG response to the $N_{NTD\text{-}LKR\text{-}CTD\text{-}Carm}$ tends to follow the same trend as that of N_{CTD} (**Figure 5D**). When we assessed the kinetics of N_{CTD} and the $N_{NTD\text{-}LKR\text{-}CTD\text{-}Carm}$ responses, we find that the magnitude of the N_{CTD} IgG detection tends to reach a similar level to that of $N_{NTD\text{-}LKR\text{-}CTD\text{-}Carm}$ at convalescent time-points (after day 14) (**Figure 5E**). The ELISA ratio of $N_{CTD}/N_{NTD\text{-}LKR\text{-}CTD\text{-}Carm}$ demonstrates this finding, pointing to a maturation of the humoral immune response towards the N_{CTD} with time after infection (**Figure 5F**, $p < 0.0001$ for acute versus convalescent time-points). The early dominance of the $N_{NTD\text{-}LKR\text{-}CTD\text{-}Carm}$ IgG response may reflect the recruitment of a cross-reactive pre-existing N-specific response. This response becomes more specific with time for the N_{CTD} domain as a *de novo* antibody response is made.

Given that some RNA viral N proteins are known immune antagonists, including prior studies of SARS-CoV, we hypothesized that N from SARS-CoV-2 may also suppress the type-I interferon (IFN) signaling pathway (Messaoudi et al., 2015). Using an IFNβ promoter reporter assay, we showed that N has a role in suppressing IFN signaling pathway when stimulated by Sendai virus (SeV) infection (**Figure 5G**). N_{WT} can inhibit IFNβ promoter activity, although not as well as Měnglà virus (MLAV) VP35, a potent inhibitor of IFN signaling (Williams et al., 2020). Both $N_{Narm\text{-}NTD\text{-}LKR}$ and N_{NTD} show modest inhibition at the highest concentration tested. However, $N_{CTD\text{-}Carm}$ shows similar levels of inhibition as N_{WT}, and N_{CTD} displays the highest inhibition even at lower concentrations. In summary, N is a potent inhibitor of IFN signaling and the N_{CTD} appears to be the region critical for mediating this function.

Discussion

SARS-CoV-2 N protein is a core viral protein produced by the subgenomic RNA, positioned proximal to the 3' end of genome, displays high transcription levels, and is in high abundance in virions. N is prone to forming higher-order oligomers that is impacted by the presence of different domains. We found that the isolated N_{NTD} and N_{CTD} form well-behaved monomers and dimers, respectively. Linkage of the two domains through the LKR and addition of the Narm and Carm contribute to N oligomerization, demonstrated in our DLS studies. It also contributes to enhancement of RNA binding to a continuous platform that were quantitatively measured using a sensitive fluorescence

polarization assay. Using HDX-MS for the $N_{NTD-LKR}$ phosphomimic, we confirmed a major positively charged groove as an RNA binding region in the N_{NTD} but also revealed that unstructured regions flanking the SR motif in the LKR are important for RNA interactions, which was not included in previous studies and crystal structures. The high extent of HDX observed for the LKR domain in the absence of RNA is consistent with minimal hydrogen bonding (i.e. minimal defined structure) and thus with a lack of structural information from other techniques; RNA-induced decreased HDX in the LKR domain could indicate RNA binding-induced secondary or tertiary structure, direct interactions with the RNA, or oligomerization. Overall, these results demonstrate the functional importance of the LKR domain for oligomerization and RNA binding. Thus, the coupling of oligomerization and RNA binding likely provides multilayered regulation that are important for immune evasion, viral replication, and nucleocapsid assembly.

N is also regulated by post-translational modifications and phosphorylation appears to play an important role that modulates RNA binding and changes the physicochemical properties of N. Our data revealed that mutation of S176, S188, and S206 in the SR motif to generate N phosphomimics resulted in decreased binding to RNA and a shift in protein-RNA populations with different solution properties.

In addition to characterizing the domain-specific biochemical and biophysical properties of N, we also gained insight into the antigenicity of individual domains of N and their potential utility in serological studies. Our data reveal that N_{CTD} acts more specifically in

detecting infection of SARS-CoV-2, from patient plasma in comparison to N_{WT}, consistent with predictions that the N_{CTD} region encompasses major antigenic sites of N (Bussmann et al., 2006; Liang et al., 2005). Interestingly, 2 out of 10 individuals assessed in this study of longitudinal donors unexposed to SARS-CoV-2 by Edridge *et al*., produced broadly reactive antibodies towards SARS-CoV-2 N_{WT}. The possibility of broadly reactive antibodies in unexposed individuals highlights the need for domain-specific serology, such as our use of the N_{CTD} for increased sensitivity to discriminate COVID-19 cases, while reducing the false-positive rate from cross-reactive antibodies generated by infections of the common-cold coronaviruses. Current sero-diagnostic assays to identify COVID-19 positive individuals are based on the detection of antibodies against N due to its abundant expression and corresponding high immune response (Chew et al., 2020; Tang et al., 2020a, b). However, these N-directed serological assays are highly variable, and their sensitivity depend on the sampling time-points, ranging from 0% to 93.75% (Liu et al., 2020; Tang et al., 2020a, b), suggesting that serological markers for SARS-CoV-2 infection can be further improved to include the N_{CTD}.

In conclusion, we describe our efforts to characterize how different domains contribute to the biochemical and physicochemical properties of SARS-CoV-2 N. Our results advance the understanding of the different levels of regulation involved in modulating viral replication and highlight the utility of using the N_{CTD} as a highly specific and sensitive diagnostic marker of COVID-19.

Limitations of the study

This study describes domain-specific insights into oligomerization, RNA binding, and phosphorylation of SARS-CoV-2 N protein. Generated reagents were then used to demonstrate that N_{CTD} is a more sensitive and specific serology marker. However, the underlying mechanisms of how these biochemical properties contribute to the observed immunodominance of the CTD of N are not clear and warrant further studies.

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Gaya Amarasinghe (gamarasinghe@wustl.edu).

Materials Availability

Plasmids in this study are available with a completed Materials Transfer Agreement Request for these reagents by submitting to Dr. Gaya Amarasinghe (gamarasinghe@wustl.edu).

Data and Code Availability

HDX-MS peptides table and all kinetic plots are accessible in Mendeley (doi:10.17632/sv8r6phkzt.1).

STAR Methods

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human Plasma. The collection of patient plasma was approved by the Human Research Protection Office at Washington University in St. Louis (IRB reference number 202007097) and the Institutional Review Board of The Hong Kong University and the Hong Kong Island West Cluster of Hospitals (IRB reference number UW20-169).

METHOD DETAILS

Patients and sample collection. Our study enrolled a total of 67 patients with RT-PCR confirmed COVID-19 infection: with 45 patients from St. Louis, MO, USA, and 23 patients from Hong Kong, PRC. The negative samples (n = 28) used in this study were from St. Louis, USA, and were obtained from patients following the start of the pandemic. Plasma samples were obtained from patients at Barnes-Jewish Hospital (St. Louis, MO, USA) and the Hong Kong Island West Cluster of Hospitals (Hong Kong, PRC). Both hospital systems are urban, tertiary-care, academic medical centers. Positive and negative patients from all cohorts were confirmed using standard of care, RT-PCR based methods. Plasma samples were collected from heparinized blood. Sample day was defined as days post-symptom onset.

Enzyme-linked immunosorbent assay (ELISA). ELISA assays were performed with N proteins made in house, as described below. Briefly, recombinant N proteins were coated on 96 well flat bottom immunosorbent plates (Nunc Immuno MaxiSorp) at a concentration of 500 ng/mL, in 100 µL coating buffer (PBS with 53% Na_2CO_3 and 42% $NaHCO_3$, pH 9.6) at 4°C overnight. An additional plate coated with a non-specific

protein (blocking buffer, PBS with 5% fetal bovine serum (FBS)) was used to measure the background binding of each plasma sample. Following FBS blocking and thorough washing, diluted plasma samples (1:100) were bound for 2 hours, further washed, and then detected by an anti-human IgG secondary antibody labelled with HRP (Invitrogen), and absorbance detected at 450 nm on a spectrophotometer (Wallac).

Protein Expression and Purification. SARS-CoV-2 N constructs were expressed as His-tag fusion proteins in BL21 (DE3) *E. coli* cells (Novagen). At OD_{600} of 0.6-0.7, recombinant protein expression was induced with 0.5 mM isopropyl β-d-1-thiogalactopyranoside (IPTG) for 12-14 h at 18˚C. Cells were harvested and resuspended in lysis buffer containing 20 mM Tris (pH 7.5), 1 M NaCl, 20 mM imidazole, 5 mM 2-mecaptoethanol (BME). Cells were lysed using an EmulsiFlex-C5 homogenizer (Avestin) and lysates were clarified by centrifugation at 30,000 x *g* at 4 ˚C for 40 min. N proteins were purified using affinity tag and gel filtration columns. Purity of N proteins were determined by Coomassie staining of SDS-PAGE. RNA-free species are used for oligomerization, RNA binding, and ELISA studies.

Negative Staining EM. 2 µL of N sample at a concentration of 1 mg/mL was applied to a glow-discharged copper grid (Ted Pella), washed twice with water before staining with 2% uranyl acetate for 30 s, and air dried. Grids were imaged using a JEOL JEM-1400plus Transmission Electron Microscope operating at 120 kV and recorded with an AMT XR111 high-speed 4k x 2k pixel phosphor-scintillated 12-bit CCD camera.

Dynamic Light Scattering (DLS). DLS experiments were performed on a DynaPro-PlateReader II (Wyatt Technologies Corporation). Measurements of N samples in triplicates (1 mg/mL) were obtained at 25 ºC and analyzed using Dynamics software (Wyatt).

Fluorescence Polarization Assay (FPA). FPA experiments were performed on a Cytation5 plate reader (BioTek) operating on Gen5 software. Excitation and emission wavelengths were set to 485 and 528 nm, respectively, with a bandpass of 20 nm. Read height and G factor were set to 8.5 mm and 1.26, respectively using the autogain function. For RNA binding experiments, fluorescein isothiocyanate (FITC) labelled 20-nt ssRNA or 19-nt slRNA at a final concentration of 1 nM was loaded on N samples (in 20 mM HEPES (pH 7.5), 150 mM NaCl, 2 mM TCEP, 5% glycerol) at concentrations ranging from 0.4 nM to 10 μM in a 96-well plate. After 10 min of incubation, fluorescence polarization signals were read. The fluorescence polarization values were then plotted against N concentrations to fit the dissociation constant, K_D, using ORIGIN software. For anisotropy plots, anisotropy values were converted from polarization according to previous research (Kozlov et al., 2012).

LC-MS Analysis. Unless otherwise indicated, all chemical reagents were sourced from Millipore Sigma and used without further purification. For LC-MS analyses, 30 pmol of protein in 50 μL of 1:1 solvent mixture of acetonitrile:water with 0.1% formic acid (CovaChem) was loaded onto a C8 trap (ZORBAX Eclipse XDB C8 column, 2.1 x 15 mm, Agilent), desalted for 3 min by using water/0.1% formic acid at a flow rate of 100 μL/min, and eluted using an 14 minute gradient from 0 to 80% acetonitrile/0.1% formic

acid at a flow rate of 100 µL/min. Samples were analyzed using a MaXis 4G Q-TOF (Bruker Daltonics). The mass spectrum was extracted guided by the elution peak and submitted to PMI Intact Mass and searched for M values ranging from 5-50 kDa.

HDX-MS. $N_{NTD\text{-}LKR\ S176D/S188D/S206D}$ was incubated with a 20-nt ssRNA at a 1:1 ratio. After incubation, 2 µL of 50 µM protein/protein-RNA in PBS (pH 7.4) was diluted 10-fold (v/v) with labeling buffer (PBS in D_2O, pD 7.0) (D_2O from Cambridge Isotope Laboratories), incubated for 10, 30, 300, and 3600 s on ice, quenched by using a 60% dilution with 3 M urea, PBS (pH 2.5), and flash frozen for later LC-MS analysis. A 0 s control was prepared with PBS in H_2O. Prior to incubation, each 50 µL of 2 µM sample was thawed for 1 min at 37 °C before injection into a custom-built liquid chromatography (LC) apparatus for LC-MS analysis. The labelled protein passed through two in-house packed protease columns (2 mm x 20 mm), coupled so that the first using protease from *Aspergillus saitoi* type XIII (FXIII) and the second porcine pepsin (0.1% formic acid, flow rate 200 µL/min); the resulting peptides were trapped on a ZORBAX Eclipse XDB C8 column (2.1 mm x 15 mm, Agilent), desalted for 3 min, and then separated on a XSelect CSH C18 XP column (130Å, 2.5 µm, 2.1 mm X 50 mm, Waters) with a 10.5 min linear gradient from 4 – 40% acetonitrile/0.1% formic acid (flow rate 100 µL/min). All valves, tubes, and columns (except for the protease columns, which lose activity at low temperature) were submerged in ice during the experiment to minimize back exchange. Peptides were eluted into a Bruker Maxis HM Q-TOF MS for mass analysis. Experiments were in duplicate unless otherwise indicated. The HDX data processing was performed by using HDExaminer (version 2.5.1, Sierra Analytics, Inc.).

IFNβ promoter reporter gene assay. HEK-293T cells (5 x 10^4) were co-transfected using Lipofectamine 2000 with 25 ng of an IFNβ promoter-firefly luciferase reporter plasmid, 25 ng of pRL-TK *Renilla* luciferase reporter plasmid, and 125, 12.5, and 1.25 ng of the indicated viral protein expression plasmid. Twenty-four hours post-transfection, cells were mock-treated or SeV (15 hemagglutination units / ml) infected. Eighteen hours post-treatment or post-infection, cells were lysed and analyzed for luciferase activity using a Dual-Luciferase reporter assay system (Promega). Firefly luciferase activity was normalized to *Renilla* luciferase activity. Assays were performed in triplicate; error bars indicate the standard error of the mean (SEM) for the triplicate. Viral protein expression was confirmed by Western blot analysis.

QUANTIFICATION AND STATISTICAL ANALYSIS

All HDX samples for each time point were prepared and acquired in duplicate. HDX results for peptides with m/z peaks interfering/overlapping with the theoretical centroid fit were excluded from the analysis. Error bars in the kinetic plots show the standard deviation for each protein state and time point. Error bars shown in the Woods' plot are 3 times the propagated error for each peptide across all time points (if the error bar does not cross zero then the difference is potentially significant). The global significance limit (i.e. the minimum difference in number of deuterium for a statistically significant difference) is shown in gray. Briefly, we first calculated the pooled standard deviation for each state, using standard deviations for all peptides and all time points (n = 214 peptides x 4 states = 856 total data points for each state). From the pooled standard deviations of each state, the standard error of the mean was calculated for the entire

dataset (n = 2 replicates). The global significance limit is then the confidence interval using this standard error of the mean a t-values for a two-tailed Student's t-distribution (*, $p < 0.1$; **, $p < 0.05$; ***, $p < 0.01$, 2 degrees of freedom (2 states + 2 replicates – 2)). In the Woods' plot, peptide HDX differences are considered significant if the difference exceeds this limit. Peptides are only considered statistically significant if both the conditions for the propagated error and the global significance limit are met.

For ELISA data, the cut-off is represented by the dotted line and calculated as the mean + 3 standard deviations of the negative population, used to calculate sensitivity and specificity. Experiments were repeated twice. Statistical significance was calculated by unpaired Student's t-test using GraphPad Prism.

For IFN-β promoter assay, statistical significance was determined by performing a one-way ANOVA followed with Tukey multiple comparison as compared to Sendai virus-infected control using GraphPad Prism.

Acknowledgements

We thank Dr. N. Krogan (UCSF) for sharing SARS-CoV-2 plasmids, Drs. A. Holehouse and A. Soranno (Washington University School of Medicine) for providing critical feedback, and R. Ridings for coordinating studies between WUSM and external groups. We would like to thank Bruker for mass spectrometry technical and instrument support, and Protein Metrics (R42GM1213302 to H.W.R) and Sierra Analytics for providing proteomic and HDX data analysis software. Research was supported by Fast Grant ##2161 (Emergent Ventures) to G.K.A., by Fast Grant Fast Grants Award # 2158 to

C.F.B, and NIH grants (P01AI120943, R01AI123926 to G.K.A. and C.F.B; R01AI107056 to D.W.L.; P41GM103422 and R24GM136766 to M.L.G.; R01AI143292 and R01AI148663 C.F.B). S.A.V was supported by COVID190115 and COVID190126 Health and Medical Research Fund, Food and Health Bureau, Hong Kong and NIH/NIAID CEIRS contract HHSN272201400006C. A.J.Q. is supported by a NIH T32 training grant (T32CA009547).

Author contributions: CW, GKA, and DWL conceived the overall project. All authors were integral to the design and execution of the study. CW, AQ, GKA, and DWL wrote the initial draft with significant input from all authors.

Declaration of interests

The authors declare no competing interests.

Inclusion and diversity

We worked to ensure sex balance in the selection of non-human subjects. One or more of the authors of this paper self-identifies as living with a disability. One or more of the authors of this paper received support from a program designed to increase minority representation in science.

References

Arragain, B., Reguera, J., Desfosses, A., Gutsche, I., Schoehn, G., and Malet, H. (2019). High resolution cryo-EM structure of the helical RNA-bound Hantaan virus nucleocapsid reveals its assembly mechanisms. eLife *8*, e43075.

Bar-On, Y.M., Flamholz, A., Phillips, R., and Milo, R. (2020). SARS-CoV-2 (COVID-19) by the numbers. eLife *9*, e57309.

579 Bharat, T.A.M., Noda, T., Riches, J.D., Kraehling, V., Kolesnikova, L., Becker, S., Kawaoka, Y., and Briggs,
580 J.A.G. (2012). Structural dissection of Ebola virus and its assembly determinants using cryo-electron
581 tomography. Proceedings of the National Academy of Sciences *109*, 4275-4280.
582 Bouhaddou, M., Memon, D., Meyer, B., White, K.M., Rezelj, V.V., Correa Marrero, M., Polacco, B.J.,
583 Melnyk, J.E., Ulferts, S., Kaake, R.M.*, et al.* (2020). The Global Phosphorylation Landscape of SARS-CoV-2
584 Infection. Cell *182*, 685-712.e619.
585 Bussmann, B.M., Reiche, S., Jacob, L.H., Braun, J.M., and Jassoy, C. (2006). Antigenic and cellular
586 localisation analysis of the severe acute respiratory syndrome coronavirus nucleocapsid protein using
587 monoclonal antibodies. Virus Res *122*, 119-126.
588 Carlson, C.R., Asfaha, J.B., Ghent, C.M., Howard, C.J., Hartooni, N., Safari, M., Frankel, A.D., and Morgan,
589 D.O. (2020). Phosphoregulation of phase separation by the SARS-CoV-2 N protein suggests a biophysical
590 basis for its dual functions. Molecular Cell.
591 Chang, C.-k., Hou, M.-H., Chang, C.-F., Hsiao, C.-D., and Huang, T.-h. (2014). The SARS coronavirus
592 nucleocapsid protein – Forms and functions. Antiviral Research *103*, 39-50.
593 Chang, C.-K., Hsu, Y.-L., Chang, Y.-H., Chao, F.-A., Wu, M.-C., Huang, Y.-S., Hu, C.-K., and Huang, T.-H.
594 (2009). Multiple Nucleic Acid Binding Sites and Intrinsic Disorder of Severe Acute Respiratory Syndrome
595 Coronavirus Nucleocapsid Protein: Implications for Ribonucleocapsid Protein Packaging. Journal of
596 Virology *83*, 2255-2264.
597 Chew, K.L., Tan, S.S., Saw, S., Pajarillaga, A., Zaine, S., Khoo, C., Wang, W., Tambyah, P., Jureen, R., and
598 Sethi, S.K. (2020). Clinical evaluation of serological IgG antibody response on the Abbott Architect for
599 established SARS-CoV-2 infection. Clinical Microbiology and Infection *26*, 1256.e1259-1256.e1211.
600 Cubuk, J., Alston, J.J., Incicco, J.J., Singh, S., Stuchell-Brereton, M.D., Ward, M.D., Zimmerman, M.I.,
601 Vithani, N., Griffith, D., Wagoner, J.A.*, et al.* (2021). The SARS-CoV-2 nucleocapsid protein is dynamic,
602 disordered, and phase separates with RNA. Nature Communications *12*, 1936.
603 Dinesh, D.C., Chalupska, D., Silhan, J., Koutna, E., Nencka, R., Veverka, V., and Boura, E. (2020a).
604 Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. PLOS Pathogens *16*,
605 e1009100.
606 Dinesh, D.C., Chalupska, D., Silhan, J., Veverka, V., and Boura, E. (2020b). Structural basis of RNA
607 recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. bioRxiv, 2020.2004.2002.022194.
608 Ding, B., Qin, Y., and Chen, M. (2016). Nucleocapsid proteins: roles beyond viral RNA packaging. Wiley
609 Interdiscip Rev RNA *7*, 213-226.
610 Fung, T.S., and Liu, D.X. (2019). Human Coronavirus: Host-Pathogen Interaction. Annual Review of
611 Microbiology *73*, 529-557.
612 Grossoehme, N.E., Li, L., Keane, S.C., Liu, P., Dann, C.E., Leibowitz, J.L., and Giedroc, D.P. (2009).
613 Coronavirus N Protein N-Terminal Domain (NTD) Specifically Binds the Transcriptional Regulatory
614 Sequence (TRS) and Melts TRS-cTRS RNA Duplexes. Journal of Molecular Biology *394*, 544-557.
615 Gui, M., Liu, X., Guo, D., Zhang, Z., Yin, C.-C., Chen, Y., and Xiang, Y. (2017). Electron microscopy studies
616 of the coronavirus ribonucleoprotein complex. Protein & Cell *8*, 219-224.
617 Hachim, A., Kavian, N., Cohen, C.A., Chin, A.W.H., Chu, D.K.W., Mok, C.K.P., Tsang, O.T.Y., Yeung, Y.C.,
618 Perera, R.A.P.M., Poon, L.L.M.*, et al.* (2020). ORF8 and ORF3b antibodies are accurate serological
619 markers of early and late SARS-CoV-2 infection. Nature Immunology *21*, 1293-1301.
620 Iserman, C., Roden, C., Boerneke, M., Sealfon, R., McLaughlin, G., Jungreis, I., Park, C., Boppana, A.,
621 Fritch, E., Hou, Y.J.*, et al.* (2020). Specific viral RNA drives the SARS CoV-2 nucleocapsid to phase
622 separate. bioRxiv, 2020.2006.2011.147199.
623 Jack, A., Ferro, L.S., Trnka, M.J., Wehri, E., Nadgir, A., Costa, K., Schaletzky, J., and Yildiz, A. (2020). SARS
624 CoV-2 nucleocapsid protein forms condensates with viral genomic RNA. bioRxiv,
625 2020.2009.2014.295824.

Kang, S., Yang, M., Hong, Z., Zhang, L., Huang, Z., Chen, X., He, S., Zhou, Z., Zhou, Z., Chen, Q.*, et al.* (2020). Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. Acta Pharmaceutica Sinica B *10*, 1228-1238.

Keane, S.C., Liu, P., Leibowitz, J.L., and Giedroc, D.P. (2012). Functional Transcriptional Regulatory Sequence (TRS) RNA Binding and Helix Destabilizing Determinants of Murine Hepatitis Virus (MHV) Nucleocapsid (N) Protein. Journal of Biological Chemistry *287*, 7063-7073.

Kim, D., Lee, J.-Y., Yang, J.-S., Kim, J.W., Kim, V.N., and Chang, H. (2020). The Architecture of SARS-CoV-2 Transcriptome. Cell *181*, 914-921.e910.

Kozlov, A.G., Galletto, R., and Lohman, T.M. (2012). SSB–DNA Binding Monitored by Fluorescence Intensity and Anisotropy. In Single-Stranded DNA Binding Proteins: Methods and Protocols, J.L. Keck, ed. (Totowa, NJ: Humana Press), pp. 55-83.

Le Bert, N., Tan, A.T., Kunasegaran, K., Tham, C.Y.L., Hafezi, M., Chia, A., Chng, M.H.Y., Lin, M., Tan, N., Linster, M.*, et al.* (2020). SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls. Nature *584*, 457-462.

Leung, D.W., and Amarasinghe, G.K. (2016). When your cap matters: structural insights into self vs non-self recognition of 5' RNA by immunomodulatory host proteins. Curr Opin Struct Biol *36*, 133-141.

Liang, Y., Wan, Y., Qiu, L.-w., Zhou, J., Ni, B., Guo, B., Zou, Q., Zou, L., Zhou, W., Jia, Z.*, et al.* (2005). Comprehensive Antibody Epitope Mapping of the Nucleocapsid Protein of Severe Acute Respiratory Syndrome (SARS) Coronavirus: Insight into the Humoral Immunity of SARS. Clinical Chemistry *51*, 1382-1396.

Lin, S.-Y., Liu, C.-L., Chang, Y.-M., Zhao, J., Perlman, S., and Hou, M.-H. (2014). Structural Basis for the Identification of the N-Terminal Domain of Coronavirus Nucleocapsid Protein as an Antiviral Target. Journal of Medicinal Chemistry *57*, 2247-2257.

Liu, W., Liu, L., Kou, G., Zheng, Y., Ding, Y., Ni, W., Wang, Q., Tan, L., Wu, W., Tang, S.*, et al.* (2020). Evaluation of Nucleocapsid and Spike Protein-Based Enzyme-Linked Immunosorbent Assays for Detecting Antibodies against SARS-CoV-2. Journal of Clinical Microbiology *58*, e00461-00420.

Lu, S., Xie, X.-x., Zhao, L., Wang, B., Zhu, J., Yang, T.-r., Yang, G.-w., Ji, M., Lv, C.-p., Xue, J.*, et al.* (2021). The immunodominant and neutralization linear epitopes for SARS-CoV-2. Cell Reports *34*, 108666.

Lu, S., Ye, Q., Singh, D., Villa, E., Cleveland, D.W., and Corbett, K.D. (2020). The SARS-CoV-2 Nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA and the membrane-associated M protein. bioRxiv, 2020.2007.2030.228023.

Luo, H., Ye, F., Chen, K., Shen, X., and Jiang, H. (2005). SR-Rich Motif Plays a Pivotal Role in Recombinant SARS Coronavirus Nucleocapsid Protein Multimerization. Biochemistry *44*, 15351-15358.

Luo, M., Terrell, J.R., and Mcmanus, S.A. (2020). Nucleocapsid Structure of Negative Strand RNA Virus. Viruses *12*, 835.

Mavrakis, M., Kolesnikova, L., Schoehn, G., Becker, S., and Ruigrok, R.W.H. (2002). Morphology of Marburg Virus NP–RNA. Virology *296*, 300-307.

McBride, R., Zyl, M.v., and Fielding, B. (2014). The Coronavirus Nucleocapsid Is a Multifunctional Protein. Viruses *6*, 2991-3018.

Messaoudi, I., Amarasinghe, G.K., and Basler, C.F. (2015). Filovirus pathogenesis and immune evasion: insights from Ebola virus and Marburg virus. Nature Reviews Microbiology *13*, 663-676.

Patel, A., Lee, Hyun O., Jawerth, L., Maharana, S., Jahnel, M., Hein, Marco Y., Stoynov, S., Mahamid, J., Saha, S., Franzmann, Titus M.*, et al.* (2015). A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. Cell *162*, 1066-1077.

Perlman, S., and Netland, J. (2009). Coronaviruses post-SARS: update on replication and pathogenesis. Nature Reviews Microbiology *7*, 439-450.

Raymond, D.D., Piper, M.E., Gerrard, S.R., and Smith, J.L. (2010). Structure of the Rift Valley fever virus nucleocapsid protein reveals another architecture for RNA encapsidation. Proceedings of the National Academy of Sciences *107*, 11769-11774.

Savastano, A., Ibáñez de Opakua, A., Rankovic, M., and Zweckstetter, M. (2020). Nucleocapsid protein of SARS-CoV-2 phase separates into RNA-rich polymerase-containing condensates. Nature Communications *11*, 6041.

Sola, I., Almazán, F., Zúñiga, S., and Enjuanes, L. (2015). Continuous and Discontinuous RNA Synthesis in Coronaviruses. Annual Review of Virology *2*, 265-288.

Su, Z., Wu, C., Shi, L., Luthra, P., Pintilie, G.D., Johnson, B., Porter, J.R., Ge, P., Chen, M., Liu, G.*, et al.* (2018). Electron Cryo-microscopy Structure of Ebola Virus Nucleoprotein Reveals a Mechanism for Nucleocapsid-like Assembly. Cell *172*, 966-978.e912.

Takeda, M., Chang, C.-k., Ikeya, T., Güntert, P., Chang, Y.-h., Hsu, Y.-l., Huang, T.-h., and Kainosho, M. (2008). Solution Structure of the C-terminal Dimerization Domain of SARS Coronavirus Nucleocapsid Protein Solved by the SAIL-NMR Method. Journal of Molecular Biology *380*, 608-622.

Tang, M.S., Hock, K.G., Logsdon, N.M., Hayes, J.E., Gronowski, A.M., Anderson, N.W., and Farnsworth, C.W. (2020a). Clinical Performance of the Roche SARS-CoV-2 Serologic Assay. Clinical Chemistry *66*, 1107-1109.

Tang, M.S., Hock, K.G., Logsdon, N.M., Hayes, J.E., Gronowski, A.M., Anderson, N.W., and Farnsworth, C.W. (2020b). Clinical Performance of Two SARS-CoV-2 Serologic Assays. Clinical Chemistry *66*, 1055-1062.

Wan, W., Kolesnikova, L., Clarke, M., Koehler, A., Noda, T., Becker, S., and Briggs, J.A.G. (2017). Structure and assembly of the Ebola virus nucleocapsid. Nature *551*, 394-397.

Williams, C.G., Gibbons, J.S., Keiffer, T.R., Luthra, P., Edwards, M.R., and Basler, C.F. (2020). Impact of Měnglà Virus Proteins on Human and Bat Innate Immune Pathways. Journal of Virology *94*, e00191-00120.

Wu, C.-H., Chen, P.-J., and Yeh, S.-H. (2014). Nucleocapsid Phosphorylation and RNA Helicase DDX1 Recruitment Enables Coronavirus Transition from Discontinuous to Continuous Transcription. Cell Host & Microbe *16*, 462-472.

Wu, C.-H., Yeh, S.-H., Tsay, Y.-G., Shieh, Y.-H., Kao, C.-L., Chen, Y.-S., Wang, S.-H., Kuo, T.-J., Chen, D.-S., and Chen, P.-J. (2009). Glycogen Synthase Kinase-3 Regulates the Phosphorylation of Severe Acute Respiratory Syndrome Coronavirus Nucleocapsid Protein and Viral Replication. Journal of Biological Chemistry *284*, 5229-5239.

Ye, Q., West, A.M.V., Silletti, S., and Corbett, K.D. (2020). Architecture and self-assembly of the SARS-CoV-2 nucleocapsid protein. Protein Science *29*, 1890-1901.

Figure legends

Figure 1. Characterization of N oligomerization with dynamic light scattering. A. Domain architecture of N. N has two structural domains: NTD and CTD. The sequence between NTD and CTD is a linker region (LKR) containing a serine-arginine rich motif. The Narm, LKR, and Carm are predicted to be disordered based upon sequence

analysis. **B.** Sequence identity between N of SARS-CoV-2 and common cold coronaviruses and other epidemic severe coronaviruses MERS-CoV and SARS-CoV. **C.** Measurements of N oligomerization with dynamic light scattering. Measured hydrodynamic radii, R_h, are reported in D. **D.** Table summarizes the DLS data for all constructs. Numbers are reported as average and standard deviation of three experiments. See also Figure S1 and S2.

Figure 2. SARS-CoV-2 nucleocapsid protein binds RNA with high affinity. A. Principles of fluorescence polarization assay to measure RNA binding. Increasing concentrations of N was titrated into 1 nM of FITC-labeled RNA. Protein binding to FITC-RNA leads to slower tumbling of FITC-RNA, resulting in increased fluorescence polarization. **B.** Fluorescence polarization binding curves of N constructs to a 20-nt ssRNA. The fitted K_D values are 0.007 ± 0.001 μM (N_{WT}, black square), 16 ± 12 μM (N_{NTD}, red circle), 13 ± 5 μM (N_{CTD}, blue up triangle), 0.006 ± 0.002 μM ($N_{NTD\text{-}LKR\text{-}CTD}$, magenta down triangle), and 0.006 ± 0.002 μM ($N_{NTD\text{-}LKR\text{-}CTD\text{-}Carm}$, green diamond). **C.** Fluorescence polarization binding curves of N constructs to a 20-nt ssRNA. The fitted K_D values are 0.006 ± 0.002 nM ($N_{NTD\text{-}LKR\text{-}CTD}$, magenta down triangle), 13 ± 5 μM (N_{CTD}, blue circle), 0.50 ± 0.08 μM ($N_{NTD\text{-}LKR}$, orange star), and 0.44 ± 0.04 μM ($N_{NTD\text{-}LKR}$ + N_{CTD}, purple pentagon). **D.** Fluorescence polarization binding curves of N constructs to a 20-nt ssRNA. The fitted K_D values are 0.006 ± 0.002 μM ($N_{NTD\text{-}LKR\text{-}CTD}$, magenta down triangle), 16 ± 12 μM (N_{NTD}, red circle), 0.35 ± 0.04 μM ($N_{LKR\text{-}CTD}$, orange up triangle), and 0.72 ± 0.09 μM (N_{NTD} + $N_{LKR\text{-}CTD}$, purple down triangle). **E.** Fluorescence polarization binding curves of N constructs to a slRNA. The fitted K_D values are 0.051 ±

0.004 μM (N_{WT}, black square), 124 ± 84 μM (N_{NTD}, red circle), 65 ± 44 μM (N_{CTD}, blue up triangle), 2.5 ± 0.5 μM ($N_{NTD\text{-}LKR}$, magenta down triangle), 0.22 ± 0.02 μM ($N_{NTD\text{-}LKR\text{-}CTD}$, green diamond), and 0.10 ± 0.01 μM ($N_{NTD\text{-}LKR\text{-}CTD\text{-}Carm}$, navy left triangle). **F**. Table summarizes K_D values (μM) for key constructs binding to ssRNA and slRNA. Numbers are reported as average and standard deviation of two experiments. See also Figure S3.

Figure 3. HDX-MS mapping of RNA binding to $N_{NTD\text{-}LKR\ S176D/S188D/S206D}$. A. Woods' plot showing cumulative differential HDX and validating differences using global significance limits. The horizontal bars depict the cumulative HDX differences between the RNA-bound and unbound $N_{NTD\text{-}LKR\ S176D/S188D/S206D}$. Standard deviations are shown for each peptide. Peptides showing statistically significant differences are differentiated by global significance limit using this standard error of the mean a t-values for a two-tailed Student's t-distribution (*, $p < 0.1$; **, $p < 0.05$; ***, $p < 0.01$). The blue shade of the peptide bar indicates differing statistical significance (light blue, medium blue, and navy, respectively); gray peptide bars depict peptides where statistically significant differences in HDX were not observed. Vertical bars show previously reported binding sites (residues reported for RNA-binding CoV2 N-protein (Dinesh et al., 2020b; Ye et al., 2020), AMP-binding HCoV-OC43 (Lin et al., 2014; Ye et al., 2020), and for both are shown in red, yellow, and orange, respectively). Secondary structure (PDB 6M3M) is shown above. **B**. Representative kinetic plots showing peptide level HDX as a function of exchange time (unbound, black; bound to RNA, red). **C**. Sites of protection measured by HDX mapped on the N_{NTD} structure (PDB 6M3M). Statistically significant HDX protection, regions of no difference in HDX, and regions where lacking proteolytic

coverage results in no data are shown in teal, light gray, and dark gray, respectively. Those residues unresolved in the structure are shown as a dashed line, with the exception of those reporting a statistically significant difference in teal. **D**. Electrostatic potential calculated with APBS mapped on to the N_{NTD} structure (PDB 6M3M) shows a major positive charge groove. Red and blue represent negative and positive electrostatic potential. The color scale is in kTe^{-1} units. See also Figure S4 and Mendeley dataset (doi:10.17632/sv8r6phkzt.1).

Figure 4. N-RNA forms liquid droplets and phosphorylation modulates N-RNA interactions. A. Size exclusion chromatography of N constructs (N_{WT}, black; $N_{NTD\text{-}LKR\text{-}CTD\text{-}Carm}$, red; $N_{NTD\text{-}LKR\text{-}CTD}$, blue) in 25 mM HEPES, 500 mM NaCl, 2 mM TCEP, 5% glycerol. Samples from peak 1 (p1) and p2 contain RNA whereas p3 are RNA-free based upon absorbance from the 260/280 ratio. **B.** Negative stain electron microscopy (EM) image of p1 and p2 for N_{WT}, $N_{NTD\text{-}LKR\text{-}CTD\text{-}Carm}$, and $N_{NTD\text{-}LKR\text{-}CTD}$ in 150 mM NaCl. Samples were diluted into 150 mM NaCl before negative-staining fixation by uranyl acetate. **C.** Size exclusion chromatography of N constructs (N_{WT}, black; $N_{S188D/S206D}$, blue; $N_{S176D/S188D/S206D}$, red) in 25 mM HEPES, 500 mM NaCl, 2 mM TCEP, 5% glycerol. **D**. Negative stain electron microscope image of $N_{S188D/S206D}$ and $N_{S176D/S188D/S206D}$ in 150 mM NaCl. **E.** Fluorescence polarization binding curves of N mutants to a 20-nt ssRNA. The fitted K_D values are 0.007 ± 0.001 μM (N_{WT}, black square), 0.015 ± 0.002 μM ($N_{S176D/S188D/S206D}$, black circle), 0.505 ± 0.075 μM ($N_{NTD\text{-}LKR}$, orange up triangle), and 1.1 ± 0.2 μM ($N_{NTD\text{-}LKR\ S176D/S188D/S206D}$, orange down triangle). Numbers are reported as average and standard deviation of two experiments. See also Figure S5.

Figure 5. The CTD of N is a highly sensitive serological marker. A. ELISA data of N_{WT} screened against plasma of COVID-19 positive and negative individuals from a combined Hong Kong, PRC and St. Louis, MO, USA cohort. Black solid line indicates the mean OD_{450} value for each population. **** $p < 0.0001$. **B.** ELISAs with the various N constructs for patient IgG. ELISAs were performed on plasma samples from COVID-19 patients (n = 68) and negative controls (n = 28). The cut-off is represented by the dotted line and calculated as the mean + 3 standard deviations of the negative population. Mean values ± standard deviation of COVID-19 and negative groups are shown. **C.** Sensitivity and specificity for each of the N domains calculated from the ELISA results. **D.** Heat-map of ELISA results for $N_{NTD\text{-}LKR\text{-}CTD\text{-}Carm}$, N_{CTD}, and N_{NTD} constructs from COVID-19 samples (n = 67). Each column represents an individual sample. **E.** Maturation of the N_{CTD} and $N_{NTD\text{-}LKR\text{-}CTD\text{-}Carm}$ IgG response over time (n = 67). **F.** Ratio of OD_{450} for N_{CTD} and $N_{NTD\text{-}LKR\text{-}CTD\text{-}Carm}$ for acute and convalescent time-points. Mean values ± standard deviation of acute and convalescent COVID-19 samples are shown. Experiments were repeated twice. Statistical significance was calculated by unpaired Student's t-test, ****$p < 0.0001$. **G**. Inhibition of SeV-induced IFNβ promoter activation by N constructs. Fold changes are relative to vector-only (V) transfections without SeV infection. MLAV VP35 served as a positive control for inhibition. Three transfection concentrations were used: 1.25, 12.5, and 125 ng/well. Statistical significance was determined by performing a one-way ANOVA followed with Tukey multiple comparison as compared to Sendai virus-infected control; **** $p < 0.0001$, *** $p < 0.0002$, ** $p < 0.0021$, * $p < 0.0332$.

804

TABLE FOR AUTHOR TO COMPLETE

Please upload the completed table as a separate document. ***Please do not add subheadings to the key resources table.*** *If you wish to make an entry that does not fall into one of the subheadings below, please contact your handling editor.* ***Any subheadings not relevant to your study can be skipped.*** *(**NOTE:** For authors publishing in Cell Genomics, Cell Reports Medicine, Current Biology, and Med, please note that references within the KRT should be in numbered style rather than Harvard.)*

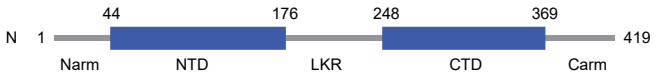
KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-human IgG secondary Ab labeled with HRP	Southern BioTech	Cat#905209
Bacterial and virus strains		
E. coli BL21(DE3)	Agilent	Cat#200131
SeV	Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY USA	N/A
Biological samples		
Human Plasma, Patient with Confirmed SARS-CoV-2 Infection	Barnes-Jewish Hospital, St. Louis, MO, USA	N/A
Human Plasma, Patient with Confirmed SARS-CoV-2 Infection	Hong Kong University and HK Island West Cluster of Hospitals, Hong Kong, PRC	N/A
Human Plasma, Patient with negative test	Barnes-Jewish Hospital, St. Louis, MO, USA	N/A
Chemicals, peptides, and recombinant proteins		
LC-MS grade formic acid	CovaChem	Cat# PI 85171; CAS 64-18-6
Phosphate buffered saline tablets	Millipore Sigma	Cat# P4417
Deuterium oxide (D, 99.9%)	Cambridge Isotope Laboratories	Cat# DLM-4-100; CAS 7789-20-0;
Urea	Millipore Sigma	Cat# U4883; CAS: 57-13-6
2% Uranyl Acetate	Ted Pella	19481
Critical commercial assays		
Dual-Glo luciferase kit	Promega	Cat#E2920
Deposited data		
Experimental models: Cell lines		
Human: HEK293T cells	ATCC	CRL-3216

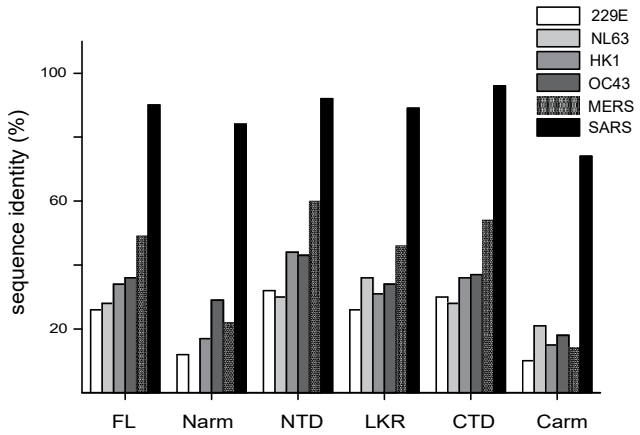
Experimental models: Organisms/strains		
Oligonucleotides		
20nt ssRNA sequence: UUUCACCUCCCUUUCAGUUU	GenScript	N/A
19nt slRNA sequence: GGAAGAUUAAUAAUUUUCC	GenScript	N/A
Recombinant DNA		
Plasmid pET28a-CterTEV His6 SARS-CoV-2 N 1-419	This work	N/A
Plasmid pET28a-CterTEV His6 SARS-CoV-2 N 44-419	This work	N/A
Plasmid pET28a-CterTEV His6 SARS-CoV-2 N 44-369	This work	N/A
Plasmid pET28a-CterTEV His6 SARS-CoV-2 N 44-247	This work	N/A
Plasmid pET28a-CterTEV His6 SARS-CoV-2 N 44-176	This work	N/A
Plasmid pET28a-CterTEV His6 SARS-CoV-2 N 248-369	This work	N/A
Plasmid pCAGGS MLAV VP35	This work	N/A
Plasmid pCAGGS SARS-CoV-2 N 1-419	This work	N/A
Plasmid pCAGGS SARS-CoV-2 N 44-369	This work	N/A
Plasmid pCAGGS SARS-CoV-2 N 44-176	This work	N/A
Plasmid pCAGGS SARS-CoV-2 N 248-419	This work	N/A
Plasmid pCAGGS SARS-CoV-2 N 248-369	This work	N/A
IFN-β promoter-firefly luciferase reporter plasmid	Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY USA	N/A
pRL-TK Renilla luciferase reporter plasmid	Promega	E2231
Software and algorithms		
Intact Mass™	Protein Metrics Incorporated	Version 3.11
Byonic™	Protein Metrics Incorporated	Version 3.11
Byologic™	Protein Metrics Incorporated	Version 3.11
DataAnalysis v 4.4	Bruker Daltonics	Version 4.4
HDExaminer	Sierra Analytics Incorporated	Version 2.5.1
Origin	OriginLab	Version 7
UCSF Chimera	Pettersen, *et al.*, 2004	http://www.cgl.uscf.edu/chimera/
PRISM	GraphPad	Version 7
Dynamics Software	Wyatt	Version 7
Gen5 Software	BioTek	Version 3
Other		
ZORBAX Eclipse XDB C8 trap column (2.1 x 15 mm)	Agilent Technologies	Cat# 975700-936
XSelect CSH C18 XP (130 Å, 2.5 µm, 2.1 x 50 mm)	Waters Corporation	Cat# 186006101
MaXis II 4G Q-TOF	Bruker Daltonics	N/A

MaXis II HM Q-TOF	Bruker Daltonics	N/A
EmulsiFlex-C5 homogenizer	Avestin	N/A
Glow-Discharged Copper Grid, 200 mesh	Ted Pella	01840-F
JEM-1400 plus TEM	JEOL	N/A
Phosphor-Scintillated 12-bit CCD Camera	AMT	XR111
DynaPro-PlateReader II	Wyatt Technologies Corporation	N/A
Cytation5 Plate Reader	BioTek	N/A

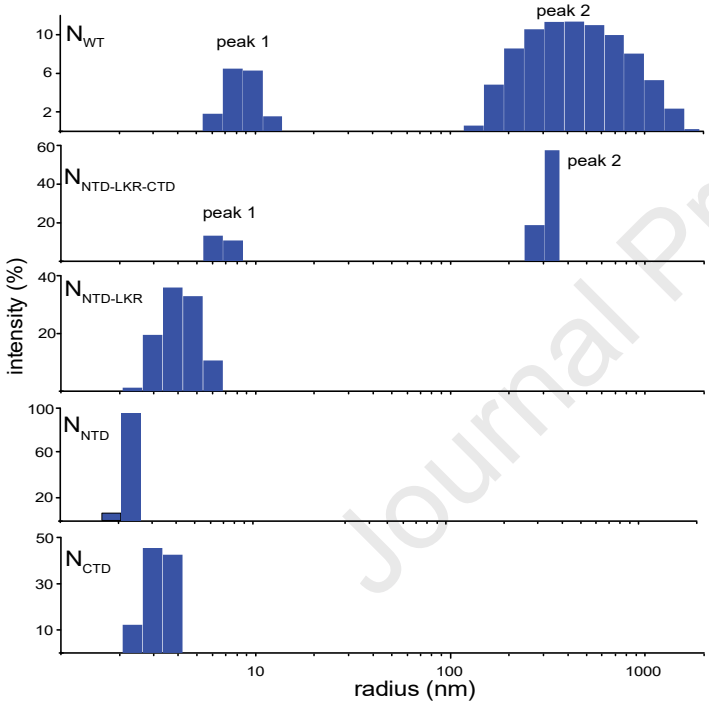
A



B



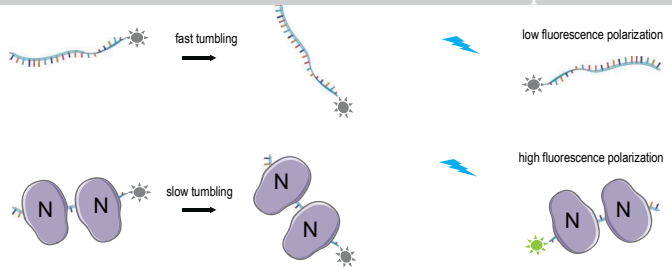
C



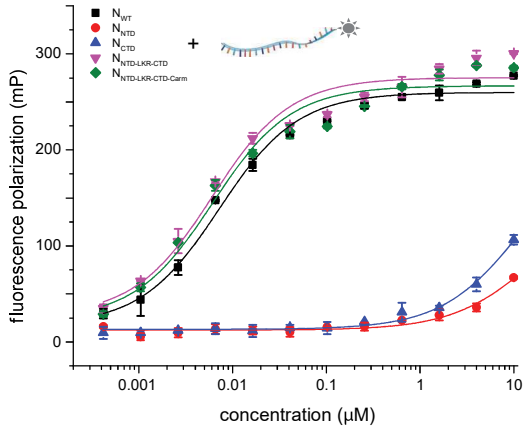
D

	MW (kDa)	Rh (nm)	
		peak 1	peak 2
N_{WT}	46	8.9 ± 0.2	450 ± 100
$N_{NTD\text{-}LKR\text{-}CTD}$	35	6.5 ± 0.4	360 ± 20
$N_{NTD\text{-}LKR}$	22	3.9 ± 0.1	
N_{NTD}	14	2.3 ± 0.1	
N_{CTD}	14	3.5 ± 0.1	

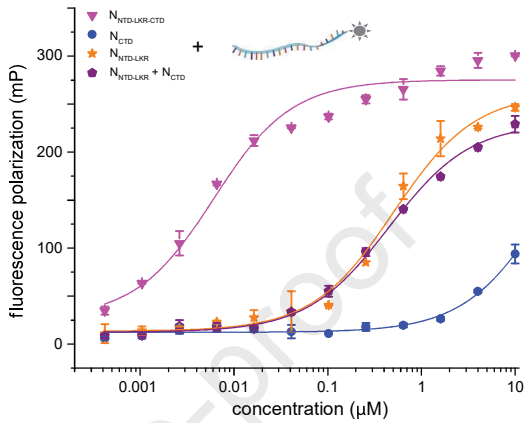
A



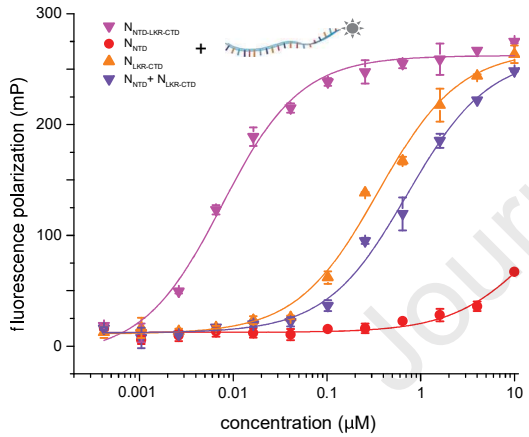
B



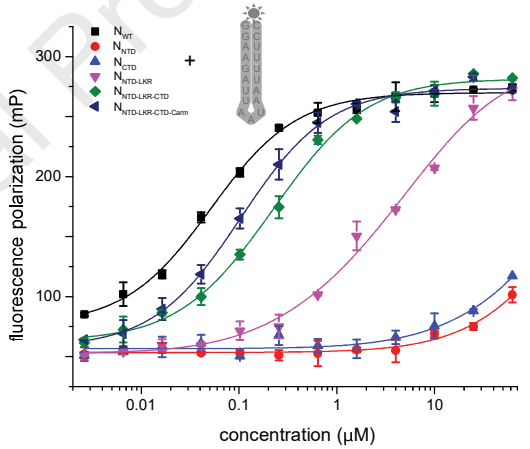
C



D

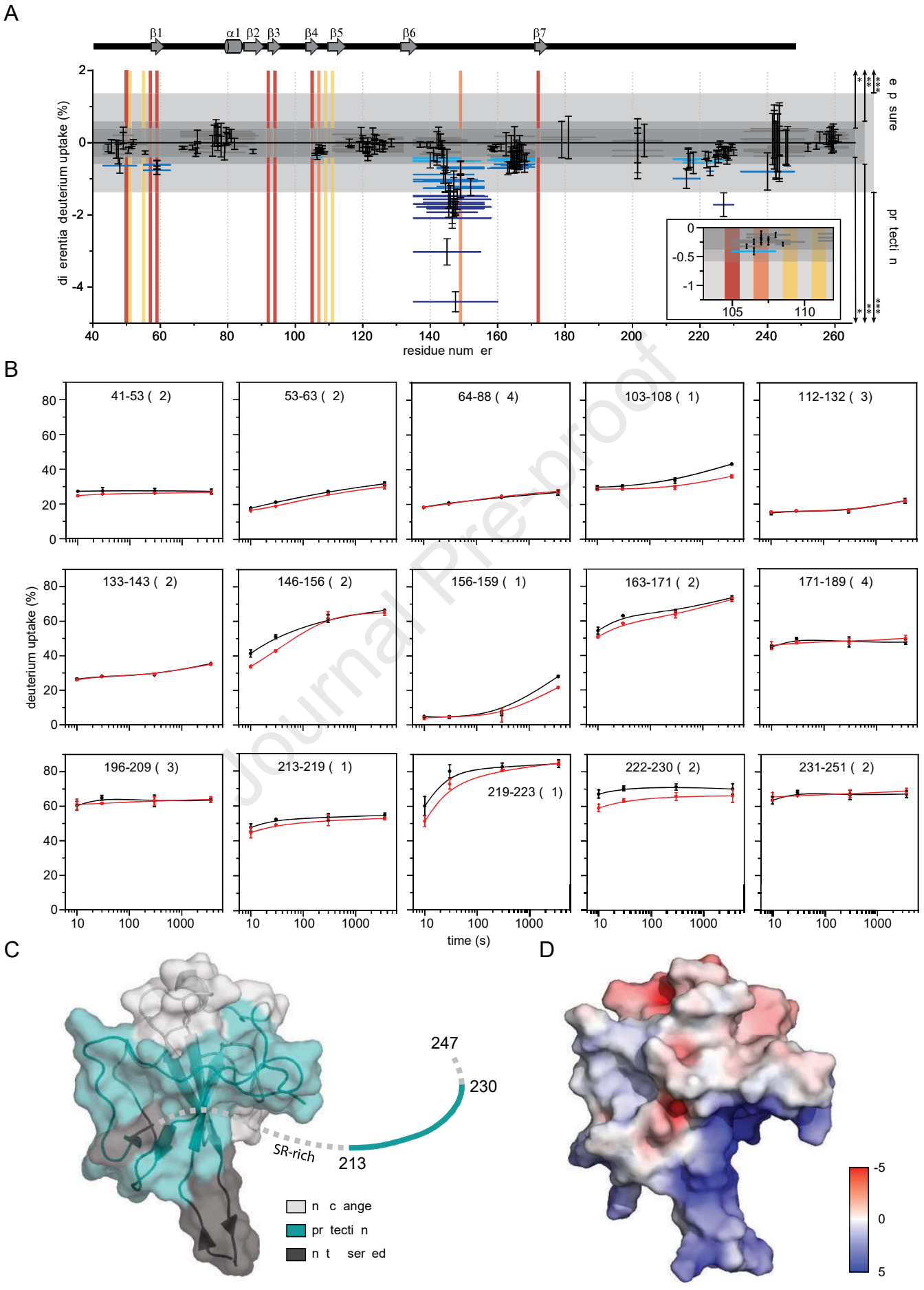


E

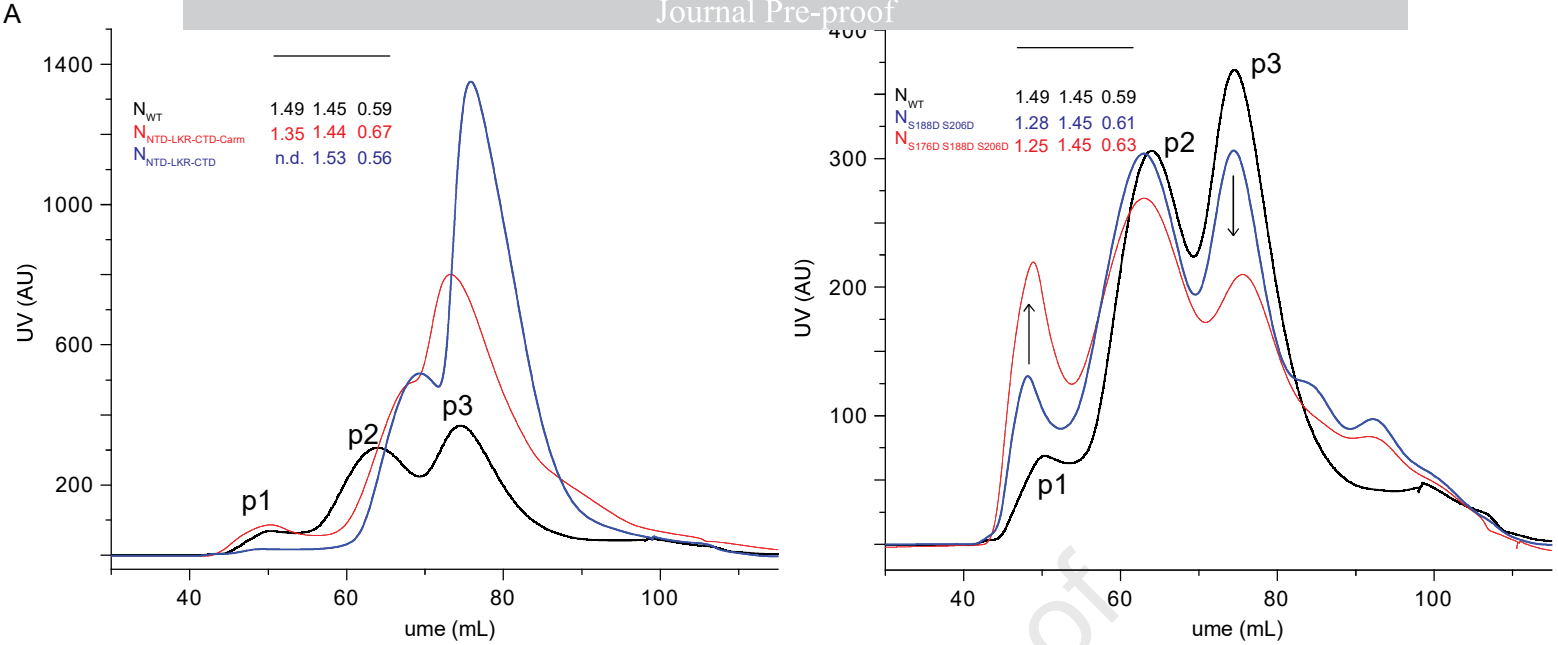


F

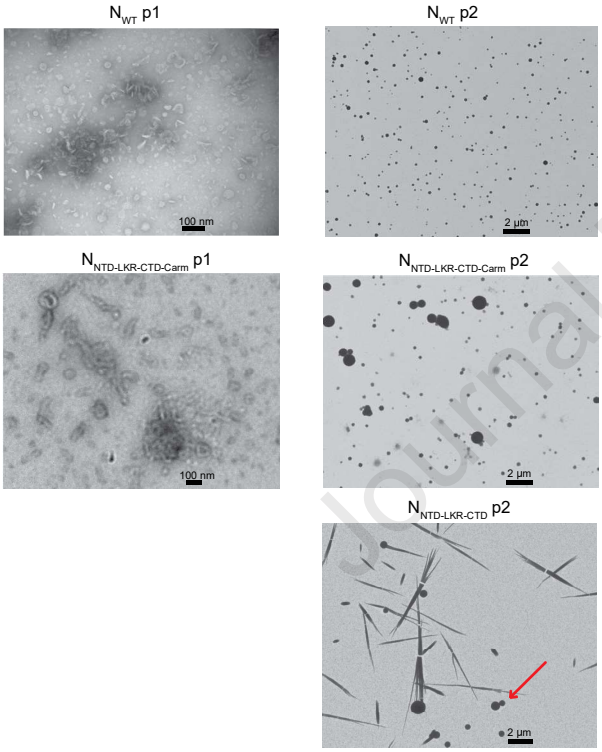
	N_{WT}	$N_{NTD-LKR-CTD-Carm}$	$N_{NTD-LKR-CTD}$	$N_{NTD-LKR}$	N_{NTD}	N_{CTD}
	0.007 ± 0.001	0.006 ± 0.002	0.006 ± 0.002	0.50 ± 0.08	20 ± 10	13 ± 5
	0.051 ± 0.004	0.10 ± 0.01	0.22 ± 0.02	2.5 ± 0.5	120 ± 80	60 ± 40



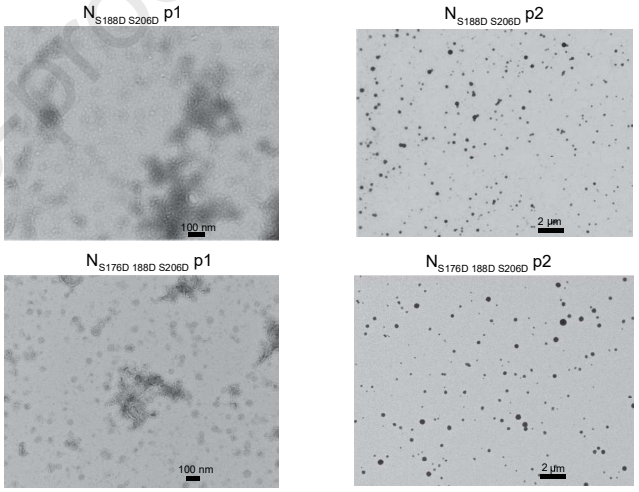
A



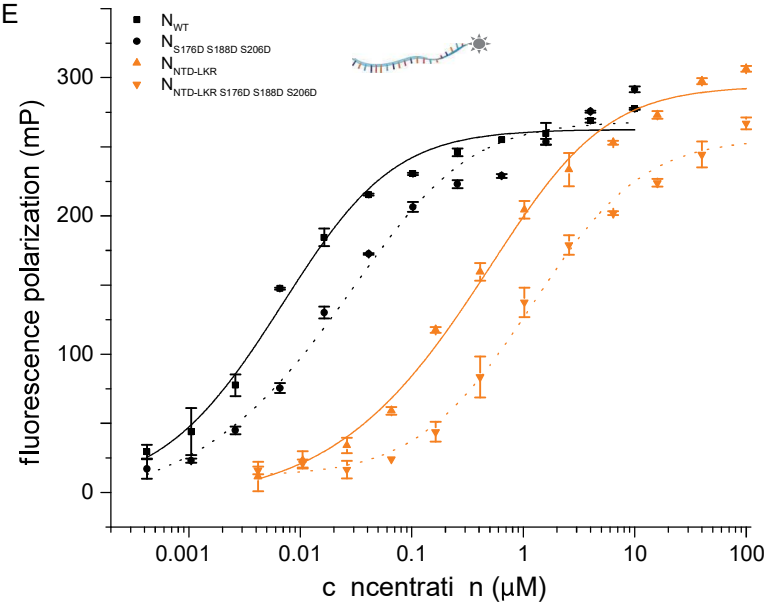
B



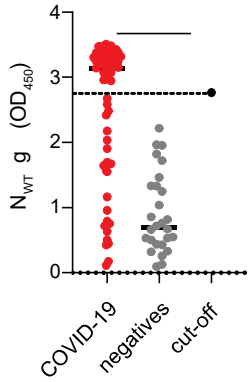
D



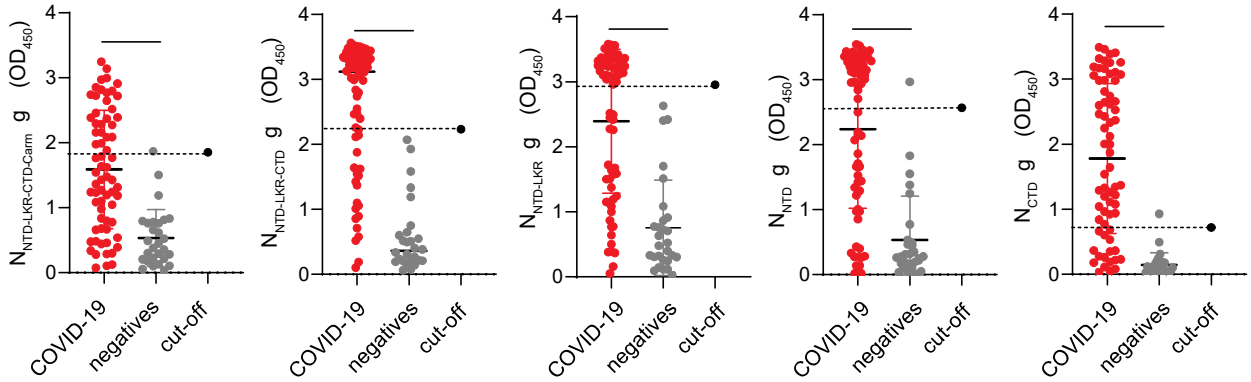
E



A



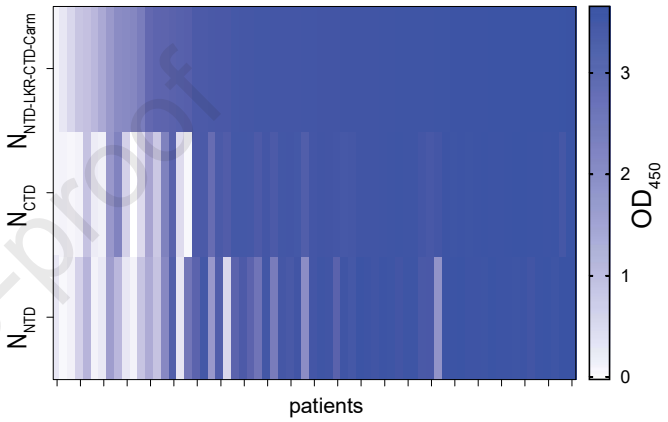
B



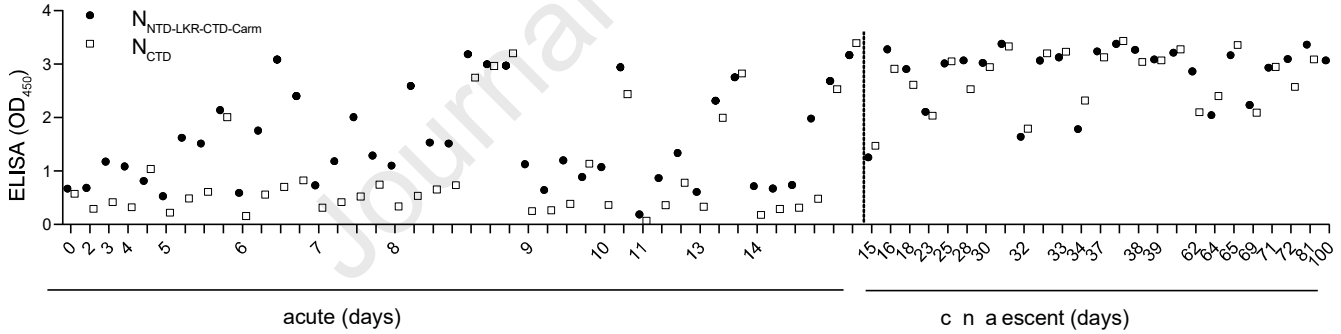
C

	sensiti ity (%)	speci icity (%)
N_{WT}	66.2	100
$N_{NTD-LKR-CTD-Carm}$	41.2	96.4
$N_{NTD-LKR-CTD}$	70.6	100
$N_{NTD-LKR}$	54.4	100
N_{NTD}	54.4	96.4
N_{CTD}	75.0	96.4

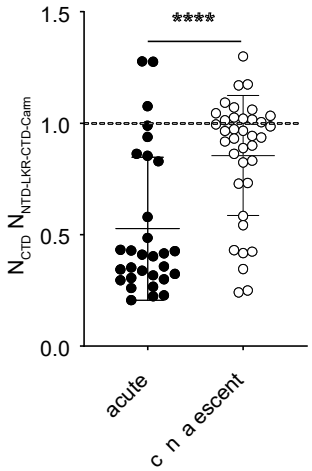
D

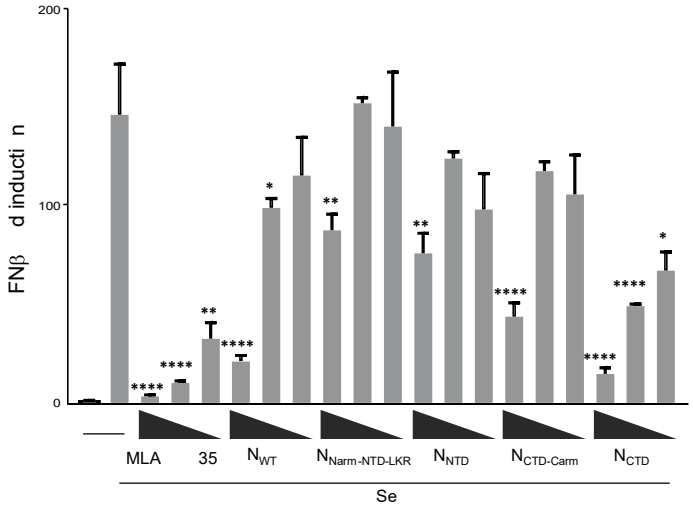


E



F





- Domain specific contributions of SARS-CoV-2 N to oligomerization and RNA binding.
- Phosphorylation of N modulate RNA binding affinity and N-RNA solution properties.
- HDX-MS results probe RNA binding interface of SARS-CoV-2 N.
- CTD may be a more sensitive and specific serological marker of infection.