

Deep mutational scanning of whole SARS-CoV-2 spike in an inverted infection system

Shunta Taminishi^{1,10}, Songling Li^{2,3,10}, Yusuke Higuchi^{1,10}, Yuhei Kirita⁴, Daisuke Motooka^{5,6}, Yuki Ozaki⁵, Takao Arimori⁷, Nariko Ikemura¹, Yumi Itoh⁸, Satoaki Matoba¹, Toru Okamoto⁸, Junichi Takagi^{6,7,9}, Daron M Standley^{2,3,9,*}, Atsushi Hoshino^{1,11,*}

¹ Department of Cardiovascular Medicine, Graduate School of Medical Science, Kyoto Prefectural University of Medicine, Kyoto, Japan.

² Department of Genome Informatics, Research Institute for Microbial Diseases, Osaka University, Osaka, Japan.

³ Department of Systems Immunology, Immunology Frontier Research Center (IFReC), Osaka University, Osaka, Japan.

⁴ Department of Nephrology, Graduate School of Medical Science, Kyoto Prefectural University of Medicine, Kyoto, Japan.

⁵ Department of Infection Metagenomics, Research Institute for Microbial Diseases, Osaka University, Osaka, Japan.

⁶ Integrated Frontier Research for Medical Science Division, Institute for Open and Transdisciplinary Research Initiatives (OTRI), Osaka University, Osaka, Japan.

⁷ Laboratory for Protein Synthesis and Expression, Institute for Protein Research, Osaka University, Osaka, Japan.

⁸ Department of Microbiology, Juntendo University School of Medicine, Tokyo, Japan

⁹ Center for Infectious Disease Education and Research (CiDER), Osaka University, Osaka, Japan.

¹⁰These authors contributed equally

¹¹Lead contact

*Correspondence: a-hoshi@koto.kpu-m.ac.jp (A.H.); standley@biken.osaka-u.ac.jp (D.M.S.)

Summary

In order to investigate SARS-CoV-2 mutations and their impact on immune evasion and infectivity, we developed a Deep Mutational Scanning (DMS) platform utilizing an inverted infection assay to measure spike expression, ACE2 affinity, and viral infectivity in human cells. Surprisingly, our analysis reveals that spike protein expression, rather than ACE2 affinity, is the primary factor affecting viral infectivity and correlated with SARS-CoV-2 evolution. Notably, within the N-terminal domain (NTD), spike expression and infectivity-enhancing mutations are concentrated in flexible loops. We also observed that Omicron variants BA.1 and BA.2 exhibit immune evasion through receptor binding domain (RBD) mutations, although these mutations reduce structural stability. Interestingly, the NTD has evolved to increase stability, compensating for the RBD instability and resulting in heightened overall infectivity. Our findings, available in SpikeScanDB, emphasize the importance of spike expression levels and compensatory mutations in both the NTD and RBD domains for shaping Omicron variant infectivity.

Key words

SARS-CoV-2, spike mutation, deep mutational scanning, viral evolution, inverted infection assay

Introduction

Since the emergence of SARS-CoV-2, the spike gene has continued to evolve under the selection pressure of better transmissibility and immune evasion. The D614G substitution that appeared early in the COVID-19 pandemic shifted the receptor binding domain (RBD) conformation toward an open state to better bind an entry receptor, ACE2(Korber et al., 2020; Yurkovetskiy et al., 2020). The P681R mutation, highly conserved in Delta lineages, facilitated cleavage of the spike protein and enhanced viral fusogenicity and replication, which resulted in higher transmission and pathogenicity(Mlcochova et al., 2021; Saito et al., 2022). In addition to transmission fitness, SARS-CoV-2 variants have acquired resistance to neutralization from vaccinated and convalescent sera as well as therapeutic antibodies. The Beta and Mu strains exhibited some degree of immune evasion from serum neutralizing antibodies(Uriu et al., 2021). After the Delta wave, the Omicron variant emerged with substantial mutations compared to any previously described SARS-CoV-2 isolates, comprising 26 to 32 amino acid level changes in the spike protein. Among these, 15 mutations are located in the RBD and induced striking immune evasion(Ikemura et al., 2022).

Deep mutational scanning (DMS), which consists of saturation mutagenesis and deep sequencing analysis, is a powerful tool to evaluate the functional alterations in a protein due to amino acid substitutions. DMS has previously been conducted to understand the mutational effects of SARS-CoV-2 spike; however, of these studies employed yeast surface display of the RBD, not full-length spike, which limited the scope to RBD expression, ACE2-binding and antibody escape(Greaney et al., 2021; Starr et al., 2021a; Starr et al., 2021b; Starr et al., 2020). Therefore, the functional roles of other domains have remained largely unknown. Unlike previous studies, we employed a human cell expression system to fully express the entire spike protein and accurately replicate the process of SARS-CoV-2 infection using ACE2-coating pseudovirus. In this system, the spike protein interacted with ACE2 and triggered membrane fusion between the cell and the virus, albeit in an inverted orientation (Ikemura *et al.*, 2022). This approach enabled the direct evaluation of infectivity and an understanding of its relationship with spike expression and binding affinity toward ACE2. The target region could also be expanded to the whole spike, including the N-terminal domain (NTD) and the S1/S2 junction.

Through extensive DMS analysis, we discovered that heightened infectivity was primarily linked to elevated levels of spike protein expression, while enhanced receptor binding domain (RBD) affinity towards ACE2 had a relatively smaller impact on viral infection. The spike RBD is the main target of neutralizing antibodies from vaccinated and convalescent sera, but the anti-NTD antibodies can also moderately neutralize infection(Cerutti et al., 2021; Ikemura *et al.*, 2022; McCallum et al., 2021). Within the NTD, three loops--N1 (residues 14-26), N3 (residues 141-156) and N5 (residues 246-260)--form an antigenic supersite and NTD-targeting antibodies are susceptible to mutations in these loops(Cerutti *et al.*, 2021; McCallum *et al.*, 2021). Functionally, our DMS analysis further unveiled an accumulation of mutations that enhance spike expression and infectivity specifically within flexible loops positioned at the periphery of the spike protein. Domain-wise analysis of Omicron BA.1 and BA.2 spikes indicated that the Omicron RBD lost structural stability in exchange for a gain in immune evasion, but that NTD mutations compensated for the fragile Omicron RBD in order to preserve infectivity. When BA.1 NTD mutations were

categorized into 4 clusters, only G142D/ Δ 143-145 contributed to immune evasion consistent with previous NTD antigenicity analysis. The additional mutations, A67V/ Δ 69-70, T95I, and Δ 211/L212I/ins214EPE, were involved in the preservation of spike expression and infectivity. DMS in the inverted infection assay thus provided comprehensive insights into functional alterations upon spike mutation, which has further deepened our understanding of viral evolution.

Results

Deep mutational scanning in an inversely oriented infection assay

To understand the effect of SARS-CoV-2 spike mutations on its characteristics, including protein expression, binding-affinity toward ACE2 and consequent viral infectivity, we conducted deep mutational scanning of the whole spike in the context of full-length spike protein expression in human Expi293F cells(Ikemura *et al.*, 2022). Original signal sequence was replaced with HMM38 synthetic one(Barash *et al.*, 2002) and HA tag was inserted in the N-terminal of the spike. Assay was independently conducted in each domain and some restriction sites were introduced between the domains for library construction (Figure 1A). The HA-tagged spike library encompassing all 20–amino acid substitutions based on the ancestral SARS-CoV-2 Wuhan strain was transfected into Expi293F cells under conditions where cells in general expressed no more than a single variant(Chan *et al.*, 2021) (Figure 1B). The spike expression level was determined by anti-HA tag antibody and the binding affinity was evaluated by two-dimensional analysis with anti-HA tag antibody and ACE2-Fc chimera protein. For spike expression analysis, the top and bottom 25% of cells among HA-positive population were collected by fluorescence-activated cell sorting (FACS) (Figure 1C). Similarly, the highest and lowest 25% of cells in terms of the ACE2-Fc signal were gated in the distribution with spike expression level to analyze the binding affinity (Figure 1D). The DMS library expressing cells were also incubated with ACE2-coating GFP reporter pseudoviruses. This infection assay functionally reproduced the process of membrane fusion with spike and ACE2 between viruses and host cells, even though it was in an inverse orientation (Figure 1E). This inverted infection system enabled sequencing of the spike mutants in a large-scale pool library. In the RNA-based deep sequencing of DMS library, the read count ratio of two populations was normalized by the silent mutation construct and log 10 scaling was used to define the spike expression, affinity, or infectivity value. To validate the accuracy of the inverted infection assay, we compared it with the SARS-CoV-2 pseudovirus infection assay and observed high correlation of infection rate in major SARS-CoV-2 variants (Fig. 1F). GFP-positive cells with ACE2-coating pseudovirus infection and GFP-negative control cells were harvested and log 10 scaling of the read count ratio was similarly defined as infectivity value.

DMS in the N-terminal domain

In addition to the RBD, SARS-CoV-2 variants frequently acquire mutations in the spike NTD. To comprehensively understand the functional importance of these mutations, we conducted DMS for the NTD. DMS experiments were performed in duplicate and produced similar results; the R² (coefficient of determination) values were 0.6039, 0.8725, and 0.8589 for infectivity, expression, and binding affinity,

respectively (Figure S1A). An overview of how amino acid mutations affected infectivity, expression and ACE2 binding affinity was visualized as heatmaps (Figure 2A). We next performed a series of experiments to confirm the accuracy of these DMS results using NTD mutations observed in major variants, including L18F and D80A in Beta, D138Y and R190S in Gamma, T19R and G142D in Delta, and T76I in Lambda. Infectivity was evaluated by luciferase reporter pseudovirus infection in ACE2-expressing HEK293T (HEK293T/ACE2) cells. The pseudovirus harbored a small luminescent peptide tag, HiBiT to precisely normalize input virus doses (Ozono et al., 2021). There was strong correlation ($R^2=0.8599$) between DMS infectivity values and the log 10 scaled luciferase signal relative to that of parental virus. The individual mutant spike expression was determined by flow cytometry for HA-tagged spike protein. Similarly to infectivity, the log 10 scaled HA signal relative to that of parent virus was well correlated with the DMS value ($R^2=0.8914$). The individual affinity was quantified as the ACE2-Fc signal in an arbitrary range of spike expression (Figure S1B). There was reasonable correlation between the log 10 scaled relative ACE2-Fc signal and DMS affinity ($R^2=0.5993$) (Figure 2B). Together, these robust validations supported the quality of the DMS data.

Infectivity is dominantly determined by spike protein expression level

Heatmap visualization indicated a similar pattern in infectivity and spike expression (Figure 2A). Statistically, infectivity was substantially correlated with expression ($R^2=0.7206$), but not with binding-affinity ($R^2=0.0052$) (Figure 3A). Mapping to the spike trimer structure indicated that mutants with enhanced infectivity and spike expression accumulated in the five NTD-loop regions that are located at the periphery of the spike. In contrast, mutations enhancing binding affinity were enriched in the region facing the RBD or S2 (Figure 3B). Many of these mutations were associated with decreased protein expression and therefore were less infectious overall (Figure 2A). Despite the dominant impact of spike expression on infectivity, there were some exceptional residues where binding affinity determined the infectivity. Among N-glycosylation sites, mutations in N165/T167 increased the binding affinity and some amino acid substitutions resulted in higher infectivity (Figures 2A and 3B). Conversely, mutations in N234/T236 decreased the affinity and infectivity, even though they increased spike expression (Figures 2A and 3C). The structural and functional roles of spike protein glycosylation have been analyzed by structural studies and molecular dynamics (MD) simulations starting from the atomic structures (Casalino et al., 2020; Choi et al., 2021; Mori et al., 2021). When the RBD is closed, the N165 glycan is located above the RBD and N234 glycan is located below the RBD, where both glycans stabilize the closed conformation with frequent interaction with the closed-state RBD (Casalino et al., 2020; Choi et al., 2021; Mori et al., 2021). When the RBD is open, the high-mannose-type N234 glycan intrudes in a cavity between the RBD and the S2 subunit and forms hydrogen bonds with S2, reinforcing the open-state from below (Choi et al., 2021; Mori et al., 2021). Our DMS results were consistent with these mechanistic models, although one previous study showed that deletion of the N165 glycan reduced the binding affinity toward ACE2 (Casalino et al., 2020). At sites unrelated to glycosylation, increased infectivity due to allosterically enhanced binding affinity was observed in L48, S50, K113, and Q115 (Figures 2A and 3B).

Increased binding affinity contributes little to infectivity

DMS for affinity was also conducted in the RBD region and we analyzed this data together with infectivity and spike expression data that we previously reported(Ikemura *et al.*, 2022) (Figure S2A). Similarly to DMS for the NTD, dominant involvement of spike expression in the infectivity was observed in the RBD ($R^2=0.712$; Figure S2B). Previously, Starr *et al.* conducted DMS for the RBD and evaluated the total ACE2 binding, not affinity, as a surrogate for infectivity. In their data, ACE2 binding correlated well with expression ($R^2=0.4278$; Figure S2C). To further examine the contribution of affinity-enhancing RBD-ACE2 interface mutations to infectivity, we prepared cells expressing the ACE2 mutant 3N39, which has a 100-fold higher affinity (Higuchi *et al.*, 2021) and observed viral infectivity. As protein expression levels affect infectivity, HA tagged wild-type ACE2 and 3N39 were stably expressed in HEK293T cells with lentivirus and cells were sorted according to the arbitrary HA signal to equalize ACE2 expression between wild-type and 3N39 (Figure S3A). Authentic SARS-CoV-2 infection and replication were comparable between wild-type ACE2 and 3N39 expressing cells (Figure S3B). Similar infectivity was observed in a SARS-CoV-2 pseudovirus infection assay (Figure S3C). These results collectively emphasized the importance of spike protein expression levels in viral infectivity, as compared with the binding affinity toward ACE2.

DMS in the S1/S2 junction

S1/S2 junction from F541 to S750 was analyzed in duplicate with high reproducibility (Figure S4A and B). D614G is the first predominant circulating viral mutant and increased infectivity and transmissibility was reported both in vitro and in vivo(Hou *et al.*, 2020; Ozono *et al.*, 2021). The D614G mutation breaks a salt bridge between D614 and K854 in the fusion peptide proximal region, which contributes to the RBD-up conformation(Cai *et al.*, 2020). In addition, the disordered loop in the D614 S trimer wedges between domains within a protomer in the G614 spike. This interaction increases the number of functional spikes and enhances infectivity(Zhang *et al.*, 2021). Consistent with this mechanism, our DMS results indicated that the D614G mutation enhanced affinity, spike expression, and consequent infectivity (Figure S4A). The dominant contribution of spike expression to infectivity was also observed in the S1/S2 junction (Figure S4C). Interestingly, affinity-enhancing mutations were frequently observed here unlike other domains and they tended to be negatively correlated with spike expression. One exception was the furin cleavage site, R682 to V687, where both affinity and spike expression were strikingly upregulated. Furin recognition site mutations resulted in increased affinity and stabilized the S1/S2 structure through the prevention of S1/S2 cleavage; However, infectivity was little affected. There are two pathways for SARS-CoV-2 infection. One is the S2' site cleavage by TMPRSS2 and cellular entry (cell surface entry). The other is endocytosis and lysosomal cathepsin-mediated spike cleavage, which occurs when TMPRSS2 is unavailable (endosomal entry)(Jackson *et al.*, 2022; Takeda, 2022). Furin-mediated S1/S2 cleavage promotes S1 shedding and exposes the S2' site to TMPRSS2(Jackson *et al.*, 2022). It means that the loss of furin-cleavage site impairs cell surface entry but consequent structural stability and increased affinity facilitate endosomal entry. Since the HEK293T cells we used here modestly express TMPRSS2(Partridge *et al.*, 2021), the

overall infectivity was nearly unchanged.

Spike expression level is involved in SARS-CoV-2 evolution

We next examined the contribution of spike expression and ACE2-binding affinity to human SARS-CoV-2 evolution. Mutation frequency in naturally circulating SARS-CoV-2 was adopted from the previous report which included mutations expected to have ≥ 20 occurrences (Dadonaite et al., 2023). The results of DMS analyzing infectivity and spike expression were reasonably correlated with the enrichment of mutations observed in naturally circulating SARS-CoV-2, whereas DMS for affinity exhibited no association (Figures 4A-C). Consistent with the impact on infectivity, these results indicate that spike expression, not ACE2 binding affinity, is the predominant factor for SARS-CoV-2 evolution. This trend was observed similarly in NTD, RBD and S1/S2 junction (Figures S5A-C). Our DMS in the inverted infection assay was correlated with the previous DMS for pseudovirus infection (Figure 4D) and both DMS measurements comparably reflected the actual enrichment of SARS-CoV-2 mutations (Dadonaite et al., 2023). The DMS for infection reasonably correlated with viral evolution, but it was somewhat limited. The first reason is that not only infectivity but replication activity determines virus transmissibility. The P681R mutation is known to be a determinant for enhanced fusogenicity and viral replication fitness of the Delta strain (Liu et al., 2022; Saito et al., 2022); other spike mutations could similarly affect transmissibility through replication activity. The second is epistasis. It is known that effects of some mutations in Omicron spike RBD differ from those measured in the ancestral background. For instance, Q498R and N501Y substitutions contribute to ACE2 binding in Omicron with more impact than in earlier viral strains (Starr et al., 2022). Concomitant mutations could contribute nonadditively compared with single amino acid mutations. The last reason is that immune-evading mutations is a driving force for SARS-CoV-2 evolution and reduces the impact of infectivity on fitness in naturally circulating viruses.

In position-wise analysis, previous DMS studies reported a correlation between relative accessible surface area (RSA) and spike expression-affecting mutations in the RBD (Ouyang et al., 2022; Starr et al., 2020). In general, solvent inaccessible buried residues are known to be important for proper protein folding and structural stability. Here, structural flexibility was further quantified by the change of RSA of each residue position in 607 trimeric spike complex structures collected from Protein Data Bank (PDB) (Bekker et al., 2022). Spike complexes were decomposed into monomeric chains, from which residue-wise RSA by the program NACCESS (Hubbard and Thornton, 1993) and standard deviation of RSA (dRSA) values were calculated. Residues with smaller dRSA values corresponded to buried or conserved interfaces residues mediating stable interactions between spike chains. In contrast, those with higher dRSA values were more frequently found on the surface or at structurally dynamic interfaces between spike chains (Figure S6A). When dRSA values were compared with DMS expression scores, a qualitative agreement was apparent. This was most obvious within the NTD, where mutations in residues with lower dRSA values were likely to reduce protein expression, while mutations in residues with higher dRSA values were tolerated or even associated with increased expression (Figure S6B).

NTD mutations compensate for RBD instability in Omicron variants

DMS data indicated that mutations altering spike protein expression dominantly contributed to viral infectivity. Previously we and others reported that immune evasion properties of Omicron were achieved mainly by RBD mutations while NTD mutations had more modest impact (Cui et al., 2022; Ikemura *et al.*, 2022). Next, we examined the contribution of Omicron NTD mutations on infectivity and spike protein expression. We generated chimera spikes (C1-C6) with Wuhan-Hu-1 (D614G) and Omicron (BA.1) strains (Figure 5A) and examined the infectivity and spike expression. In the chimera spike with Wuhan background, Omicron NTD replacement substantially increased infectivity, whereas Omicron RBD replacement reduced it. In the case of Omicron background, Wuhan NTD replacement rendered the virus less infective, but Wuhan RBD replacement enhanced infectivity (Figure 5B). The same trend was observed in the spike protein expression on the cell surface as well as pseudovirus particles, where Omicron NTD and Wuhan RBD increased the spike protein expression, while Wuhan NTD and Omicron RBD reduced it (Figures 5C and 5D). These results indicated that Omicron RBD lost structural stability in exchange for gained immune evasion, and that the Omicron NTD evolution compensated for the RBD instability. Consistent with this hypothesis, when we generated the RBD-Fc protein, the Omicron RBD was low yield and highly prone to aggregation (Figure S7). To evaluate the stability of spike proteins, alteration of pseudovirus infectivity on plastic surface were measured over time. As expected, Wuhan with Omicron NTD (C1) exhibited stable infectivity, while Omicron with Wuhan NTD (C4) rendered the virus more fragile (Figure 5E).

To further understand the impact of each mutation on the Omicron NTD, we separated NTD mutations into 4 groups: NTD1, A67V/ Δ 69-70; NTD2, T95I; NTD3, G142D/ Δ 143-145; NTD4, Δ 211/L212I/ins214EPE (Figure 6A). When NTD group mutations were extracted from Omicron BA.1, NTD1, 2, and 4 exhibited impaired infectivity, but NTD3 did not (Figure 6B). In the case of introducing BA.1 NTD mutations in Wuhan-Hu-1 (D614G), NTD1,2, and 4 enhanced infectivity and NTD3 reduced it (Figure 6C). To further validate these observations, we also analyzed them in the chimeric Omicron spike with Wuhan NTD (C4) and confirmed no contribution of NTD3 to the enhanced infectivity (Figure 6D). Regarding spike protein expression, removal of NTD1, 2, and 4 reduced the expression, while their introduction into Wuhan NTD enhanced it in both Wuhan and C4 chimeric spike expression (Figures 6E-G). We previously reported that BA.1 NTD also modestly contributed to the escape from the neutralization of vaccinated sera (Ikemura *et al.*, 2022). In neutralization assays with chimeric spike pseudoviruses, only NTD3 was responsible for immune evasion (Figures 6H). These intensive chimeric spike experiments revealed the evolutionary significance of Omicron BA.1 NTD, where NTD1, 2, 4 mutations compensated the spike stability and the NTD3 mutation contributed to immune evasion.

We also conducted domain analysis in another Omicron variant, BA.2 with a similar lineup of chimeric spikes (Figure S8A). BA.2 NTD replacement enhanced both infectivity while spike expression and BA.2 RBD replacement conversely impaired them in Wuhan spike. Consistently, higher infectivity and spike expression were observed in BA.2 spike with Wuhan RBD, while BA.2 with Wuhan NTD exhibited the opposite phenotypes (Figures S8B and C). For further domain analysis, BA.2 NTD mutations were

separated into 3 groups; NTD1, T19I/L24S/ Δ 25-27; NTD2, G142D; NTD3, V213G (Figure S8D). BA.2 NTD2 dominantly, and NTD3 modestly, improved the expression and infectivity, while NTD1 decreased both (Figures S8E-H). Interestingly, G142D with Δ 143-145 enhanced immune evasion, but did not preserve infectivity, in BA.1; whereas, G142D alone strongly enhanced infectivity in BA.2.

SARS-CoV-1 NTD contributes to higher infectivity

Although the spike RBD of SARS-CoV-2 has higher affinity to ACE2 than that of SARS-CoV-1(Lan et al., 2020; Nguyen et al., 2020; Shang et al., 2020), SARS-CoV-2 infectivity is much lower than SARS-CoV-1, at least in pseudovirus infection assays(Peacock et al., 2021; Zeng et al., 2020). SARS-CoV-2 is characterized by the gain of a furin cleavage site in the S1/S2 junction(Lan et al., 2020; Shang et al., 2020) and exhibits increased TMPRSS2-mediated infection and fusogenicity, which results in the promotion of virus transmissibility and pathogenicity(Peacock et al., 2021; Saito et al., 2022). In contrast, the biological meaning of alteration in the NTD remains poorly understood. To investigate this question, we constructed chimeric spikes of SARS-CoV1 and SARS-CoV-2 (Figure 7A). SARS-CoV-2 exhibited lower infectivity than SARS-CoV-1, consistent with other reports, and SARS-CoV-1 NTD replacement in SARS-CoV-2 improved infectivity. Conversely, SARS-CoV-1 infectivity was strikingly attenuated with SARS-CoV-2 NTD (Figure 7B). Spike expression was also altered in parallel with the infectivity (Figure 7C). Our binding affinity assay confirmed higher ACE2 affinity in the spike RBD of SARS-CoV-2 (Figure 7D). These results suggest that the NTD of SARS-CoV-1 was more fit to infect human cells.

Next, we examined the fitness of NTD from other sarbecoviruses and major SARS-CoV-2 variants in the context of chimera spike with Wuhan (Figures S9A-B). Correlation between spike expression and infectivity was also observed here and PG-GX-P5L and PG-GD-1 NTDs exhibited better fitness (Figure 7E). In the case of SARS-CoV-2 variants, BA.1 and BA.2 NTDs achieved the highest spike expression, indicating that the Omicron NTD further evolved to compensate the fragile RBD due to accumulated mutations for immune evasion (Figure 7F).

Discussion

We have developed a DMS platform with an inverted infection assay to comprehensively evaluate viral infectivity, spike expression, affinity toward receptor and their correlation. The inverted infection assay uniquely employs ACE2-coating pseudovirus and spike-expressing cells. Naturally, SARS-CoV-2 spikes bind to cell surface ACE2 and protease-mediated cleavage of spike facilitates the membrane fusion between virus and cell. This process can be reproduced in the context of inverted positioning of spike and ACE2, and the expression of spike in cells allows to conduct high-throughput pooled library screening. Recently, pseudovirus-based spike DMS was performed to analyze infectivity and antibody escape but pseudotyped lentivirus DMS library was limited due to uncontrolled combination of mutations due to lentiviral swapping characteristics (Dadonaite et al., 2023). In comparison, our DMS assay system is all-encompassing and infectivity measurements similarly correlated with the enrichment of mutations in actual SARS-CoV-2 sequences. DMS analysis for SARS-CoV-2 has been intensively conducted by yeast surface

display of the RBD and generated much data for RBD expression and binding to ACE2 or antibodies (Greaney *et al.*, 2021; Starr *et al.*, 2021a; Starr *et al.*, 2021b; Starr *et al.*, 2020). Although these DMS experiments provided important insights into the effect of viral mutations on infectivity and immune evasion, the region of study was limited to the RBD. In the context of these DMS data, the present study provides breadth and also reveals that spike protein expression is the primarily driving force for viral infectivity.

Based on DMS in the NTD, mutations enhancing spike expression and infectivity clustered in the loop regions and structurally fragile mutations were located at the interface to the RBD and S2. It is known that N1, N3 and N5 loop regions form an antigenic supersite and are targeted by neutralizing antibodies (Cerutti *et al.*, 2021; McCallum *et al.*, 2021). As the mutation frequency of naturally circulating mutants is high in the loop regions, it was expected that these mutations were selected by avoidance of neutralization by antibodies. However, they can also surprisingly increase spike expression and thus infectivity. Actually, in the case of BA.1 NTD, A67V/Δ69-70, T95I, and Δ211/L212I/ins214EPE mutations increase infectivity while G142D/Δ143-145 mutation contributes to immune evasion. A recent study also performed DMS on the NTD for spike protein expression in the context of full-length spike expression in HEK293T cells (Ouyang *et al.*, 2022). Even though mutagenesis coverage was only 69%, this DMS similarly identified a trend wherein mutation tolerability was associated with distance to the RBD/S2. Omicron BA.1 contains several mutations in NTD loop regions and domain-wise experiments demonstrated that BA.1 NTD increased the environmental stability of the virus (Figure 5E). These results indicate that NTD loops are structurally flexible but, interestingly, a number of mutations facilitate increase in protein folding stability.

It is also surprising that spike protein expression is the predominant factor affecting viral infectivity rather than ACE2 affinity. Previously we generated high affinity ACE2 mutants to develop SARS-CoV-2 neutralizing ACE2 decoys (Higuchi *et al.*, 2021; Ikemura *et al.*, 2022). When this high affinity ACE2 was expressed in HEK 293T cells, pseudovirus infection rate was comparable to the wild-type ACE2 expression (Figure S3B). In another experiment, we isolated the mutant virus escaping wild-type ACE2 decoy and identified Y489H as the mutation responsible for this decoy escape. The Y489H mutation reduced ACE2 affinity and enabled escape from the ACE2 decoy. But this mutation also increased spike expression and overall infectivity was compensatively preserved (Urano *et al.*, 2023). Similar decoy escape was reported in the case of CD4 decoys against HIV infection and it was suggested that nanoparticle coating CD4 could overcome the escape due to the avidity effect (Hoffmann *et al.*, 2020). These results indicate the possibility that affinity between spike and ACE2 is saturated for effective infectivity and avidity effect due to higher spike expression dominantly determines viral infection activity.

The impact of NTD mutations on the spike expression and infectivity reinforces their contribution to SARS-CoV-2 evolution. The domain-wise analysis in Omicron BA.1 and BA.2 exhibited that Omicron RBD lost structural stability in exchange for immune evasion but the NTD evolved to preserve spike expression and infectivity. Interestingly, in the NTD of BA.1, only the G142D/Δ143-145 mutation exhibited no impact on the spike stability and simply contributed to antibody escape. BA.2 mutations in the N1 loop, T19I/L24S/Δ25-27, induced instability but mutations in this antigen supersite are also expected to affect

immune evasion. Another study similarly reported domain-wise functional alteration in Omicron (Javanmardi et al., 2022), which made this evidence more convincing. In addition, we also examined the NTD of other major variants and found that BA.1, BA.2, and Alpha NTDs remarkably increased spike expression and infectivity. Moreover, NTD loop regions were poorly conserved among sarbecoviruses. When the NTD of SARS-CoV-2 was replaced with SARS-CoV-1, PG-GX-P5L or PG-GD-1, the infectivity was increased, indicating that SARS-CoV-2 could further evolve by recombination with zoonotic sarbecoviruses.

The DMS in the context of ACE2-coating pseudovirus infection into spike-expressing cells we describe here enables direct assessment of the alteration of infectivity due to single missense mutations. Combining with DMS analyzing spike expression and ACE2 affinity provides important insights into virus evolution, vaccine design and drug development. This system can also be applied to understand mutational escape from neutralizing reagents (Alcantara et al., 2023; Ikemura *et al.*, 2022) and provides valuable information for antigenic surveillance.

Acknowledgements

We would like to thank Takaaki Nakaya (Kyoto Prefectural University of Medicine) for helpful discussion; Kenzo Tokunaga (Department of Pathology, National Institute of Infectious Diseases) for the kind gift of the plasmid coding psPAX2-IN/HiBiT. This work was supported by the Japan Agency for Medical Research and Development (AMED), the Research Program on Emerging and Re-emerging Infectious Diseases under JP21fk0108465, JP21fk0108481, and 22fk0108523 (to A.H., J.T. and T.O.), the Platform Project for Supporting Drug Discovery and Life Science Research (Basis for Supporting Innovative Drug Discovery and Life Science Research) under JP21am0101075(2527) (to J.T.) and JP21am0101108 (to D.M.S), a grant from the Cell Science Research Foundation (to A.H.).

Author contribution

A.H designed the research; S.T. and Y.H. conducted DMS experiments; S.M., Y.H. and N.I. performed pseudovirus neutralization assay; T.I. provided clinical serum samples from vaccinated individuals; S.L., Y.K., D.M., and Y.O. performed and analyzed next-generation sequencing; T.A., J.T. purified and prepared the proteins; S.L. and D.M.S. manage SpikeScanDB; Y.I. and T.O. conducted authentic virus experiments; A.H., S.L. and D.M.S. wrote the manuscript; all authors discussed the results and commented on the manuscript.

Declaration of interests

The authors declare no competing interests.

Figure legends

Figure 1. Deep mutational scanning in the inverted infection assay.

(A) A design of the deep mutational scanning (DMS) library for the spike of SARS-CoV-2. HA tag was introduced in the N-terminus of the spike. Restriction sites were also introduced between the domains to insert library fragments.

(B-E) Schematic of deep mutational scanning (DMS). DMS library was developed using pooled synthesized DNS with NNK mixed bases and expressed Expi293F cells in the context of full-length spike (B). Spike expression level was analyzed with the gating of the highest and lowest 25% HA signal among HA-positive cells C. For affinity analysis, library expressing cells were incubated with anti-HA antibody and ACE2-Fc. The highest and lowest 25% of cells with ACE2-Fc binding relative to HA-spike expression were collected (D). Infectivity was examined in the assay where DMS library expressing cells were incubated with ACE2-coating GFP reporter pseudovirus. GFP-positive and negative cells was sorted (E).

(F) Correlation between luciferase-reporter SARS-CoV-2 pseudoviruses and inverted infection assay among major variants. Data are mean \pm SEM from three biological replicates.

Figure 2. Deep mutational scanning for N-terminal domain.

(A) Heatmap showing how all single mutations affect the infectivity, spike expression, and affinity toward ACE2 in the N-terminal domain. Squares are colored by mutational effect according to scale bars on the bottom, with blue indicating deleterious mutations. Squares with a diagonal line through them indicate the original Wuhan strain amino acid.

(B) Validation of DMS results using NTD mutations including major variants ones. Infectivity was verified by the standard infection assay with luciferase reporter SARS-CoV-2 pseudovirus and HEK293T/ACE2 cells. Spike expression was examined by the same assay as DMS and the strategy of affinity assessment was explained in Figure S1B.

Figure 3. The spike expression mainly contributes to SARS-CoV-2 infectivity.

(A) Correlation between functional alterations of infectivity, spike expression and affinity in DMS for the NTD.

(B) Projection of functional changes due to mutations onto SARS-CoV-2 spike trimer with one RBD in the open state built via homology modelling. The values of infectivity, expression and affinity were summed up in each residue and indicated according to scale bars on the right. Sphere models are indicated residues in the results. ACE2 is colored in purple. Each spike monomer is in light, middle, and dark gray.

Figure 4. Functional effects of spike mutations on the viral transmission fitness

(A-C) Correlation between the mutation frequency of naturally circulating SARS-CoV-2 and functional alteration of spike including infectivity (A), spike expression (B), and affinity toward ACE2 (C). These correlations in each domain were presented in Figure S5.

(D) Correlation between our infectivity DMS in the inverted infection assay and Dadonaite et al. pseudotyped lentivirus-mediated DMS(Dadonaite *et al.*, 2023).

Figure 5. Omicron BA.1 NTD compensates the RBD structural instability.

(A) Schematic of chimera spike used in this study. Spikes of Wuhan (D614G) and BA.1 were mixed in each domain and resultant chimera spikes were referred to as C1 to C6.

(B) Infectivity of chimera spike-coating luciferase reporter pseudoviruses in HEK293T/ACE2 cells (left) and Calu-3 cells (right).

(C) Spike expression of each chimera spike in HEK293T cells. Tagged HA was quantified by flow cytometry for anti HA antibody.

(D) Spike protein in lentivirus particles pseudotyped with chimera spikes. The p24 lentivirus protein was used as an internal control.

(E) Stability of pseudovirus particles on the surface of polystyrene plate. Virus infectivity was analyzed in each time point. * $P=0.0004$ vs. D614G, $P=0.0002$ vs.BA.1, $P<0.0001$ vs. C4. P -values were determined by one-way ANOVA and Dunnett's multiple comparison test.

For all plots, mean \pm SD from at least two biological replicates.

Figure 6. BA.1 NTD mutation enhances structural stability or immune evasion.

(A) Schematic of BA.1 NTD mutation clusters used in this study. NTD1, A67V/ Δ 69-70; NTD2, T95I; NTD3, G142D/ Δ 143-145; NTD4, Δ 211/L212I/ins214EPE.

(B) Infectivity of luciferase reporter pseudoviruses carrying BA.1 spike with subtraction of NTD mutation cluster in HEK293T/ACE2 cells.

(C) Infectivity of D614G variant spike with BA.1 NTD mutation cluster in HEK293T/ACE2 cells.

(D) Infectivity of BA.1 chimeric spike with Wuhan NTD (C4) accompanied by BA.1 NTD mutation cluster in HEK293T/ACE2 cells.

(E) Spike expression of BA.1 spike with subtraction of NTD mutation cluster in HEK293T cells.

(F) Spike expression of D614G variant spike with BA.1 NTD mutation cluster in HEK293T cells.

(G) Spike expression of BA.1 chimeric spike with Wuhan NTD (C4) accompanied by BA.1 NTD mutation cluster in HEK293T cells. For all plots above, mean \pm SD from at least two biological replicates.

(H) Neutralization efficacy in 293T/ACE2 cells for 5 vaccinated serum samples against BA.1 spike with subtraction of NTD mutation cluster. The NT50 values are summarized in the right panels. Dashed lines indicate detection limit.

Figure 7. The fitness of NTDs from Sarbecoviruses and major SARS-CoV-2 variants.

(A) Schematic of chimera spike with SARS-CoV-2 (Wuhan strain) and SARS-CoV-1.

(B) Infectivity of each chimera spike-coating luciferase reporter pseudoviruses in HEK293T/ACE2 cells.

(C) Spike expression of each chimera spike in HEK293T cells.

(D) Spike affinity toward ACE2. Gating of flow cytometry-based affinity assessment was explained in Figure

S1B.

(E) Correlation between infectivity and spike expression in chimeric Wuhan D614G spike with Sarbecovirus NTDs.

(F) Correlation between infectivity and spike expression in chimeric Wuhan D614G spike with major SARS-CoV-2 variant NTDs. The NTD sequences of sarbecoviruses and NTD mutations of SARS-CoV-2 variants are in Figure S9.

For all plots, mean \pm SD from at least two biological replicates.

STAR Method

Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
SARS-CoV-2 Spike Protein S1/S2 Polyclonal Antibody	Thermo Scientific Fisher	Cat# PA5-112048; RRID: AB_2866784
Mouse monoclonal anti-p24 (clone 183-H12-5C)	NIH HIV Reagent Program	Cat# 183-H12-5C; RRID: AB_2819250
Anti-HA-tag mAb-Alexa Fluor 647	MBL International	Cat# M180-A64; RRID: AB_10951015
Fluorescein (FITC) AffiniPure Mouse Anti-Human IgG, Fcy fragment specific	Jackson ImmunoResearch	Cat# 209-095-098; RRID: AB_2339104
Bacterial and virus strains		
hCoV-19/Japan/TY-WK-521/2020	National Institute of Infectious Diseases	EPI_ISL_408667
ACE2-coating pseudotyped pcDNA4/TO-CMV-GFP lentivirus	This paper	N/A
SARS-CoV-2-Spike pseudotyped pcDNA4/TO-CMV-GFP lentivirus	This paper	N/A
SARS-CoV-2-Spike pseudotyped pcDNA4/TO-CMV-Firefly lentivirus	This paper	N/A
Biological samples		
SARS-CoV-2 vaccinated human sera	(Ikemura <i>et al.</i> , 2022)	N/A
Chemicals, peptides, and recombinant proteins		
ONE-Glo Luciferase Assay System	Promega	Cat# E6120
Nano-Glo HiBit Lytic Detection System	Promega	Cat# N3040
ExpiFectamine 293 Transfection Kit	Thermo Scientific Fisher	Cat# A14524
Lipofectamine 3000	Thermo Scientific Fisher	Cat# L3000015
TRIzol Reagent	Thermo Scientific Fisher	Cat# 15596026
PrimeScript II Reverse Transcriptase	Takara	Cat# 2690B
RNA lysis buffer	This paper	N/A
Superase IN	Thermo Scientific Fisher	Cat# AM2696
rProtein A Sepharose Fast Flow	Cytiva	Cat# 17-1279-xx
Wuhan RBD-Fc	This paper	N/A
Omicron (BA.1) RBD-Fc	This paper	N/A
Critical commercial assays		
Direct-zol RNA Miniprep	Zymo Research	Cat# R2052
Power SYBR Green RNA-to-CT 1-Step Kit	Thermo Scientific Fisher	Cat# 4389986
Deposited data		
Structure of SARS-CoV-2 spike trimer	This paper	PDB:
Experimental models: Cell lines		
Lenti-X 293T cells	Clontech	Cat# 632180
Expi293F cells	Thermo Scientific Fisher	Cat# A14635
Calu-3 cells	ATCC	Cat# HTB-55
293T/ACE2 cells	This paper	N/A
Vero E6/TMPRSS2 cells	JCRB Cell Bank	JCRB1819

Experimental models: Organisms/strains		
Oligonucleotides		
SARS2-Spike-library (NTD, RBD, and S1/S2 junction)	This paper	N/A
Recombinant DNA		
pcDNA3.1-SARS2-Spike	(Shang <i>et al.</i> , 2020)	Addgene Plasmid #145032
pcDNA4/TO Mammalian Expression Vector	Thermo Scientific Fisher	Cat# V102020
pLenti-CMV-GFP Puro	(Campeau <i>et al.</i> , 2009)	Addgene Plasmid #17448
pLenti-CMV-Firefly	This paper	N/A
psPAX2	psPAX2 was a gift from Didier Trono	Addgene Plasmid #12260
psPAX2-IN/HiBiT	(Ozono <i>et al.</i> , 2020)	N/A
pcDNA4/TO-HMM38-HA-Spike library (NTD, RBD, and S1/S2 junction)	This paper	N/A
pcDNA4/TO-HMM38-HA-Spike (D614G, Alpha, Beta, Gamma, Delta, Epsilon, Lambda, Mu, BA.1, BA.2, and SARS1)	This paper	N/A
pcDNA4/TO-Chimera spike with D614G and BA.1 (C1, C2, C3, C4, C5, and C6)	This paper	N/A
pcDNA4/TO-Chimera spike with D614G and BA.2 (C1, C2, C3, C4, C5, and C6)	This paper	N/A
pcDNA4/TO-Spike BA.1 NTD mutation (NTD1, A67V/Δ69-70; NTD2, T95I; NTD3, G142D/Δ143-145; and NTD4, Δ211/L212I/ins214EPE)	This paper	N/A
pcDNA4/TO-Spike BA.2 NTD mutation (NTD1, T19I/L24S/Δ25-27; NTD2, G142D; and NTD3, V213G)	This paper	N/A
pcDNA4/TO-Chimera spike with SARS-CoV-2 and SARS-CoV-1	This paper	N/A
pcDNA4/TO-Chimera spike with D614G and SARS-CoV-2 variant NTDs (Wuhan, Alpha, Beta, Gamma, Delta, Epsilon, Lambda, Mu, BA.1, and BA.2)	This paper	N/A
pcDNA4/TO-Chimera spike with D614G and sarbecoviruses NTDs (PG-GD-1, PG-GX-P5L, RaTG13, Rs4231, RsSHC, and WIV1)	This paper	N/A
pLenti-CMV-HMM38-HA-ACE2 (WT)	This paper	N/A
pLenti-CMV-HMM38-HA-ACE2 (3N39)	(Higuchi <i>et al.</i> , 2021)	N/A
Software and algorithms		
GraphPad Prism 9	GraphPad Software	https://www.graphpad.com/scientific-software/prism/
FlowJoX software	BD Biosciences	https://www.flowjo.com/solutions/flowjo/downloads
Other		
Infinite F200 pro	Tecan	https://lifesciences.tecan.com/plate_readers/infinite_200_pro?p=tab--2
ChemiDoc MP Imaging System	Bio-Rad	Cat# 12003154

MA900 Multi-Application Cell Sorter	SONY	https://www.sonybiotechnology.com/us/instruments/ma900-multi-application-cell-sorter/
Illumina NovaSeq 6000	Illumina	https://www.illumina.com/systems/sequencing-platforms/novaseq.html
AriaMx Real-time PCR System	Agilent	Cat# G8830A

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Atsushi Hoshino (a-hoshi@koto.kpu-m.ac.jp).

Material availability

All unique/stable reagents generated in this study are available from the Lead Contact with a completed Materials Transfer Agreement.

Data and code availability

- Interactive versions of DMS heatmaps are at (https://sysimm.ifrec.osaka-u.ac.jp/sarscov2_dms/data/).
- All raw sequencing data are uploaded to the NCBI Short Read Archive under BioProject ID PRJNA988964
- This paper does not report original code.

MATERIALS AND METHODS

Study Design

Research blood samples were obtained from vaccinated individuals who were enrolled in a prospective cohort study approved by the Clinical Research Review Committee in Kyoto Prefectural University of medicine (ERB-C-1954-1). All human participants provided written informed consent. We collected peripheral blood from 7 individuals at 3 months post vaccination with two doses of the Pfizer-BioNTech vaccine, BNT162b2. The study size was determined by the number of samples that were available from the cohort study and not based on any power calculations. Experiments described in this manuscript were not performed blinded.

Cell culture

Lenti-X 293T cells (Clontech) and its derivative, 293T/ACE2 cells were cultured at 37 °C with 5% CO₂ in Dulbecco's modified Eagle's medium (DMEM, WAKO) containing 10% fetal bovine serum (Gibco) and penicillin/streptomycin (100 U/ml, Invitrogen). Calu3 cells were purchased from the American Tissue Culture Collection (ATCC, Rockville, MD) and cultured at 37 °C with 5% CO₂ in DMEM/F-12 (WAKO) containing 10% fetal bovine serum (Gibco) and penicillin/streptomycin (100 U/ml, Invitrogen). All cell lines routinely tested negative for mycoplasma contamination.

Plasmids construction

Using a plasmid encoding the SARS-CoV-2 S (addgene #145032) as a template, variant S fragments were amplified by bridge PCR with mutation-containing primers and cloned into pcDNA4TO (Invitrogen). To express HA-tagged spike protein, variant S fragments were cloned into pcDNA4TO as a form of HMM38-HA-spike. The NTD of Pangolin CoV GD-1, GX-P5L, Bat CoV RsSHC014, WIV1, RaTG13, and Rs4231 were synthesized (Integrated DNA Technologies) and cloned into pcDNA4TO (Invitrogen) in the form of ΔC19(10, 39). Pangolin CoV GX-P5L, Bat CoV RaTG13, and Rs4231 RBDs were synthesized (Integrated DNA Technologies) and cloned into the NTD of HMM38-HA-spike.

Pseudotyped virus preparation and cell entry assays.

Pseudotyped reporter virus assays were conducted as previously described (Higuchi *et al.*, 2021). Spike protein-expressing pseudoviruses with a luciferase reporter gene was prepared by transfecting plasmids (pcDNA4TO spike, psPAX2-IN/HiBiT (Ozono *et al.*, 2020), and pLenti firefly) into LentiX-293T cells with Lipofectamine 3000 (Invitrogen). After 48 hours, supernatants were harvested, filtered with a 0.45 μm low protein-binding filter (SFCA), and frozen at −80 °C. The lentivirus levels in viral supernatants were measured by the HiBiT assay, as previously described (Ozono *et al.*, 2020). Briefly, a lentivirus stock with known levels of p24 antigen was serially diluted. Either the standards or viral supernatants containing pseudotyped viruses (25 μl) and LgBiT Protein (1:100)/HiBiT Lytic Substrate (1:50) in Nano-Glo HiBiT Lytic Buffer (25 μl) (Nano-Glo HiBiT Lytic Detection System; Promega) were mixed and incubated for 10min at room temperature according to the manufacturer's instructions. HiBiT-based luciferase activity in viral supernatants was determined with Infinite F200 pro system (Tecan), and translated into p24 antigen levels. Pseudoviruses with different spike proteins corresponding to 1 ng of p24 antigen were incubated in 293T/ACE2 cells or Calu-3 cell. After 48 h, the cells were lysed in 100 μl of One-Glo Luciferase Assay Reagent (Promega). Firefly luciferase activity was determined with Infinite F200 pro system (Tecan).

Pseudotyped virus neutralization assay

Pseudotyped reporter virus assays were conducted as previously described (Higuchi *et al.*, 2021). Spike protein-expressing pseudoviruses with a luciferase reporter gene was prepared by transfecting plasmids (spikeΔC19 (Giroglou *et al.*, 2004), psPAX2-IN/HiBiT (Ozono *et al.*, 2020), and pLenti firefly) into LentiX-293T cells with Lipofectamine 3000 (Invitrogen). After 48 hours, supernatants were harvested, filtered with a 0.45 μm low protein-binding filter (SFCA), and frozen at −80 °C. The 293T/ACE2 cells were seeded at

10,000 cells per well in 96-well plates. HiBiT value-matched pseudoviruses and three-fold dilution series of serum were incubated for 1 hour, then this mixture was added to 293T/ACE2 cells. After 1 hour pre-incubation, medium was changed. At 48 hours post infection, cellular expression of the luciferase reporter, indicating viral infection, was determined using ONE-Glo Luciferase Assay System (Promega). Luminescence was read on Infinite F200 pro system (Tecan). The assay for each serum sample was performed in triplicate, and the 50% neutralization titer was calculated using Prism version 9 (GraphPad Software).

Immunoblot

Pseudoviruses with different spike proteins corresponding to 1 ng of p24 antigen were loaded onto Tris-glycine sodium dodecyl sulfate-polyacrylamide gels, separated by electrophoresis, and then the proteins were transferred onto polyvinylidene difluoride membranes (Millipore, IPVH00010). The membranes were probed with an anti-SARS-CoV-2 spike protein S1/S2 polyclonal antibody (1:1,000, Thermo Fisher Scientific PA5-112048), and an anti-p24 monoclonal antibody (1:1,000). Blots were visualized with an HRP-conjugated secondary antibody, the Clarity Western ECL substrate (Bio-Rad Laboratories, 1705060), and analyzed by ChemiDoc Touch MP (Bio-Rad).

Protein synthesis and purification

Monoclonal antibodies and engineered ACE2 were expressed using the Expi293F cell expression system (Thermo Fisher Scientific) according to the manufacturer's protocol. Fc-fusion proteins were purified from conditioned media using the rProtein A Sepharose Fast Flow (Cytiva). Fractions containing target proteins were pooled and dialyzed against phosphate buffered saline (PBS).

Library construction, FACS, and Illumina sequencing analysis

Saturation mutagenesis was conducted independently in the NTD (S13 to T307), RBD (F329 to C538), and S1/S2 junction (F541 to S750) of ancestral original Wuhan strain. Pooled oligos with degenerate NNK codons were synthesized by Integrated DNA Technologies, Inc. Synthesized oligos were extended by overlap PCR and cloned into pcDNA4TO HMM38-HA-full length spike plasmids. Transient transfection conditions were used that typically provide no more than a single coding variant per cell (Chan *et al.*, 2021). Expi293F cells at 2×10^6 cells per ml were transfected with a mixture of 1 ng of library plasmid with 1 μ g of pMSCV as a junk plasmid per ml using ExpiFectamine (Thermo Fisher Scientific). At 24 hours after transfection, cells were washed twice with PBS containing 10% bovine serum albumin (BSA) and then stained for 20 minutes with anti-hemagglutinin (HA) Alexa Fluor 647 (clone TANA2, 1:4000 dilution; MBL) at 4 °C for expression assay. For affinity assay, cells were washed twice with PBS containing 10% BSA and then incubated with ACE2-Fc for stained for 20 minutes at 4 °C. Cells were again washed twice with PBS containing 10% BSA and then co-stained for 20 minutes with anti-HA Alexa Fluor 647 (clone TANA2, 1:4000 dilution; MBL) and anti-human IgG-Fc γ fragment specific Fluorescein (FITC) (209-095-098, 1:100, Jackson ImmunoResearch) for 20 minutes at 4 °C. For infection assay, cells were incubated

with ACE2-coating green fluorescent protein (GFP) reporter viruses, which were generated by transfecting pcDNA4TO ACE2, psPAX2 (addgene #12260), and pLenti GFP into LentiX-293T cells with Lipofectamine 3000 (Thermo Fisher Scientific). The viruses have ACE2 on the surface instead of glycoprotein and can infect the spike protein-expressing cells. At 24 hours after infection, cells were washed twice with PBS containing 10% BSA and then co-stained for 20 minutes with anti-HA Alexa Fluor 647 (clone TANA2, 1:4000 dilution; MBL). Cells were again washed twice before sorting on a MA900 cell sorter (Sony). Dead cells, doublets, and debris were excluded by first gating on the main population by forward and side scatter. For expression analysis, the highest and lowest 25% HA signal among HA-positive cells were sorted. For affinity analysis, the highest and lowest 25% of cells with ACE2-Fc binding relative to HA-spike expression were sorted. For infection analysis, GFP-positive and -negative cells were collected from the HA positive (Alexa Fluor 647 positive) population. The total numbers of collected cells were about 2 million cells for each group. Total RNA was extracted from collected cells TRIzol (Thermo Fisher Scientific) and Direct-zol RNA MiniPrep (Zymo Research Corporation) according to the manufacturer's protocol. First-strand complementary DNA (cDNA) was synthesized with PrimeScript II Reverse Transcriptase (Takara) primed with a gene-specific oligonucleotide. Libraries were designed for 3 sections separately in the RBD and then pooled. After cDNA synthesis, each library was amplified with specific primers. Following a second round of PCR, primers added adapters for annealing to the Illumina flow cell and sequencing primers, together with barcodes for experiment identification. The PCR products were sequenced on an Illumina NovaSeq 6000 using a 2 × 150 nucleotide paired-end protocol in Department of Infection Metagenomics, Research Institute for Microbial Diseases, Osaka University. Data were analyzed comparing the read counts with each group normalized relative to the wild type sequence read-count. Log10 enrichment ratios for all the individual mutations were calculated and normalized by subtracting the log10 enrichment ratio for the wild type sequence across the same PCR-amplified fragment.

Viruses

The SARS-CoV-2 (Wuhan: 2019-nCoV/Japan/TY/WK-521/2020) strain was isolated at National Institute of Infectious Diseases (NIID). SARS-CoV-2 viruses were propagated in Vero E6/TMPRSS2 cells. Viral supernatant was harvested at two days post-infection and viral titers were determined by plaque assay.

Quantitative RT-PCR of Viral RNA in the supernatant

The amount of RNA copies in the culture medium was determined using a qRT-PCR assay as previously described with slight modifications (Shema Mugisha et al., 2020). In brief, 5 µl of culture supernatants were mixed with 5 µl of 2 × RNA lysis buffer (2% Triton X-100, 50 mM KCl, 100 mM Tris-HCl [pH 7.4], 40% glycerol, 0.4 U/µl of Superase IN [Thermo Fisher Scientific]) and incubated at room temperature for 10 minutes, followed by addition of 90 µl of RNase free water. Next, 2.5 µl of volume of the diluted samples was added to 17.5 µl of the reaction mixture. Real-time RT-PCR was performed with the Power SYBR Green RNA-to-CT 1-Step Kit (Applied Biosystems) using a AriaMx Real-Time PCR system (Agilent).

Evaluation of virus stability on plastic surface

Virus survival was evaluated on polystyrene plate as previously described(Hirose et al., 2022). Luciferase reporter pseudoviruses with different spike proteins in 10 μ L medium were applied on the surface of polystyrene plate. Each sample was incubated in a controlled environment (25°C, 45–55% relative humidity) for 0–8 h. The virus that remained on the surface was subsequently collected in 1.0 mL of DMEM and titrated by incubation in 293T/ACE2 cells.

Statistical analysis

Neutralization measurements were done in technical triplicates and relative luciferase units were converted to percent neutralization and plotted with a non-linear regression model to determine 50% neutralization titer (NT₅₀) values using GraphPad Prism software (version 9.0.0). Comparisons between two groups were made with two-sided unpaired *t* test. More than two groups were compared by one-way ANOVA and Tukey's multiple comparison tests.

References

- Alcantara, M.C., Higuchi, Y., Kirita, Y., Matoba, S., and Hoshino, A. (2023). Deep Mutational Scanning to Predict Escape from Bebtelovimab in SARS-CoV-2 Omicron Subvariants. *Vaccines (Basel)* *11*, 10.3390/vaccines11030711.
- Barash, S., Wang, W., and Shi, Y. (2002). Human secretory signal peptide description by hidden Markov model and generation of a strong artificial signal peptide for secreted protein expression. *Biochem Biophys Res Commun* *294*, 835-842. 10.1016/S0006-291X(02)00566-1.
- Bekker, G.J., Yokochi, M., Suzuki, H., Ikegawa, Y., Iwata, T., Kudou, T., Yura, K., Fujiwara, T., Kawabata, T., and Kurisu, G. (2022). Protein Data Bank Japan: Celebrating our 20th anniversary during a global pandemic as the Asian hub of three dimensional macromolecular structural data. *Protein Sci* *31*, 173-186. 10.1002/pro.4211.
- Cai, Y., Zhang, J., Xiao, T., Peng, H., Sterling, S.M., Walsh, R.M., Jr., Rawson, S., Rits-Volloch, S., and Chen, B. (2020). Distinct conformational states of SARS-CoV-2 spike protein. *Science* *369*, 1586-1592. 10.1126/science.abd4251.
- Campeau, E., Ruhl, V.E., Rodier, F., Smith, C.L., Rahmberg, B.L., Fuss, J.O., Campisi, J., Yaswen, P., Cooper, P.K., and Kaufman, P.D. (2009). A versatile viral system for expression and depletion of proteins in mammalian cells. *PLoS One* *4*, e6529. 10.1371/journal.pone.0006529.
- Casalino, L., Gaieb, Z., Goldsmith, J.A., Hjorth, C.K., Dommer, A.C., Harbison, A.M., Fogarty, C.A., Barros, E.P., Taylor, B.C., McLellan, J.S., et al. (2020). Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent Sci* *6*, 1722-1734. 10.1021/acscentsci.0c01056.
- Cerutti, G., Guo, Y., Zhou, T., Gorman, J., Lee, M., Rapp, M., Reddem, E.R., Yu, J., Bahna, F., Bimela, J., et al. (2021). Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite. *Cell Host Microbe* *29*, 819-833 e817. 10.1016/j.chom.2021.03.005.
- Chan, K.K., Tan, T.J.C., Narayanan, K.K., and Procko, E. (2021). An engineered decoy receptor for SARS-CoV-

- 2 broadly binds protein S sequence variants. *Sci Adv* **7**, 10.1126/sciadv.abf1738.
- Choi, Y.K., Cao, Y., Frank, M., Woo, H., Park, S.J., Yeom, M.S., Croll, T.I., Seok, C., and Im, W. (2021). Structure, Dynamics, Receptor Binding, and Antibody Binding of the Fully Glycosylated Full-Length SARS-CoV-2 Spike Protein in a Viral Membrane. *J Chem Theory Comput* **17**, 2479-2487. 10.1021/acs.jctc.0c01144.
- Cui, Z., Liu, P., Wang, N., Wang, L., Fan, K., Zhu, Q., Wang, K., Chen, R., Feng, R., Jia, Z., et al. (2022). Structural and functional characterizations of infectivity and immune evasion of SARS-CoV-2 Omicron. *Cell* **185**, 860-871 e813. 10.1016/j.cell.2022.01.019.
- Dadonaite, B., Crawford, K.H.D., Radford, C.E., Farrell, A.G., Yu, T.C., Hannon, W.W., Zhou, P., Andrabi, R., Burton, D.R., Liu, L., et al. (2023). A pseudovirus system enables deep mutational scanning of the full SARS-CoV-2 spike. *Cell* **186**, 1263-1278 e1220. 10.1016/j.cell.2023.02.001.
- Giroglou, T., Cinatl, J., Jr., Rabenau, H., Drosten, C., Schwalbe, H., Doerr, H.W., and von Laer, D. (2004). Retroviral vectors pseudotyped with severe acute respiratory syndrome coronavirus S protein. *J Virol* **78**, 9007-9015. 10.1128/JVI.78.17.9007-9015.2004
- 78/17/9007 [pii].
- Greaney, A.J., Loes, A.N., Gentles, L.E., Crawford, K.H.D., Starr, T.N., Malone, K.D., Chu, H.Y., and Bloom, J.D. (2021). Antibodies elicited by mRNA-1273 vaccination bind more broadly to the receptor binding domain than do those from SARS-CoV-2 infection. *Sci Transl Med* **13**, 10.1126/scitranslmed.abi9915.
- Higuchi, Y., Suzuki, T., Arimori, T., Ikemura, N., Mihara, E., Kirita, Y., Ohgitani, E., Mazda, O., Motooka, D., Nakamura, S., et al. (2021). Engineered ACE2 receptor therapy overcomes mutational escape of SARS-CoV-2. *Nat Commun* **12**, 3802. 10.1038/s41467-021-24013-y.
- Hirose, R., Itoh, Y., Ikegaya, H., Miyazaki, H., Watanabe, N., Yoshida, T., Bandou, R., Daidoji, T., and Nakaya, T. (2022). Differences in environmental stability among SARS-CoV-2 variants of concern: both omicron BA.1 and BA.2 have higher stability. *Clin Microbiol Infect* **28**, 1486-1491. 10.1016/j.cmi.2022.05.020.
- Hoffmann, M.A.G., Bar-On, Y., Yang, Z., Gristick, H.B., Gnanapragasam, P.N.P., Vielmetter, J., Nussenzweig, M.C., and Bjorkman, P.J. (2020). Nanoparticles presenting clusters of CD4 expose a universal vulnerability of HIV-1 by mimicking target cells. *Proc Natl Acad Sci U S A* **117**, 18719-18728. 10.1073/pnas.2010320117.
- Hou, Y.J., Chiba, S., Halfmann, P., Ehre, C., Kuroda, M., Dinno, K.H., 3rd, Leist, S.R., Schafer, A., Nakajima, N., Takahashi, K., et al. (2020). SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. *Science* **370**, 1464-1468. 10.1126/science.abe8499.
- Hubbard, S.J., and Thornton, J.M. (1993). Naccess V2.1.1 - Atomic Solvent Accessible Area Calculations. <http://www.bioinf.manchester.ac.uk/naccess/>.
- Ikemura, N., Taminishi, S., Inaba, T., Arimori, T., Motooka, D., Katoh, K., Kirita, Y., Higuchi, Y., Li, S., Suzuki, T., et al. (2022). An engineered ACE2 decoy neutralizes the SARS-CoV-2 Omicron variant and confers protection against infection in vivo. *Sci Transl Med*, eabn7737. 10.1126/scitranslmed.abn7737.
- Jackson, C.B., Farzan, M., Chen, B., and Choe, H. (2022). Mechanisms of SARS-CoV-2 entry into cells. *Nat Rev Mol Cell Biol* **23**, 3-20. 10.1038/s41580-021-00418-x.
- Javanmardi, K., Segall-Shapiro, T.H., Chou, C.W., Boutz, D.R., Olsen, R.J., Xie, X., Xia, H., Shi, P.Y., Johnson, C.D., Annapareddy, A., et al. (2022). Antibody escape and cryptic cross-domain stabilization in the SARS-CoV-

- 2 Omicron spike protein. *Cell Host Microbe* **30**, 1242-1254 e1246. 10.1016/j.chom.2022.07.016.
- Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E.E., Bhattacharya, T., Foley, B., et al. (2020). Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* **182**, 812-827 e819. 10.1016/j.cell.2020.06.043.
- Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., and Wang, X. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215-220. 10.1038/s41586-020-2180-5.
- Liu, Y., Liu, J., Johnson, B.A., Xia, H., Ku, Z., Schindewolf, C., Widen, S.G., An, Z., Weaver, S.C., Menachery, V.D., et al. (2022). Delta spike P681R mutation enhances SARS-CoV-2 fitness over Alpha variant. *Cell Rep* **39**, 110829. 10.1016/j.celrep.2022.110829.
- McCallum, M., De Marco, A., Lempp, F.A., Tortorici, M.A., Pinto, D., Walls, A.C., Beltramello, M., Chen, A., Liu, Z., Zatta, F., et al. (2021). N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *Cell* **184**, 2332-2347 e2316. 10.1016/j.cell.2021.03.028.
- Mlcochova, P., Kemp, S.A., Dhar, M.S., Papa, G., Meng, B., Ferreira, I., Datir, R., Collier, D.A., Albecka, A., Singh, S., et al. (2021). SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. *Nature* **599**, 114-119. 10.1038/s41586-021-03944-y.
- Mori, T., Jung, J., Kobayashi, C., Dokainish, H.M., Re, S., and Sugita, Y. (2021). Elucidation of interactions regulating conformational stability and dynamics of SARS-CoV-2 S-protein. *Biophys J* **120**, 1060-1071. 10.1016/j.bpj.2021.01.012.
- Nguyen, H.L., Lan, P.D., Thai, N.Q., Nissley, D.A., O'Brien, E.P., and Li, M.S. (2020). Does SARS-CoV-2 Bind to Human ACE2 More Strongly Than Does SARS-CoV? *J Phys Chem B* **124**, 7336-7347. 10.1021/acs.jpcc.0c04511.
- Ouyang, W.O., Tan, T.J.C., Lei, R., Song, G., Kieffer, C., Andrabi, R., Matreyek, K.A., and Wu, N.C. (2022). Probing the biophysical constraints of SARS-CoV-2 spike N-terminal domain using deep mutational scanning. *Sci Adv* **8**, eadd7221. 10.1126/sciadv.add7221.
- Ozono, S., Zhang, Y., Ode, H., Sano, K., Tan, T.S., Imai, K., Miyoshi, K., Kishigami, S., Ueno, T., Iwatani, Y., et al. (2021). SARS-CoV-2 D614G spike mutation increases entry efficiency with enhanced ACE2-binding affinity. *Nat Commun* **12**, 848. 10.1038/s41467-021-21118-2.
- Ozono, S., Zhang, Y., Tobiume, M., Kishigami, S., and Tokunaga, K. (2020). Super-rapid quantitation of the production of HIV-1 harboring a luminescent peptide tag. *J Biol Chem* **295**, 13023-13030. 10.1074/jbc.RA120.013887.
- Partridge, L.J., Urwin, L., Nicklin, M.J.H., James, D.C., Green, L.R., and Monk, P.N. (2021). ACE2-Independent Interaction of SARS-CoV-2 Spike Protein with Human Epithelial Cells Is Inhibited by Unfractionated Heparin. *Cells* **10**. 10.3390/cells10061419.
- Peacock, T.P., Goldhill, D.H., Zhou, J., Baillon, L., Frise, R., Swann, O.C., Kugathasan, R., Penn, R., Brown, J.C., Sanchez-David, R.Y., et al. (2021). The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets. *Nat Microbiol* **6**, 899-909. 10.1038/s41564-021-00908-w.
- Saito, A., Irie, T., Suzuki, R., Maemura, T., Nasser, H., Uriu, K., Kosugi, Y., Shirakawa, K., Sadamasu, K., Kimura,

- I., et al. (2022). Enhanced fusogenicity and pathogenicity of SARS-CoV-2 Delta P681R mutation. *Nature* *602*, 300-306. 10.1038/s41586-021-04266-9.
- Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A., and Li, F. (2020). Structural basis of receptor recognition by SARS-CoV-2. *Nature* *581*, 221-224. 10.1038/s41586-020-2179-y.
- Shema Mugisha, C., Vuong, H.R., Puray-Chavez, M., Bailey, A.L., Fox, J.M., Chen, R.E., Wessel, A.W., Scott, J.M., Harastani, H.H., Boon, A.C.M., et al. (2020). A Simplified Quantitative Real-Time PCR Assay for Monitoring SARS-CoV-2 Growth in Cell Culture. *mSphere* *5*. 10.1128/mSphere.00658-20.
- Starr, T.N., Czudnochowski, N., Liu, Z., Zatta, F., Park, Y.J., Addetia, A., Pinto, D., Beltramello, M., Hernandez, P., Greaney, A.J., et al. (2021a). SARS-CoV-2 RBD antibodies that maximize breadth and resistance to escape. *Nature* *597*, 97-102. 10.1038/s41586-021-03807-6.
- Starr, T.N., Greaney, A.J., Addetia, A., Hannon, W.W., Choudhary, M.C., Dingens, A.S., Li, J.Z., and Bloom, J.D. (2021b). Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. *Science* *371*, 850-854. 10.1126/science.abf9302.
- Starr, T.N., Greaney, A.J., Hilton, S.K., Ellis, D., Crawford, K.H.D., Dingens, A.S., Navarro, M.J., Bowen, J.E., Tortorici, M.A., Walls, A.C., et al. (2020). Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* *182*, 1295-1310 e1220. 10.1016/j.cell.2020.08.012.
- Starr, T.N., Greaney, A.J., Stewart, C.M., Walls, A.C., Hannon, W.W., Veessler, D., and Bloom, J.D. (2022). Deep mutational scans for ACE2 binding, RBD expression, and antibody escape in the SARS-CoV-2 Omicron BA.1 and BA.2 receptor-binding domains. *PLoS Pathog* *18*, e1010951. 10.1371/journal.ppat.1010951.
- Takeda, M. (2022). Proteolytic activation of SARS-CoV-2 spike protein. *Microbiol Immunol* *66*, 15-23. 10.1111/1348-0421.12945.
- Urano, E., Itoh, Y., Suzuki, T., Sasaki, T., Kishikawa, J.-i., Akamatsu, K., Higuchi, Y., Sakai, Y., Okamura, T., Mitoma, S., et al. (2023). An engineered ACE2 decoy broadly neutralizes Omicron subvariants and shows therapeutic effect in SARS-CoV-2-infected cynomolgus macaques. *bioRxiv*, 2022.2012.2029.522275. 10.1101/2022.12.29.522275.
- Uriu, K., Kimura, I., Shirakawa, K., Takaori-Kondo, A., Nakada, T.A., Kaneda, A., Nakagawa, S., Sato, K., and Genotype to Phenotype Japan, C. (2021). Neutralization of the SARS-CoV-2 Mu Variant by Convalescent and Vaccine Serum. *N Engl J Med* *385*, 2397-2399. 10.1056/NEJMc2114706.
- Yurkovetskiy, L., Wang, X., Pascal, K.E., Tomkins-Tinch, C., Nyalile, T.P., Wang, Y., Baum, A., Diehl, W.E., Dauphin, A., Carbone, C., et al. (2020). Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. *Cell* *183*, 739-751 e738. 10.1016/j.cell.2020.09.032.
- Zeng, C., Evans, J.P., Pearson, R., Qu, P., Zheng, Y.M., Robinson, R.T., Hall-Stoodley, L., Yount, J., Pannu, S., Mallampalli, R.K., et al. (2020). Neutralizing antibody against SARS-CoV-2 spike in COVID-19 patients, health care workers, and convalescent plasma donors. *JCI Insight* *5*. 10.1172/jci.insight.143213.
- Zhang, J., Cai, Y., Xiao, T., Lu, J., Peng, H., Sterling, S.M., Walsh, R.M., Jr., Rits-Volloch, S., Zhu, H., Woosley, A.N., et al. (2021). Structural impact on SARS-CoV-2 spike protein by D614G substitution. *Science* *372*, 525-530. 10.1126/science.abf2303.

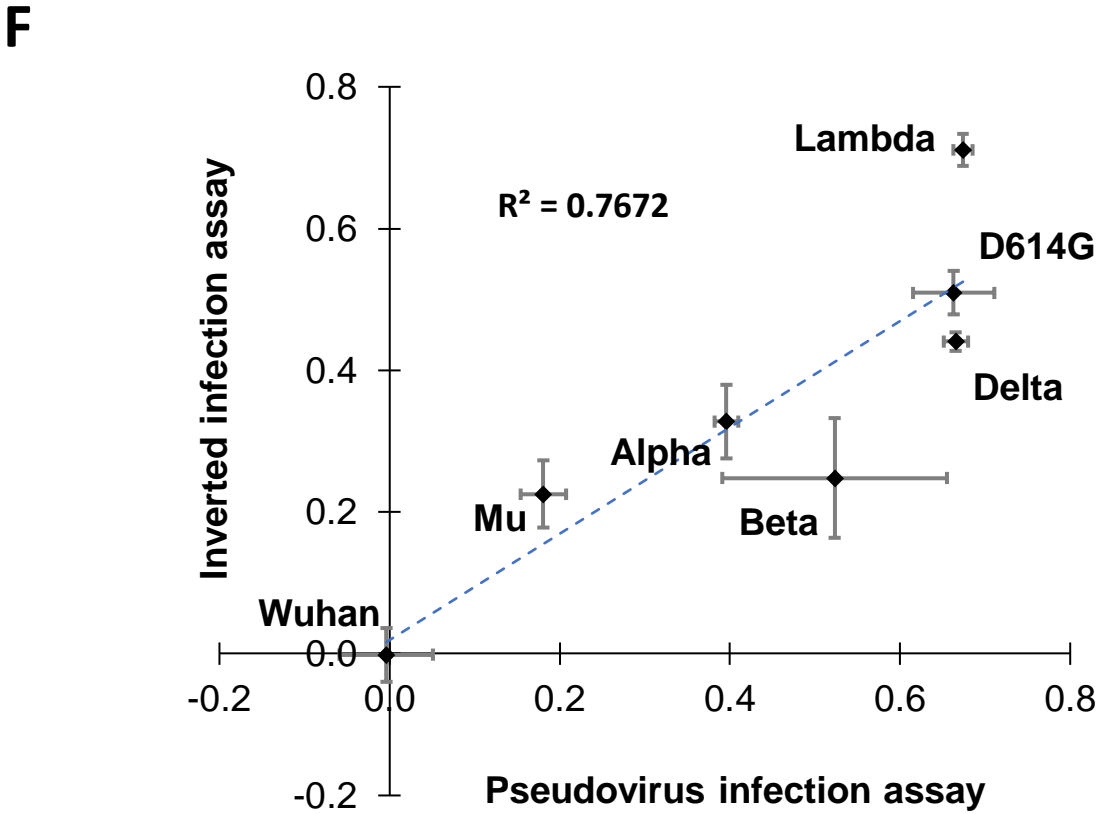
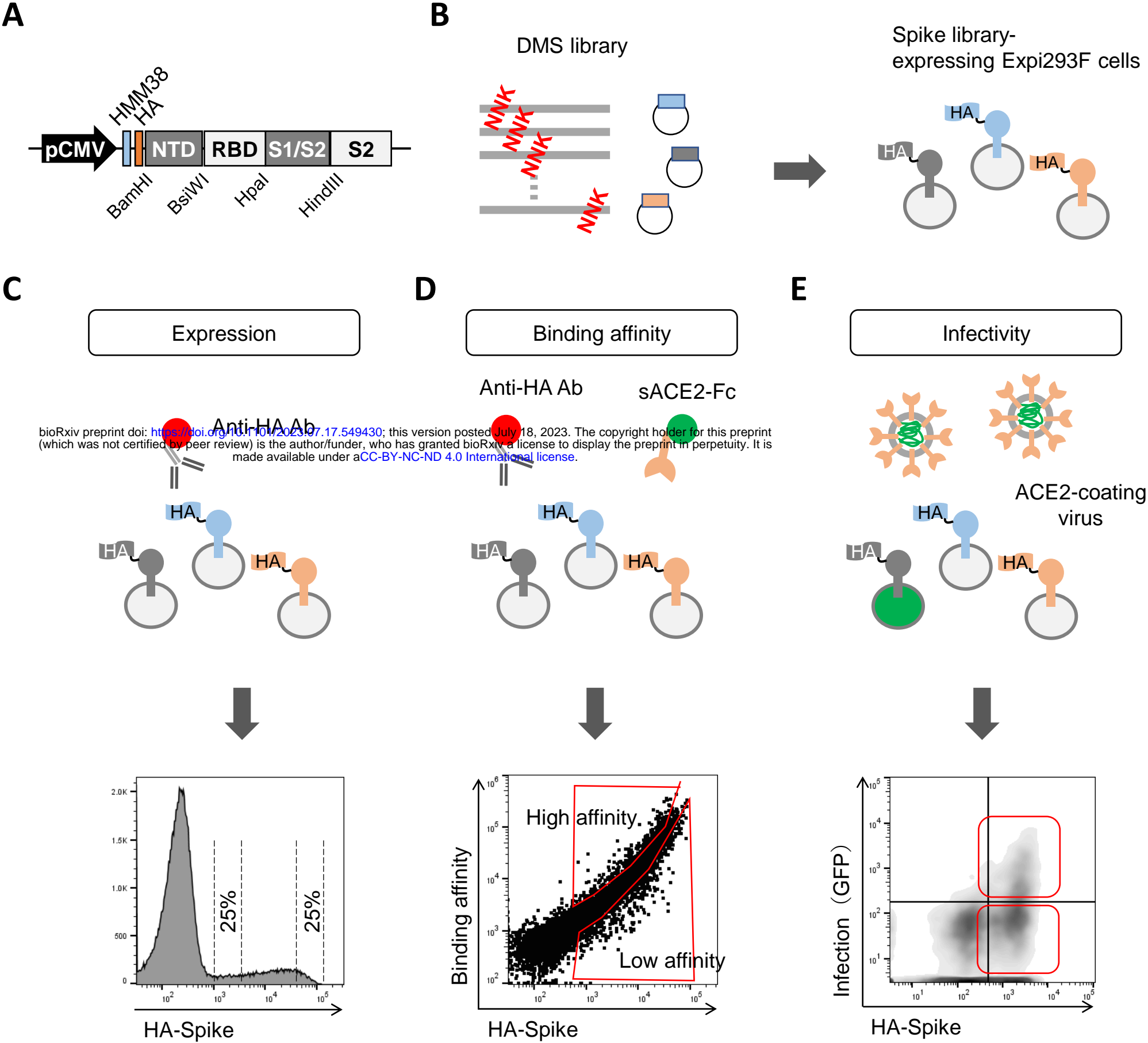


Figure 1

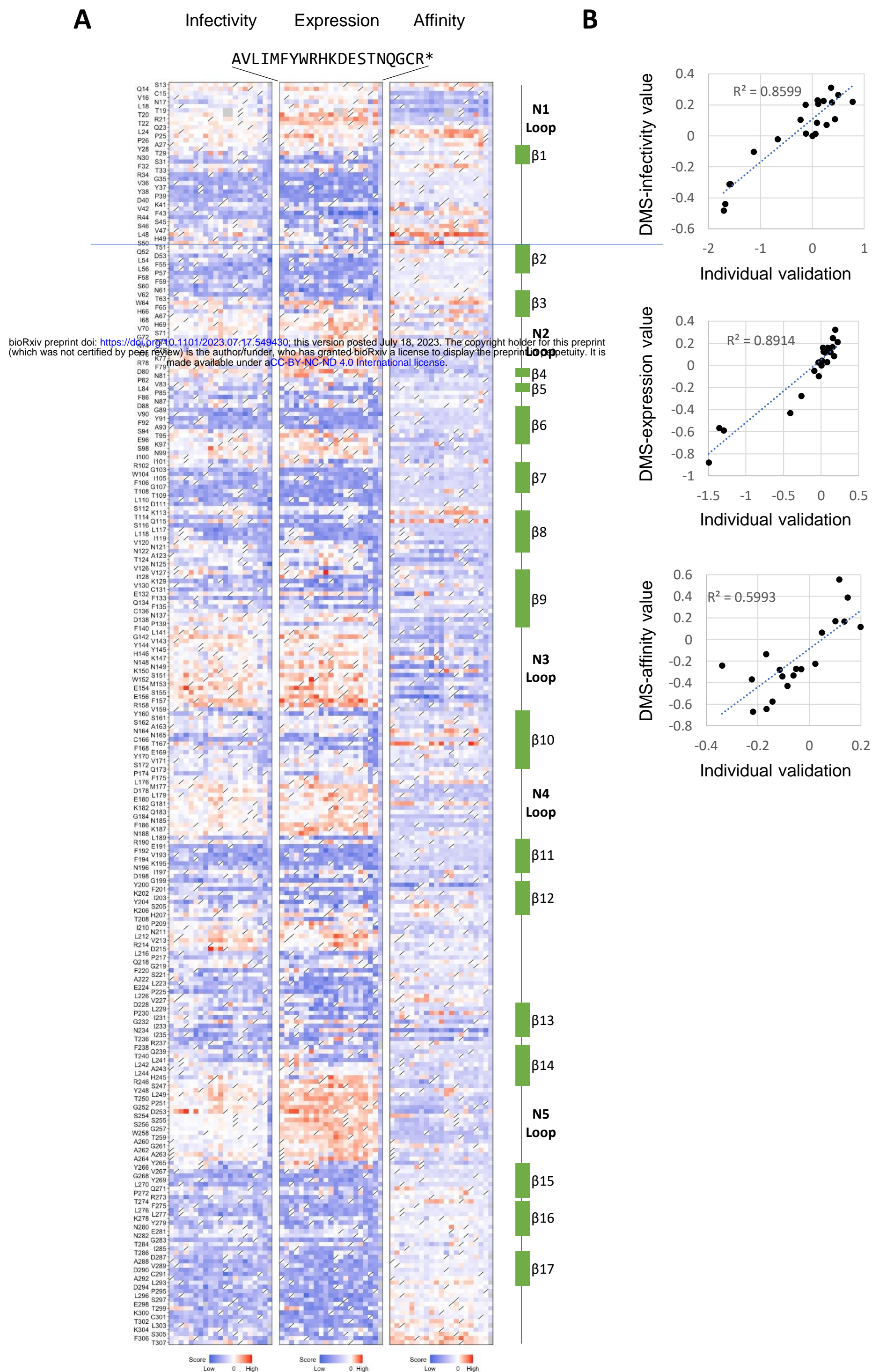
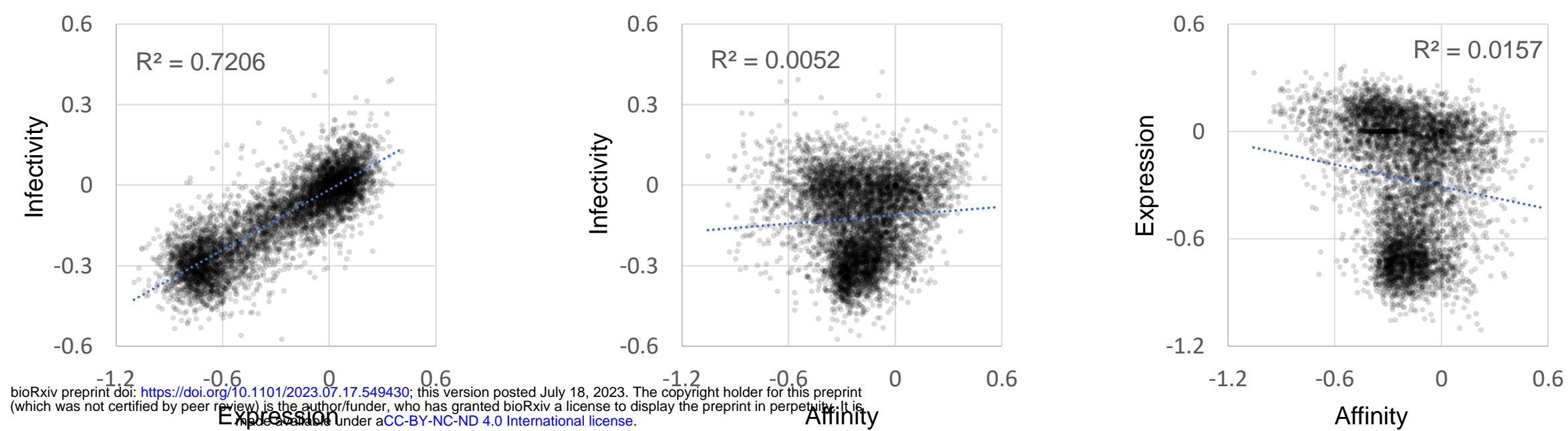
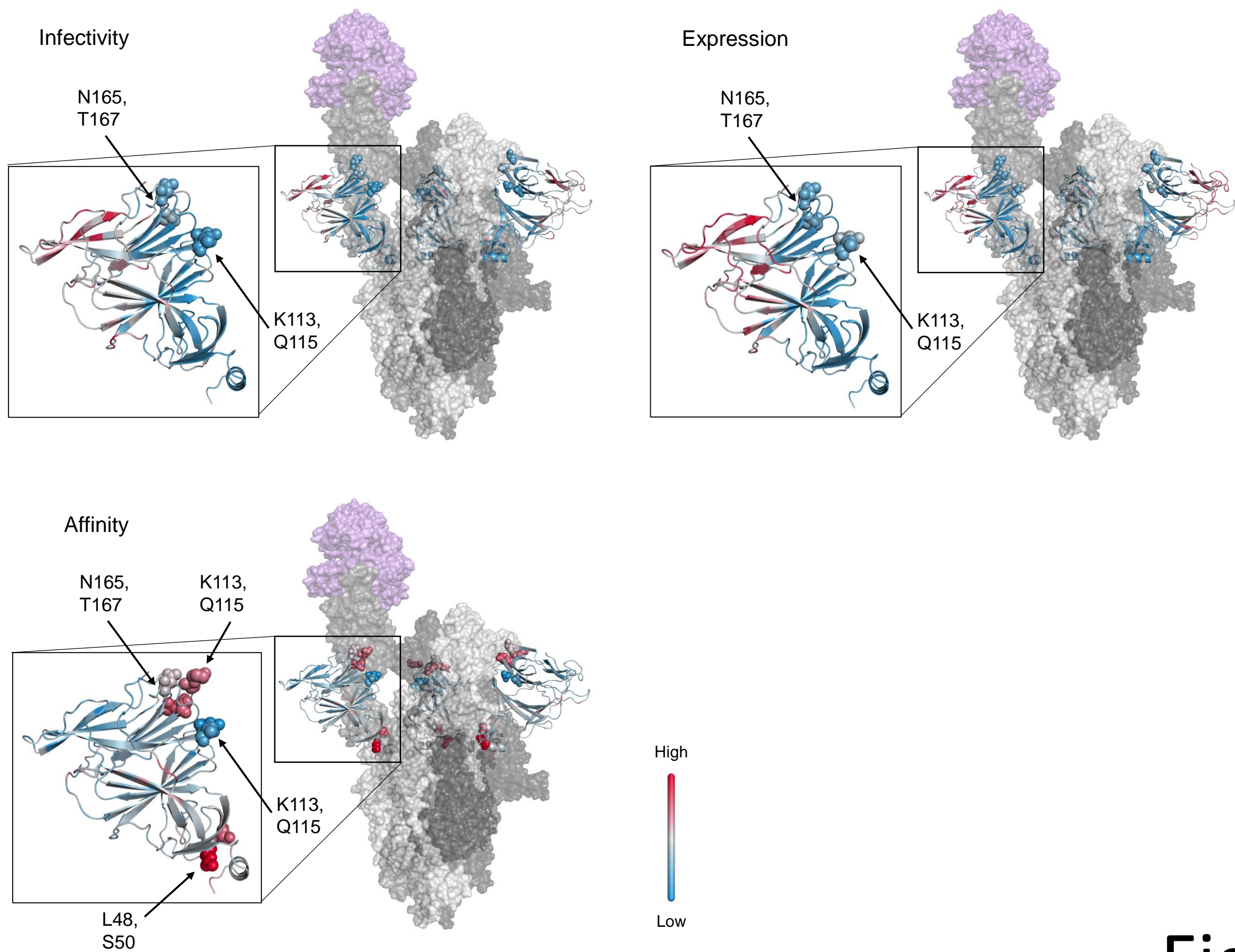


Figure 2

A**B****Figure 3**

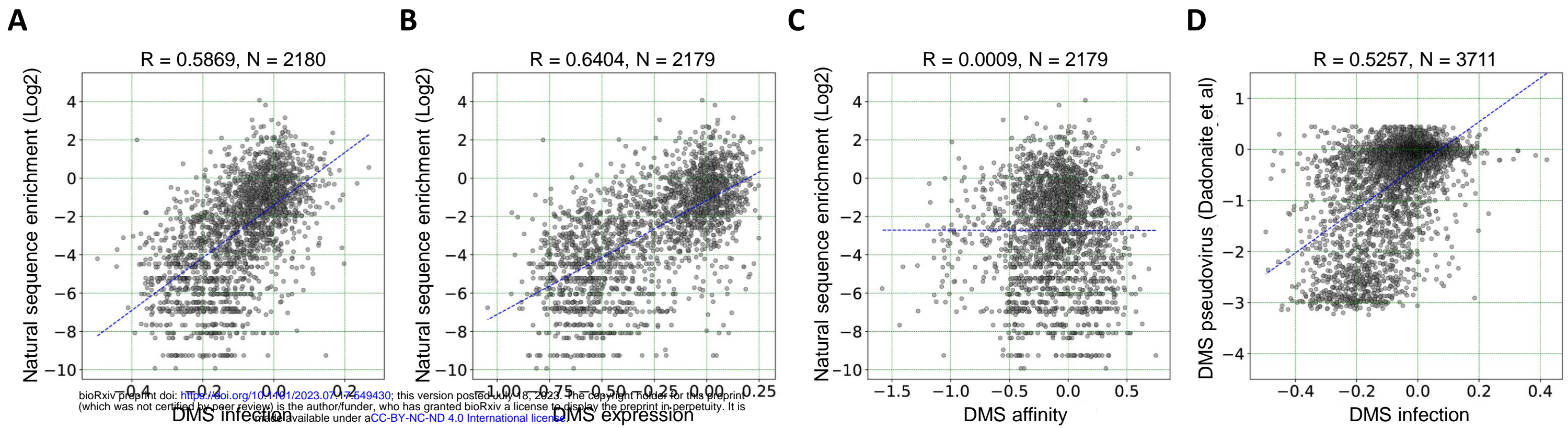


Figure 4

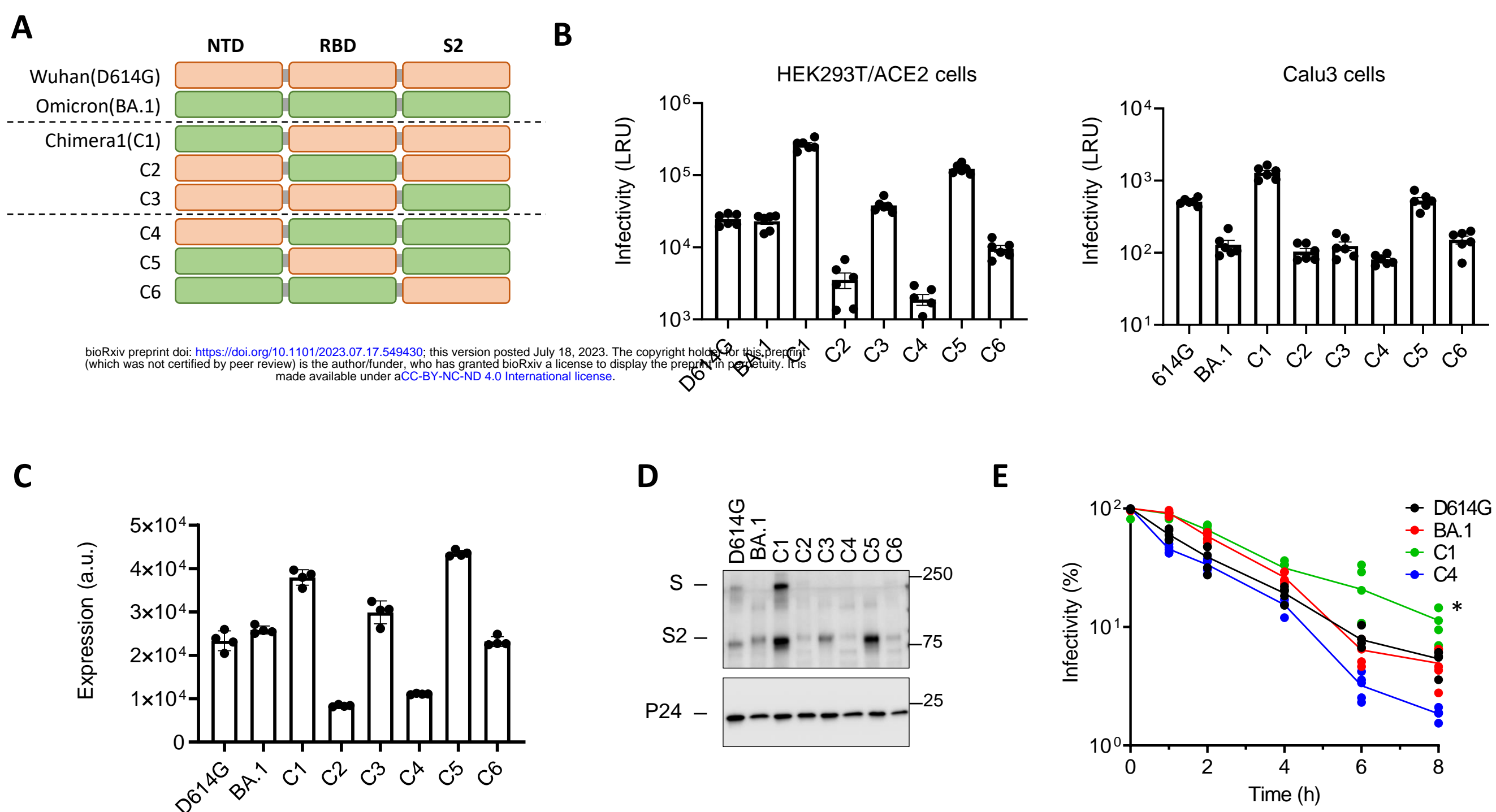


Figure 5

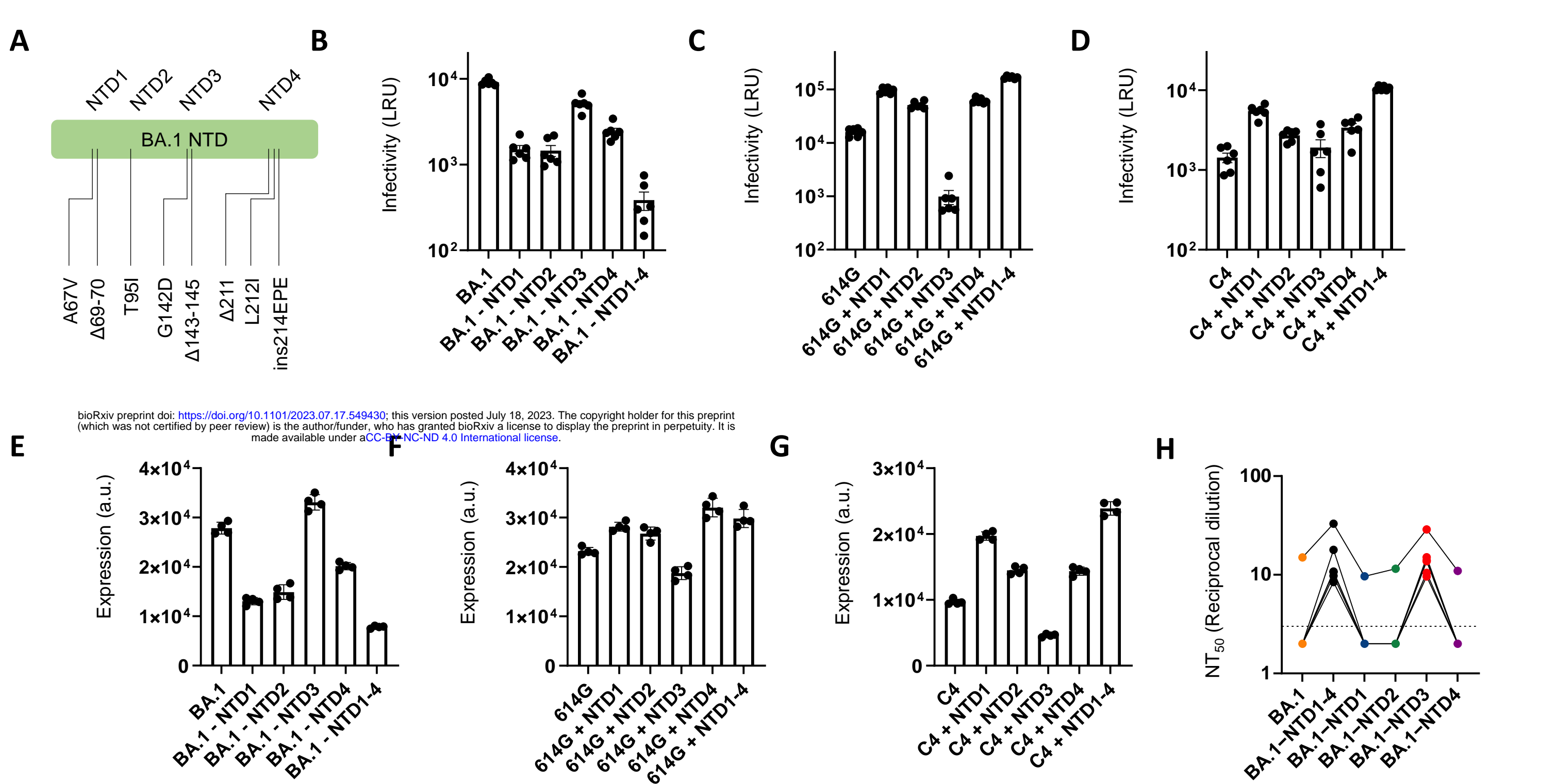
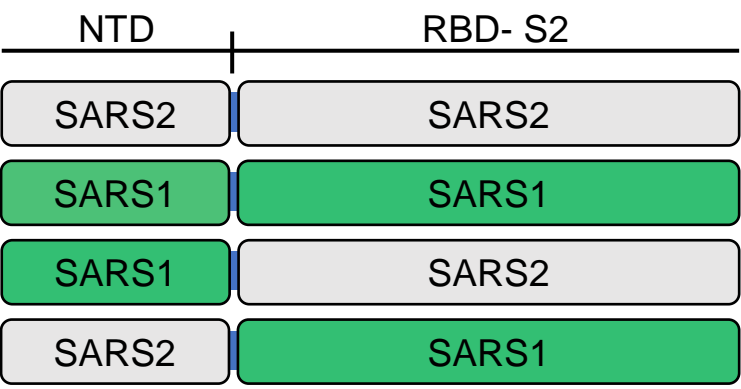
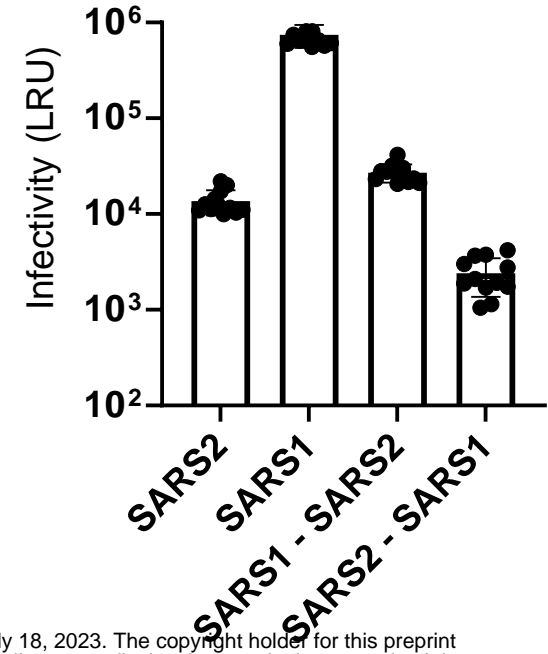


Figure 6

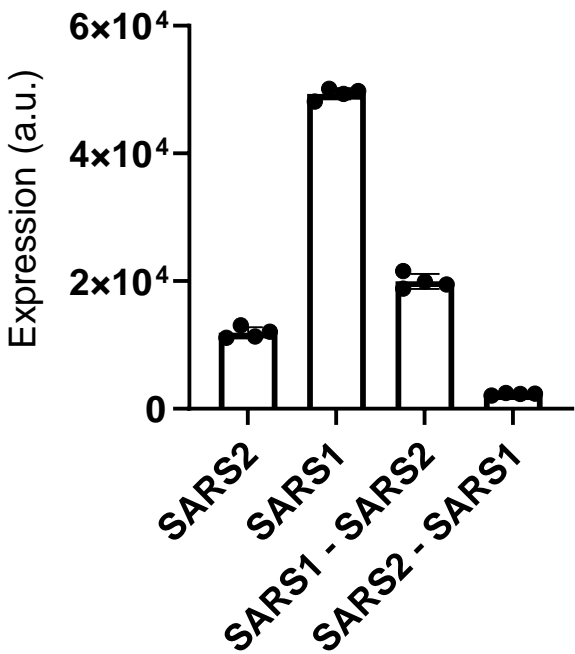
A



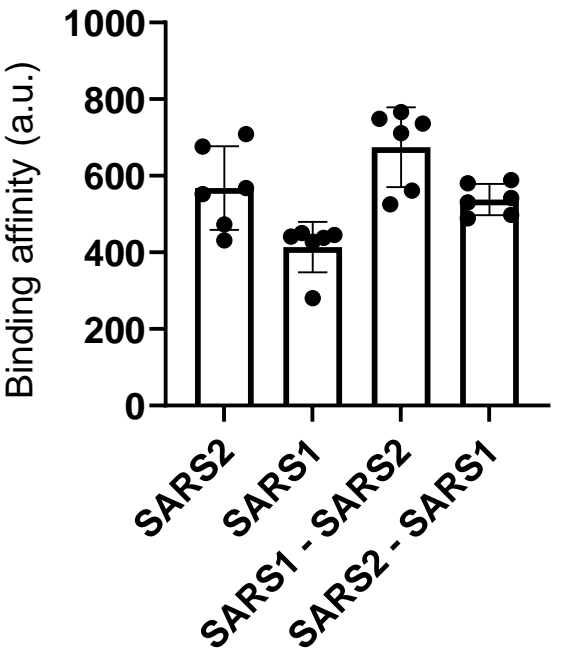
B



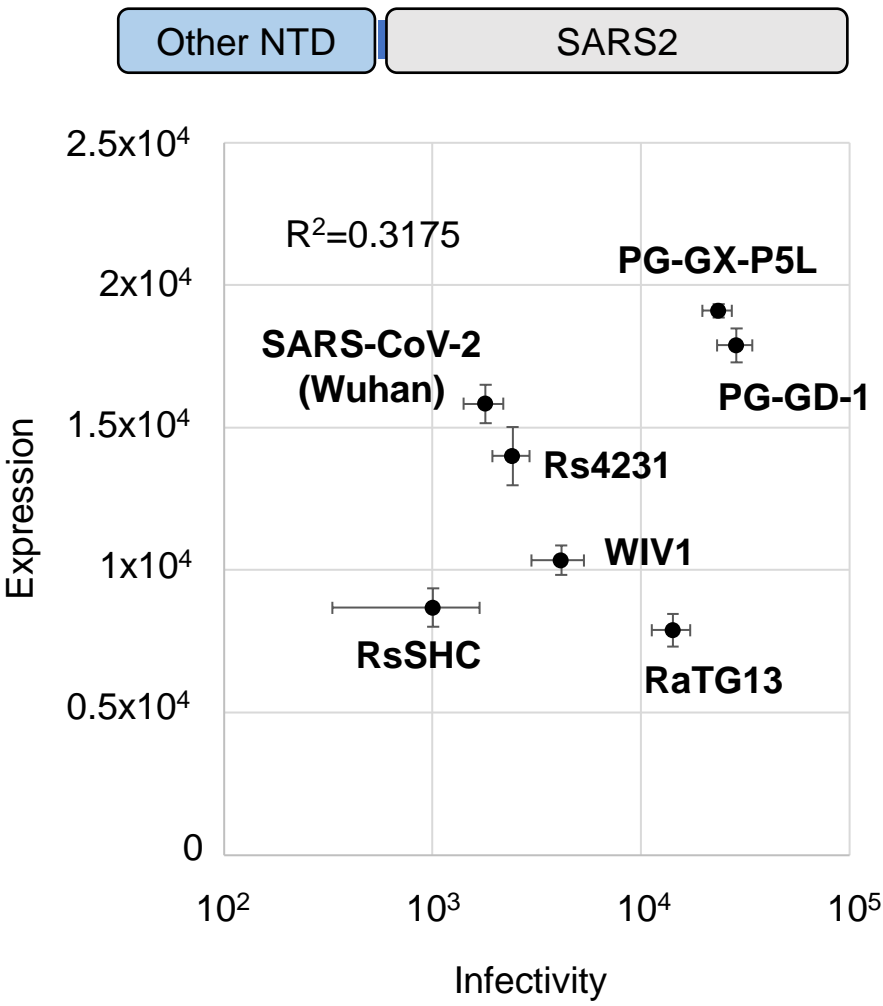
C



D



E



F

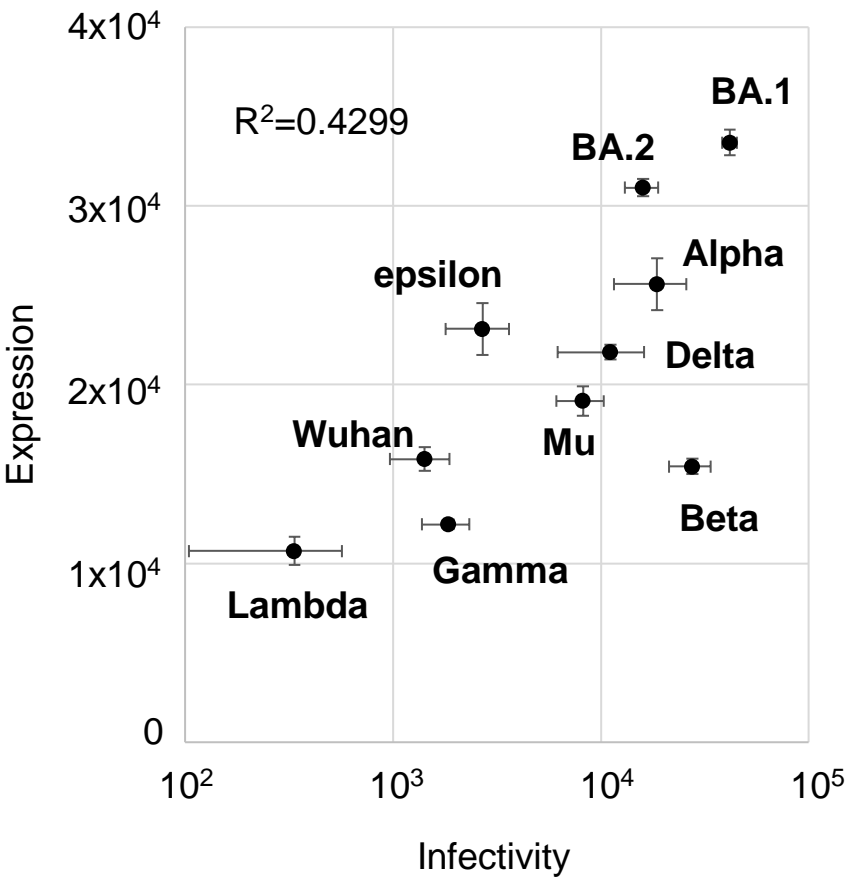


Figure 7