

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

3

4 Chao Wu¹, Abraham J. Qavi^{1*}, Asmaa Hachim^{2*}, Niloufar Kavian^{2,3,4*}, Aidan R. Cole^{1*},
5 Austin B. Moyle^{6*}, Nicole D. Wagner^{6*}, Joyce Sweeney-Gibbons^{7*}, Henry W. Rohrs⁶,
6 Michael L. Gross⁶, J. S. Malik Peiris^{2,8}, Christopher F. Basler⁷, Christopher W.
7 Farnsworth^{1#}, Sophie A. Valkenburg^{2#}, Gaya K. Amarasinghe^{#1}, and Daisy W. Leung^{1,5#},

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

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

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

15 ⁴Institut Cochin, INSERM U1016, Université Paris Descartes, Sorbonne Paris Cité, Paris.

16 ⁵Department of Internal Medicine, Washington University School of Medicine in St. Louis, St.
17 Louis, MO, USA.

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

19 ⁷Center for Microbial Pathogenesis, Institute for Biomedical Sciences, Georgia State University,
20 Atlanta, GA, USA

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

23 *equal contributions

24 #corresponding authors

25

26 Key words: SARS-CoV-2, COVID-19, RNA binding, oligomerization, LLPS, nucleoprotein,
27 serology

28

29 **Summary**

30 Nucleocapsid protein (N) is the most abundant viral protein encoded by SARS-CoV-2, the
31 causative agent of COVID-19. N plays key roles at different steps in the replication cycle
32 and is used as a serological marker of infection. Here we characterize the biochemical
33 properties of SARS-CoV-2 N. We define the N domains important for oligomerization and
34 RNA binding that are associated with spherical droplet formation and suggest that N
35 accessibility and assembly may be regulated by phosphorylation. We also map the RNA
36 binding interface using hydrogen-deuterium exchange mass spectrometry. Finally, we
37 find that the N protein C-terminal domain is the most immunogenic by sensitivity, based
38 upon antibody binding to COVID-19 patient samples from the US and Hong Kong.
39 Together, these findings uncover domain-specific insights into the significance of SARS-
40 CoV-2 N and highlight the diagnostic value of using N domains as highly specific and
41 sensitive markers of COVID-19.

42

43

44

45 Introduction

46 Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a novel coronavirus
47 and the causative agent of COVID-19. Coronavirus has a single-stranded, positive-sense
48 RNA genome. The genome encodes four major structural proteins: spike (S), envelope
49 (E), membrane (M), and nucleocapsid (N). The N protein is the second most proximal to
50 the 3' end of the genome and is one of the most abundantly expressed viral proteins given
51 the multifunctional roles N has during viral replication and assembly (Fung and Liu, 2019;
52 Kim et al., 2020; McBride et al., 2014; Perlman and Netland, 2009). It is estimated that
53 1,000 copies of N are incorporated into each virion versus only 100 copies of S (Bar-On
54 et al., 2020). Whereas N exists mostly in a phosphorylated state in the cytoplasm, it is
55 predominantly in a dephosphorylated state in mature virions, suggesting that genome
56 packaging is regulated by the phosphorylation state of N (Wu et al., 2014; Wu et al., 2009).

57 A critical function of N is to encapsidate the ssRNA viral genome to evade immune
58 detection and to protect the viral RNA from degradation by host factors (Chang et al.,
59 2014; McBride et al., 2014). N has two structural domains (**Figure 1A**): a N-terminal
60 domain (N_{NTD}; amino acid residues 44-176) and a C-terminal domain (N_{CTD}; residues 248-
61 369). N_{NTD} is often referred to as the RNA-binding domain (RBD), although the exact
62 regions involved in RNA binding are not yet well defined (Chang et al., 2014; Grosseohme
63 et al., 2009; Gui et al., 2017; Kang et al., 2020; McBride et al., 2014). The isolated N_{CTD}
64 exists as a dimer in solution and is potentially involved in RNA binding as well (Bouhaddou
65 et al., 2020; Gui et al., 2017; Takeda et al., 2008). A conserved serine/arginine rich-linker
66 region (N_{LKR}) connects the N_{NTD} and the N_{CTD}. Phosphorylation of residues in the LKR is
67 believed to regulate discontinuous transcription, particularly for shorter subgenomic

68 mRNA closer to the 3' end during early stages of replication (Wu et al., 2014; Wu et al.,
69 2009). The LKR, along with Narm and Carm, have been shown to be intrinsically
70 disordered (Chang et al., 2014; Cubuk et al., 2020). However, the LKR is similarly
71 conserved as the NTD and CTD (**Figure 1B** and **Supp. Figure 1B**), supporting its
72 essential functional role across strains of coronaviruses (Chang et al., 2014), while the
73 Narm and Carm are the least conserved regions.

74 Given its abundant expression and conservation within the genome, N has been used as
75 an antigen for serology tests (Chew et al., 2020; Tang et al., 2020a, b), such as the widely
76 used Roche Elecsys Anti-SARS-CoV-2 assay and Abbott SARS-CoV-2 IgG assay (Tang
77 et al., 2020a, b). Comparison of the serological response to the entire proteome of the
78 SARS-CoV-2 virus using a luciferase immunoprecipitation assay (LIPS) (Hachim et al.,
79 2020) showed that N-specific antibodies dominated the overall antibody response.
80 Furthermore, the T cell responses directed towards N are highly immunodominant in
81 SARS-CoV and SARS-CoV-2 infection, with N-specific memory T cell responses evident
82 17 years after the initial SARS-CoV infection (Le Bert et al., 2020). Due to epitope
83 conservation of the N protein, there is some T cell cross-reactivity towards the N of the
84 SARS-CoV-2 (Le Bert et al., 2020). The N protein stability, RNA binding characteristics,
85 abundance, and conservation altogether impact immunogenicity for T and B cell
86 immunities.

87 Previous studies, including our own, have examined fundamental properties of
88 nucleocapsid or nucleoproteins (N) from RNA viruses, which revealed overlapping and
89 unique functions (Arragain et al., 2019; Ding et al., 2016; Lu et al., 2020; Luo et al., 2020;
90 Raymond et al., 2010; Su et al., 2018; Wan et al., 2017). These include the insights into

91 RNA binding, oligomerization, as well as potential roles in RNA synthesis and immune
92 evasion. However, such studies for SARS-CoV-2 N protein remain incomplete. To
93 address this gap, here we use a series of biochemical and biophysical assays to probe
94 essential functions of N. Our results reveal that oligomeric N provides a continuous RNA
95 binding platform and that N-RNA has different morphologies, including spherical droplets.
96 Our data also suggest how phosphorylation could modulate these processes, which can
97 be exploited by drugs targeting of these cellular processes (Bouhaddou et al., 2020).
98 Given the recent progress in SARS-CoV-2 spike-based vaccines, knowledge of N will
99 likely provide the basis to differentiate individuals with immunity due to natural infections
100 from those that are immunized. Such tools will play a critical role in managing herd
101 immunity. In addition, our domain-specific insights into the immunogenicity of N provide
102 opportunities to further enhance sensitivity and specificity of serology testing. Therefore,
103 insights gained here regarding N protein function and regulation are critical for improving
104 diagnostic testing for the ongoing COVID-19 pandemic and future outbreaks.

105

106 **Results**

107 **Multiple domains within N protein are important for oligomerization.** To characterize
108 how the domains of N contribute to oligomerization, we used dynamic light scattering
109 (DLS) to determine the hydrodynamic properties of isolated N domains. The results reveal
110 that there are two major oligomeric species for wildtype (WT) N (N_{WT} ; 46 kDa), one with
111 a hydrodynamic radius (R_h) of 8.9 nm and another of 450 nm (**Figure 1C and 1D**). For
112 comparison, the R_h values for maltose binding protein (44 kDa) and bovine serum albumin
113 (66 kDa) are 2.9 nm and 3.7 nm, respectively. Removal of the Narm and Carm ($N_{NTD-LKR}$ -

114 CTD) generates two major species, similar to N_{WT}. However, both N_{NTD-LKR-CTD} populations
115 display reduced polydispersity (peak width, **Supp. Figure 2A**) in the high order oligomeric
116 species, suggesting that both arms contribute to N oligomerization. Removal of the CTD
117 (N_{NTD-LKR}) results in a single peak representing a dimeric species ($R_h = 3.9$ nm), but with
118 considerable polydispersity. N_{NTD} and N_{CTD} alone are stable domains; N_{NTD} is a monomer
119 ($R_h = 2.3$ nm) whereas N_{CTD} is a dimer in solution ($R_h = 3.5$ nm). Exact mass measurement
120 by denaturing mass spectrometry yields values corresponding to the expected sequence
121 (± 1 Da) and supports the identity of the constructs used here (**Supp. Figure 2B-E**).

122

123 **N provides an oligomerization platform for high affinity RNA binding.** N
124 encapsidates viral genomic ssRNA in a sequence-independent manner (Chang et al.,
125 2014). To gain insight into how each N domain contributes to ssRNA binding, we
126 measured fluorescence polarization (FP) upon binding of a FITC-labeled 20-nt ssRNA
127 (Sequence: UUUCACCUCCCUUUCAGUUU) (**Figure 2A**). We find that N_{WT} binds the
128 20-nt ssRNA with high affinity ($K_D = 0.007 \pm 0.001$ μ M). Removal of the Narm and Carm
129 do not impact ssRNA binding ($K_D = 0.006 \pm 0.002$ and 0.006 ± 0.002 μ M for N_{NTD-LKR-CTD-}
130 _{Carm} and N_{NTD-LKR-CTD}, respectively) (**Figure 2B**). The individual domains N_{NTD} and N_{CTD}
131 have low affinity binding ($K_D = 20 \pm 10$ and 13 ± 5 μ M, respectively), both of which bind
132 with increased affinity after inclusion of the LKR region (**Figure 2B-2D**). Interestingly,
133 addition of CTD that dimerizes onto NTD-LKR improves the binding affinity to the single
134 digit nM range (**Figure 2C**). The affinity increase does not happen when CTD is provided
135 *in trans* (compare N_{NTD-LKR} + N_{CTD} with N_{NTD-LKR-CTD}). This affinity increase also occurs
136 when NTD is added to LKR-CTD *in cis* (**Figure 2D**, compare N_{LKR-CTD} + N_{NTD} with N_{NTD-}

137 LKR-CTD). Similar binding curves and K_D values were obtained when fluorescence
138 anisotropy values were converted from polarization (**Supp. Figure 3A**). Collectively, our
139 data show that NTD, CTD, and LKR all contribute to ssRNA binding, and the presence of
140 three domains in tandem confers N with high affinity binding.

141

142 N also binds dsRNA and is proposed to disrupt dsRNA structures formed by transcription
143 regulatory sequences (TRS) during discontinuous transcription (Grossoehme et al., 2009;
144 Keane et al., 2012; Sola et al., 2015). We next tested N binding to a stem-loop RNA
145 (siRNA) (Sequence: GGAAGAUUAAUAAUUUCC) (**Figure 2E**). We find that N_{WT} binds
146 with relatively high affinity ($K_D = 0.051 \pm 0.004 \mu\text{M}$) whereas both N_{NTD} and N_{CTD} alone
147 have very weak binding affinities ($K_D = 120 \pm 80$ and $60 \pm 40 \mu\text{M}$, respectively). The
148 addition of LKR significantly improves binding, consistent with binding results for ssRNA.
149 Overall, binding of N to siRNA appears similar to binding to ssRNA, but at an order of
150 magnitude lower than that of ssRNA (**Supp. Figure 3B**). This may be due, in part, to the
151 energetic penalty of unfolding the stem loop structure. Furthermore, it seems that Narm
152 and Carm contribute more to siRNA binding than ssRNA because the reduction of K_D for
153 siRNA are more pronounced after removing the Narm and Carm (**Supp. Figure 3C**).

154

155 **N-RNA forms spherical droplets, requiring multiple N domains.** Purification of
156 recombinant N protein on size exclusion chromatography revealed three populations,
157 including two RNA-bound states (p1 and p2) and an RNA-free state (p3) despite
158 purification in high salt (500 mM) to eliminate RNA binding (**Figure 3A**). Truncation of the

159 Narm results in an increase of the RNA-free peak (p3), suggesting that N truncations can
160 alter the structure of N and correspondingly impact N-related properties, such as RNA
161 binding and oligomerization, thereby potentially impacting the physiological function of N.
162 We observe an even greater shift to p3 when both Narm and Carm were removed,
163 suggesting that both arms contribute to RNA binding interactions. To gain additional
164 insight into the two RNA-bound populations p1 and p2, we visualized these samples by
165 using negative-stain electron microscopy (EM) in near-physiological salt concentrations
166 (150 mM). N_{WT} p1 contains N-RNA with a loose-coil appearance (**Figure 3B, top left**),
167 similar to that observed for other RNA-bound nucleocapsids (Bharat et al., 2012; Mavrakis
168 et al., 2002). Other recent studies have also observed that N protein undergoes liquid-
169 liquid phase separation (LLPS) when mixed with RNA (Carlson et al., 2020; Cubuk et al.,
170 2020; Iserman et al., 2020; Jack et al., 2020; Savastano et al., 2020). In agreement with
171 these results, for N_{WT} p2, we mostly observe spheres corresponding to liquid droplets
172 separated from the surrounding buffer (**Figure 3B, top right**).

173

174 We next examined the role of Narm and Carm on high-order assembly. N_{NTD-LKR-CTD-Carm}
175 behaves similarly to N_{WT} and has loose coils in p1 (**Figure 3B, middle left**) and forms
176 spherical liquid droplets in p2 (**Figure 3B, middle right**). However, examination of p2
177 from N_{NTD-LKR-CTD} (**Figure 3B, bottom right**) revealed a much smaller population of liquid
178 droplets (red arrow) and mostly crystal-like needle aggregates, suggesting that the Carm
179 is important for droplet formation. A transition from spherical liquids to needle-like solids
180 is consistent with the liquid-to-solid transitions observed for other proteins that undergo
181 phase separation (Patel et al., 2015).

182 **Phosphorylation of LKR modulates RNA binding and higher-order assembly.**

183 Previous studies have shown that N is dephosphorylated in mature virions of a closely
184 related strain, SARS-CoV, and predominantly phosphorylated during active transcription
185 and replication in the cytoplasm (Wu et al., 2014). For SARS-CoV, glycogen synthase
186 kinase (GSK)-3 was shown to phosphorylate N at Ser177 (corresponding to Ser176 in
187 SARS-CoV-2 N) (Wu et al., 2009). Phosphorylation of Ser177 is preceded by
188 phosphorylation of Ser189 and Ser207 (Ser188 and 206 in SARS-CoV-2 N) by other
189 priming kinases (Wu et al., 2009). Moreover, N protein phosphorylation has been
190 qualitatively shown to modulate both RNA binding and phase separation (Carlson et al.,
191 2020; Lu et al., 2020; Savastano et al., 2020). To test if phosphorylation impacts RNA
192 binding and the phase separation of SARS CoV-2 N-RNA, we generated a set of N
193 phosphomimics by mutating specific serine residues to aspartate residues. Size exclusion
194 chromatography showed that, compared to N_{WT}, N_{S188D/S206D} produced a reduced RNA-
195 free peak (p3) and an increased RNA-bound peak (p1) (**Figure 3C**). Introduction of
196 S176D to generate N_{S176D/S188D/S206D} resulted in an even greater shift in p1 and p3
197 distributions, showing how phosphorylation can affect N interactions with RNA. The height
198 of p2 remains relatively the same for all preparations. Examination of these protein peaks
199 using electron microscopy revealed that N_{S188D/S206D} displays similar loose coils in p1
200 (**Figure 3D, top left**) and spherical droplets in p2 (**Figure 3D, top right**) for the RNA-
201 bound species. Similar observations were made for N_{S176D/S188D/S206D} (**Figure 3D**). To
202 describe this interaction further, we measured ssRNA binding to the N phosphomimics
203 (**Figure 3E** and **Supp. Figure 4A**). N_{S176D/S188D/S206D} displays ~5-fold lower binding affinity
204 to ssRNA compared to N_{WT} binding, a result consistent with previous work examining the

205 impact of LKR phosphorylation on RNA binding(Savastano et al., 2020). We observed a
206 similar trend for the N_{NTD-LKR} construct. Furthermore, binding to siRNA is also affected by
207 these mutations (**Supp. Figure 4B**). Collectively, our data suggest that phosphorylation
208 of the LKR region modulates N interactions with RNA, causing changes in solution
209 properties. Interestingly, there are 14 serine residues in the LKR, of which 13 are found
210 in SARS-CoV, and an increase in phosphorylation in this region may further enhance
211 these changes for RNA interaction and subsequent viral replication and immunogenicity.

212

213 **Hydrogen-deuterium exchange mass spectrometry (HDX-MS) locates the RNA**
214 **binding interface of NTD-LKR.** We performed HDX-MS to locate regions that become
215 protected upon RNA binding. Using N_{NTD-LKR-CTD}, we observed phase-separation and
216 aggregation upon RNA binding, causing a 100-fold mass spectrometric signal loss for the
217 bound state possibly attributable to phase separation or aggregation inducing poor
218 digestion. However, this problem does not pertain to N_{NTD-LKR S176D/S188D/S206D} where
219 digestion yielded 152 peptides after sequential FXIII and peptic digestion and provided
220 93.3% sequence coverage (**Supp. Figure 5A**).

221 HDX analysis of N_{NTD-LKR S176D/S188D/S206D} shows clear protection in four distinct regions of
222 upon RNA binding (aa residues 41-63, 105-108, 146-171, and 213-230) (**Figure 4A** and
223 **4B**). Residues 133-143 are not perturbed by RNA binding, but peptides covering 146-171
224 show clear protection. The largest differences in HDX are observed where 50-80% of the
225 residues of unbound peptides undergo a burst phase of HDX in the first 10 s (146-156,
226 163-171, and 213-230 aa), i.e., the peptides cover regions of little hydrogen bonding in
227 the unbound state. When bound to RNA, the fraction of residues participating in the burst

228 phase decreases, resulting in observed protection. Then, HDX either converges over time
229 (146-156, 163-171, and 219-223 aa) consistent with protein conformation or RNA binding
230 dynamics, or the HDX never converges in the timescale of the experiment (222-230 aa),
231 consistent with relatively static binding. Interestingly, peptides covering 103-108 aa and
232 156-159 aa undergo very little HDX throughout the experiment, consistent with either
233 hydrogen bonding of secondary and tertiary structure or a hydrophobic pocket. Of note,
234 HDX decreases for the bound state in these peptides only after 1 h. The low initial HDX
235 limits the dynamic range of binding-induced protection from HDX, but statistically
236 significant protection is still observed.

237 Further, HDX analysis revealed that the protected regions (**Figure 4C**) overlap well with
238 a basic patch groove in the N_{NTD} structure (**Figure 4D**). In addition, a region (213-230 aa)
239 within the LKR domain, after the structured domain and SR-motif, shows statistically
240 significant HDX protection. Interestingly, we did not detect HDX protection in the SR-motif,
241 which was proposed to bind RNA. This may be due to the Ser-to-Asp mutations
242 introduced into this region, changing the RNA binding patterns. Altogether, HDX results
243 along with our biochemical data define an RNA binding interface within the NTD-LKR
244 region.

245

246 **N_{CTD} is a sensitive serological marker.** Current sero-diagnostic assays to identify
247 COVID-19 positive individuals are based on the detection of antibodies against N due to
248 its abundant expression and corresponding high immune response (Chew et al., 2020;
249 Tang et al., 2020a, b). However, these N-directed serological assays are highly variable
250 and their sensitivity depend on the sampling time-points, ranging from 0% to 93.75% (Liu

251 et al., 2020; Tang et al., 2020a, b), suggesting that serological markers for SARS-CoV-2
252 infection can be further improved.

253

254 Given that the domains of N impact the various essential properties of N, we next
255 assessed which domains contribute to the observed N immunodominance in RT-PCR
256 confirmed COVID-19 patients. Plasma samples were collected from two cohorts, one in
257 St. Louis, USA (n = 45) and one in Hong Kong (n = 23), at different time points of infection
258 (Supplementary Table 1). Using these samples, we performed enzyme-linked
259 immunosorbent assays (ELISAs) to detect IgG present in COVID-19 patient plasma to
260 different N domains. First, we confirmed that purified N_{WT} is a sensitive serological marker
261 to differentiate between COVID-19 positive and negative individuals (**Figure 5A**). As
262 shown in **Figure 5B**, antibodies against all five N constructs were detected in the COVID-
263 19 cohort ($p < 0.0001$ versus negative controls for all). A cut-off based on the mean of the
264 negatives plus three standard deviations allowed us to assess the performance of each
265 N construct at detecting IgG antibodies in COVID-19 positive individuals (**Figure 5C**). We
266 find that N_{NTD-LKR-CTD-C_{arm}} shows the lowest sensitivity (41.2%), whereas the truncated
267 N_{NTD-LKR-CTD} can detect more COVID-19 positive individuals (70.6%). Furthermore, N_{CTD}
268 shows the highest combination of sensitivity (75%) and specificity (96.4%) over the other
269 N constructs tested. This is demonstrated by the lowest cut-off score for the N_{CTD} for
270 negative control samples, despite a comparable level of amino acid sequence
271 conservation of the N_{CTD} (29-41%) to the N_{NTD} (32-48%) and N_{LKR} (28-42%) domains with
272 common cold corona viruses (Figure 1B).

273

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

286

287 Given that some RNA viral N proteins are known immune antagonists, including prior
288 studies of SARS-CoV, we hypothesized that N from SARS-CoV-2 may also suppress the
289 type-I interferon (IFN) signaling pathway (Messaoudi et al., 2015). Using an IFN- β
290 promoter assay, we showed that N has a role in suppressing IFN signaling pathway when
291 stimulated by Sendai virus (SeV) infection (**Figure 5G**). N_{WT} can inhibit IFN- β promoter
292 activity, although not as well as M \ddot{e} ngl \grave{a} virus (MLAV) VP35, a potent inhibitor of IFN
293 signaling (Williams et al., 2020). Both N_{Nam-NTD-LKR} and N_{NTD} show modest inhibition at
294 the highest concentration tested. However, N_{CTD-Carm} shows similar levels of inhibition as
295 N_{WT}, and N_{CTD} displays the highest inhibition even at lower protein concentrations. In

296 summary, N is a potent inhibitor of IFN signaling and the CTD domain appears to be the
297 region critical for mediating this function.

298

299 **Discussion**

300 SARS-CoV-2 N protein is a core viral protein produced by the subgenomic RNA,
301 positioned proximal to the 3' end of genome, display high transcription levels, and is in
302 high abundance in virions. N is prone to forming higher-order oligomers that have a role
303 in binding to the RNA genome of SARS-CoV-2 and nucleocapsid formation. N is also
304 regulated by post-translational modifications, including phosphorylation, which changes
305 the physiochemical properties of N and likely directs its multiple roles at different stages
306 of the viral replication cycle.

307

308 Here we used a series of N constructs to dissect how each domain contributes to
309 oligomerization and RNA binding, and how phosphorylation can modulate these
310 properties. N oligomerization and RNA binding are likely linked; N oligomerization
311 provides a platform for high affinity RNA interactions whereas genomic RNA serves as a
312 string connecting N proteins. The modular domains of N provide multiple regulatory layers
313 for genomic access. For example, phosphorylation of N in the LKR region reduces RNA
314 binding to N and modify distribution of N-RNA species.

315

316 We also gained insight into the antigenicity of individual domains of N and its potential
317 utility in serological studies. We data revealed that N_{CTD} acts more specifically in detecting
318 infection of SARS-CoV-2, the causative agent of COVID-19, from plasma in comparison
319 to N_{WT}. Consistent with our observation, the N_{CTD} region is also predicted to encompass
320 major antigenic sites of N (Bussmann et al., 2006; Liang et al., 2005). Owing to the relative
321 conservation of N within coronaviruses, it is crucial to understand how N-directed
322 antibodies generated by different coronaviruses are cross-reactive with those that are
323 derived when exposed to SARS-CoV-2. The common cold coronavirus protective
324 immunity is short-lasting (Edridge et al., 2020); reinfection with the same seasonal
325 coronavirus occurred frequently at 12 months after the initial infection using an ELISA
326 approach against the N C-terminal region (coronavirus NL63, 215-377 aa; 229E, 213-389
327 aa; OC43, 328-428 aa; HKU1, 326-441 aa), which is similar to the N_{CTD} region we used
328 (SARS-CoV-2, 248-369 aa). Interestingly, 2 out of 10 individuals assessed in this study
329 of longitudinal donors unexposed to SARS-CoV-2 by Edridge *et al.*, produced broadly
330 reactive antibodies towards SARS-CoV-2 N_{WT}. The possibility of broadly reactive
331 antibodies in unexposed individuals highlights the need for domain specific serology, such
332 as our use of the N_{CTD} for increased sensitivity to discriminate COVID-19 cases, while
333 reducing the false-positive rate from cross reactive antibodies generated by infections of
334 the common cold coronaviruses.

335

336 In conclusion, we describe our efforts to characterize domain specific insights into
337 essential biochemical and serological properties of SARS-CoV-2 N. Our results advance
338 the understanding of viral replication processes and highlight the diagnostic value of using

339 N domains as a highly specific and sensitive markers of COVID-19. While our study
340 highlights the important functions associated with N_{CTD}, much remains to be characterized,
341 including the mechanistic link between the high immunogenicity of N_{CTD} and the physical
342 properties such as RNA binding and oligomerization.

343

344

345 **Methods**

346 **Patients and sample collection.** Our study enrolled a total of 67 patients with RT-PCR
347 confirmed COVID-19 infection: with 45 patients from St. Louis, MO, USA, and 23 patients
348 from Hong Kong, PRC. The negative samples (n = 28) used in this study were from St.
349 Louis, USA, and were obtained from patients following the start of the pandemic. Plasma
350 samples were obtained from patients at Barnes-Jewish Hospital (St. Louis, MO, USA)
351 and the Hong Kong Island West Cluster of Hospitals (Hong Kong, PRC). Both hospital
352 systems are urban, tertiary-care, academic medical centers. Positive and negative
353 patients from all cohorts were confirmed using standard of care, RT-PCR based methods.
354 The collection of patient plasma was approved by the Human Research Protection Office
355 at Washington University in St. Louis (IRB reference number 202007097) and the
356 Institutional Review Board of The Hong Kong University and the Hong Kong Island West
357 Cluster of Hospitals (IRB reference number UW20-169). Plasma samples were collected
358 from heparinized blood. Sample day was defined as days post-symptom onset.

359

360 **Enzyme-linked immunosorbent assay (ELISA).** ELISA assays were performed with
361 Nucleoprotein (N) proteins made in house, as described below. Briefly, recombinant N
362 proteins were coated on 96 well flatbottom immunosorbent plates (Nunc Immuno
363 MaxiSorp) at a concentration of 500 ng/mL, in 100 μ L coating buffer (PBS with 53%
364 Na_2CO_3 and 42% NaHCO_3 , pH 9.6) at 4°C overnight. An additional plate coated with a
365 non - specific protein (blocking buffer, PBS with 5% fetal bovine serum (FBS)) was used
366 to measure the background binding of each plasma sample. Following FBS blocking and
367 thorough washing, diluted plasma samples (1:100) were bound for 2 hours, further
368 washed and then detected by an anti - human IgG secondary antibody labelled with HRP
369 (Invitrogen), and absorbance detected at 450nm on a spectrophotometer (Wallac).

370

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

380

381 **Negative Staining EM.** 2 μ L of N sample at a concentration of 1 mg/mL was applied to
382 a glow-discharged copper grid (Ted Pella), washed twice with water before staining with

383 2% uranyl acetate for 30 s and air dried. Grids were imaged using a JEOL JEM-1400plus
384 Transmission Electron Microscope operating at 120 kV and recorded with an AMT XR111
385 high-speed 4k x 2k pixel phosphor-scintillated 12-bit CCD camera.

386

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

391

392 **Fluorescence Polarization Assay (FPA).** FPA experiments were performed on a
393 Cytation5 plate reader (BioTek) operating on Gen5 software. Excitation and emission
394 wavelengths were set to 485 and 528 nm, respectively, with a bandpass of 20 nm. Read
395 height and G factor were set to 8.5 mm and 1.26, respectively using the autogain function.
396 For RNA binding experiments, fluorescein isothiocyanate (FITC) labelled 20 nt-ssRNA or
397 19 nt siRNA at a final concentration of 1 nM was loaded on N samples (in 20 mM HEPES
398 (pH 7.5), 150 mM NaCl, 2 mM TCEP, 5% glycerol) at concentrations ranging from 0.4 nM
399 to 10 μ M in a 96-well plate. After 10 min of incubation, fluorescence polarization signals
400 were read. The fluorescence polarization values were then plotted against N
401 concentrations to fit the dissociation constant, K_D , using ORIGIN software. For anisotropy
402 plots, anisotropy values were converted from polarization according to previous research
403 (Kozlov et al., 2012).

404

405 **LC-MS Analysis.** Unless otherwise indicated, all chemical reagents were sourced from
406 Millipore Sigma and used without further purification. For LC-MS analyses, 30 pmol of

407 protein in 50 μ L of 1:1 solvent mixture of acetonitrile:water with 0.1% formic acid
408 (CovaChem) was loaded onto a C8 trap (ZORBAX Eclipse XDB C8 column, 2.1 x 15 mm,
409 Agilent), desalted for 3 min by using water/0.1% formic acid at a flow rate of 100 μ L/min,
410 and eluted using an 14 minute gradient from 0 to 80% acetonitrile/0.1% formic acid at a
411 flow rate of 100 μ L/min. Samples were analyzed using a MaXis 4G Q-TOF (Bruker
412 Daltonics). The mass spectrum was extracted guided by the elution peak and submitted
413 to PMI Intact Mass and searched for M values ranging from 5-50 kDa.

414

415 **HDX-MS.** N_{NTD}-LKR S176D/S188D/S206D was incubated with a 20-nt ssRNA at a 1:1 ratio. After
416 incubation, 2 μ L of 50 μ M protein/protein-RNA in PBS (pH 7.4) was diluted 10-fold (v/v)
417 with labeling buffer (PBS in D₂O, pD 7.0) (D₂O from Cambridge Isotope Laboratories),
418 incubated for 10, 30, 300, and 3600 s on ice, quenched by using a 60% dilution with 3 M
419 urea, PBS (pH 2.5), and flash frozen for later LC-MS analysis. A 0 s control was prepared
420 with PBS in H₂O. Prior to incubation, each 50 μ L of 2 μ M sample was thawed for 1 min at
421 37 °C before injection into a custom-built liquid chromatography (LC) apparatus for LC-
422 MS analysis. The labelled protein passed through two in-house packed protease columns
423 (2 mm x 20 mm), coupled so that the first using protease from *Aspergillus saitoi* type XIII
424 (FXIII) and the second porcine pepsin (0.1% formic acid, flow rate 200 μ L/min); the
425 resulting peptides were trapped on a ZORBAX Eclipse XDB C8 column (2.1 mm x 15 mm,
426 Agilent), desalted for 3 min, and then separated on a Hypersil Gold C18 column (2.1 x
427 50 mm, Thermo Fisher) with a 10.5 min linear gradient from 4 – 40% acetonitrile/0.1%
428 formic acid (flow rate 100 μ L/min). All valves, tubes, and columns (except for the protease
429 columns, which lose activity at low temperature) were submerged in ice during the

430 experiment to minimize back exchange. Peptides were eluted into a Bruker Maxis HM Q-
431 TOF MS for mass analysis. Experiments were in duplicate unless otherwise indicated.
432 The HDX data processing was performed by using HDEaminer (version 2.5.1, Sierra
433 Analytics, Inc.).

434

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

445

446 **Acknowledgements**

447 We thank Dr. N. Krogan (UCSF) for sharing SARS-CoV-2 plasmids, Drs. A. Holehouse
448 and A. Soranno (Washington University School of Medicine) for providing critical
449 feedback, and R. Ridings for coordinating studies between WUSM and external groups.
450 Research was supported by Fast Grant ##2161 (Emergent Ventures) to G.K.A. and NIH
451 grants (P01AI120943, R01AI123926 to C.F.B, G.K.A. and C.F.B; R01AI107056 to D.W.L.;
452 P41GM103422 and R24GM136766 to M.L.G.; P01AI120943, R01AI143292
453 and R01AI148663 C.F.B). S.A.V was supported by COVID190115 and COVID190126

454 Health and Medical Research Fund, Food and Health Bureau, Hong Kong and NIH/NIAID
455 CEIRS contract HHSN272201400006C. A.J.Q. is supported by a NIH T32 training grant
456 (T32CA009547).

457

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

461

462 **Figures**

463 **Figure 1. Characterization of N oligomerization with dynamic light scattering. A.**

464 Domain architecture of N. N has two structural domains: NTD and CTD. The sequence
465 between NTD and CTD is a linker region (LKR) containing a serine-arginine rich motif.
466 The Narm, LKR, and Carm are predicted to be disordered based upon sequence analysis.

467 **B.** Sequence identity between N of SARS-CoV-2 and common cold coronaviruses and
468 other epidemic severe coronaviruses MERS-CoV and SARS-CoV. **C.** Measurements of

469 N oligomerization with dynamic light scattering. Measured hydrodynamic radii, R_h , are
470 reported in D. **D.** Table summarizes the DLS data for all constructs. Numbers are reported
471 as average and standard deviation of three experiments.

472

473 **Figure 2. SARS-CoV-2 nucleocapsid protein binds ssRNA with high affinity. A.**

474 Principles of fluorescence polarization assay to measure RNA binding. Increasing

475 concentrations of N was titrated into 1nM of labeled RNA. Protein binding to labeled RNA
476 leads to slower tumbling of labeled RNA, resulting in increased fluorescence polarization.
477 **B.** Fluorescence polarization binding curves of N constructs to a 20-nt ssRNA. The fitted
478 K_D values are $0.007 \pm 0.001 \mu\text{M}$ (N_{WT} , black square), $16 \pm 12 \mu\text{M}$ (N_{NTD} , red circle), $13 \pm$
479 $5 \mu\text{M}$ (N_{CTD} , blue up triangle), $0.006 \pm 0.002 \mu\text{M}$ ($N_{NTD-LKR-CTD}$, magenta down triangle),
480 and $0.006 \pm 0.002 \mu\text{M}$ ($N_{NTD-LKR-CTD-Carm}$, green diamond). **C.** Fluorescence polarization
481 binding curves of N constructs to a 20-nt ssRNA. The fitted K_D values are 0.006 ± 0.002
482 nM ($N_{NTD-LKR-CTD}$, magenta down triangle), $13 \pm 5 \mu\text{M}$ (N_{CTD} , blue circle), $0.50 \pm 0.08 \mu\text{M}$
483 ($N_{NTD-LKR}$, orange star), and $0.44 \pm 0.04 \mu\text{M}$ ($N_{NTD-LKR} + N_{CTD}$, purple pentagon). **D.**
484 Fluorescence polarization binding curves of N constructs to a 20-nt ssRNA. The fitted K_D
485 values are $0.006 \pm 0.002 \mu\text{M}$ ($N_{NTD-LKR-CTD}$, magenta down triangle), $16 \pm 12 \mu\text{M}$ (N_{NTD} ,
486 red circle), $0.35 \pm 0.04 \mu\text{M}$ ($N_{LKR-CTD}$, orange up triangle), and $0.72 \pm 0.09 \mu\text{M}$ ($N_{NTD} +$
487 $N_{LKR-CTD}$, purple down triangle). **E.** Fluorescence polarization binding curves of N
488 constructs to a siRNA. The fitted K_D values are $0.051 \pm 0.004 \mu\text{M}$ (N_{WT} , black square),
489 $124 \pm 84 \mu\text{M}$ (N_{NTD} , red circle), $65 \pm 44 \mu\text{M}$ (N_{CTD} , blue up triangle), $2.5 \pm 0.5 \mu\text{M}$ (N_{NTD-}
490 LKR , magenta down triangle), $0.22 \pm 0.02 \mu\text{M}$ ($N_{NTD-LKR-CTD}$, green diamond), and $0.10 \pm$
491 $0.01 \mu\text{M}$ ($N_{NTD-LKR-CTD-Carm}$, navy left triangle). **F.** Table summarizes K_D values (μM) for key
492 constructs binding to ssRNA and siRNA. Numbers are reported as average and standard
493 deviation of two experiments.

494

495 **Figure 3. N-RNA phase separates and phosphorylation modulates N-RNA. A.** Size
496 exclusion chromatography of N constructs (N_{WT} , black; $N_{NTD-LKR-CTD-Carm}$, red; $N_{NTD-LKR-CTD}$,
497 blue) in 25 mM HEPES, 500 mM NaCl, 2 mM TCEP, 5% glycerol from a 2 L pellet.

498 Samples from peak 1 (p1) and p2 contain RNA whereas p3 are RNA-free based upon
499 absorbance from the 260/280 ratio. **B.** Negative stain electron microscopy (EM) image of
500 p1 and p2 for N_{WT}, N_{NTD-LKR-CTD-Carm}, and N_{NTD-LKR-CTD} in 150 mM NaCl. Samples were
501 diluted into 150 mM NaCl before negative-staining fixation by uranyl acetate. **C.** Size
502 exclusion chromatography of N constructs (N_{WT}, black; N_{S188D/S206D}, blue;
503 N_{S176D/S188D/S206D}, red) in 25 mM HEPES, 500 mM NaCl, 2 mM TCEP, 5% glycerol. **D.**
504 Negative stain electron microscope image of N_{S188D/S206D} and N_{S176D/S188D/S206D} in 150 mM
505 NaCl. **E.** Fluorescence polarization binding curves of N mutants to a 20-nt ssRNA. The
506 fitted K_D values are $0.007 \pm 0.001 \mu\text{M}$ (N_{WT}, black square), $0.015 \pm 0.002 \mu\text{M}$
507 (N_{S176D/S188D/S206D}, black circle), $0.505 \pm 0.075 \mu\text{M}$ (N_{NTD-LKR}, orange up triangle), and 1.1
508 $\pm 0.2 \mu\text{M}$ (N_{NTD-LKR S176D/S188D/S206D}, orange down triangle). Numbers are reported as
509 average and standard deviation of two experiments.

510

511 **Figure 4. HDX-MS mapping of RNA binding to N_{NTD-LKR S176D/S188D/S206D}.** **A.** Woods' plot
512 showing cumulative differential HDX and validating differences using global significance
513 limits. The horizontal bars depict the cumulative HDX differences between the RNA-
514 bound and unbound N_{NTD-LKR S176D/S188D/S206D}. Standard deviations are shown for each
515 peptide. Peptides showing statistically significant differences are differentiated by global
516 significance limit (*, $p < 0.1$; **, $p < 0.05$; ***, $p < 0.01$). The blue shade of the peptide bar
517 indicates differing statistical significance (light blue, medium blue, and navy, respectively);
518 gray peptide bars depict peptides where statistically significant differences in HDX were
519 not observed. Vertical bars show previously reported binding sites (residues reported for
520 RNA-binding CoV2 N-protein (Dinesh et al., 2020; Ye et al., 2020), AMP-binding HCoV-

521 OC43 (Lin et al., 2014; Ye et al., 2020), and for both are shown in red, yellow, and orange,
522 respectively). Secondary structure (PDB 6M3M) is shown above. **B.** Representative
523 kinetic plots showing peptide level HDX as a function of exchange time (unbound, black;
524 bound to RNA, red). **C.** Sites of protection measured by HDX mapped on the N_{NTD}
525 structure (PDB 6M3M). Statistically significant HDX protection, regions of no difference in
526 HDX, and regions where lacking proteolytic coverage results in no data are shown in teal,
527 light gray, and dark gray, respectively. Those residues unresolved in the structure are
528 shown as a dashed line, with the exception of those reporting a statistically significant
529 difference in teal. **D.** Electrostatic potential calculated with APBS mapped on to the N_{NTD}
530 structure (PDB 6M3M) shows a major positive charge groove. Red and blue represent
531 negative and positive electrostatic potential. The color scale is in kT/e units.

532

533 **Figure 5. The CTD of N is a highly sensitive serological marker. A.** ELISA data of
534 N_{WT} screened against plasma of COVID-19 positive and negative individuals from a
535 combined Hong Kong, PRC and St. Louis, MO, USA cohort. Black solid line indicates the
536 mean OD₄₅₀ value for each population. **** p < 0.0001. **B.** ELISAs with the various N
537 constructs for patient IgG. ELISAs were performed on plasma samples from COVID-19
538 patients (n = 68) and negative controls (n = 28). The cut-off is represented by the dotted
539 line and calculated as the mean + 3 standard deviations of the negative population. Mean
540 values ± standard deviation of COVID-19 and negative groups are shown. **C.** Sensitivity
541 and specificity for each of the N domains calculated from the ELISA results. **D.** Heat-map
542 of ELISA results for N_{NTD}-LKR-CTD-C_{arm}, N_{CTD}, and N_{NTD} constructs from COVID-19 samples
543 (n = 67). Each column represents an individual sample. **E.** Maturation of the N_{CTD} and

544 N_{NTD-LKR-CTD-Carm} IgG response over time (n = 67). **F.** Ratio of OD₄₅₀ for N_{CTD} and N_{NTD-LKR-}
545 CTD-Carm for acute and convalescent time-points. Mean values ± standard deviation of
546 acute and convalescent COVID-19 samples are shown. Experiments were repeated twice.
547 Statistical significance was calculated by unpaired Student's t-test, ****p < 0.0001. **G.**
548 Inhibition of SeV-induced IFNβ promoter activation by N constructs. Fold changes are
549 relative to vector-only (V) transfections without SeV infection. MLAV VP35 served as a
550 positive control for inhibition. Three transfection concentrations were used: 1.25, 12.5,
551 and 125 ng/well. Statistical significance was determined by performing a one-way ANOVA
552 followed with Tukey multiple comparison as compared to Sendai virus-infected control;
553 **** p < 0.0001, *** p < 0.0002, ** p < 0.0021, * p < 0.0332.

554

555 **Supplementary Figure Legends**

556 **Supplementary Figure 1, related to Figure 1. Multiple sequence alignment of**
557 **coronavirus nucleocapsids. A.** Multiple sequence alignment of coronavirus
558 nucleocapsids. Sequences were aligned using Clustal Omega. Accession numbers used
559 are 229E (APT69891.1), NL63 (YP_003771.1), HK1 (AAT98585.1), OC43 (AAR01019.1),
560 MERS (AKL80590.1), SARS (AAP30037.1), SARS2 (YP_009724397.2). Alignments
561 were analyzed using ESPript3. The three serines (176, 188, and 206) are labeled with
562 red arrows. **B.** Sequence identity between SARS-CoV-2 N and that of common cold
563 coronaviruses and MERS and SARS. FL, full length. All units are in %. Percent identity
564 matrixes for corresponding domains of N are generated using Clustal2.1.

565

566 **Supplementary Figure 2, related to Figure 1. Denaturing mass spectra of N protein**
567 **truncations N_{NTD}, N_{CTD}, N_{NTD-LKR}, and N_{NTD-LKR-CTD}. A.** DLS polydispersity table for N
568 constructs. Higher values indicate broader size distributions. Numbers are reported as
569 average and standard deviation of three experiments. Deconvolution yields experimental
570 M values of 16,881 Da, 16,078 Da, 24,155 Da, and 37,829 Da for **(B)** N_{NTD}, **(C)** N_{CTD}, **(D)**
571 N_{NTD-LKR}, and **(E)** N_{NTD-LKR-CTD} respectively, matching theoretical values within 1 Da, based
572 on protein sequence. Deconvoluted mass spectra (right) and adduct series corresponding
573 to pervasive trifluoroacetic adducts (delta mass 114 Da, circle) and α-N-gluconoylation
574 (delta mass 178 Da, star). TFA adducts are introduced by the ion pairing reagent in
575 solvent, while α-N-gluconoylation is a common modification occurring on His-tagged
576 proteins. Native spray of NTD (not pictured) yielded no peaks with delta mass 114 Da,
577 but retained a single delta mass 178 Da, confirming transient TFA adducts are an artifact
578 of the denaturing experiment, but the α-N-gluconoylation of the His-tag is covalent.

579

580 **Supplementary Figure 3, related to Figure 2. Nucleocapsid binds stem-loop RNA**
581 **with reduced affinity. A.** Fluorescence anisotropy binding curves of N constructs to a
582 20-nt ssRNA. Anisotropy values were converted from polarization according to previous
583 research (Kozlov et al., 2012). The fitted K_D values are 0.007 ± 0.001 μM (N_{WT}, black
584 square), 0.006 ± 0.002 μM (N_{NTD-LKR-CTD}, magenta circle), 14 ± 5 μM (N_{CTD}, blue up
585 triangle) and 18 ± 14 μM (N_{NTD}, red down triangle). These values are very close to those
586 of polarization. In this system, binding monitored by anisotropy is similar to that of
587 polarization. **B.** Fitted K_D values for N constructs binding to ssRNA (black) and siRNA
588 (grey). **C.** Ratio of K_D of siRNA over that of ssRNA for N constructs. The reduced binding

589 to siRNA is around 5-fold for most N constructs. The reduction is higher for those of N_{NTD}-
590 LKR-CTD-Carm and N_{NTD}-LKR-CTD, suggesting Narm and Carm are more involved in siRNA
591 binding.

592

593 **Supplementary Figure 4, related to Figure 3. Phosphorylation mimetics of N reduce**
594 **RNA binding. A.** Fluorescence polarization binding curves of N constructs to a 20-nt
595 ssRNA. The fitted K_D values are $0.007 \pm 0.001 \mu\text{M}$ for N_{WT}, $0.015 \pm 0.002 \mu\text{M}$ for
596 N_{S188D/S206D}, and $0.023 \pm 0.006 \mu\text{M}$ for N_{S176D/S188D/S188D}. **B.** Fluorescence polarization
597 binding curves of N constructs to a 19-nt siRNA. The fitted K_D values are $1.3 \pm 0.3 \mu\text{M}$
598 (N_{LKR-CTD}, black square), $3.0 \pm 0.5 \mu\text{M}$ (N_{NTD-LKR}, red circle), and $2.9 \pm 1.4 \mu\text{M}$ (N_{NTD-LKR}
599 S_{176D/S188D/S206D}, blue up triangle).

600

601 **Supplementary Figure 5, related to Figure 4. Sequence coverage of N_{NTD-LKR}**
602 **S_{176D/S188D/S206D} in HDX-MS and HDX of the unbound state. A.** Protein coverage map of
603 unbound state N_{NTD-LKR} S_{176D/S188D/S206D} HDX yielding 152 peptides with 93.3% sequence
604 coverage. Peptide bars are colored according to their average %HDX relative to the color
605 bar, where cooler colors depict low average %HDX and warmer colors depict high
606 average %HDX. The secondary structure reported by PDB 6M3M is shown above the
607 sequence. Overall, the HDX of the unbound state is largely consistent with the reported
608 secondary structure and a well-ordered tertiary structure; regions outside of the reported
609 structure undergo relatively rapid HDX, consistent with a lack of backbone hydrogen
610 bonding. Interestingly, despite a lack of reported secondary structure in the region of 155-

611 160, relatively low HDX was observed, consistent with either hydrogen bonding of
612 secondary/tertiary structure or a hydrophobic pocket. SR-motif in LKR are boxed in red.
613 **B.** All kinetic plots used in the peptide-level difference plot in **Figure 5A** show peptide
614 level HDX as a function of exchange time (unbound, black; bound to RNA, red).

615

616

617

618

619

620

621

622

623 **References**

- 624 Arragain, B., Reguera, J., Desfosses, A., Gutsche, I., Schoehn, G., and Malet, H. (2019).
625 High resolution cryo-EM structure of the helical RNA-bound Hantaan virus nucleocapsid
626 reveals its assembly mechanisms. *eLife* 8, e43075.
- 627 Bar-On, Y.M., Flamholz, A., Phillips, R., and Milo, R. (2020). SARS-CoV-2 (COVID-19)
628 by the numbers. *eLife* 9, e57309.
- 629 Bharat, T.A.M., Noda, T., Riches, J.D., Kraehling, V., Kolesnikova, L., Becker, S.,
630 Kawaoka, Y., and Briggs, J.A.G. (2012). Structural dissection of Ebola virus and its
631 assembly determinants using cryo-electron tomography. *Proceedings of the National*
632 *Academy of Sciences* 109, 4275-4280.
- 633 Bouhaddou, M., Memon, D., Meyer, B., White, K.M., Rezelj, V.V., Correa Marrero, M.,
634 Polacco, B.J., Melnyk, J.E., Ulferts, S., Kaake, R.M., *et al.* (2020). The Global
635 Phosphorylation Landscape of SARS-CoV-2 Infection. *Cell* 182, 685-712.e619.
- 636 Bussmann, B.M., Reiche, S., Jacob, L.H., Braun, J.M., and Jassoy, C. (2006). Antigenic
637 and cellular localisation analysis of the severe acute respiratory syndrome coronavirus
638 nucleocapsid protein using monoclonal antibodies. *Virus Res* 122, 119-126.
- 639 Carlson, C.R., Asfaha, J.B., Ghent, C.M., Howard, C.J., Hartooni, N., Safari, M., Frankel,
640 A.D., and Morgan, D.O. (2020). Phosphoregulation of phase separation by the SARS-
641 CoV-2 N protein suggests a biophysical basis for its dual functions. *Molecular Cell*.
- 642 Chang, C.-k., Hou, M.-H., Chang, C.-F., Hsiao, C.-D., and Huang, T.-h. (2014). The SARS
643 coronavirus nucleocapsid protein – Forms and functions. *Antiviral Research* 103, 39-50.
- 644 Chew, K.L., Tan, S.S., Saw, S., Pajarillaga, A., Zaine, S., Khoo, C., Wang, W., Tambyah,
645 P., Jureen, R., and Sethi, S.K. (2020). Clinical evaluation of serological IgG antibody

646 response on the Abbott Architect for established SARS-CoV-2 infection. *Clinical*
647 *Microbiology and Infection* 26, 1256.e1259-1256.e1211.

648 Cubuk, J., Alston, J., Incicco, J., Singh, S., Stuchell-Brereton, M., Ward, M., Zimmerman,
649 M., Vithani, N., Griffith, D., Wagoner, J., *et al.* (2020). The SARS-CoV-2 nucleocapsid
650 protein is dynamic, disordered, and phase separates with RNA (bioRxiv).

651 Dinesh, D.C., Chalupska, D., Silhan, J., Veverka, V., and Boura, E. (2020). Structural
652 basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. *bioRxiv*,
653 2020.2004.2002.022194.

654 Ding, B., Qin, Y., and Chen, M. (2016). Nucleocapsid proteins: roles beyond viral RNA
655 packaging. *Wiley Interdiscip Rev RNA* 7, 213-226.

656 Edridge, A.W.D., Kaczorowska, J., Hoste, A.C.R., Bakker, M., Klein, M., Loens, K.,
657 Jebbink, M.F., Matser, A., Kinsella, C.M., Rueda, P., *et al.* (2020). Seasonal coronavirus
658 protective immunity is short-lasting. *Nature Medicine*.

659 Fung, T.S., and Liu, D.X. (2019). Human Coronavirus: Host-Pathogen Interaction. *Annual*
660 *Review of Microbiology* 73, 529-557.

661 Grosseohme, N.E., Li, L., Keane, S.C., Liu, P., Dann, C.E., Leibowitz, J.L., and Giedroc,
662 D.P. (2009). Coronavirus N Protein N-Terminal Domain (NTD) Specifically Binds the
663 Transcriptional Regulatory Sequence (TRS) and Melts TRS-cTRS RNA Duplexes.
664 *Journal of Molecular Biology* 394, 544-557.

665 Gui, M., Liu, X., Guo, D., Zhang, Z., Yin, C.-C., Chen, Y., and Xiang, Y. (2017). Electron
666 microscopy studies of the coronavirus ribonucleoprotein complex. *Protein & Cell* 8, 219-
667 224.

668 Hachim, A., Kavian, N., Cohen, C.A., Chin, A.W.H., Chu, D.K.W., Mok, C.K.P., Tsang,
669 O.T.Y., Yeung, Y.C., Perera, R.A.P.M., Poon, L.L.M., *et al.* (2020). ORF8 and ORF3b
670 antibodies are accurate serological markers of early and late SARS-CoV-2 infection.
671 *Nature Immunology* 21, 1293-1301.

672 Iserman, C., Roden, C., Boerneke, M., Sealfon, R., McLaughlin, G., Jungreis, I., Park, C.,
673 Boppana, A., Fritch, E., Hou, Y.J., *et al.* (2020). Specific viral RNA drives the SARS CoV-
674 2 nucleocapsid to phase separate. *bioRxiv*, 2020.2006.2011.147199.

675 Jack, A., Ferro, L.S., Trnka, M.J., Wehri, E., Nadgir, A., Costa, K., Schaletzky, J., and
676 Yildiz, A. (2020). SARS CoV-2 nucleocapsid protein forms condensates with viral
677 genomic RNA. *bioRxiv*, 2020.2009.2014.295824.

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

682 Keane, S.C., Liu, P., Leibowitz, J.L., and Giedroc, D.P. (2012). Functional Transcriptional
683 Regulatory Sequence (TRS) RNA Binding and Helix Destabilizing Determinants of Murine
684 Hepatitis Virus (MHV) Nucleocapsid (N) Protein. *Journal of Biological Chemistry* 287,
685 7063-7073.

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

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

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

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

698 Lin, S.-Y., Liu, C.-L., Chang, Y.-M., Zhao, J., Perlman, S., and Hou, M.-H. (2014).
699 Structural Basis for the Identification of the N-Terminal Domain of Coronavirus
700 Nucleocapsid Protein as an Antiviral Target. *Journal of Medicinal Chemistry* 57, 2247-
701 2257.

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

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

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

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

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

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

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

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

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

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

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

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

734 Takeda, M., Chang, C.-k., Ikeya, T., Güntert, P., Chang, Y.-h., Hsu, Y.-l., Huang, T.-h.,
735 and Kainosho, M. (2008). Solution Structure of the C-terminal Dimerization Domain of

736 SARS Coronavirus Nucleocapsid Protein Solved by the SAIL-NMR Method. *Journal of*
737 *Molecular Biology* 380, 608-622.

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

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

744 Wan, W., Kolesnikova, L., Clarke, M., Koehler, A., Noda, T., Becker, S., and Briggs, J.A.G.
745 (2017). Structure and assembly of the Ebola virus nucleocapsid. *Nature* 551, 394-397.

746 Williams, C.G., Gibbons, J.S., Keiffer, T.R., Luthra, P., Edwards, M.R., and Basler, C.F.
747 (2020). Impact of Měnglà Virus Proteins on Human and Bat Innate Immune Pathways.
748 *Journal of Virology* 94, e00191-00120.

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

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

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

758

Figure 1. Characterization of N oligomerization with dynamic light scattering

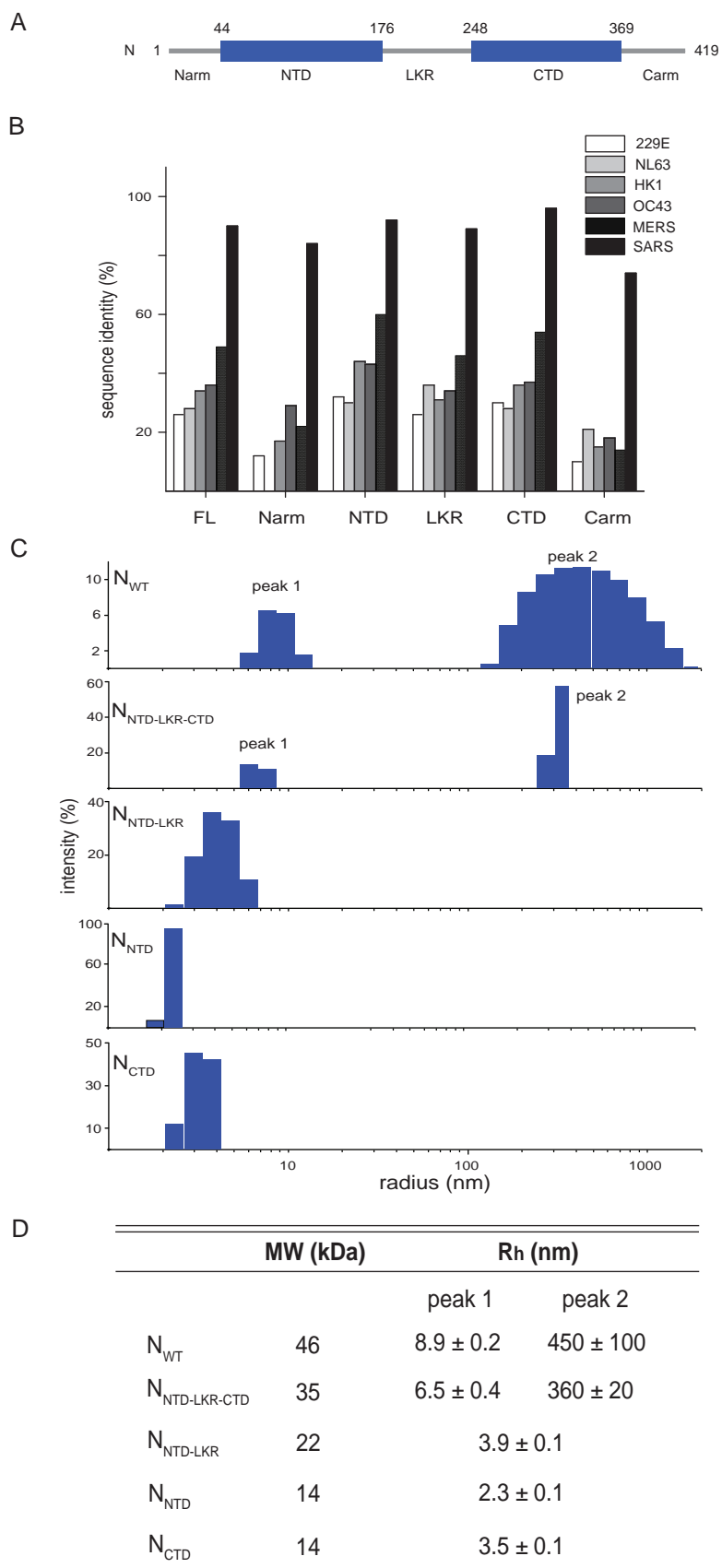


Figure 3 | N-RNA phase separates and phosphorylation modulates N-RNA

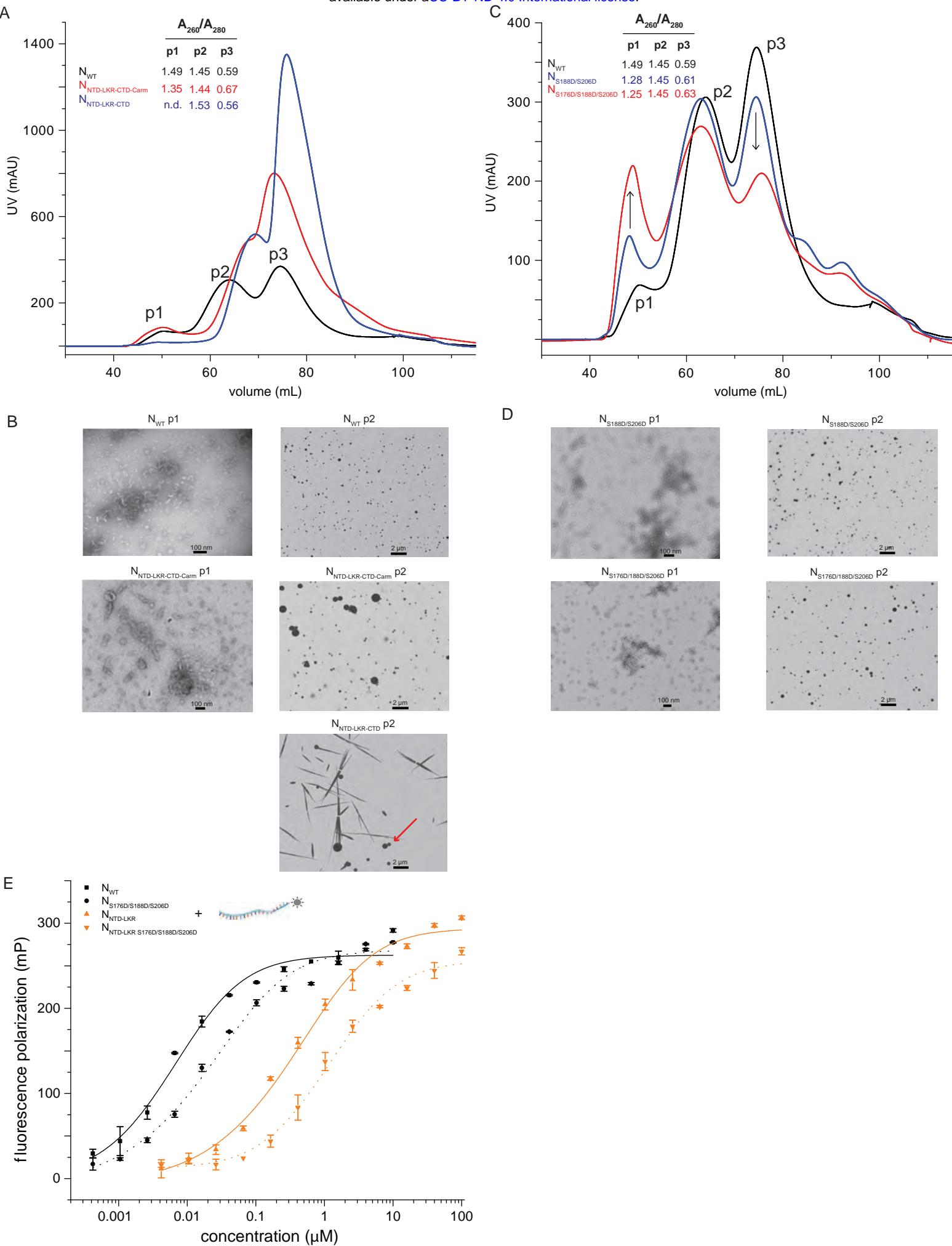


Figure 4. HDX-MS mapping of RNA binding interface within N

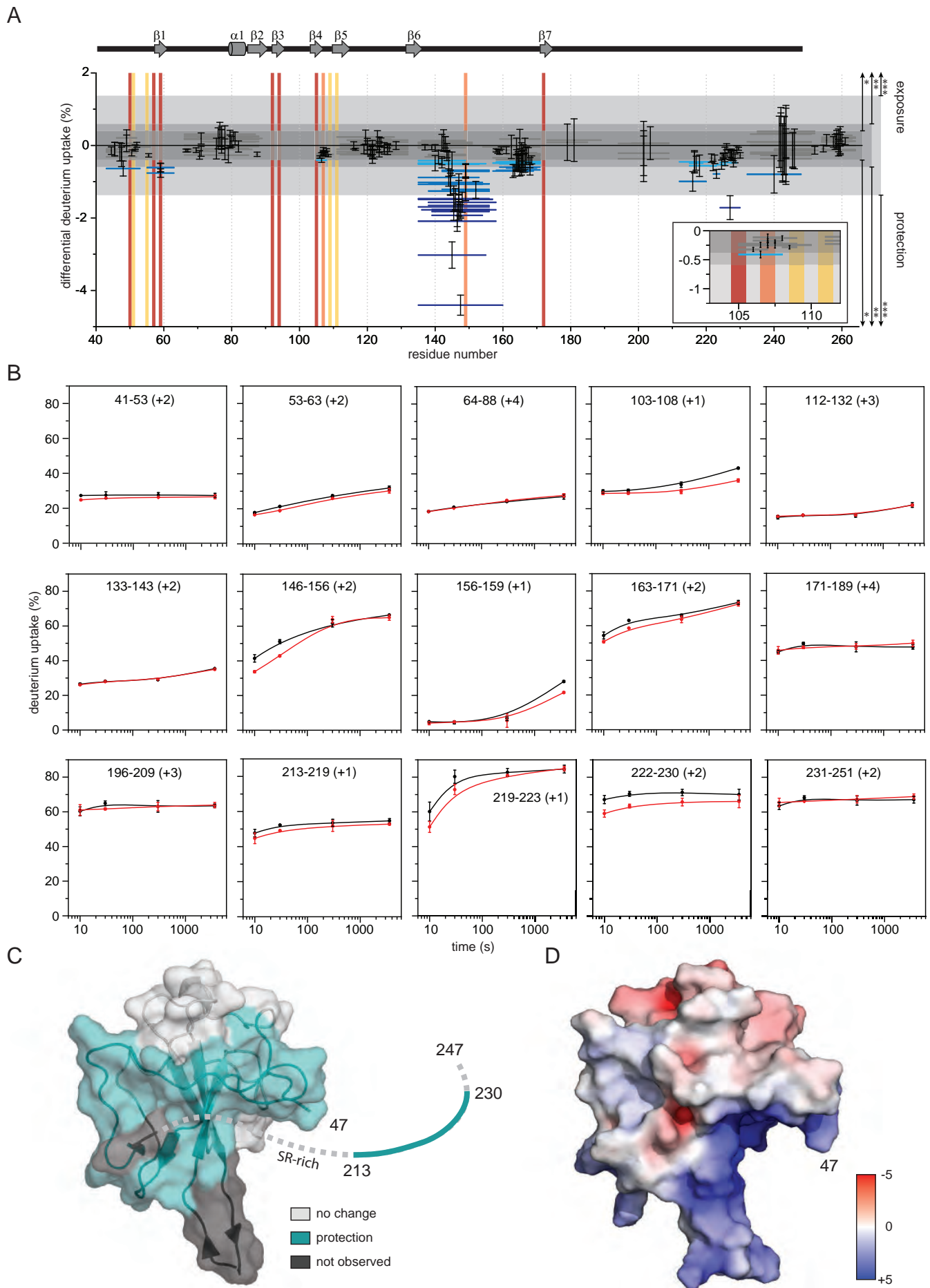
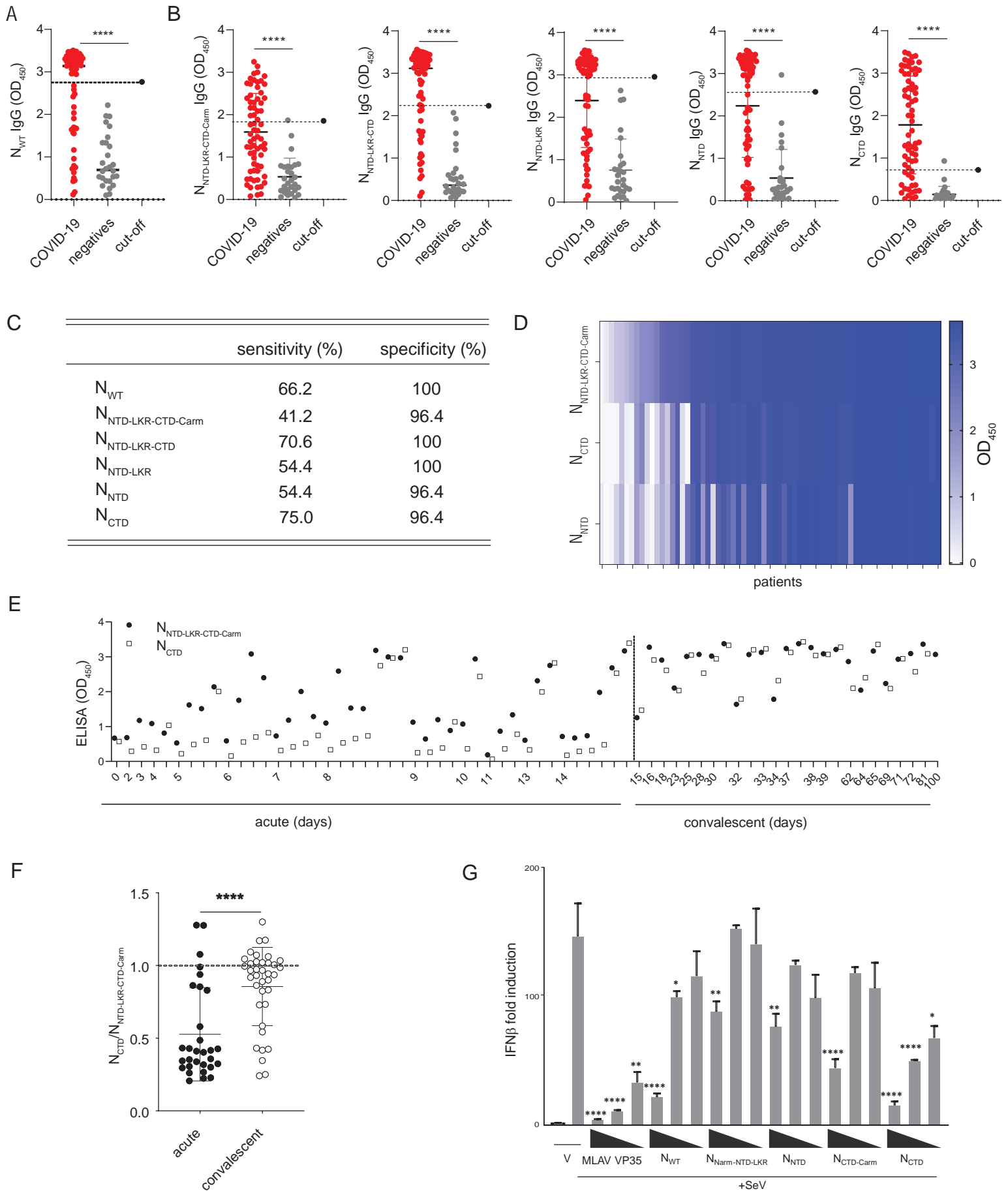


Figure 5. The CTD of N is a highly sensitive serological marker



A



B

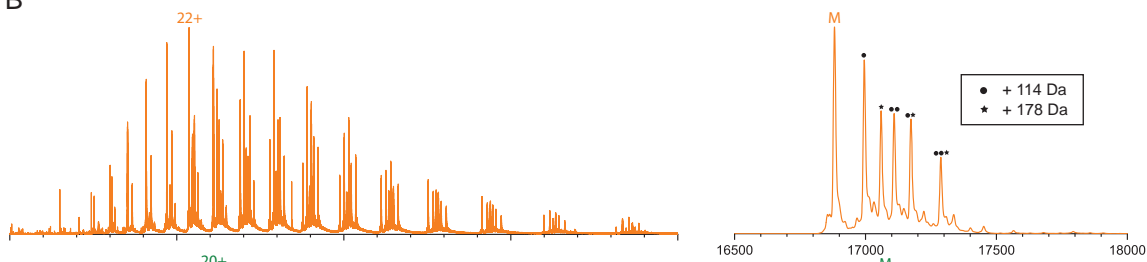
	229E	NL63	HK1	OC43	MERS	SARS
FL	26	28	34	36	49	90
Narm	12	0	17	29	22	84
NTD	32	30	44	43	60	92
LKR	26	36	31	34	46	89
CTD	30	28	36	37	54	96
Carm	10	21	15	18	14	74

Supplementary Figure 2, related to Figure 1. Determined mass spectra of N protein truncations.

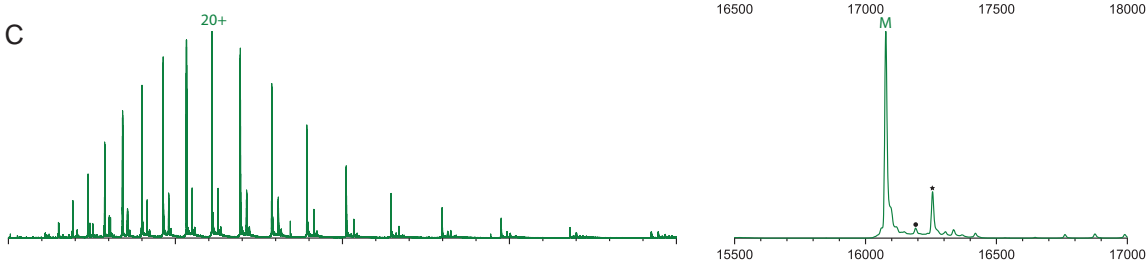
A

	MW (kDa)	polydispersity (%)	
		peak 1	peak 2
N_{WT}	46	20 ± 1	40 ± 20
$N_{NTD-LKR-CTD}$	35	17 ± 7	30 ± 20
$N_{NTD-LKR}$	22	23 ± 3	
N_{NTD}	14	7 ± 2	
N_{CTD}	14	15 ± 1	

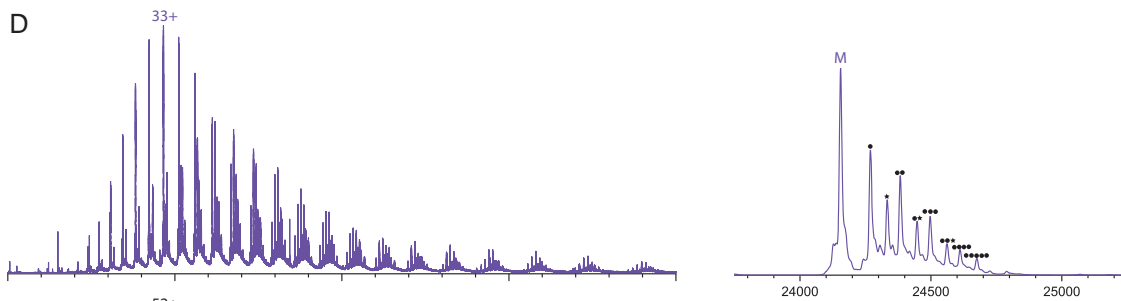
B



C



D



E

