Development of high affinity monobodies recognizing SARS-CoV-2 antigen

Yushen Du
Department of Molecular and Medical Pharmacology, University of California, Los Angeles, ZJU-UCLA Joint Center for Medical Education and Research, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Cancer Institute, The Second Affiliated Hospital, Zhejiang University School of Medicine

Tian-hao Zhang
Molecular Biology Institute, University of California, Los Angeles

Xiangzhi Meng
Department of Microbiology, Immunology, & Molecular Genetics, University of California, Los Angeles

Yuan Shi
Department of Molecular and Medical Pharmacology, University of California, Los Angeles

Menglong Hu
School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong

Shuofeng Yuan
State Key Laboratory of Emerging Infectious Diseases, Carol Yu Centre for Infection, Department of Microbiology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong

Chit Ying La
School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong

Shu-xing Li
Center of Excellence in NanoBiophysics, Department of Biological Sciences, University of Southern California, Los Angeles

Siwei Liu
State Key Laboratory of Emerging Infectious Diseases, Carol Yu Centre for Infection, Department of Microbiology, Li Ka Shing Faculty of Medicine, The University of Hong Kong

Jiayan Li
State Key Laboratory of Emerging Infectious Diseases, Carol Yu Centre for Infection, Department of Microbiology, Li Ka Shing Faculty of Medicine, The University of Hong Kong

Haigen Huang
Department of Molecular and Medical Pharmacology, University of California, Los Angeles

Hong Jian
Department of Molecular and Medical Pharmacology, University of California, Los Angeles

Dongdong Chen

ZJU-UCLA Joint Center for Medical Education and Research, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Cancer Institute, The Second Affiliated Hospital, Zhejiang University School of Medicine

Li Sheng

Department of Molecular and Medical Pharmacology, University of California, Los Angeles

Mengying Hong

Department of Molecular and Medical Pharmacology, University of California, Los Angeles, ZJU-UCLA Joint Center for Medical Education and Research, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Cancer Institute, The Second Affiliated Hospital, Zhejiang University School of Medicine

Anders Olson

Department of Molecular and Medical Pharmacology, University of California, Los Angeles

Hsiang-I Liao

Department of Molecular and Medical Pharmacology, University of California, Los Angeles

Xiaojiang Chen

Center of Excellence in NanoBiophysics, Department of Biological Sciences, University of Southern California, Los Angeles

Xinmin Li

Department of Pathology and Laboratory Medicine, David Geffen School of Medicine at UCLA

Gexin Zhao

Department of Pathology and Laboratory Medicine, David Geffen School of Medicine at UCLA

Richard W. Roberts

Department of Chemistry, Chemical Engineering and Biology, University of Southern California, Los Angeles,

Jasper F W Chan

State Key Laboratory of Emerging Infectious Diseases, Carol Yu Centre for Infection, Department of Microbiology, Li Ka Shing Faculty of Medicine, The University of Hong Kong

Dong-Yan Jin

School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong

Irvin SY Chen

Department of Microbiology, Immunology, & Molecular Genetics, University of California, Los Angeles

Honglin Chen

State Key Laboratory of Emerging Infectious Diseases, Carol Yu Centre for Infection, Department of Microbiology, Li Ka Shing Faculty of Medicine, The University of Hong Kong

Kwok-Yung Yuen

State Key Laboratory of Emerging Infectious Diseases, Carol Yu Centre for Infection, Department of Microbiology, Li Ka Shing Faculty of Medicine, The University of Hong Kong

Quan Hao

School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong

Ren Sun (rsun@mednet.ucla.edu)
Department of Molecular and Medical Pharmacology, University of California, Los Angeles and Molecular Biology Institute, University of California, Los Angeles

Research Article

Keywords: monobody development, SARS-CoV-2, nucleocapsid (N) protein, binding affinity, machine learning, ELISA

Posted Date: April 28th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-25828/v1

Abstract

The coronavirus disease 2019 (COVID-19) pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has been a threat to global public health. Prompt patient identification and quarantine is the most effective way to control its rapid transmission, which can be facilitated by early detection of viral antigens. Here we present a platform to develop and optimize the fibronectin-based affinity-enhanced antibody mimetics (monobodies) for recognizing viral antigens. Specifically, we developed monobodies targeting SARS-CoV-2 nucleocapsid (N) protein. We showed that two monobodies, NN2 and NC2, bind to N protein's N- and C-terminal domains respectively with a Kd in nM range.

The specificity of the recognition was confirmed with co-immunoprecipitation and immunofluorescence assays. Furthermore, we demonstrated that one round of *in vitro* maturation using mRNA display can improve the binding affinity of monobodies. Machine learning algorithms were integrated with deep sequencing data for selecting candidates that improve the detection sensitivity of N. Using this pair of monobodies, we have developed an enzyme-linked immunosorbent assay (ELISA) for viral detection. We were able to detect recombinant N at 4 pg/ml and detect N in viral culture supernatant, with no cross-reactivity with other CoV. Integrating high-dense mutagenesis, mRNA display, deep sequencing and machine learning, this platform can be applied through iterations to identify and optimize monobodies against emerging viral antigens, potentiating point-of-care detection of communicable diseases in a cost- and time-sensitive manner.

Authors Yushen Du, Tian-hao Zhang, Xiangzhi Meng, Yuan Shi, and Menglong Hu contributed equally to this work.

Introduction

Since December 2019, there has been an ongoing global outbreak of coronavirus disease 2019 (COVID-19), a highly contagious viral pneumonia causing fever and shortness of breath (Chan et al., 2020; Heymann and Shindo, 2020; Huang et al., 2020; Wang et al., 2020a; WHO, 2020). The disease has quickly spread to more than 100 countries and regions within 3 months, developing into a pandemic as announced by the World Health Organization (WHO) in early March 2020. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of COVID-19, has an estimated reproduction number (R0) of around 2-2.5 without effective control, higher than seasonal influenza virus (Kucharski et al., 2020; Liu et al., 2020; Zhao et al., 2020). To break the chain of infection, prompt diagnosis and quarantine of patients, especially at the early phase of infection, is crucial for reducing chances of community transmission (Wang et al., 2020b; Wilder-Smith and Freedman, 2020). Nevertheless, current diagnostic techniques have various limitations. (Huang et al., 2020; Kim et al., 2020; Thevarajan et al., 2020; Wang et al., 2020c; Wölfel et al., 2020; Xie et al., 2020), while serological tests are ineffective in detecting active and recent infections before antibody responses occur (Amanat et al., 2020; Pan et al.,

2020; Xia et al., 2020; Xiang et al., 2020). There is hence an urgent need to improve the development of diagnostic techniques to achieve high sensitivity and speed for effective disease control of COVID-19.

A sensitive viral antigen detection reagent is one way of advancing diagnostics for emerging diseases. An ideal reagent for antigen recognition is expected to fulfill several criteria, including (1) efficacy (high sensitivity and specificity), (2) scalability (production at low cost and high speed), (3) customizability (selection of desired properties at a defined condition *in vitro*) and (4) small size (compatibility with sensitive nano-platforms and *in vivo* imaging). To date, antibodies are the most widely used reagents in conventional assays for viral antigen detection, including ELISA, chemiluminescent enzyme immunoassay (CLIEA) and lateral flow immunoassay (LFIA) (Chevaliez and Pawlotsky, 2008; Young et al., 2000). Despite being effective, antibody-based detections have a couple of limitations. Firstly, it takes at least a few months to develop specific antibodies, which often involves immunizing animal hosts and selection of monoclones. Secondly, generation of antibodies is relatively high-cost and subject to batch-to-batch differences (Sajid et al., 2015). Thirdly, the application of antibodies is usually limited to close to physiological conditions due to the lack of resistance to high temperature or pH extremes. Lastly, the large size of common antibodies restricts their applications in nanobiosensors, such as field-effect transistor (FET) nanosensors, where signal readings are limited by the Debye screening length (Vu and Chen, 2019). Establishment of a sensitive and small recognition reagent that overcomes the above limitations of antibodies will be useful in facilitating population-wide screening of COVID- 19.

Previously, using mRNA display method, we developed a pair of fibronectin-based antibody mimetics (monobodies), NN1 and NC1 for detecting nucleocapsid (N) protein of SARS-CoV (SARS) with a K_d in sub-nM range (Liao et al., 2009). Evolved from the 10th fibronectin type III domain (10Fn3), these monobodies can be complementary to conventional antibodies and overcome some of the above limitations. They comprise of a beta sheet sandwich structure that is energy-favorable, and can be easily scaled up in *E. coli* for time-sensitive mass production at a low cost (Olson and Roberts, 2007). With a small size of 3 nm in diameter, they are also suitable to be used in conjunction with nanotechnology.

We demonstrated that the monobodies can be conjugated onto single welled carbon nanotube, and facilitate the large-scale parallel viral detection at a high speed (Ishikawa et al., 2009).

Thus, these monobodies may complement antibodies as detection reagents by having (1) high binding affinity and specificity, (2) low cost, and (3) small protein size.

Here we presented a platform that can rapidly and iteratively evolve the binding affinity of monobody against SARS-CoV-2 N. By mutagenizing the parental sequences, we evolved a pair of affinity-enhanced monobodies that could efficiently recognize SARS-CoV- 2 N *in vitro* and *in vivo*. The two monobodies, namely N-NTD-2 (NN2) and N-CTD-2 (NC2), bind to N-terminal domain (NTD) and C-terminal domain (CTD) of SARS-CoV-2 N respectively with a high affinity demonstrated with SPR and ITC. NN2 and NC2 can form a protein complex with N, and can specifically recognize endogenous N in viral infected cells. We also demonstrated the feasibility of using one-round *in vitro* maturation and machine learning

algorithms to further improve the binding affinity. Additionally, we demonstrated the potential diagnostic utility of monobodies with ELISA-based detection for N of SARS- CoV-2 with sensitivity at 4 pg/ml. The ELISA can also be used to detect N in viral supernatant with no cross-reactivity against other CoV. This platform can be generally applied to develop detection reagents for emerging viral antigens,in a time and cost-effective manner.

Results

Development of fibronectin-based monobodies against SARS-CoV-2-N

By randomizing BC and FG loops, the 10Fn3 has been developed as antibody mimetic recognize different targets (Koide et al., 1998). It is topologically analogous to the immunoglobulin VH domain, and expresses well in both eukaryotic and prokaryotic cells (Olson and Roberts, 2007). Previously using mRNA display of a randomized 10Fn3 library, we generated two monobodies for SARS N with a Kd in nM range (Liao et al., 2009). The monobody binding to the SARS N at N-terminal domain (NTD) with a Kd=72 nM is called NN1, and the monobody binding to the N protein C-terminal domain (CTD) with a Kd=1.7 nM is called NC1.

To evaluate potential binding of NN1 and NC1 to SARS-CoV-2 N and examine the cross-reactivity with other CoV, we first performed sequence comparison of SARS-CoV-2 , SARS, MERS-CoV (MERS), and other CoV causing mild upper respiratory disease (OC43, 229E and NL63) (Stothard, 2000). SARS-CoV-2 N is highly similar to SARS N, with 91% identity and 94% similarity (Figure. 1A). The NTD and CTD of SARS-CoV-2 N are almost identical to that of SARS N (96.9% and 98.3% similarity, respectively), suggesting that our monobodies (NN1 and NC1) may also recognize SARS-CoV-2. Unlike SARS, other CoV are dissimilar from SARS-CoV-2 N, which has 50% identity with MERS and 28-35% identity with the others (Fig.1A). Phylogenetic tree also suggested that SARS-CoV-2 is closely related to SARS, but distant from other CoV (Supplementary Figure. 1A) (Stecher et al., 2020). Moreover, the sequence of N, particularly NTD and CTD that potentially recognized by our monobodies, is highly conserved among reported SARS-CoV-2 sequences. This indicates that NTD and CTD of N are robust targets for SARS-CoV-2 detection (Supplementary Figure. 1B).

We performed *in vitro* maturation using a doped library to further improve the affinity and stability of monobodies (Figure. 1B). Using NN1 and NC1 as backbone, we generated mutations in all 17 amino acids across BC and FG loops. The diversity was generated to be biased towards the parental monobodies using doped nucleotide mixtures (Supplementary Figure. 1C&D). The resulting library contained mutations towards all 19 other amino acids on each position, with 23-37% remaining to be the parental amino acid (Supplementary Figure. 1E). As SARS-CoV-2 N sequence was not available at the time of performing the experiment, SARS N was used as the bait to select the mutant libraries through mRNA display. After 6 rounds of selection with an increased binding temperature (Method, Extended Data Figure. 2), we identified and confirmed two monobodies with increased binding capacity to SARS N (NN2

and NC2, Figure. 1C). The enhancement in affinity of NN2 and NC2 was quantified by surface plasmon resonance (SPR). NN2 and NC2 were expressed in bacteria and purified by affinity chromatography, while SARS N, expressed and purified from 293T cells, were immobilized on Biacore chips as ligands. NN2 achieves 22- fold increase in binding affinity, resulting in a Kd=3.3 nM. NC2 has a binding constant of 390 pM with SARS N, corresponding to about 4-fold affinity improvement compared with parental NC1 (Supplementary Figure. 2A&B).

After the sequence of SARS-CoV-2 is available, we tested whether these binders have affinity to SARS-CoV-2 N, prompted by the above sequence analysis (Figure. 1D). By quantifying their binding affinity to SARS-CoV-2 N with SPR, we observed that NN2 has an on-rate of 3.1 x 105 M-1 s-1 and an off-rate of 8.5 x 10-3 s-1, resulting in a Kd=27.6 nM. NC2 has on-rate of 5.7 x 105 M-1 s-1 an off-rate of 3.9 x 10-3 s-1, resulting in a Kd=6.7 nM. These data suggested that these monobodies bind with SARS-CoV-2 N *in vitro*, with high affinity.

Monobodies recognize endogenous SARS-CoV-2-N *in vivo*

Following verification of the high-affinity binding between the two monobodies with SARS-CoV-2 N *in vitro*, we examined if they can specifically bind to endogenous N in cells. Specifically, we examined if both remain binding to their respective domains of SARS-CoV- 2 N. Co-immunoprecipitation (co-IP) experiment was performed in 293T cells using truncated fragments of SARS-CoV-2 N (Figure. 2A). Consistent with parental NN1 and NC1, NN2 showed binding to full-length N, but lost interaction with CTD only or when NTD was deleted (N-deltaN), suggesting its specific binding to NTD of SARS-CoV-2 N. On the other hand, NC2 remained binding with CTD, suggesting that it is a CTD binder. Using sequential IP, we further demonstrated that NN2, NC2 and SARS-CoV-2 N formed a protein complex in cells, indicating that the bindings of these two monobodies to SARS-CoV-2 N are not competitive (Figure. 2B). Independent binding of two monobodies was also supported by SPR analysis. When combining both monobodies at a concentration exceeding saturation, the response is additive with totaling 94% of the separate binding responses, demonstrating that neither of the two molecules obscure one another (Figure. 2C). The above results, consistent with the fact that the CTD and NTD are separated by a flexible region on N protein, support the feasibility of using this pair of monobodies simultaneously to establish sandwich ELISA for antigen detection.

After investigating the physical interaction between the two binders and SARS-CoV- 2 N, we validated their binding *in situ* by immunofluorescence assay (IFA). Fibronectin- based monobodies are shown to localize in nucleus upon overexpression in mammalian cells (Liao et al., 2009). However, with the expression of SARS-CoV-2 N in A549 cells, the localization of monobodies became cytosolic (Supplementary Figure. 2C). Both NN2 and NC2 co-localized significantly with antibodies against N in SARS-CoV-2-infected cells (Figure. 2D, Supplementary Figure. 2D), confirming that these monobodies can bind and recognize endogenous N. They also showed high specificity by interacting only with SARS- and SARS-CoV-2-infected cells (Supplementary Figure. 2E&2F), but not with mock- or MERS-infected cells

(Supplementary Figure. 2D). Notably, NC2 exhibited higher intensity than NN2 in SARS-CoV-2 infected cells, probably due to the higher binding affinity (Figure 2D, Supplementary Figure. 2D).

Further affinity maturation of SARS-CoV-2 N-targeting monobodies using single round selection of a high-complexity variant library

We next sought to further improve the affinity of monobodies specifically towards SARS-CoV-2 N by in vitro maturation. NN2 and NC2 were used as parentals and large variant libraries for each of them were generated. As we previously showed that functional interactions usually occur in nearby residues (Olson et al., 2014), we chose to cover all possible variations at two adjacent residues in BC and FG loops. The two flexible loops were diversified independently and combined orthogonally, resulting in about 10.5 million amino acid variants of each monobody (Figure. 3A). This diversification strategy gives us the flexibility to discover residue interactions at nearby positions and between the two loops, while controlling the library size to enable single round of selection. The libraries also had ~0.1% spike-in of parental NN2 or NC2 as the internal positive control. Following library construction, strep tagged-SARS-CoV-2 N were expressed in 293T cells, purified and immobilized onto beads as the bait. Single round of mRNA display of variant libraries was performed with N at two amounts, low (~1 ug) and high (~10 ug), representing different selection pressures. We also included NP of IAV as a bait as the negative control. Lastly, deep sequencing was used to identify variants and quantify the relative frequency of each variant in the libraries.

Sequencing data revealed even representation of variations for all positions in both loops in the input library (Supplementary Figure. 3A&B). 1.74 million NN2 variants and 1.82 million NC2 variants with read depth more than 5 in the input library were considered as candidates of high confidence, and included in downstream analysis. The input read depth of variants correlated well with the ones post SARS-CoV-2 N selection, but not with those post IAV NP selection (Supplementary Figure. 3C). Quantitatively, 24.9-32.4% NN2 variants and 25.1-33.5% of NC2 variants were identified after selection with SARS-CoV-2 N, while only 8.4% of NN2 variants and 1.3% NC2 variants were identified with IAV NP. This indicated efficient and specific selection with SARS-CoV-2 N for both monobody libraries, but not with IAV NP. To quantify the relative binding affinity for each variant, we calculated the relative enrichment (RE) score as the ratio of its frequency in the selected library to that in the input library, normalizing to the relative frequency of its corresponding parental sequence. As expected, histogram of RE scores showed clusters of enriched variants with SARS-CoV-2 N as bait, but not with IAV NP (Figure. 3B). The RE scores correlated well between the two selection conditions of N (Figure. 3C). NN2 library contained more highly enriched variants (RE score > 10 for both conditions) comparing with NC2, probably due to the larger space for improvement, when starting with NN2 with lower affinity (Figure. 3C). To provide a global visualization of amino acid preference in the selected libraries, we calculated the averaged RE scores of all variants with certain substitutions on each amino acid position (Figure. 3D). The averaged RE scores were normalized

across all randomized positions, to reveal the preference of substitutions on different residues. In general, selection with low and high amount of SARS-CoV-2 N provided similar heatmaps. NN2 and NC2 showed different preference on variated residues. For NN2, the highly enriched variations clustered mostly at the N terminal of BC loop and relatively dispersed on FG loop. For NC2, the N terminus of BC loop and the C terminus of FG loop showed more enrichment upon substitutions. Our results suggested efficient affinity enhancement of monobodies by one- round of in vitro maturation.

Incorporating machine learning for the validation of 3rd generation monobodies

The high throughput measurement enables us to survey a large number of variants.

However, there are uncertainties due to many factors, including sampling bias and sequencing errors. Artificial Neural Network (ANN) models are generic classifier machines that can extract underlying patterns in different types of data, minimizing the uncertainties. They have been widely adapted to facilitate sequence-to-structure or sequence-to-function prediction (Fernandez-de-Cossio-Diaz et al., 2020; Girija, 2016; Liu et al., 2018; Senior et al., 2020). To identify the sequence pattern more likely responsible for the increased binding affinity and to rule out stochastic errors for candidate selection, we trained an ANN model (Figure. 4A) to predict RE scores of variants based on their amino acid sequences. 90% of the experimental data was fed to the model and the remaining 10% was used for validation. The 17-residue mutagenized region was parsed into a 357-dimension boolean array, and the RE scores were parsed into 10 quantiles as output. The predicted RE score was calculated as the weighted average of the output array. The predictions faithfully captured the patterns of sequences with increased affinity (Figure. 4B). It correlated significantly with experimental enrichment with SARS-CoV-2 N as bait for the 10% validation data, as well as the whole dataset, but not with IAV NP (Supplementary Figure. 3D&3E).

As NN2 had lower affinity to N (Figure. 1D), and had more highly-enriched candidates during the *in vitro* maturation (Figure. 3B&3C), we expect that it has a larger room for improving binding affinity. We hence focused on NN2 for validating potential 3rd generation NTD binders. First, by considering only the experimental RE scores, we identified 1170 candidates with RE scores >10 in both N selection conditions, and RE score <1 under IAV NP selection. Second, by integrating with ANN prediction results (top 10%), we narrow the lists down to 134 candidates (Supplementary Table 1). We picked 4 clones as the representatives of the 3rd generation NTD binders (NN3-1, NN3-2, NN3-3 and NN3-4), with a bias towards hydrophilic residues on the loops (Supplementary Figure. 3F). All 4 monobodies were expressed and purified. NN3-1 showed the highest expression level in *E. coli* and improved solubility comparing with NN2. SPR analysis suggested that NN3-1 showed improvement in both on-rate (4.1 x 105 M-1 s-1) and off-rate (7.4 x 10-3 s-1) when binding with SARS-CoV-2 N. This results in a Kd=18.1 nM, corresponding to 1.5-fold increase with parental NN2 (Figure. 4E). To more accurately measure the binding affinity between NN3 with N-NTD in solution, we purified NTD from *E. coli* and performed isothermal titration calorimetry (ITC). Similar to the results from SPR, NN2 showed to have a Kd of 17.36 nM, while NN3 improved to be

14.43 nM (Figure. 4F). Consistently, when comparing the staining intensity of monobodies with anti-SARS-N antibody in viral infected cells, we noticed that NN2 showed weaker signal comparing with antibody in SARS-CoV-2- infected cells, but similar intensity in SARS-infected cells. However, staining with NN3-1 showed much stronger signal in SARS-CoV-2-infected cell and decreased signal in SARS- infected cells, supporting that NN3-1 improved the affinity and selectivity against SARS- CoV-2.

Development of ELISA for N detection using monobodies

To demonstrate the potential usage of monobodies in viral detection, we developed sandwich ELISA for SARS-CoV-2 N detection. To assess feasibility of the assay, we first utilized monobodies as a capture protein and a commercial anti-SARS N antibody for detection (i.e. single monobody-based assay). As capture proteins usually require a higher amount than detection proteins, the ease and efficiency of monobody production in *E. coli* makes them an appealing replacement for traditional antibody-based capture in ELISA. The standard curve was established with purified recombinant N diluted in BSA. It was found that using colorimetric TMB substrate, NN2-coated plate could detect N at a limit of 1 ng/ml, while NC2-coated plate had a sensitivity of 360 pg/ml (Supplementary Figure. 4A&4B). The detection sensitivity could be improved with a chemiluminescence substrate, reaching 320 pg/ml for NN2 and 64 pg/ml for NC2 (Figure. 5A&5B, Supplementary Figure. 4C&4D). The assay was reproducible with replicates performed independently at different time. To further improve the sensitivity, we utilized Tyramide signal amplification (TSA) to push the limit to 4 pg/ml (Figure. 5C). Such sensitivity did not decrease when diluting N in human sera, suggesting potential assay compatibility with clinical samples (Figure. 5D).

To examine the detection sensitivity and specificity of the above single monobody- based ELISA for N in cell culture, we obtained viral culture supernatants of SARS-CoV-2, MERS, and IAV (strain: A/Cal/04/09). The viral supernatants were inactivated in 0.5% Triton X-100, and serially diluted in culture media plus 0.5% Triton X-100. From the standard curve constructed with recombinant N, we estimated that the concentration of N was about 26.4 ng/ml in the supernatant (Figure. 5E). Regarding specificity, no cross-reactivity was observed with MERS viral culture, which showed the highest similarity to SARS-CoV-2 among other CoV (Figure. 1A, Figure. 5E). The ELISA could also distinguish between SARS-CoV-2 and IAV, a virus causing similar clinical manifestations (Figure. 5E).

The above results illustrated that our monobodies could be paired with antibody to detect circulating N. We then demonstrated that the ELISA could work by pairing the two monobodies NC2 and NN2 (i.e. dual monobody-based assay). Firstly, we coated HIS-tagged NC2 onto a plate and detected serially diluted recombinant N using FLAG-tagged NN2 (Figure. 5F). The anti-FLAG antibody was used for detection. Using chemiluminescence substrate, we could detect N at a concentration of 10 pg/ml. Secondly, to simplify the experimental procedure, we constructed a NN2 protein with three cysteines at the C terminus, enabling direct HRP conjugation through activated sulfhydryl groups. Using NN2-HRP as the detection agent, we could detect recombinant N at 1.5 ng/ml with TMB substrate, and around 10 pg/ml with chemiluminescence substrate (Figure. 5G, Supplementary Figure. 4E).

As previously we demonstrated the 3rd generation monobody NN3-1 has a higher affinity to N comparing with NN2, we examined if the usage of NN3-1 can improve the detection sensitivity of N. As expected, the single monobody-based assay using NN3-1 as the capture protein improved the sensitivity of ELISA from 1 ng/ml to 316 pg/ml, using TMB substrate.

Together, we presented a platform for identifying and evolving monobodies against SARS-CoV-2 N (Figure 6). We first created millions of variants on the backbone of parental monobody through high-dense mutagenesis. Followed by mRNA display and deep sequencing, we were able to systematically profile the relative affinity of these variants against viral antigen. Machine learning can be integrated to facilitate the validation of positive hits. The binding affinity and specificity can be determined by various biological and biochemical assays. As demonstrated by the high sensitivity of ELISA developed from the monobody pair, the combined advantages of high binding affinity and specificity, low cost, and small protein size in monobodies support their use as antigen detection reagents for developing diagnostic methods for emerging infectious diseases, including but not limited to COVID-19.

Discussion

COVID-19 is a rapidly evolving global pandemic. An effective and scalable detection method for early identification and isolation of infected individuals is hence critical for controlling the spread of COVID-19. In view of the limitations of current diagnostic techniques with high cost, low sensitivity, and long turn-around assay time, we presented a platform for developing high-affinity monobodies recognizing viral antigen. Particularly, we developed affinity-enhanced monobodies that specifically recognize N both *in vitro* and *in vivo*. The two monobodies have a high affinity to SARS-CoV-2 N with Kd at nM range, and can be mass-produced from *E. coli*. They can recognize and, together, form a protein complex with endogenous N. The potential utility of this pair of monobodies was shown with an ELISA that has relatively high sensitivity and specificity.

Comparing with antibodies, our monobody-based platform has the following advantages. First, we can rapidly and iteratively evolve monobodies to target selected viral antigen. Coupling mRNA display with high-throughput sequencing, the screening and validation process takes about two weeks, faster than traditional antibody production involving animal immunization, isolation and purification of antibodies. It is therefore suitable for emerging situations like COVID-19. Second, it is relatively low-cost through mass-production in *E. coli*, demonstrating great scalability. Third, in addition to sensitivity and specificity, monobodies with desired properties can be generated through different selection pressure. For example, performing the selection at an elevated temperature or with protease challenge generate monobodies with higher thermal stability and protease resistance. Such selection is hardly customizable in antibodies where production is restricted at physiological conditions. Last, monobodies are significantly smaller and less complex than multi-domain antibodies, possibly achieving higher density in nanodevices to improve sensitivity, and higher permeability to cells for imaging. Thus, complementary to antibodies, monobodies can be potentially applied to developing diagnostic assays and enable population-wide screening in a cost-and-time effective manner.

As a proof of concept, we presented the application of monobodies through the establishment of ELISA. Although RT-qPCR is the current standard test for diagnosing COVID-19, it has specific equipment requirement and laborious procedures that are prone to contamination without dedicated facilities. For the diagnosis of SARS, we and others have developed ELISA or chemiluminescent enzyme immunoassay (CLIEA) using antibodies to detect circulating viral N with sensitivity of 80-100% (Che et al., 2004b, 2004a; Fujimoto et al., 2008; Li et al., 2005). In SARS-infected patients, a high concentration of viral N has been detected in multiple body fluids as early as 1 day after the onset of symptoms, making it an ideal target for early viral detection The concentration of N in patient sera is estimated to be around 100 pg/ml to 3.2 ng/ml (Che et al., 2004a). Given the high similarity of viral genome and clinical course between SARS and SARS-CoV-2, we postulate that the concentration range of SARS-CoV-2 N in patient sera will be similar to that of SARS. For our established ELISA using monobodies, we achieve a detection sensitivity of around 10 pg/ml, which is the estimated lower boundary of N concentration in SARS-CoV-2 patients. To make it clinically applicable, we are optimizing the conditions to, on one hand, further improve the detection limit by reducing background noises, and on the other hand, enable long-term storage of monobody-coated ELISA plates. Different formulations of dispensing solution will be tested for their ability to immobilize the monobodies to surfaces with long-term stability. Parameters include but are not limited to the following: varying the buffer type, pH, the addition of stabilizing sugars, glycerol, ethanol, surfactants and polymer additives.

Besides ELISA, our monobody-based system may be compatible with other point-of- care detection platforms, including LFIA strips and nanobiosensors. By conjugating the monobodies with gold nanoparticles, LFIA strips can detect the existence of N in biological samples through simple visualization. Miniature devices like nanobiosensors have also attracted great attention due to their short detection time, high sensitivity and accuracy for real time analysis with small amounts of samples (Lee et al., 2008; Noah and Ndangili, 2019; Singh et al., 2016; Vu and Chen, 2019). The number of immobilized binders is a critical factor determining the performance of sensors. Ascribing to their small size, higher density and larger surface-to-volume ratio of monobodies may enable the detection of antigens at low concentrations. In fact, we previously demonstrated that our monobody can be coated onto nanobiosensor and detect circulating N at nM concentration, suggesting the potential of using monobody as detection reagents on nanosensors (Ishikawa et al., 2009). Furthermore, our small monobodies can possibly act as *in vivo* imaging probes for basic and clinical research, compatible with photon emission computed tomography (SPECT) or positron emission tomography (PET). Molecular imaging has been developed for monitoring inflammation of lung diseases and viral dynamics in primates (Gordon et al., 2019; Lepin et al., 2010; Santangelo et al., 2015), which often uses antibody as detection probe. Despite their high sensitivity, antibodies are large in size and thus with low tissue penetration and low clearance time. As an alternative, the monobodies may be able to increase tissue permeability and clearance time, thereby reducing background noises and allowing early imaging. Moreover, the small monobody can be efficiently packed into nanocapsule and potentially delivered to target cells, eventually allowing intracellular tracing of endogenous antigens (Abellan-Pose et al., 2016; Wen et al., 2019).

In summary, we have presented a platform for developing and evolving monobodies as a viral antigen recognition element, including high-dense mutagenesis, mRNA display, deep sequencing and machine learning. It can be potentially applied to developing various diagnostic techniques for screening viral detection reagents against emerging communicable diseases, including but not limited to COVID-19.

Declarations

Acknowledgement:

The work on COVID-19 and SARS-CoV-2 is supported by UCLA Innovation Fund and UCLA AIDS Institute Seed Fund. We appreciate the generous help from colleagues at UCLA and HKU. The use of human serum specimens for testing was approved by the Institutional Review Board of The University of Hong Kong / Hospital Authority Hong Kong West Cluster.

Author Contributions

YD, THZ, XM, YS and RS conceived the project; YD, THZ, XM, YS, MH, XJC, RWR, DYJ, KYY, QH, IC, and RS designed the experiments; YD, THZ, XM, YS, MH, SY, CYL, SXL,SWL, JL, HH, HJ, DC, LS, MH, AO and HL performed the experiments; XL, GZ performed the high throughput sequencing; YD, THZ, XM and RS analyzed data; YD, CYL, THZ, XM, and RS drafted and/or revised the manuscripts.

Conflict of Interests

Authors declare no conflict of interests.

Methods

Sequence analysis

Sliding window similarity was calculated using Sequence Manipulation Suite 2, with a window size of 40 residues. Residue 45-175 were highlighted as NTD and residue 248-365 were highlighted as CTD.

The average distance tree was constructed using MEGA. The distance matrix for constructing the tree was BLOSUM62. The sequences used in the tree were reference genome sequences from NCBI nucleotide database, with the following accession number (HKU1: KF686346.1, 229E: KU291448.1, NL63: AY567487.2, OC43: NC_006213.1, MERS-CoV: NC_019843.3, SARS-CoV: NC_004718.3, SARS-CoV-2: NC_045512.2).

Entropy was calculated using GISAID SARS-CoV-2 data. All viral sequences were first aligned by MAFFT, followed by annotation and extraction of the sequenced coding N genes for comparison. The entropy on each site was calculated using scipy.

Expression and purification of monobodies

The DNA encoding monobodies were cloned into pET11 plasmid (HIS tag) or pAO9 plasmid (MBP tag), and transformed in BL21 competent cells. BL21 cells containing the plasmid were inoculated into 5 ml LB containing 50 ug/ml ampicillin or kanamycin and cultured overnight at 37oC and 200 rpm. The cultured bacteria were transferred to fresh LB media with Ampicillin or Kanamycin at 1:100 ratio. OD600 was monitored until it reaches 0.4. IPTG was added to the culture to a final concentration of 200 uM, and then cultured overnight at 18oC (~20 hours). Cells were pelleted by centrifugation at 8000 g for 20 minutes, and lysed by B-PER buffer with protease inhibitors. Affinity purification was performed using glutathione Sepharose 4B resin for MBP-tagged monobody, or Ni-Column for HIS- tagged monobody. Protein fraction was dialyzed to ion-exchange A buffer (20 mM Tris-HCl pH8.0; 75 mM NaCl; 6 mM b-ME) overnight under 4oC. After dialysis, the protein sample was loaded on Q-HP column (GE) and eluted in a linear gradient way to B buffer (20 mM Tris-HCl pH8.0; 1000 mM NaCl; 6 mM b-ME). Target protein fractions were collected and concentrated. Gel-filtration was performed using HiLoad Superdex 75 prep grade (GE) equilibrated by a buffer (20 mM Tris-HCl pH8.0; 150 mM NaCl; 2 mM DTT). Aggregation was removed. Target protein was eluted and pooled. Protein sample was concentrated and stored at -80oC. Protein concentration was measured by Commassie Plus Reagent.

Expression and purification of viral N

Open reading frames of viral proteins were cloned into pcDNA3 mammalian expression plasmid with a strep-tag at the C terminal. HEK293T cells were transfected with the plasmid by lipofectamine 3000. Cells were collected 24 hours post transfection and lysed with lysis buffer (20 mM Tris pH8.0, 500 mM NaCl, 1 mM EDTA, 1 mM DTT, 0.5% Triton X-100, supplemented with protease inhibitor cocktail, 1 mM NaF, 2 mM Na2V3O4 and 2 mM beta-Glycerophosphate) along with sonication. Cell lysates were incubated with MagStrep "type3" XT beads overnight at 4°C with constant agitation, and washed with lysis buffer for 5 times. Proteins were eluted with Elution Buffer BXT (IBA) for 30 minutes at 37oC and dialyzed against HBS-E buffer (10 mM HEPES pH7.5, 150 mM NaCl, 1 mM EDTA).

Surface plasmon resonance (SPR)

Binding kinetics between monobodies and N was measured with a Biacore T100 instrument. Various concentrations of monobodies were flowed over a blank CM5 chip and an N-bound chip at 100 ul per minute for 120 seconds. Dissociation was monitored for 10 minutes before regeneration by washing with 5 mM sodium hydroxide at 30 ul per minute for 10 seconds. Kinetics data was obtained by fitting with the Biacore evaluation software. For the double binding experiments, purified monobodies were flowed at 100 ul per minute at increasing concentrations. The two monobodies were flowed for 120 seconds separately at 300 nM and then together at 300 nM each, with regeneration in between injections.

Isothermal titration calorimetry (ITC)

The binding affinity between affibody and N_Nter protein was measured under 25 degrees through isothermal titration calorimeter (MicroCal iTC200, Malvern Panalytical). The buffers of affibody and N_Nter protein were the same (20mM Tris-HCl PH8.0, 150mM NaCl, 2mM DTT). The protein concentration in syringe ranged from 0.3 to 0.6mM while in reaction cell ranged from 0.015 to 0.06mM. After excluding the first injection, all titration data was calculated and analyzed by MicroCal ITC-ORIGIN Analysis Software (Malvern Panalytical).

Enzyme-linked immunosorbent assay (ELISA)

Monobody protein stock was diluted in 1X coating buffer (Bio-Rad) to a working concentration of 6.4 ug/ml. 50 ul of working solution was added into each well and the plate was incubated overnight at 4oC to coat the binder to the plate. After coating, the binder working solution was discarded and the plate was washed with washing buffer (PBS, 0.05% Tween-20) twice. For blocking, 100 ul of blocking buffer was added to each well and the plate was incubated for 2 hours at 37oC. Then the blocking buffer was removed, and the plate was washed with washing buffer for 3 times. For N spike-in experiments, sample solution was diluted in 1% BSA/washing buffer containing 0.5% Triton X-100 or human serum specimens; for viral culture, sample solution was diluted in viral medium (1% FBS in DMEM medium) containing 0.5% Triton X-100. Next, 100 ul of sample solution was added into each well and incubated for 1 hour at 37oC. Afterwards, the plate was washed with washing buffer 5 times. Primary antibody (anti-SARS N rabbit polyclonal antibody, Rockland) was diluted 1:500 in washing buffer. 50 ul of primary antibody solution was added into each well and incubated for 1 hour at 37oC. Then the plate was washed again with washing buffer for 5 times. Secondary antibody (anti-rabbit IgG-HRP, Cell Signaling Technology) was diluted 1:5000 in washing buffer. 50 ul of secondary antibody solution was added into each well and incubated for 30 minutes at 37oC. After 5 times of washing with washing buffer, 50 ul of chemiluminescence substrate solution (SuperSignal™ ELISA Pico Substrate, Thermo Scientific) was added into each well and incubated for 1 minute at room temperature. The chemiluminescence signal was read using plate reader (Varioskan LUX Multimode Reader). For readings at OD450, 100 ul of colorimetric TMB substrate (1-Step™ Ultra TMB-ELISA Substrate Solution, ThermoFisher) was added into each well and incubated for 15 minutes at room temperature. The reaction was stopped using 2 M sulfuric acid.

For full monobody-based ELISA, NC2 was used to coat the plate at 640 ng/well. NN1 with C-terminal FLAG tag was used at 1 ng/ml, as primary antibody. Then DYK rabbit polyclonal antibody was used to detect FLAG tag, followed by the addition of secondary antibody (anti-rabbit IgG-HRP, Cell Signaling Technology).

The use of human serum specimens for testing was approved by the Institutional Review Board of The University of Hong Kong / Hospital Authority Hong Kong West Cluster.

HRP conjugation of monobody

NC2 monobody with three Cystines on the C terminus was used for HRP conjugation.

100 µL of Maleimide Conjugation Buffer was added to 2-MEA, and incubated for 90 minutes at 37°C. Activated HRP was conjugated to NC2 monobody by incubating for 2 hours at room temperature.

Single-walled carbon nanotube nanobiosensor detection

Our carbon nanotube devices were prepared as follows. First, carbon nanotubes were grown on a Si/SiO2 substrate by chemical vapor deposition (CVD) using iron nanoparticles as catalysts. Metal electrodes were then defined by depositing Cr/Au through a shadow mask. Nanotubes located outside the channel were removed by an oxygen plasma treatment while protecting nanotubes inside the channel with polymethyl methacrylate. The surface of nanotubes was functionalized with linker molecules by soaking the devices in a 0.1 mM solution of 1-pyrenebutanoic acid succinimide ester in dry methoanol for 1 h, followed by an extensive washing with methanol. This linker molecule is a bifunctional molecule that has a pyrene derivative at one terminus and a succinimide ester group at the other terminus. The pyrene group is known to bind to the nanotube surface by p-p interaction, and the succinimide ester group is reactive towards nucleophilic groups such as amines. The device was then submerged in a 0.01 mM aqueous solution of BMPA for 2 hr, which resulted in adding maleimide terminals on the nanotube surface. NN2 was subcloned into pAO9 with C- terminal cysteine and purified as described above. NN2-Cys was immobilized on the nanotube over-night in PBS. After extensive washing, a baseline was established using increasing concentrations of BSA in 0.01% PBS. N protein was prepared as described previously, and buffer exchanged into 0.01% PBS (PD-10, GE Healthcare) before application to the nanobiosensor.

Immunofluorescence assay (IFA)

Vero E6 cells were used for infection. Cells were fixed with 4% PFA/PBS for 5 minutes at room temperature 24 hours post infection. After fixation, cells were permeabilized by 0.3% Triton X-100/PBS for 10 minutes and blocked by 5% BSA/TBST (20 mM Tris 2.0, 140 mM NaCl, 0.05% Tween-20) for 1 hour. MBP-NN2 or MBP-NC2 was diluted in 1% BSA/TBST to 2 ug/ml and cells were incubated with the binder solution for 1 hour. Then, anti-MBP monoclonal antibody (NEB, 1:2500), anti-SARS N rabbit polyclonal antibody (1:2500) and anti-MERS N guinea pig antibody (1:1000) were added to the slide chamber and incubated for 1 hour. Anti-mouse Alexa Fluor-594 and anti-rabbit Alexa Fluor-488 (or anti-guinea pig Alexa Fluor-647) were diluted to 2 ug/ml to label the corresponding primary antibodies. Nuclei were stained by Hoechst 33324 and slides were fixed again by 1% PFA/PBS for 5 minutes. Finally, slides were mounted by ProLong™ Diamond Antifade Mountant (Thermo Fisher) and cured for 24 hours at room temperate before image acquisition.

Immunoprecipitation (IP)

Immunoprecipitation experiments were performed with HIS-, Strep- and FLAG- tagged proteins expressed in 293T cells. Briefly, cells were transfected with the corresponding expression plasmids with Lipofectamine 2000 reagents (Invitrogen), and lysed at 2 days post-transfection with RIPA buffer (50 mM Tris-HCl pH 7.4, 0.5% NP-40, 150 mM NaCl, 1 mM EDTA and protease inhibitors). Cell lysates were incubated with Strep-Tactin XT beads (IBA) or anti-M2 FLAG beads (Sigma-Aldrich) overnight at 4° C with constant agitation, washed with RIPA buffer 5 times and eluted with 60 µl of SDS-PAGE sample buffer. All samples were subjected to SDS-PAGE and western blotting analysis. For sequential IP assay, protein complexes were eluted with 3XFLAG peptide (Sigma), then subjected to the second immunoprecipitation with Strep-Tactin XT beads, and finally eluted with 60 µl SDS-PAGE sample buffer for western blotting.

RNA extraction, reverse transcription and real-time PCR

Viral RNA was extracted using the QIAamp Viral RNA Mini Kit (Qiagen Sciences), and reverse transcribed by Superscript III Reverse Transcriptase (Thermo Fisher) Quantitative real-time PCR was performed using Taq polymerase and SYBG. For viral copy number, a standard curve ranging from 1x103 to 1x108 was used for quantification.

Construction of doped library

The NN1 and NC1 doped libraries were constructed similar to the naïve library previously described (Liao et al., 2009). For the NC1 library, Fnoligo3 was replaced with (5’-CCAGCCTCCTGATCAGCTGG78S557586866857557 68SCGCTACTACCGCATCACCTACG) and Fnoligo7 was replaced with (5’-CGGTAGTTGATGGAGATCGGS77S67S65788S65775555S65S76788CGTGACGGCGTACACCGTGA). For the NN1 library, Fnoligo3 was replaced with (5’-CCAGCCTCCTGATCAGCTGG77S67S557857857557587CGCTACTACCGCATCACCTACG) and Fnoligo7 was replaced with (5’-CGGTAGTTGATGGAGATCGGS76775S56S65685758S76775786S68CGTGACGGCGTACACCGTGA). Nucleotide positions labeled “5” were mixed with 67% A, 11% G, 12% C, and 10% T. Nucleotide positions labeled “6” were mixed with 16% A, 59% G, 14% C, and 11% T. Nucleotide positions labeled “7” were mixed with 16% A, 12% G, 62% C, and 10%T. Nucleotide positions labeled “8” were mixed with 17% A, 13% G, 15% C, and 55% T. Nucleotide positions labeled “S” were mixed with 50% G, 50% C. These frequencies were used as an attempt to compensate for varying reaction rates of each nucleotide.

Construction of variant library for the 3rd generation binders

Based on the parental sequences (NN2 and NC2), we introduced all possible random mutations to each two adjacent codons, using synthesized oligos containing NNSNNS sequence. N stands for a uniform mixture of A/T/C/G and S stands for G or C. We used 8 oligos to cover the full length of 10Fn3, and assembled them by PCR. Each library was expected to contain 56.6 million nucleotide variants, i.e. 10.5 million amino acid variants.

mRNA display and sequencing

10Fn3 library (DNA templates) was transcribed by T7 polymerase (NEB), and ligated to the pF30P linker (Phospho-AAAAAAAAAAAAAAAAAAAA-spacer9-spacer9- spacer9-ACC-puromycin) via the splint oligonucleotide by T4 DNA ligase (NEB). After purification and isolation of ligated mRNA templates, *in vitro* translation was performed using reticulocyte lysate (Promega) followed by incubation with KCl (500 mM) and MgCl2 (60 mM) for 30 minutes at room temperature to enhance fusion formation. The mRNA-protein fusions were then affinity-purified using M2 anti-Flag beads (Sigma-Aldrich). After elution with 3XFLAG peptides, the fusions were reverse transcribed with Superscript III (Invitrogen) and a fraction of the purified sample was reserved to determine the frequencies of each coding sequence in the input library. The purified fusion sample was incubated with bait protein for 2 hours at room temperature. After extensive washing (6 rounds of washing at room temperature), the immobilized fusion samples were eluted by heat (95°C) and PCR amplified. The amplified DNA fragments from input and post selection were then prepared for high throughput sequencing using NovaSeq 6000 (PE250).

Sequencing data analysis

Data were analyzed by customized bash and python scripts. Paired-end fastq reads were de-multiplexed into corresponding samples. Reads were mapped to corresponding parental monobody sequences. We reached ~50 million depth for the input libraries. The distribution of read depth centered around 10 reads, indicating that most variants in the library were sufficiently covered. Variants that have read depth more than 5 in the input library were considered as candidates of high confidence, and included in downstream analysis. The relative enrichment (RE) score for each variant was calculated as its relative frequency in the selected library to that in the input library, normalizing to the relative frequency of its corresponding parental sequence.

$$RE\ score\ = \frac{\frac{Frequency\ (mut, selected)}{Frequency\ (mut, input)}}{\frac{Frequency\ (WT, selected)}{Frequency\ (WT, input)}}$$

Deep learning

Artificial neural network model was constructed using TensorFlow 2.1.0. The 4-layer network was compiled by a 357-node input layer, a 64-node dense layer, a 64-node dense layer and a 10 nodes output layer. All input layers and middle layers used leaky-ReLu as activation function. The model used Adam optimizer with a learning rate of 0.01. Loss function MSE was used to optimize accuracy. We parsed each mutant sequence into a 357- dimension Boolean matrix as input. We then categorized the RE scores into 10 quantiles as output array. We fed 90% RNA display data into the network to train the model. We trained the model for each library and selection conditions. The accuracy of the models was 0.63-0.74 for N selected libraries. We then used the model to predict the rest 10% data. The predicted RE scores were calculated as the weighted average of the output array.

Data Availability

The sequencing data were deposited to NIH Short Read Archive (SRA) with access numbers PRJNA615649. Custom python scripts for further data analysis were deposited to github: https://github.com/Tian-hao/mRNAdisplay.

References

Abellan-Pose, R., Teijeiro-Valiño, C., Santander-Ortega, M.J., Borrajo, E., Vidal, A., Garcia-Fuentes, M., Csaba, N., and Alonso, M.J. (2016). Polyaminoacid nanocapsules for drug delivery to the lymphatic system: Effect of the particle size. Int. J. Pharm. *509*, 107–117.

Amanat, F., Nguyen, T.H.O., Chromikova, V., Strohmeier, S., and Stadlbauer, D. (2020). A serological assay to detect SARS-CoV-2 seroconversion in humans. MedRxiv.

Chan, J.F.-W., Yuan, S., Kok, K.-H., To, K.K.-W., Chu, H., Yang, J., Xing, F., Liu, J., Yip, C.C.-Y., and Poon, R.W.-S. (2020). A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet *395*, 514–523.

Che, X.Y., Qiu, L.W., Pan, Y.X., Wen, K., Hao, W., Zhang, L.Y., Wang, Y. Di, Liao, Z.Y., Hua, X., Cheng, V.C.C., et al. (2004a). Sensitive and specific monoclonal antibody- based capture enzyme immunoassay for detection of nucleocapsid antigen in sera from patients with severe acute respiratory syndrome. J. Clin. Microbiol. *42*, 2629–2635.

Che, X.Y., Hao, W., Wang, Y., Di, B., Yin, K., Xu, Y.C., Feng, C. Sen, Wan, Z.Y.,

Cheng, V.C.C., and Yuen, K.Y. (2004b). Nucleocapsid protein as early diagnostic marker for SARS. Emerg. Infect. Dis. *10*, 1947–1949.

Chevaliez, S., and Pawlotsky, J.-M. (2008). Diagnosis and management of chronic viral hepatitis: antigens, antibodies and viral genomes. Best Pract. Res. Clin. Gastroenterol. *22*, 1031–1048.

Fernandez-de-Cossio-Diaz, J., Uguzzoni, G., and Pagnani, A. (2020). Unsupervised inference of protein fitness landscape from deep mutational scan. BioRxiv.

Fujimoto, K., Chan, K.H., Takeda, K., Lo, K.F., Leung, R.H.K., and Okamoto, T. (2008). Sensitive and specific enzyme-linked immunosorbent assay using chemiluminescence for detection of severe acute respiratory syndrome viral infection. J. Clin. Microbiol. *46*, 302–310.

Girija, S.S. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. Softw. Available from Tensorflow. Org *39*.

Gordon, O., Ruiz-Bedoya, C.A., Ordonez, A.A., Tucker, E.W., and Jain, S.K. (2019). Molecular imaging: A novel tool to visualize pathogenesis of infections in situ. MBio *10*, 1– 11.

Heymann, D.L., and Shindo, N. (2020). COVID-19: what is next for public health? Lancet *6736*, 19–21.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet *395*, 497–506.

Ishikawa, F.N., Chang, H.-K., Curreli, M., Liao, H.-I., Olson, C.A., Chen, P.-C., Zhang, R., Roberts, R.W., Sun, R., and Cote, R.J. (2009). Label-free, electrical detection of the SARS virus N-protein with nanowire biosensors utilizing antibody mimics as capture probes. ACS Nano *3*, 1219–1224.

Kim, S., Kim, D.-M., and Lee, B. (2020). Insufficient Sensitivity of RNA Dependent RNA Polymerase Gene of SARS-CoV-2 Viral Genome as Confirmatory Test using Korean COVID-19 Cases. Preprints 1–4.

Koide, A., Bailey, C.W., Huang, X., and Koide, S. (1998). The fibronectin type III domain as a scaffold for novel binding proteins. J. Mol. Biol. *284*, 1141–1151.

Kucharski, A.J., Russell, T.W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., Eggo, R.M., Sun, F., Jit, M., and Munday, J.D. (2020). Early dynamics of transmission and control of COVID-19: a mathematical modelling study. Lancet Infect. Dis.

Lee, J.-O., So, H.-M., Jeon, E.-K., Chang, H., Won, K., and Kim, Y.H. (2008). Aptamers as molecular recognition elements for electrical nanobiosensors. Anal. Bioanal. Chem. *390*, 1023–1032.

Lepin, E.J., Leyton, J. V., Zhou, Y., Olafsen, T., Salazar, F.B., McCabe, K.E., Hahm, S., Marks, J.D., Reiter, R.E., and Wu, A.M. (2010). An affinity matured minibody for PET imaging of prostate stem cell antigen (PSCA)-expressing tumors. Eur. J. Nucl. Med. Mol. Imaging *37*, 1529–1538.

Li, Y.H., Li, J., Liu, X.E., Wang, L., Li, T., Zhou, Y.H., and Zhuang, H. (2005). Detection of the nucleocapsid protein of severe acute respiratory syndrome coronavirus in serum: Comparison with results of other viral markers. J. Virol. Methods *130*, 45–50.

Liao, H.I., Olson, C.A., Hwang, S., Deng, H., Wong, E., Baric, R.S., Roberts, R.W., and Sun, R. (2009). mRNA display design of fibronectin-based intrabodies that detect and inhibit severe acute respiratory syndrome coronavirus nucleocapsid protein. J. Biol. Chem. *284*, 17512–17520.

Liu, Y., Palmedo, P., Ye, Q., Berger, B., and Peng, J. (2018). Enhancing evolutionary couplings with deep convolutional neural networks. Cell Syst. *6*, 65–74.

Liu, Y., Gayle, A.A., Wilder-Smith, A., and Rocklöv, J. (2020). The reproductive number of COVID-19 is higher compared to SARS coronavirus. J. Travel Med. 1–4.

Noah, N.M., and Ndangili, P.M. (2019). Current Trends of Nanobiosensors for Point- of-Care Diagnostics. J. Anal. Methods Chem. *2019*.

Olson, C.A., and Roberts, R.W. (2007). Design, expression, and stability of a diverse protein library based on the human fibronectin type III domain. Protein Sci. *16*, 476–484.

Olson, C.A., Wu, N.C., and Sun, R. (2014). A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain. Curr. Biol. *24*, 2643–2651.

Pan, Y., Li, X., Yang, G., Fan, J., Tang, Y., Zhao, J., Long, X., Guo, S., Zhao, Z., and Liu, Y. (2020). Serological immunochromatographic approach in diagnosis with SARS-CoV- 2 infected COVID-19 patients. MedRxiv.

Sajid, M., Kawde, A.-N., and Daud, M. (2015). Designs, formats and applications of lateral flow assay: A literature review. J. Saudi Chem. Soc. *19*, 689–705.

Santangelo, P.J., Rogers, K.A., Zurla, C., Blanchard, E.L., Gumber, S., Strait, K., Connor-Stroud, F., Schuster, D.M., Amancha, P.K., Hong, J.J., et al. (2015). Whole-body immunoPET reveals active SIV dynamics in viremic and antiretroviral therapy-treated macaques. Nat. Methods *12*, 427–432.

Senior, A.W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Žídek, A., Nelson, A.W.R., and Bridgland, A. (2020). Improved protein structure prediction using potentials from deep learning. Nature 1–5.

Singh, P., Pandey, S.K., Singh, J., Srivastava, S., Sachan, S., and Singh, S.K. (2016).

Biomedical perspective of electrochemical nanobiosensor. Nano-Micro Lett. *8*, 193–203.

Stecher, G., Tamura, K., and Kumar, S. (2020). Molecular evolutionary genetics analysis (MEGA) for macOS. Mol. Biol. Evol. *37*, 1237–1239.

Stothard, P. (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques *28*, 1102–1104.

Thevarajan, I., Nguyen, T.H.O., Koutsakos, M., Druce, J., Caly, L., van de Sandt, C.E., Jia, X., Nicholson, S., Catton, M., and Cowie, B. (2020). Breadth of concomitant immune responses prior to patient recovery: a

case report of non-severe COVID-19. Nat. Med. 1–3.

Vu, C.-A., and Chen, W.-Y. (2019). Field-effect transistor biosensors for biomedical applications: recent advances and future prospects. Sensors *19*, 4214.

Wang, C., Horby, P.W., Hayden, F.G., and Gao, G.F. (2020a). A novel coronavirus outbreak of global health concern. Lancet *395*, 470–473.

Wang, H., Wang, Z., Dong, Y., Chang, R., Xu, C., Yu, X., Zhang, S., Tsamlag, L., Shang, M., Huang, J., et al. (2020b). Phase-adjusted estimation of the number of Coronavirus Disease 2019 cases in Wuhan, China. Cell Discov. *6*, 4–11.

Wang, Y., Kang, H., Liu, X., and Tong, Z. (2020c). Combination of RT-qPCR Testing and Clinical Features For Diagnosis of COVID-19 facilitates management of SARS- CoV-2 Outbreak. J. Med. Virol.

Wen, J., Wu, D., Qin, M., Liu, C., Wang, L., Xu, D., Vinters, H. V., Liu, Y., Kranz, E., Guan, X., et al. (2019). Sustained delivery and molecular targeting of a therapeutic monoclonal antibody to metastases in the central nervous system of mice. Nat. Biomed. Eng. *3*, 706–716.

WHO (2020). https://www.who.int/emergencies/diseases/novel-coronavirus-2019.

Wilder-Smith, A., and Freedman, D.O. (2020). Isolation, quarantine, social distancing and community containment: pivotal role for old-style public health measures in the novel coronavirus (2019-nCoV) outbreak. J. Travel Med. 1–4.

Wölfel, R., Corman, V.M., Guggemos, W., Seilmaier, M., Müller, M.A., Niemeyer, D., Kelly, T.C.J., Vollmar, P., Hoelscher, M., Bleicker, T., et al. (2020). Virological assessment of hospitalized cases of coronavirus disease 2019. MedRxiv.

Xia, N., Wang, G., and Gong, W. (2020). Serological test is an efficient supplement of RNA detection for confirmation of SARS - CoV - 2 infection. Preprints 1–6.

Xiang, J., Yan, M., Li, H., Liu, T., Lin, C., Huang, S., and Shen, C. (2020). Evaluation of Enzyme-Linked Immunoassay and Colloidal Gold- Immunochromatographic Assay Kit for Detection of Novel Coronavirus (SARS-Cov-2) Causing an Outbreak of Pneumonia (COVID-19). MedRxiv 2020.02.27.20028787.

Xie, C., Jiang, L., Huang, G., Pu, H., Gong, B., Lin, H., Ma, S., Chen, X., Long, B., Si, G., et al. (2020). Comparison of different samples for 2019 novel coronavirus detection by nucleic acid amplification tests. Int. J. Infect. Dis.

Young, P.R., Hilditch, P.A., Bletchly, C., and Halloran, W. (2000). An antigen capture enzyme-linked immunosorbent assay reveals high levels of the dengue virus protein NS1 in the sera of infected patients. J. Clin. Microbiol. *38*, 1053–1057.

Zhao, S., Lin, Q., Ran, J., Musa, S.S., Yang, G., Wang, W., Lou, Y., Gao, D., Yang, L., He, D., et al. (2020). Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. Int. J. Infect. Dis. *92*, 214–217.

Figures

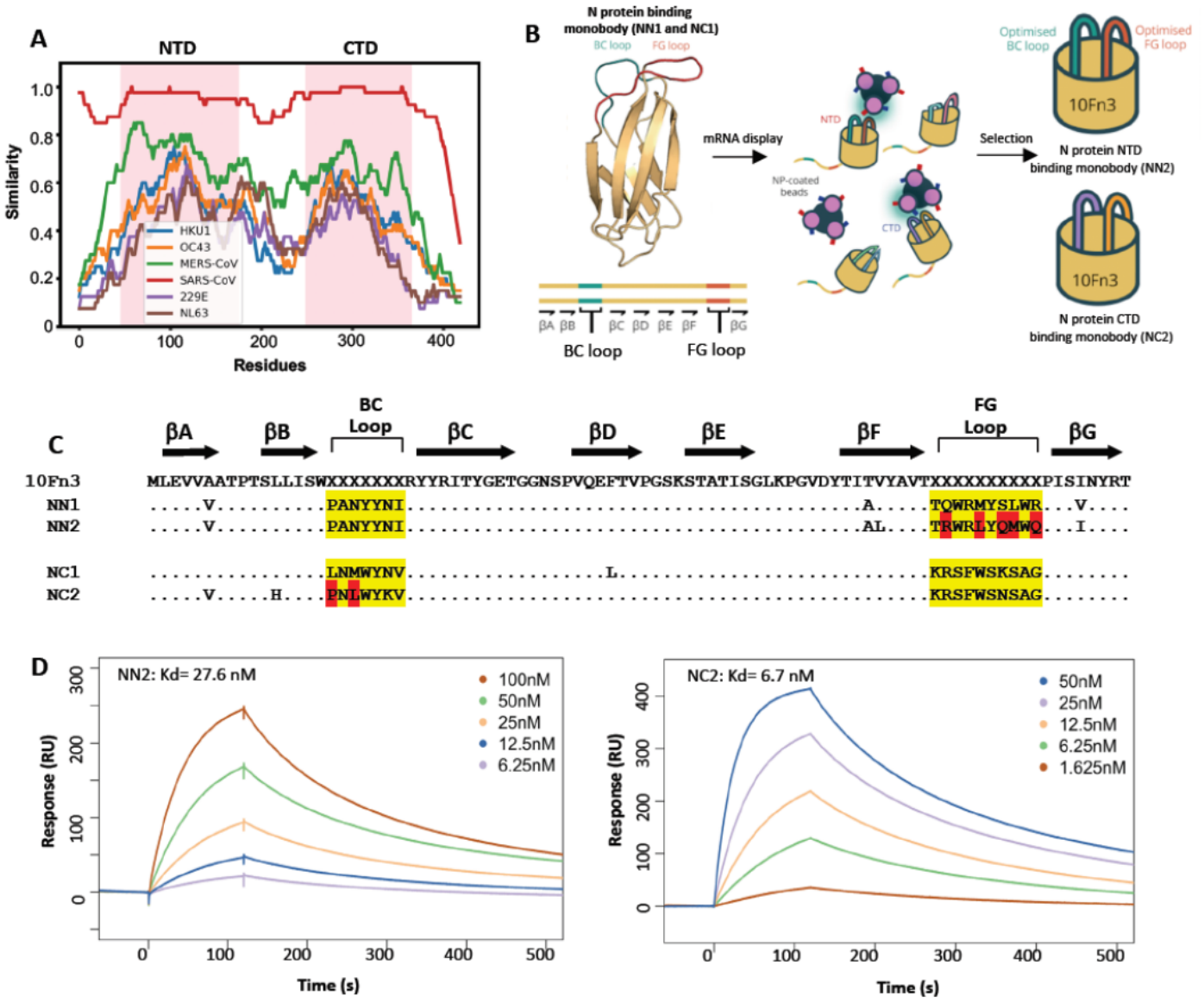

Figure 1

Development of fibronectin-based monobodies against SARS-CoV-2 N (A) : Similarity of N from indicated coronaviruses is shown. Similarity was calculated using Sequence Manipulation Suite, with a sliding window size of 40 residues. Residues 45-175 are highlighted as N-terminal domain (NTD) and residues 248-365 are highlighted as C-terminal domain (CTD) of N. (B) : Schematic plot shows the in vitro maturation of NN1 and NC1 through mRNA display. FC and BG loops were randomized using a doped strategy, and selected against purified SARS N. Six rounds of selection with an increased binding

temperature were performed. The last two pools of each selection were cloned and sequenced. The monobodies with increased binding affinity were identified and confirmed. (C) : Amino acid sequences of the two confirmed monobodies with improved affinity are shown (NN2 and NC2). The sequences were compared with the parental sequences (NN1 and NC1). Yellow shades indicate the residues within the BC and FG loops, and red shades mark the different residues. (D) : SPR analysis of binding affinity of NN2 and NC2 to SARS-CoV-2 N.

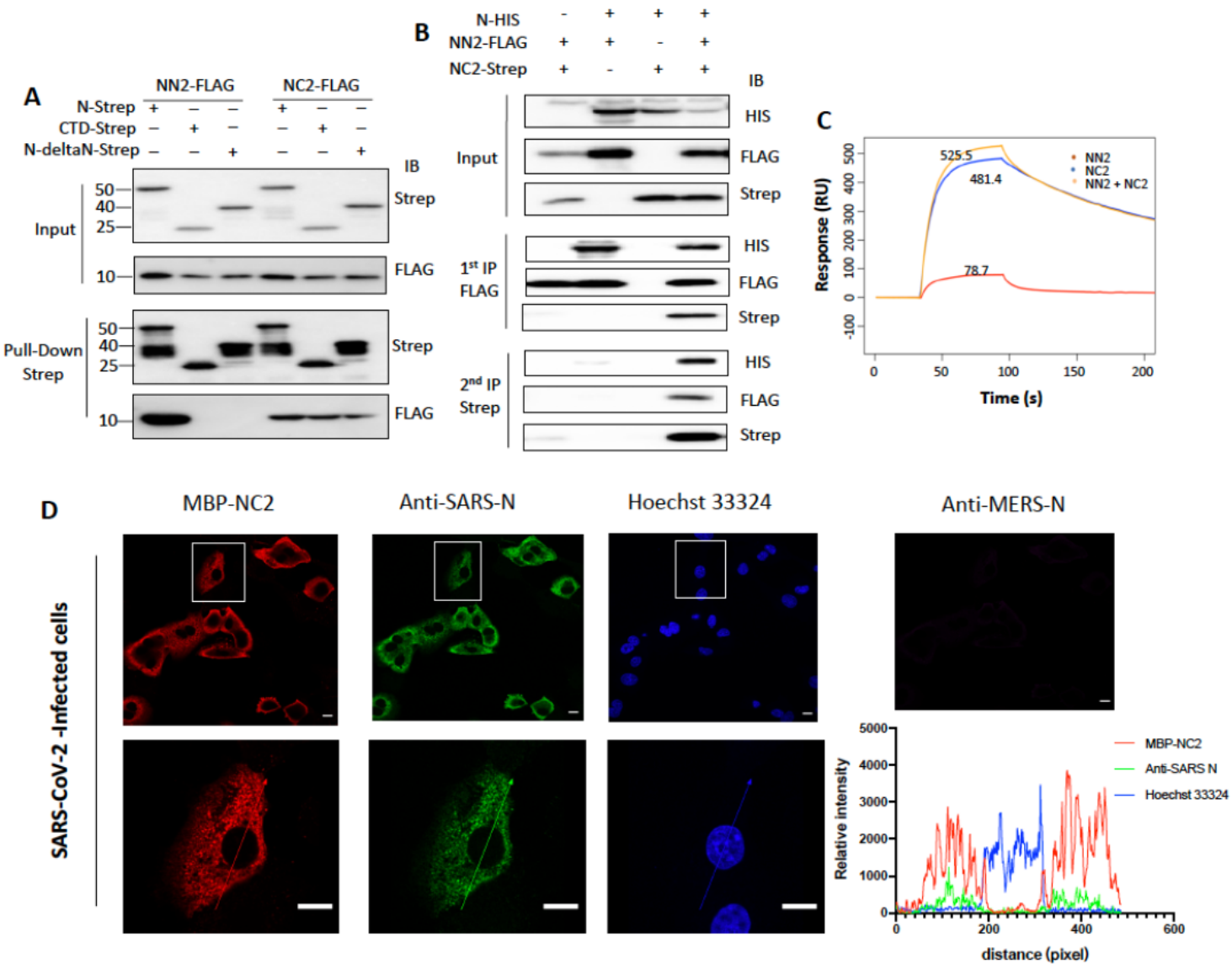

Figure 2

Monobodies recognize endogenous SARS-CoV-2-N in vivo (A) : IP examined the interactions between the indicated domains of SARS-CoV-2 N and monobodies (NN2 and NC2). (B) : Sequential IP indicated that N, NN2 and NC2 form a complex in cells. (C) : Simultaneous binding of NN2 and NC2 at saturating concentrations (200 nM each), examined by SPR. (D) : Immunofluorescence assay (IFA) examined the co-localization of monobodies with N in infected cells. Scale bar represents 10 uM.

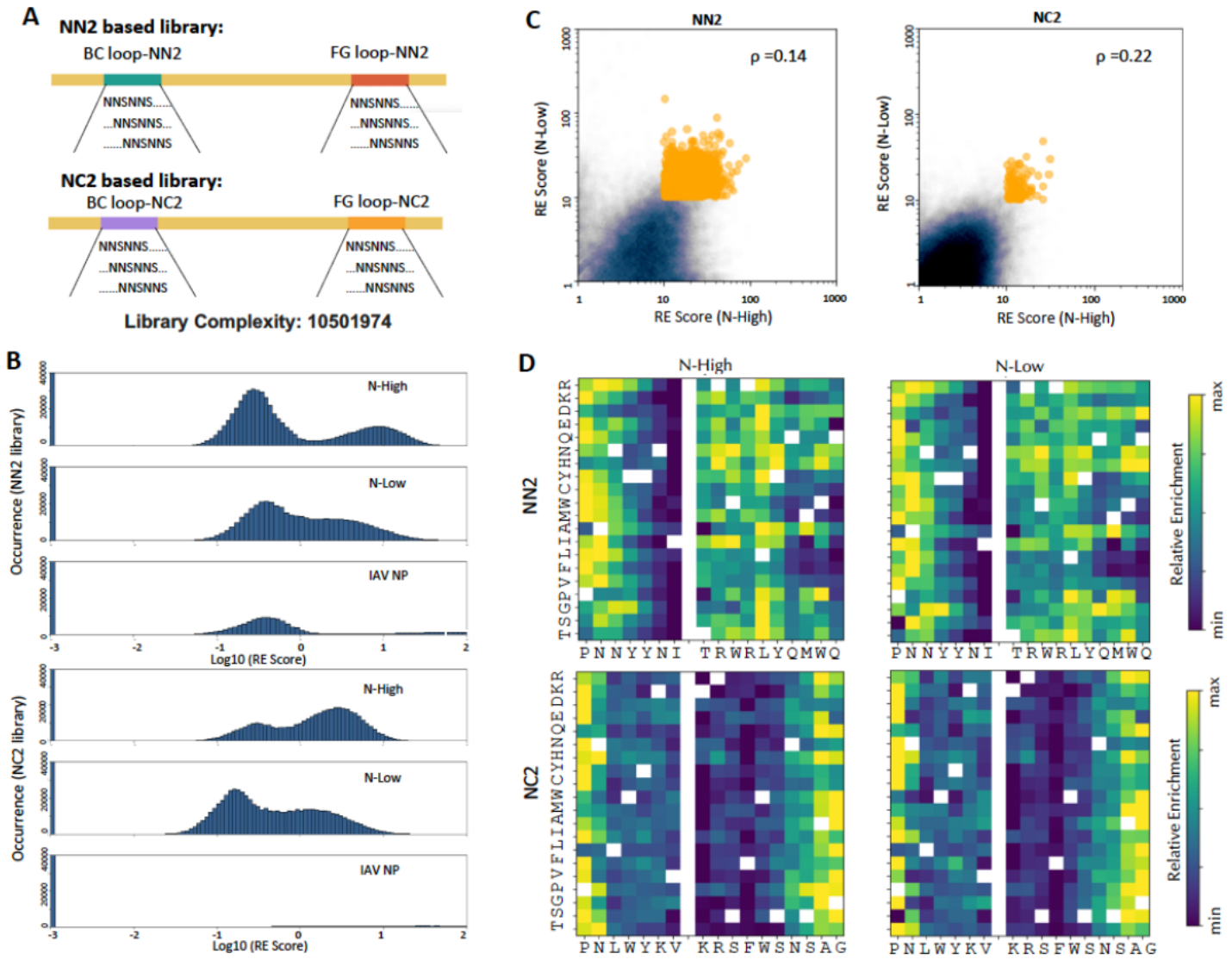

Figure 3

Single-round affinity maturation using mRNA display for the generation of 3rd generation monobodies (A): Schematic plot shows the variation strategy for single-round affinity maturation. (B): Histograms show the distribution of RE scores under indicated conditions. (C) : Scatter plots show the RE scores of individual variants selected with SARS-CoV-2 N at two concentrations. Top candidates with RE scores >10 are colored in orange. □ indicates Spearman Correlations. (D) : Heatmaps show RE scores of amino acid substitutions under indicated conditions, averaged among all genetic backgrounds. Each type of substitution was normalized independently across all randomized positions, showing the maximum and minimum relative enrichment within linear color space. White blocks indicate parental amino acids.

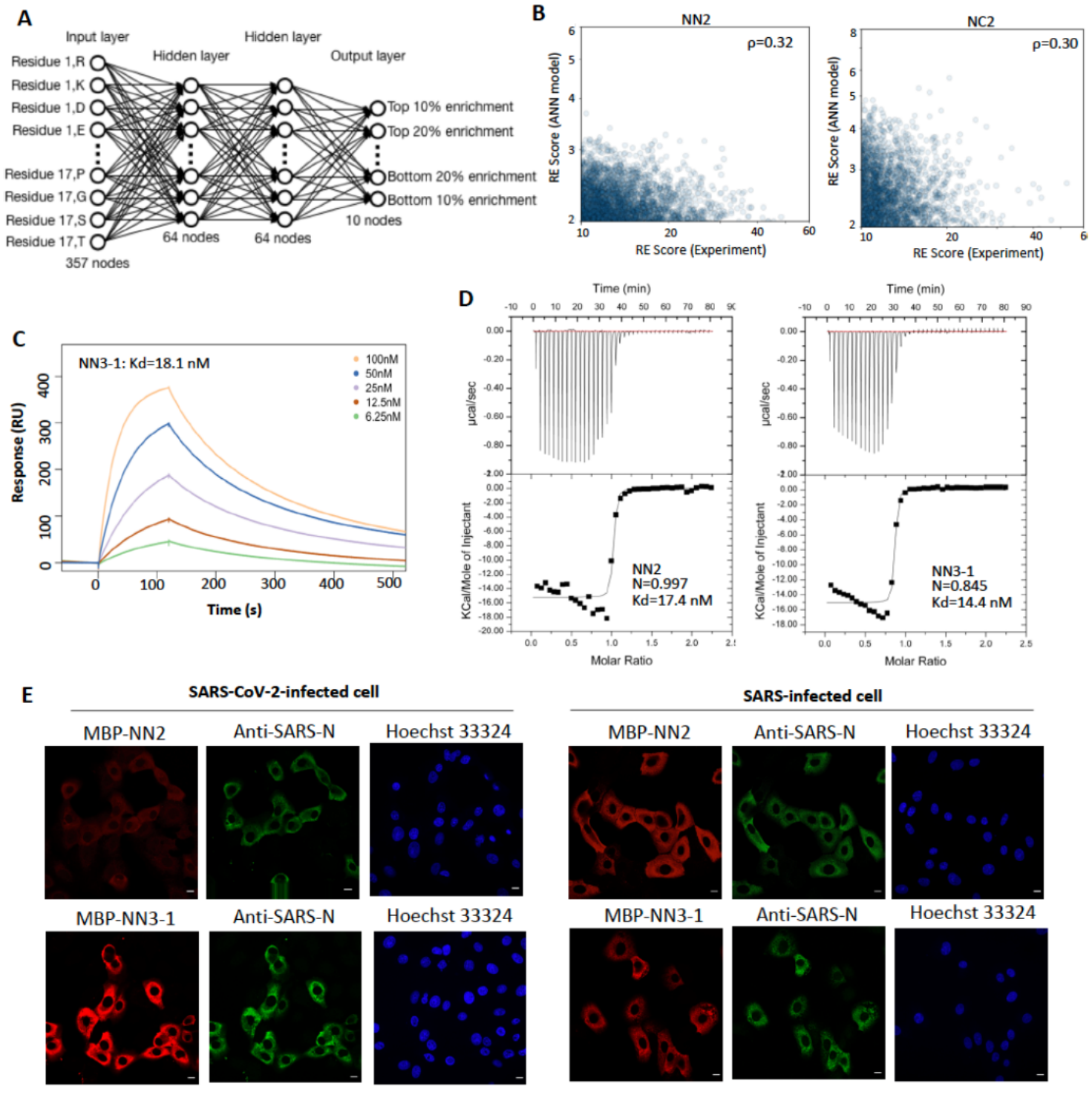

Figure 4

Incorporating machine learning for the validation of 3rd generation monobodies (A) : Diagram shows the applied artificial neural network (ANN) model. (B) : Scatter plots show the correlation between experimental and predicated RE scores of top candidates (experimental RE score > 10 for both conditions, ρ indicates spearman correlation). (C) : SPR analysis of the binding affinity of NN3 to SARS-CoV-2 N. (D) : ITC analysis of the binding affinity of NN2 and NN3 to SARS-CoV-2 N-NTD. (E) : Comparison of IFA in viral infected cells using NN2 and NN3-1. Scale bar represents 10 uM.

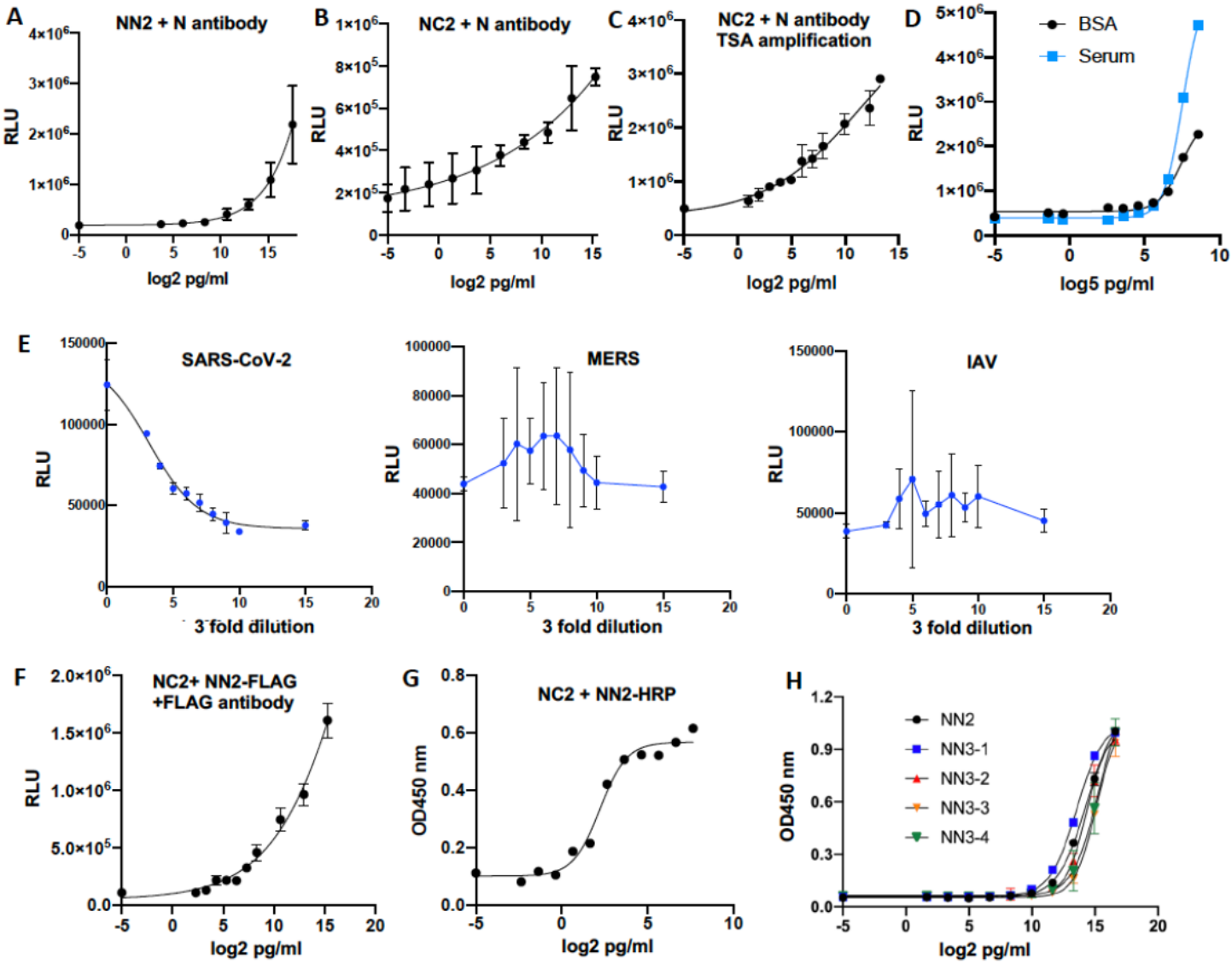

Figure 5

Development of ELISA for N detection using monobodies (A-C): Single monobody-based ELISA for N detection. MBP-tagged NC2 or MBP-tagged NN2 was coated onto ELISA plates, and anti-SARS N antibody was utilized for detection Recombinant N were serially diluted in BSA. TSA was used in panel C (A&B:N=3, C:N=4). (D): Comparison of ELISA detection sensitivity using recombinant N diluted in BSA or serum. (E): Detection of N in viral culture supernatant using ELISA with SARS-CoV-2, MERS or IAV (strain: A/Cal/04/09). Viral culture was serially diluted 1:3 in viral medium/0.5% TritonX-100 (N=3). (F&G) Dual monobody-based ELISA for N detection. MBP-tagged NC2 was coated onto ELISA plates for capturing N. (E) FLAG-tagged NN2 anti-FLAG mouse antibody were used for detection. (F) HRP-conjugated NN2 for detection. (H): Comparison of ELISA detection sensitivity using the 3rd generation monobodies (N=3). Error bars indicate SD for all panels. Detection sensitivity was calculated as the concentration of N that reached statistically different signal comparing to BSA control. Error bars indicate SD. Negative control with no addition of N is showed with -5 on the curve.

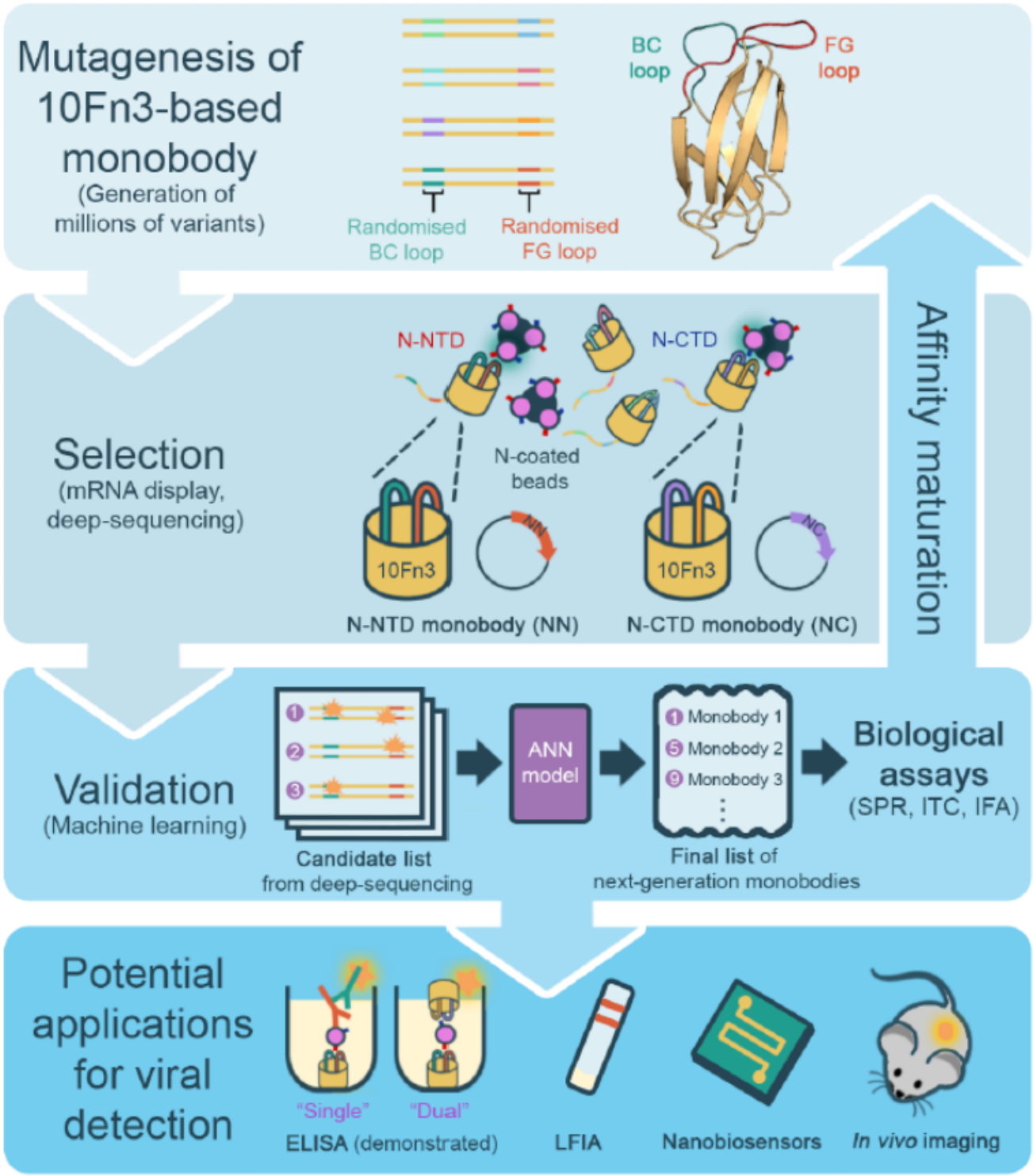

Figure 6

Developing and optimizing monobodies against SARS-CoV2-N Diagram shows the developed platform for identifying, validating and evolving monobodies against SARS-CoV-2 N for potential diagnostic use.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- 2540300supp4569674q95vyl.xlsx
- Supplementalinformation.pdf