

Article

De Novo Isolation of an Antigen-Binding Antibody by Targeted CDR-H3 Diversification of a Germline-Derived Scaffold in a Site-Specifically Integrated Mammalian Display Platform: A Proof-of-Principle Study

Ryoichi Sakurai ¹, Nanami Hashimoto ¹, Narumi Uno ¹, Akitsu Hotta ², Yasuhiro Kazuki ^{3,4,5} and Kazuma Tomizuka ^{1,*}

¹ Laboratory of Bioengineering, School of Life Sciences, Tokyo University of Pharmacy and Life Sciences, 1432-1 Horinouchi, Hachioji 192-0392, Tokyo, Japan

² Hotta Laboratory, Center for iPS Cell Research and Application (CiRA), Kyoto University, 53 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan

³ Department of Chromosome Biomedical Engineering, School of Life Science, Faculty of Medicine, Tottori University, 86 Nishi-cho, Yonago 683-8503, Tottori, Japan

⁴ Chromosome Engineering Research Center, Tottori University, 86 Nishi-cho, Yonago 683-8503, Tottori, Japan

⁵ Chromosome Engineering Research Group, The Exploratory Research Center on Life and Living Systems (ExCELLS), National Institutes of Natural Sciences, 5-1 Higashiyama, Myodaiji, Okazaki 444-8787, Aichi, Japan

* Correspondence: tomizuka@toyaku.ac.jp

Abstract

Background/Objectives: Mammalian display enables direct screening of full-length IgG antibodies, but genetically controlled library construction remains technically demanding. Here, we developed a mammalian display platform in which a target-naïve Δ DH antibody scaffold is first established at a defined human artificial chromosome (HAC) acceptor site, and CDR-H3 diversity is then introduced directly into that fixed genomic background by CRISPR-Cas9-mediated homology-directed repair (HDR). **Methods:** A site-specifically integrated display cell line carrying the Δ DH scaffold on an HAC was generated by Cre-mediated recombination. Randomized CDR-H3 libraries were introduced into this defined genomic background by HDR and screened by magnetic- and fluorescence-activated cell sorting together with orthogonal antigen-detection formats. Binding of isolated clones was characterized by flow cytometry and enzyme-linked immunosorbent assay. **Results:** Although early selection enriched clones reactive with a fluorescent reagent rather than the intended antigen, refinement of the selection scheme enabled isolation of AFFVVF. This selected antibody variant recognized multiple SARS-CoV-2 receptor-binding domain (RBD) variants by flow cytometry, with no detectable binding to unrelated antigens. As a recombinant protein, AFFVVF bound the JN.1 and XEC spike trimers by ELISA, but not monomeric RBD, including the XEC RBD used for selection, indicating that the form of antigen recognized differed between the cell-based and plate-based assays. Exploratory selections against epidermal growth factor receptor (EGFR) and glycoprotein A33 (GPA33) also produced antigen-dependent population shifts and enrichment of discrete sequence variants, although individual candidates were not independently validated as recombinant antibodies. **Conclusions:** These results provide proof of principle that targeted CDR-H3 diversification within a pre-established, genetically defined, target-naïve full-length IgG scaffold can yield a new antigen-binding antibody without starting from a pre-existing antigen-specific antibody or repertoire.



Academic Editor: Itai Benhar

Received: 11 September 2026

Revised: 16 October 2026

Accepted: 19 October 2026

Published: 22 October 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: mammalian display; human artificial chromosome; CRISPR-Cas9; homology-directed repair; site-specific integration; antibody library screening; genome editing; SARS-CoV-2 receptor-binding domain

1. Introduction

Antibodies constitute one of the largest and most rapidly expanding classes of therapeutics, with more than 200 antibody-based drugs approved worldwide and their applications continuing to expand across diverse molecular formats and disease areas [1]. The development of a novel therapeutic antibody can be broadly divided into two stages: the discovery of antigen-specific binders and the subsequent optimization of their affinity, specificity, and developability [2]. Initial binder discovery typically relies on either in vitro display technologies (using naïve or synthetic libraries) or in vivo immunization [2–4]. Phage and yeast displays accommodate large, highly diverse libraries (on the order of 10^{11} and 10^9 variants, respectively) and are therefore well suited to de novo discovery; however, antibodies are commonly screened as fragments such as scFv or Fab and are subsequently reformatted into full-length IgG for further characterization [3]. In vivo immunization yields antibodies shaped by physiological B-cell selection and maturation but can be constrained by immune tolerance, particularly toward self-antigens or antigens highly conserved between the host and target species. This can complicate the generation of cross-species-reactive antibodies, which are often required for preclinical studies [5].

Mammalian display offers a complementary approach in which antibodies are screened directly as full-length IgG within a mammalian cellular environment. This permits selection in the context of mammalian protein folding, assembly, and post-translational processing and enables early assessment of expression and other developability-relevant properties [3]. Historically, however, mammalian display has been limited by the difficulty of generating large libraries while maintaining reliable genotype–phenotype linkage and uniform antibody expression. Conventional transfection- or viral-vector-based delivery can introduce variation in genomic integration site, copy number, and expression level, potentially confounding antigen-binding selection [6,7]. Although recent genome-engineering approaches have substantially increased achievable library sizes, mammalian display libraries (typically 10^5 – 10^7 variants for genome-engineered systems) generally remain smaller and more technically demanding to construct than those used in phage or yeast display [3,6,7].

To address these limitations, site-specific integration and genome-editing strategies have been developed to confine antibody expression to defined genomic loci and reduce expression heterogeneity [8]. CRISPR-Cas9- or TALEN-mediated approaches have enabled mammalian antibody libraries ranging from more than 10^5 to as many as 10^7 variants [6,7,9] and have been applied successfully to affinity maturation, deep mutational scanning, and optimization of pre-existing antibodies [6,7,9]. Other mammalian display strategies, including somatic hypermutation and repertoire-based approaches, have also supported antibody discovery or maturation [10–13]. Substantially larger naïve libraries (up to $\sim 10^9$ variants) have also been generated using lentivirus-mediated random integration, which tolerates variations in copy number per cell and typically screens antibodies as secreted single-chain fragments detected indirectly on antigen-expressing cells; reformatting to full-length IgG after selection can then alter or abolish binding activity [14]. One strategy for generating genetically uniform libraries is to first establish a clonal cell line carrying an antibody-expression cassette at a defined site and then introduce diversity at defined positions within that fixed genomic background through homology-directed repair (HDR) using synthetic

oligonucleotide donors [7]. In this configuration, library members can share the same integration site, expression architecture, and antibody framework. Nevertheless, direct generation and screening of diversity from a genetically defined, target-naïve full-length IgG scaffold, without relying on a pre-existing binder or an immune-derived repertoire, remains relatively underexplored.

We reasoned that CDR-H3 is a particularly suitable target for this purpose. Among the six complementarity-determining regions, CDR-H3 exhibits exceptional diversity in sequence, length, and structure and occupies a central position within the antigen-binding site, making it a major determinant of antigen recognition [15,16]. Critically, diversity restricted to the heavy-chain CDR3 has been shown to be sufficient to generate antibodies capable of recognizing a broad range of hapten and protein antigens [17]. Consistent with this, fully synthetic Fab libraries built on fixed germline-based frameworks, with germline-derived CDR-H1/H2 and randomization concentrated in CDR-H3 and CDR-L3, have yielded binders against diverse antigens [18]. At the same time, CDR-H3 diversity is not strictly required for antigen recognition, since functional antibodies against multiple antigens have also been obtained from synthetic repertoires carrying a fixed CDR-H3 and diversity in CDR-L3 [19]. We therefore selected CDR-H3-only diversification as a simple and genetically tractable starting point, without assuming that it is sufficient for every target. We hypothesized that extensive diversification of CDR-H3, rather than local mutagenesis of an existing binder, could generate new antigen-binding specificities from a target-naïve scaffold. To implement this concept, we designed a Δ DH scaffold based on human germline antibody sequences and introduced randomized CDR-H3 segments at a single defined position by CRISPR-Cas9-mediated HDR [20]. Because diversification is performed within a pre-established, site-specifically integrated display line, library generation is decoupled from initial cassette integration, thereby preserving a common genomic and expression background across library members.

In this study, we established a mammalian display workflow in which a target-naïve full-length IgG scaffold is first fixed at a defined human artificial chromosome (HAC) acceptor site and CDR-H3 diversity is then introduced directly into that same genomic background by CRISPR-Cas9-mediated HDR. We optimized library construction and selection and combined cell-surface IgG display with a workflow integrating magnetic-activated cell sorting (MACS) and fluorescence-activated cell sorting (FACS). Using the SARS-CoV-2 receptor-binding domain (RBD) as a model antigen, we isolated AFFVVF, a newly selected antibody variant that recognized monomeric RBD when displayed on the cell surface, and we additionally obtained antigen-dependent enriched populations in exploratory screens against EGFR and GPA33. Together, these results support genetically defined, scaffold-first mammalian display as a proof-of-principle route for generating new full-length IgG binders without starting from an antigen-specific antibody.

2. Materials and Methods

2.1. Cell Culture Conditions

The parental HEK293T cell line (RCB2202) was obtained from the RIKEN BioResource Research Center (Tsukuba, Japan). The CHO21-HAC2 donor cell line carrying the 21HAC2 human artificial chromosome [21] was provided by the laboratory of Y. Kazuki (Tottori University, Tottori, Japan). HEK293T cells harboring 21HAC2 were established in our laboratory by microcell-mediated chromosome transfer from CHO21-HAC2 and have not been described previously. Neither this cell line nor the HAC/293T- Δ DH display cell line generated in this study (Figure S1) has been deposited in a public cell bank.

HEK293T cells harboring the HAC were cultured in D-MEM (high glucose) (containing L-glutamine and phenol red) (FUJIFILM Wako Pure Chemical Corporation, Os-

aka, Japan) supplemented with 10% fetal bovine serum (FBS) (Nichirei Biosciences Inc., Tokyo, Japan) and 1% penicillin-streptomycin solution ($\times 100$) (FUJIFILM Wako Pure Chemical Corporation).

Prior to genetic modification within the HAC, 8 $\mu\text{g}/\text{mL}$ Blastidicin S (FUJIFILM Wako Pure Chemical Corporation) was added to the culture medium to suppress HAC loss during cell culture. Following genetic modification, Blastidicin S was omitted during drug selection, and 2% HAT Supplement ($50\times$) (Gibco, Thermo Fisher Scientific, Waltham, MA, USA) was added instead to suppress the proliferation of unmodified cells. All cells used in this study were maintained in an incubator at 37 °C with 5% CO_2 .

2.2. Design of Antibody Genes for Genome Editing

The ΔDH scaffold was designed from human germline-derived heavy-chain elements by directly joining IGHV3-23 and IGHJ4 after removal of the D-segment-equivalent central region. Junctional residues were adjusted to create a defined genome-editing site while maintaining an in-frame CDR-H3 architecture.

At the IGHV3-23 end of the junction, lysine (K) was changed to arginine (R) to establish the desired gRNA/protospacer adjacent motif (PAM) configuration for Cas9 cleavage.

The residues flanking the insertion junction were set to serine (S) and glycine (G) to generate a compact in-frame junction surrounding the randomized insertion site. The human IgG1 heavy-chain constant region was used, followed by the membrane-anchoring module described below.

The light-chain variable region was constructed by joining IGKV3-20 and IGKJ1, generating a germline-derived synthetic light-chain sequence that was not based on a specific known antigen-binding antibody.

2.3. Design of gRNA and ssODN

The efficiency of amino acid insertion via HDR is generally known to decrease with increasing distance from the cleavage site. Therefore, an sgRNA (custom-made by Integrated DNA Technologies, Coralville, IA, USA) was designed to cleave within the serine codon (TC/C) located at the center of CDR-H3. Scoring of this sgRNA sequence using CRISPOR yielded a high specificity score of 94, with 47 predicted off-target sites (0/0/1/1/45 at 0/1/2/3/4 mismatches, respectively).

With the cleavage site designed here, direct insertion would cause a frameshift in the reading frame. To maintain the correct reading frame of the antibody sequence, the template was designed so that the reading frame of the cleaved site matches that of the inserted amino acid codons. Single-stranded oligodeoxynucleotides (ssODNs) were used as the HDR template and, like the standard PCR primers used in this study, were custom-made by Eurofins Genomics. The ssODNs were designed to be complementary to the antisense strand and have 50-nt homology arms flanking the inserted amino acid sequence. Because insertion of multiple amino acid sequences after genome editing results in loss of the gRNA recognition sequence, the disruption of the gRNA recognition sequence is expected to reduce the likelihood of re-cleavage.

For the ssODN design used to introduce the CDR3 sequences of known anti-SARS-CoV-2 antibodies, sequences were selected from the CoV-AbDab database (<https://opig.stats.ox.ac.uk/webapps/covabdab/>, accessed on 18 October 2026) in which the heavy-chain germline (VH3-23, JH4) and light-chain germline (KV3-20, KJ1) matched those of the ΔDH antibody.

The gRNA and ssODN sequences used are shown in Source Data.

2.4. Construction of the Donor Plasmid Vector and Integration into the HAC

The antibody gene within the vector is located downstream of the CAG promoter, with expression of the light chain and heavy chain (including the Fc region) controlled by a single promoter. A Furin-T2A sequence is inserted between the light chain and heavy chain, inducing peptide cleavage. The cleaved light-chain fragment associates with the heavy chain to form a full-length IgG-format antibody. Downstream of the heavy-chain Fc sequence is a sequence encoding the transmembrane (TM) domain of the platelet-derived growth factor (PDGF) receptor, which anchors the expressed antibody to the cell-surface membrane.

The HAC/293T-ΔDH cell line was established using the 21HAC2 vector [21]. Upstream of the antibody sequence are a loxP sequence and a 3' HPRT sequence. The loxP sequence can be integrated by Cre protein into a gene containing the same loxP sequence, allowing the antibody gene cassette to be introduced into the loxP site retained in the HAC. Upon Cre-loxP-mediated recombination, the 5' HPRT sequence in the HAC fuses with the 3' HPRT sequence of the donor vector to form a complete HPRT gene, conferring HAT resistance to cells in which the antibody gene has been correctly integrated.

Electroporation was performed using 8 µg of the constructed plasmid and 2 µg of the Cre expression vector pBS185 (a gift from Brian Sauer; Addgene plasmid #11916; RRID: Addgene_11916), using a NEPA21 Super Electroporator (Nepa Gene Co., Ltd., Ichikawa, Japan) under conditions of 150 V and 5.0 ms, with 3×10^6 cells. Following electroporation, cells were allowed to recover for 48 h at 37 °C with 5% CO₂, and were subsequently incubated for 2 weeks in HAT-supplemented medium under the same conditions. Antibody expression was confirmed using a FACS Aria III (BD Biosciences, San Jose, CA, USA). After bulk selection of the antibody-positive cell population, single-cell cloning was performed. Targeted integration at the HAC acceptor site was assessed based on a relatively uniform antibody-expression profile together with PCR using previously described primers [22] (HPRT_junc_sp_F, 5'-CGGCTTCCTCCTCCTGAACAA-3'; HPRT_junc_sp_R, 5'-TCCATAAGACAGAATGCTATGCAACC-3') spanning the junction formed between the 5' and 3' HPRT sequences upon Cre-loxP-mediated recombination.

2.5. Genome Editing by Electroporation

The mammalian display library was constructed using the MaxCyte ATx electroporation system by incorporating random sequences into the heavy-chain CDR3 region of the antibody gene.

Optimization of reagent ratios and pulse conditions. To identify optimal electroporation conditions, small-scale reactions were performed using the OC-25 × 3 processing assembly. Cells were seeded at a density of 0.5×10^6 cells/mL 1 day prior to electroporation, harvested by centrifugation (200 × g, 5 min), and resuspended in MaxCyte Electroporation Buffer at 1×10^6 cells per reaction. sgRNA was prepared by annealing 200 µM Alt-R CRISPR-Cas9 tracrRNA and crRNA (6 µL each) at 95 °C for 5 min, followed by cooling to room temperature. The RNP complex was formed by incubating the annealed sgRNA with 16 µL of 62 µM Alt-R S.p. Cas9 Nuclease V3 at room temperature for 20 min. To prepare the final electroporation mixture, 8 µL of 300 µM ssODN and 9.6 µL of Alt-R Cas9 Electroporation Enhancer were added to the RNP complex. The concentrations of Cas9, sgRNA, and ssODN were varied across three reagent conditions, each tested with two electroporation pulse programs (Opt4 and Opt7), yielding six conditions in total, designated EP1–EP6 (Table S1).

Large-scale library construction. Based on the optimized conditions identified as EP1, large-scale electroporation was performed by scaling the EP1 reagent ratio proportionally to the cell number, using the OC-400 or R-1000 processing assembly (MaxCyte). For the large-

scale 6NNB library, six OC-400 reactions (1.6×10^7 cells each) were performed, yielding a total of 9.6×10^7 cells. For the 10NNB and CX4C libraries, two OC-400 reactions and one R-1000 reaction (4.0×10^7 cells) were performed, yielding a total of 7.2×10^7 cells.

Following electroporation, cells were transferred to pre-warmed culture dishes and allowed to recover for 20 min at 37 °C with 5% CO₂. They were then cultured in medium supplemented with 1 µM HDR Enhancer V2. After 48 h, the medium was completely replaced to remove the enhancer, at which point all suspended cells were collected and re-seeded.

All CRISPR reagents except the ssODNs were purchased from Integrated DNA Technologies. Details are provided in Source Data.

2.6. Flow Cytometry Analysis and Cell Sorting

Flow cytometry analysis and cell sorting were performed using a BD FACS Melody (BD Biosciences). After removal of the culture medium, cells were washed with PBS and harvested by trypsinization. Harvested cells were pelleted by centrifugation ($\sim 200 \times g$, 5 min) and resuspended in PBS containing 1% FBS (FACS buffer). After cell counting, each sample was adjusted to 1×10^6 cells.

To analyze antibody expression, fluorescently labeled antibodies were added and cells were incubated on ice in the dark for 30 min. To analyze antigen binding, antigen protein (0.3 µg/ 10^6 cells in a total staining volume of 50 µL) was added as the primary stain and incubated on ice for 30 min; after washing, a fluorescently labeled secondary antibody was added and incubated for a further 30 min on ice in the dark. Cells were washed with FACS buffer after each staining step.

Cells were finally resuspended in FACS buffer, passed through a filter (~ 35 µm), and subjected to flow cytometric analysis or fluorescence-activated cell sorting. Details of the antigen proteins, labeling reagents, and concentrations used are provided in Source Data. The amounts of each reagent and reaction volume were adjusted according to the manufacturer's instructions.

Flow cytometry data were acquired as FCS files and analyzed using FlowJo 10 or 11 (BD Biosciences). Gating was performed by first selecting the main cell population based on forward and side scatter (FSC-A vs. SSC-A), followed by doublet exclusion using FSC-W vs. FSC-H and SSC-A vs. SSC-H, followed by evaluation of fluorescence intensity. No viability dye was used.

Because the gating strategy restricted analysis to cells within the IgG-positive population, the reported binding-positive percentages reflect antigen binding among surface-IgG-expressing cells rather than the total cell population. As the CDR-H3 sequence can influence surface IgG expression level (see Section 3.4), these percentages should be interpreted as a measure of binding specificity within the display-competent population, rather than as an absolute value directly comparable across clones with differing expression levels.

Flow cytometry experiments presented in this study generally represent individual measurements performed once per condition, rather than assays independently repeated across separate biological replicates. The exception is the binding of AFFVVF to XEC RBD, which was examined in independent experiments using two different detection systems (anti-His tag with Alexa Fluor 647, and biotinylated antigen with APC-streptavidin; Section 3.8).

For detection with biotinylated antigen, cells expressing AFFVVF or the parental ΔDH antibody (4×10^5 cells per tube) were incubated with biotinylated XEC RBD, biotinylated CD19 (an unrelated antigen), or no antigen (0.3 µg per 50 µL), and then stained with APC-streptavidin and anti-human IgG-Fc (PE/Dazzle 594).

2.7. Magnetic-Activated Cell Sorting

MACS was performed using a MidiMACS Separator and LS Columns (Miltenyi Biotec, Bergisch Gladbach, Germany). Cells were washed with MACS buffer (PBS, 2 mM EDTA, 0.5% BSA) and incubated with labeled antibody or antigen protein on ice in the dark for 20 min. Cells were then washed twice with MACS buffer and incubated with the corresponding microbeads for a further 20 min under the same conditions. After incubation, cells were washed and resuspended in 1 mL of MACS buffer.

The resuspended cells were applied to a magnetic column pre-equilibrated with MACS buffer and washed three times with 3 mL of MACS buffer. After magnetic separation, the column was removed from the separator, and labeled cells were eluted with 3 mL of MACS buffer. Eluted cells were collected by centrifugation ($\sim 200 \times g$, 5 min) and cultured according to their number.

The amount of each reagent and the reaction volume were adjusted according to the manufacturer's instructions, proportionally scaled based on a reference of 1×10^6 cells.

2.8. Large-Scale Antibody Library Construction and Culture

Under the optimized genome editing conditions, the number of cells recoverable at the 10^6 -cell scale was insufficient to construct a library approaching the targeted scale of 10^6 – 10^7 variants. Electroporation was therefore performed using approximately 1×10^8 cells in an attempt to compensate for this limitation and isolate novel binding antibodies. Three templates (6NNB, 10NNB, and CX4C) were used, and cells were processed in multiple batches of either 1.6×10^7 or 4×10^7 cells per round. Reagent conditions optimized at the 10^6 -cell scale were maintained and scaled up in proportion to cell number.

When culturing a cell population carrying an antibody library over an extended period, differences in proliferation rate between individual cells introduce bias, preventing proper enrichment. To standardize the culture duration of cells following electroporation, the genome-edited cell population was cryopreserved once at the timepoint corresponding to 48 h of HDR Enhancer V2 treatment plus 72 h of recovery culture; subsequent procedures were then performed on all cells collected together within the same tube at that time point. To efficiently recover antibody-expressing cells, antibody-positive cells were enriched using an anti-human Fc antibody and MACS beads.

2.9. DNA Extraction from Cells

Genomic DNA was extracted from cells using the Wizard Genomic DNA Purification Kit (Promega, Madison, WI, USA) in accordance with the manufacturer's protocol for tissue culture cells, with the following modifications: cells were pelleted at $200 \times g$ for 5 min; the centrifugation steps following protein precipitation, isopropanol precipitation, and ethanol washing were each performed at $15,000 \times g$ for 5 min; and the DNA pellet was redissolved in 50 μ L of TE (10 mM Tris-HCl [pH 8.0], 1 mM EDTA-2Na [pH 8.0]) by incubating at 65 °C for 1 h, followed by storage at 4 °C.

2.10. PCR Amplification

PCR was performed using TaKaRa Ex Premier DNA Polymerase Dye Plus (TaKaRa Bio, Kusatsu, Japan). Reactions were prepared in a total volume of 10 μ L, consisting of 5 μ L of Polymerase mix, forward and reverse primers each at a final concentration of 200 nM, 50–100 ng of template DNA, and sterile water. Primers were custom-made by Eurofins Genomics.

PCR was performed with initial denaturation at 94 °C for 1 min, followed by 30 cycles of 98 °C for 10 s, 55–60 °C for 15 s, and 68 °C for 30 s/kb, with a final hold at 4 °C. The identity of amplified products was confirmed by agarose gel electrophoresis. Gel images

were acquired as JPEG files directly from the gel documentation system. For presentation, images were cropped, and where brightness and contrast were adjusted, the adjustments were applied uniformly across the entire image; no feature-specific enhancement was performed. The primer sequences used are shown in Source Data.

For long-read PCR to amplify the full-length antibody sequence, primers annealing to the CAG promoter and hGH poly(A) region flanking the antibody coding sequence on either side were used.

2.11. Sample Preparation for Next-Generation Sequencing (NGS)

For short-amplicon sequencing around the CDR-H3 editing site, samples were prepared using primers with adapter sequences added in addition to the primer sequences used for standard PCR. PCR reactions were performed in a total volume of 40 µL. A portion of the amplified product (10 µL) was confirmed by agarose gel electrophoresis, and samples yielding the expected band were used as the primary PCR product for subsequent analysis. Sequencing of the resulting PCR products was outsourced to Biotech Co., Ltd. (Sagami-hara, Japan), and performed using a NextSeq 1000 system with the NextSeq 1000/2000 P1 XLEAP-SBS Reagent kit (600 cycles) or a MiSeq system with the MiSeq Reagent Kit v3 (Illumina, San Diego, CA, USA), in a 2 × 300 bp paired-end configuration. Raw reads were quality-filtered using sickle (v.1.33), discarding bases with a quality score below 20 and removing any read (and its paired mate) whose length fell below 40 bp after trimming. Paired-end reads were then merged using FLASH (v.1.2.11) with the default parameters. Merged reads were retained for downstream analysis only if they exactly matched the expected primer sequences.

For long-amplicon sequencing of the full-length antibody-coding region, the target region was amplified by 2-step PCR (TaKaRa Ex Premier DNA Polymerase) with initial denaturation at 94 °C for 1 min, followed by 10 cycles of 98 °C for 10 s, 60 °C for 15 s, and 68 °C for 30 s. Libraries were prepared using the SMRTbell prep kit 3.0 (PacBio, Menlo Park, CA, USA) and sequenced on a Revio system (PacBio) with the Revio SPRQ polymerase kit, through the same outsourced sequencing service. Circular consensus sequences were generated using SMRT Link (v.25.2.0.266456); reads with a mean quality score below 20 were discarded to yield HiFi reads. Reads were demultiplexed by index using lima (v.2.12.0), which also removed adapter and index sequences; when the number of reads per sample exceeded approximately 40,000, reads were downsampled to that level.

2.12. Next-Generation Sequencing (NGS) Data Analysis

For short-amplicon NGS analysis, reads were extracted that exactly matched the 5-bp flanking sequences (GATCC and GGCTA) on either side of the editing site. Reads with no sequence alterations were classified as “Unedited (WT).” Reads containing frameshift-free insertions corresponding to the designed random sequences (6NNB, CX4C, 10NNB) were classified accordingly. All remaining reads, including those with mismatches in the flanking regions or frameshift mutations, were classified as “Indel.” Given the sequencing quality obtained ($Q_{30} \geq 93.8\%$ across all samples), the expected contribution of sequencing error to flanking-region mismatches was minor relative to the observed indel frequencies.

For long-amplicon reads, sequences outside the inserted region were further compared with the pre-editing reference to subclassify reads by mutation type. Reads identical to the reference across the entire analyzed region were classified as “No mutation”; reads containing amino acid substitutions without insertions or deletions were classified as “Substitution”; reads shorter than the reference, consistent with a deletion, were classified as “Deletion”; reads containing an insertion within the expected 10NNB length range were classified as “Partial 10NNB insertion” if the sequence outside the inserted CDR-H3

contained additional amino acid substitutions or indels, or as “Complete 10NNB insertion” if that sequence was completely identical to the reference (the latter considered sequence-intact); reads containing an insertion outside the expected 10NNB length range were classified as “Insertion of other length”; and reads lacking a definable start or stop position within the antibody gene were classified as “Other.”

The number of unique CDR-H3 sequences observed at each selection round was calculated using a custom script written in Python 3.13 and implemented in the Google Colaboratory environment (Code S1), with the read counts in Data S6 as input.

The raw sequencing data generated in this study have been deposited in the DDBJ Sequence Read Archive (DRA) under BioProject accession number PRJDB45658 (Run accessions DRR1092909–DRR1092925).

2.13. CDR-H3 Hydrophobicity Analysis

The hydrophobicity of each unique CDR-H3 sequence was calculated using the hydrophathy scale of Kyte and Doolittle [23], applied to the randomized six-amino-acid segment within the CDR-H3. For each population at a given selection round, a read-weighted mean hydrophobicity score was calculated by weighting each unique sequence’s score by its relative read frequency within that population. Calculations were performed using the same Python 3.13 script (Code S1), with Data S6 as input; sequences containing stop codons were excluded from this analysis.

2.14. ELISA for Binding Activity

ELISA was performed using Nunc MaxiSorp 96-well plates (Thermo Fisher Scientific). The plates were coated with antigen protein (listed in Source Data) dissolved in carbonate buffer (50 mM, pH 9.6) at 50 ng/100 µL/well, and incubated for 1 h at room temperature or overnight at 4 °C.

After coating, the plates were washed four times with PBS-T and blocked with 200 µL/well of TBS-T (TaKaRa Bio) containing 1% bovine serum albumin (Proliant Health & Biologicals, Feilding, New Zealand) for 30 min at room temperature. After washing, recombinant antibodies (produced by Biointron, Shanghai, China) were serially diluted (0–1000 ng/mL or 0–50,000 ng/mL), added at 100 µL/well, and incubated for 1 h at room temperature. Following another wash with PBS-T, HRP-conjugated goat anti-human IgG (Fc-specific) secondary antibody (Sigma-Aldrich, St. Louis, MO, USA; Cat# A0170) diluted 1:60,000 was added at 200 µL/well and incubated for 1 h at room temperature.

After washing, TMB substrate (TMB Microwell Peroxidase Substrate System) was added at 100 µL/well and allowed to develop for 15 min at room temperature, after which the reaction was stopped with sulfuric acid. Absorbance was measured at 450 nm using a Nivo Multi-Mode Plate Reader (Revvity).

In a separate set of experiments, binding of recombinant AFFVVF to the XEC spike trimer, the XEC spike RBD, the JN.1 spike trimer, and the His-tagged extracellular domain of EGFR was examined using the same coating, blocking, detection, and development conditions. Two plates were processed in parallel on the same day, using a single antibody dilution series, a single dilution of secondary antibody, and a single preparation of substrate. The first plate contained the JN.1 and XEC spike trimers probed with both AFFVVF and sotrovimab [24], the latter serving as a reference antibody to confirm that each antigen had been immobilized in a functional form. The second plate contained AFFVVF alone, probed against the XEC spike RBD, EGFR, and the two spike trimers; these last two conditions were common to both plates and allowed the responses obtained on the two plates to be compared directly. All antigens were coated at the same mass concentration (50 ng/100 µL/well); because the antigens differ in molecular mass, absorbance values

obtained with different antigens were not compared quantitatively. Each condition was tested in triplicate across seven antibody concentrations (12.21–50,000 ng/mL) together with wells receiving no antibody, which defined the background for that condition.

The resulting absorbance data were fitted to a four-parameter logistic (Rodbard) equation $[y = d + (a - d)/(1 + (x/c)^b)]$ using the Curve Fitting tool in ImageJ (v.1.53t, NIH). For each dataset, technical replicates that deviated by more than the typical spread observed among the remaining replicates at the same concentration were excluded prior to curve fitting; excluded and retained values are provided in Data S2 and S7. Because binding signals for several RBD variants were weak and close to the assay background, outlier identification in these datasets was necessarily more sensitive to small absolute differences in absorbance; raw data for all technical replicates, including those identified as outliers, are provided in Data S7 to allow independent evaluation. Raw data and curve-fitting parameters for the ELISA experiments shown in Figure S6 are provided in Data S8.

2.15. Use of Artificial Intelligence

During the preparation of this manuscript, the authors used Claude (Anthropic, San Francisco, CA, USA), accessed between June and October 2026, to assist with organizing content and translating and refining the original Japanese draft into English, as well as with aspects of NGS data analysis, including selection of diversity-estimation approaches and generation of associated plots, the design of the hydrophobicity analysis approach, and the design of the ELISA curve-fitting analysis workflow. No primary experimental data were generated, altered, or manipulated by the AI tool; all analyses were performed using author-verified scripts or established software. The author-verified analysis scripts used for these analyses are provided as Code S1. All text and analyses were reviewed and critically evaluated by the authors, and the authors take full responsibility for the content of this publication.

3. Results

3.1. Design and Establishment of a Defined Mammalian Antibody-Display Scaffold

For the mammalian antibody-display scaffold, we selected an IGHV3-23/IGHJ4 heavy chain paired with an IGKV3-20/IGKJ1 light chain. These germline elements are frequently used in the human antibody repertoire, and IGHV3-23/IGKV3-20-based scaffolds have been reported to exhibit favorable biophysical properties in synthetic antibody libraries [25], motivating their use in our platform. The light chain and membrane-anchored heavy chain were co-expressed from a single transcript separated by a furin cleavage site and a T2A self-cleaving peptide. The heavy chain consisted of the VH domain fused to the human IgG1 heavy-chain constant region (CH1–CH3), followed by a C-terminal PDGFR transmembrane domain to enable cell-surface display of full-length IgG (Figures 1A and S1C).

To enable targeted diversification of CDR-H3, we designed a Δ DH scaffold carrying a minimized CDR-H3 sequence, CARSGYFDY, in which the D-segment-equivalent central region was replaced by a short in-frame bridging sequence between IGHV3-23 and IGHJ4 (Figure 1A). This bridging sequence contains a Cas9 cleavage site within the serine codon (TCC), allowing additional CDR-H3 sequences to be introduced at a defined position by HDR using synthetic oligonucleotide donors (Figure 1B). We used an insertion-based diversification strategy so that HDR products carrying the designed insert could be distinguished from the parental Δ DH sequence by amplicon length and subsequently characterized by sequence analysis. This scaffold therefore provided a defined genetic background for introducing either predetermined or randomized CDR-H3 sequences.

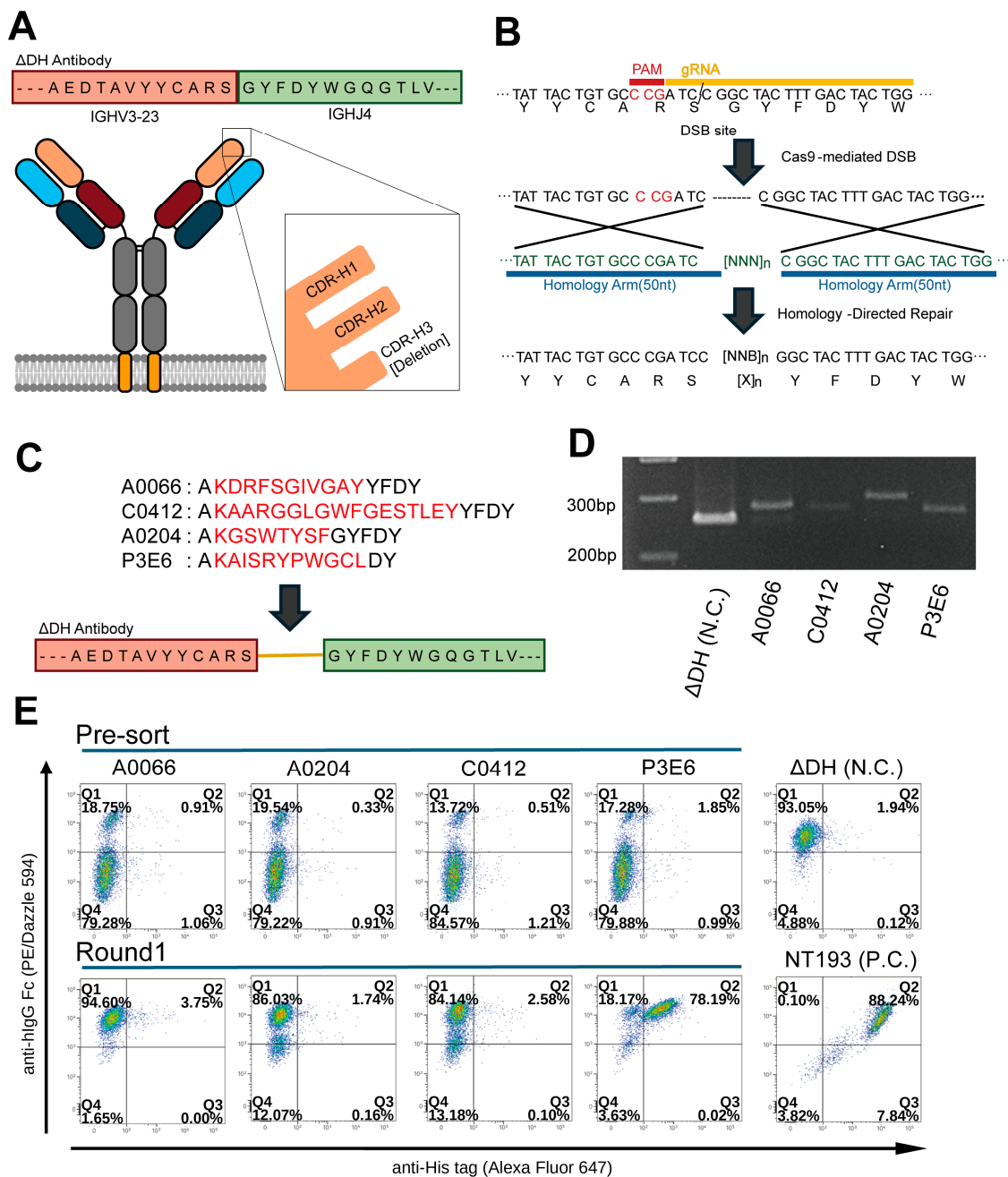


Figure 1. Insertion of known anti-SARS-CoV-2 antibody CDR-H3 sequences into the Δ DH scaffold. (A) Schematic of the Δ DH antibody scaffold based on the IGHV3-23/IGHJ4 germline framework. The D segment-equivalent region of CDR-H3 was deleted and replaced with a minimal in-frame bridging sequence, generating a defined single-cleavage site for Cas9-mediated homology-directed repair (HDR)-based insertion. (B) Strategy for HDR-mediated insertion of randomized or defined CDR-H3 sequences. A Cas9 ribonucleoprotein complex generates a double-strand break (DSB) at the bridging sequence (PAM and gRNA target site indicated). Repair using a donor template flanked by 50-nt homology arms results in insertion of the desired sequence ([NNB]_n or a defined [X]_n sequence) at the cleavage site. (C) Four CDR-H3 sequences derived from known anti-SARS-CoV-2 antibodies (A0066, C0412, A0204, P3E6), selected from the CoV-AbDab database for compatibility with the IGHV3-23/IGKV3-20 germline background, were each inserted into the Δ DH scaffold at the engineered cleavage site within the bridging region. (D) PCR analysis spanning the Δ DH locus, showing a band shift for each of the four candidate sequences relative to the unmodified Δ DH scaffold (negative control, N.C.). (E) Flow-cytometric analysis of antigen binding before (Pre-sort) and after one round of enrichment (Round 1) using fluorescently labeled SARS-CoV-2 spike protein

(Alpha variant). Surface-IgG-positive cells were gated into Q1 (IgG⁺/antigen[−]) and Q2 (IgG⁺/antigen⁺) based on antigen-binding signal. ΔDH (N.C.) and NT193 (P.C.) serve as negative and positive controls, respectively. Percentages of the antigen-binding-positive population are indicated for each condition. Colors denote antibody domains: VH (orange), VL (light blue), CH1 (dark red), CL (dark teal), CH2–CH3 (gray), and the transmembrane anchor (amber). In (A), the box and connecting lines indicate an enlarged view of the variable region. In (B), the vertical arrows indicate successive reaction steps and the crossed lines indicate homologous recombination between the donor template and the target locus. In (C), the arrow indicates insertion of the listed CDR-H3 sequences at the position shown.

To establish a defined genomic background for mammalian display, we used 293T cells harboring an HAC introduced by microcell-mediated chromosome transfer (Figure S1A). The HAC contains a single loxP acceptor site adjacent to a 5′ HPRT fragment, whereas the donor vector carries a loxP site, the complementary 3′ HPRT fragment, and the antibody expression cassette. Cre-mediated recombination at the HAC acceptor site reconstituted a functional HPRT gene, enabling selection of targeted recombinants in HAT-containing medium (Figure S1B). The integrated expression cassette encoded the light chain and membrane-anchored heavy chain as a single expression unit (Figure S1C). To generate a stable display cell line, a donor vector encoding the ΔDH scaffold was co-transfected with a Cre recombinase expression vector into HAC/293T cells. Following HAT selection and cloning, cell-surface antibody expression was assessed by flow cytometry, and a clone showing strong and relatively uniform surface IgG expression was selected and designated HAC/293T-ΔDH (Figure S1D). PCR using primer pairs spanning the reconstituted HPRT junction and the IGHV3-23/IGHJ4 region confirmed integration of the antibody cassette into the HAC acceptor site (Figure S1E). HAC/293T-ΔDH was used as the parental cell line for subsequent CDR-H3 editing.

3.2. Targeted Insertion of a Defined CDR-H3 Sequence Restored Antigen Binding

We first examined whether insertion of a defined CDR-H3 sequence could confer antigen-binding activity on the ΔDH scaffold. Four CDR-H3 sequences derived from reported anti-SARS-CoV-2 antibodies with an IGHV3-23/IGKV3-20 germline background were selected from the CoV-AbDab database and introduced individually into 1×10^6 HAC/293T-ΔDH cells using Cas9 ribonucleoprotein complexes and sequence-specific single-stranded oligodeoxynucleotide (ssODN) donors (Figure 1C). PCR analysis across the ΔDH locus 10 days after electroporation showed larger amplicons for all four edited populations than for the parental ΔDH locus, with size shifts consistent with the respective inserted sequences (Figure 1D).

Antigen binding was assessed by flow cytometry using His-tagged SARS-CoV-2 Alpha spike protein detected with an Alexa Fluor 647-conjugated anti-His antibody, while surface IgG expression was measured in parallel (Figure 1E). Before antigen-based enrichment, approximately 14–20% of cells in the four edited populations were positive for surface IgG (Q1 + Q2), compared with approximately 93% of the parental ΔDH cells. The parental ΔDH cells showed essentially no antigen-binding population despite strong IgG expression on their surface.

In the P3E6-edited population, 1.85% of cells were positive for both surface IgG expression and antigen binding before enrichment. After one round of antigen-based enrichment, this proportion increased markedly to 78.2%, approaching that of the positive-control antibody NT193 [26] (88.2%). In contrast, the corresponding fractions for A0066, A0204, and C0412 remained low after enrichment (3.8%, 1.7%, and 2.6%, respectively). These results show that insertion of an appropriate CDR-H3 sequence into the ΔDH scaffold

can restore antigen-binding activity and that cells displaying the resulting antibody can be efficiently enriched by mammalian display selection.

3.3. Pilot Screening of a Randomized 6NNB Library Selected an Alexa Fluor 647-Reactive Clone

We next tested whether antigen-binding antibodies could be isolated from a randomized CDR-H3 library. ssODN donors containing six NNB codons flanked by 50-nt homology arms were designed to introduce a randomized six-amino-acid segment at the Cas9 cleavage site of the Δ DH scaffold. The resulting sequences corresponded to CDR-H3 regions of 14 amino acids as defined by IgBLAST. NNB codons were selected because they encode all 20 amino acids while reducing the frequency of stop codons relative to other degenerate codon schemes such as NNK and NNS, thereby increasing the proportion of potentially productive CDR-H3 variants [27].

Cas9/gRNA ribonucleoprotein complexes and the randomized donor were co-electroporated into 1×10^6 HAC/293T- Δ DH cells. Following genome editing, surface hIgG-Fc-positive cells were enriched and subjected to antigen-binding selection using His-tagged SARS-CoV-2 Alpha spike protein detected with an Alexa Fluor 647-conjugated anti-His antibody. The binding-positive population increased after successive rounds of selection (Figure 2A). However, a comparable signal was observed in the negative-control (OVA antigen) condition, with 86.8% and 85.0% of cells falling within the double-positive gate in the spike and negative-control conditions, respectively. A true antigen-free condition was additionally tested and showed a comparably high background signal (~83.8%); the corresponding raw flow cytometry data are provided as Source Data.

Furthermore, when the same selected population was analyzed using a PE/Dazzle 594-conjugated anti-His antibody instead of the Alexa Fluor 647-conjugated reagent, the signal decreased to the level observed in parental Δ DH cells (Figure 2B). These results indicated that enrichment was dependent on the Alexa Fluor 647-based detection reagent rather than on the intended spike antigen.

NGS analysis showed progressive enrichment of a single inserted CDR-H3 sequence, LFVSVS. After two rounds of selection, LFVSVS accounted for 52,176 of 69,846 reads (74.7%), whereas no other 6NNB sequence showed comparable enrichment (Figure 2C, Table 1, and Data S1).

Table 1. NGS-based sequence composition of the 6NNB library across the screening series.

(top)				
	Pre-sort	hIgG-Fc-positive	1st round	2nd round
Unedited	16,653	13,695	10,161	2015
6NNB	23,527	58,069	48,307	13,178
LFVSVS	2	406	7137	52,176
Indel	22,694	1397	3506	2477
Total	62,876	73,567	69,111	69,846
(bottom)				
	Pre-sort	hIgG-Fc-positive	1st round	2nd round
Unedited	26.49%	18.62%	14.70%	2.88%
6NNB	37.42%	78.93%	69.90%	18.87%
LFVSVS	0.00%	0.55%	10.33%	74.70%
Indel	36.09%	1.90%	5.07%	3.55%

Read counts (top) and the corresponding percentage of total reads (bottom) for each sequence category (unedited, 6NNB [non-LFVSVS], LFVSVS, and indel) at each stage of the screening series: pre-sort, hIgG-Fc-positive population, 1st round, and 2nd round.

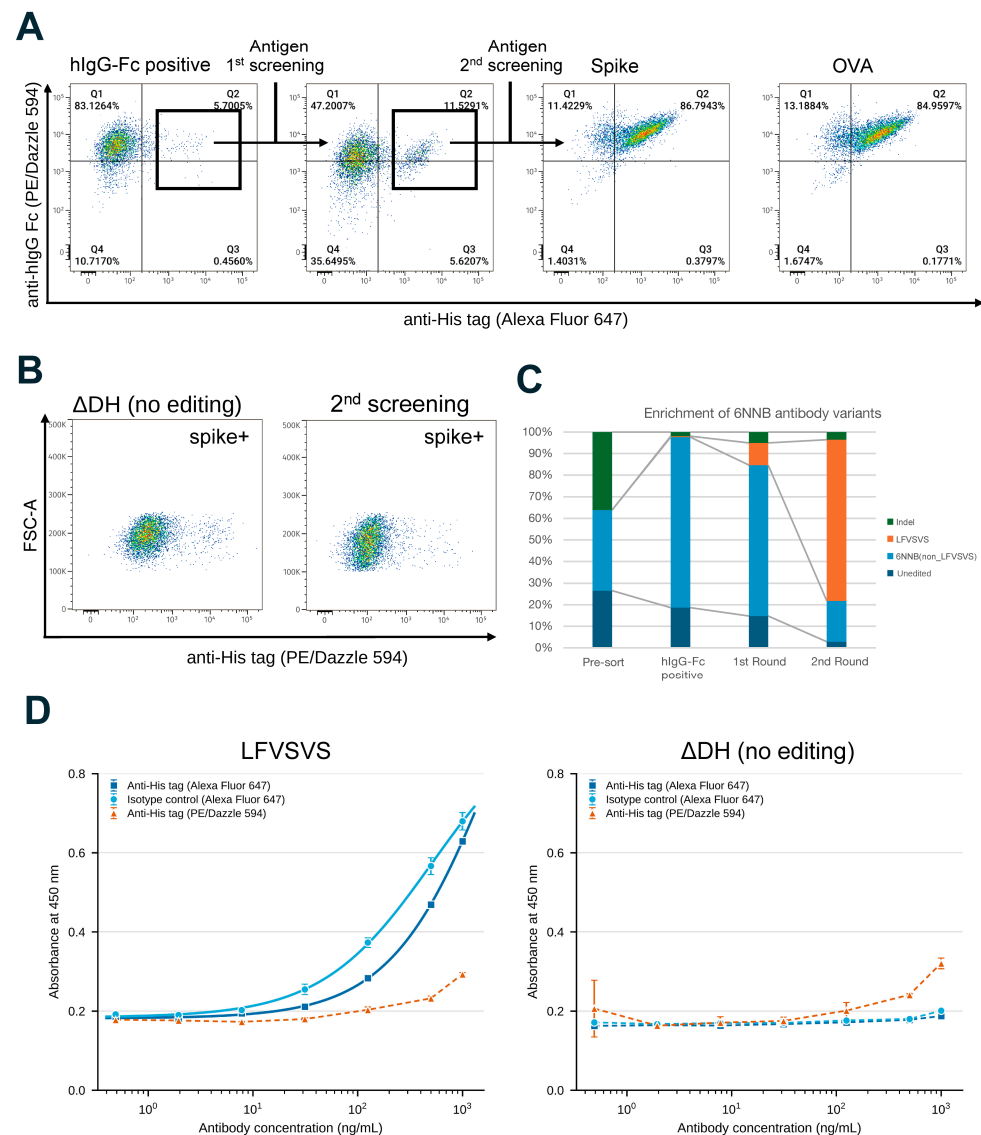


Figure 2. Feasibility of antibody isolation from a randomized 6NNB library at a small scale. **(A)** Flow-cytometric enrichment of antigen-binding cells from the hIgG-Fc-positive 6NNB library (6NNB Ab+) using fluorescently labeled SARS-CoV-2 spike protein (Alpha variant), shown before and after the 1st and 2nd rounds of enrichment ("Spike"), alongside a negative-control condition using an unrelated antigen (OVA). **(B)** Flow-cytometric analysis of the unmodified Δ DH scaffold (no editing) and the 2nd-screening population, both stained with spike protein detected via a PE/Dazzle 594-conjugated anti-His tag antibody (switched from the Alexa Fluor 647-conjugated antibody used in A). **(C)** NGS-based tracking of sequence composition across enrichment steps (6NNB, Ab+, 1st screening, 2nd screening), classified as unedited, 6NNB (non-LFVSVS), LFVSVS, or indel. **(D)** Indirect ELISA comparing the LFVSVS clone (left) with the unmodified Δ DH antibody (right) using an isotype control antibody and anti-His tag antibodies conjugated to Alexa Fluor 647 or PE/Dazzle 594. Data are mean $\pm$ SD of technical triplicates, except PE/Dazzle 594 anti-His tag at 1000 and 500 ng/mL for the Δ DH (unedited) antibody, where one outlier replicate was excluded ($n = 2$). Solid curves show four-parameter logistic fits; dashed lines connect the means of conditions for which a satisfactory fit could not be obtained.

To determine the binding specificity of this clone, recombinant LFVSVS and parental Δ DH antibodies were analyzed by ELISA. LFVSVS bound both the Alexa Fluor 647-conjugated anti-His antibody used during screening and an Alexa Fluor 647-conjugated mouse IgG2a/ κ isotype control, whereas its binding to the PE/Dazzle 594-conjugated anti-His antibody was substantially weaker (Figure 2D). The parental Δ DH antibody showed

no comparable binding to the Alexa Fluor 647-conjugated reagents. These results identify LFVSVS as an Alexa Fluor 647-reactive clone.

Thus, the pilot experiment demonstrated strong enrichment of a rare reagent-reactive variant while also revealing limitations in the sampled library diversity and in the use of a single detection reagent for target-directed selection.

3.4. Optimization of CDR-H3 Insertion and Enrichment of Surface-IgG-Positive Library Members

The pilot experiment indicated that library construction required improved recovery of cells carrying the designed CDR-H3 insert. We therefore optimized the genome-editing conditions by varying the amounts of Cas9, gRNA, and ssODN donor together with the electroporation pulse program. A 10NNB insertion design, corresponding to a 30-bp insertion, was used to facilitate detection of HDR products.

Cells were cultured for 48 h in the presence of HDR Enhancer V2 after electroporation and analyzed by PCR and flow cytometry (Figure S2). PCR detected an insertion-sized product under all conditions tested (Figure 3A). Among these conditions, the combination of Cas9:gRNA:ssODN at 10 µg (62 pmol):1.3 µg (75 pmol):6 µg (150 pmol) with pulse condition Opt4, designated EP1, showed the strongest enrichment of the 281-bp insertion product. In the hIgG-Fc-positive population obtained under EP1, the parental 251-bp band was nearly absent and the 281-bp band predominated.

Despite the high frequency of insertion-sized products detected by PCR, flow-cytometric analysis showed that only approximately 5–10% of cells retained surface hIgG-Fc expression comparable to that of the unedited parental cells (Figures 3B and S3). The remaining cells showed either middle or undetectable hIgG-Fc expression. These populations were therefore separated into hIgG-Fc-positive, -middle, and -negative fractions and analyzed independently.

For context, in earlier nuclease-directed mammalian display systems, antibody-expressing cells were typically enriched by drug selection after integration; for example, 77–78% of cells were positive for Fc expression and antigen binding after 14 days of blasticidin selection in single-antibody cell lines [6]. In the present system, no drug selection was applied after HDR-mediated insertion, and the edited population was heterogeneous, comprising cells with correct insertions, unedited cells, and cells carrying indels or stop-codon-containing inserts (Table 2). The surface-IgG-positive fraction before sorting was therefore lower (approximately 5–10%), and enrichment depended on physical sorting; sorting nevertheless yielded a population in which 96.25% of reads were non-stop 10NNB sequences, compared with 66.29% before sorting (Table 2). Because the starting populations and readouts differ among studies (HDR frequency, Fc-positive fraction, or antigen-positive fraction), these values are compared only qualitatively.

Short-amplicon NGS of the EP1 pre-sort population showed that 33,150 of 39,713 reads (83.47%) contained a 10NNB insertion, including 26,325 non-stop sequences (66.29%) and 6825 sequences containing a stop codon (17.19%) (Figure 3C, Table 2, and Data S3). Unedited sequences represented 1.22% of reads, and indels represented 15.31%. In the hIgG-Fc-positive fraction, non-stop 10NNB sequences were strongly enriched, accounting for 96.25% of reads. In contrast, the middle and negative fractions were dominated by stop-codon-containing 10NNB sequences (57.27% and 57.73%, respectively) and indels (35.69% and 35.13%, respectively). The high proportion of apparently non-stop 10NNB sequences in the pre-sort population contrasted with the much smaller proportion of hIgG-Fc-positive cells. This discrepancy indicated that short-amplicon analysis around the CDR-H3 insertion site did not fully account for the loss of antibody expression and prompted analysis of the complete antibody-coding region.

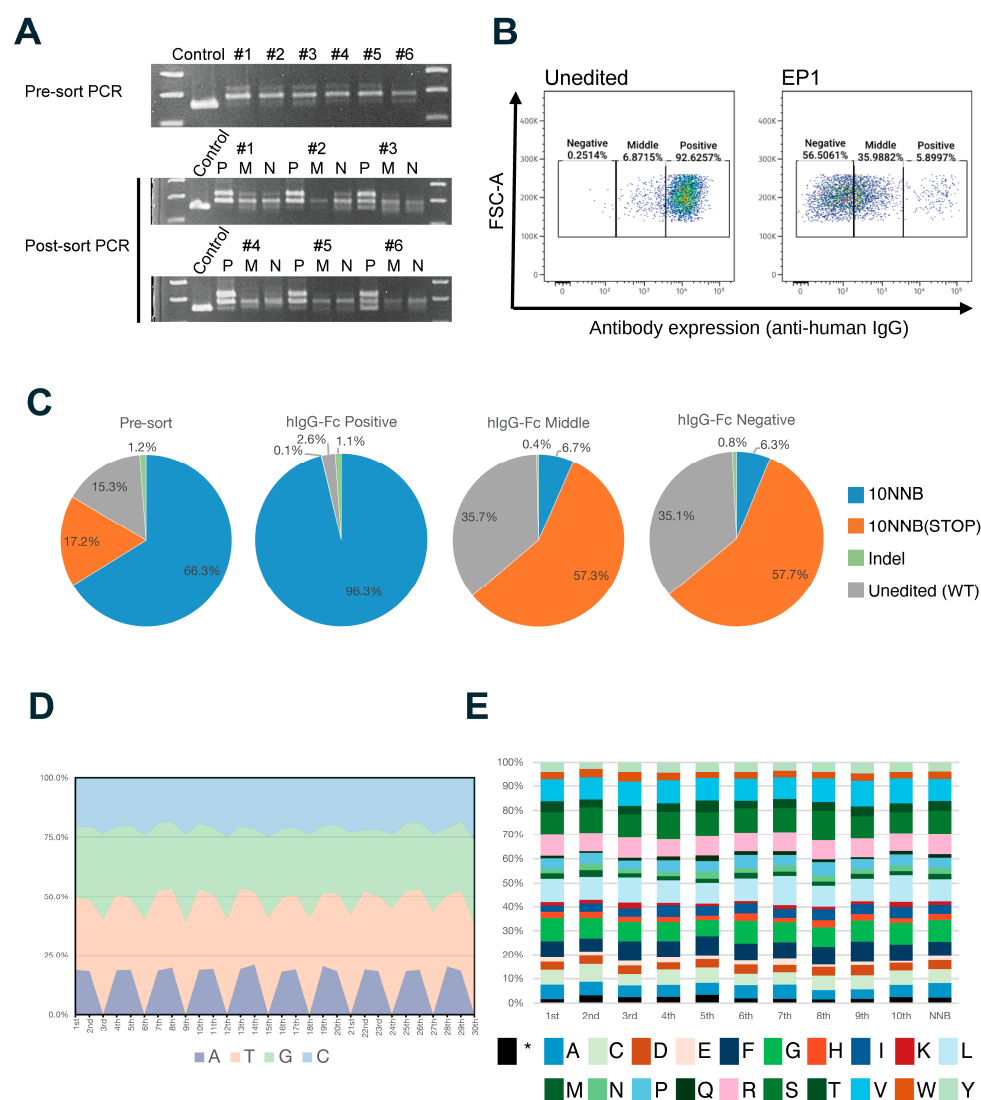


Figure 3. Randomization of CDR-H3 and optimization of genome-editing conditions. (A) PCR analysis of the target locus before (top) and after (bottom) FACS sorting, under the optimized condition (EP1) and five additional conditions tested (EP2–EP6). Post-sort PCR was performed on the hIgG-Fc-positive (P), -middle (M), and -negative (N) populations sorted from each condition. (B) Flow-cytometric analysis of surface antibody expression (anti-human IgG) for the unedited Δ DH scaffold and the EP1 condition, showing the gating strategy used to sort cells into negative, middle, and positive populations. These two panels are also shown in Figure S3, together with the corresponding isotype controls and the remaining editing conditions (EP2–EP6). (C) NGS-based classification of sequences (10NNB [in-frame], 10NNB containing a stop codon, indel, or unedited [WT]) in the pre-sort population and in the hIgG-Fc-positive, -middle, and -negative populations sorted from the EP1 condition. (D) Nucleotide frequency (A/T/G/C) at each of the 30 positions spanning the ten randomized NNB codons of the 10NNB insert, prior to selection. The periodic pattern reflects the defined base composition of the N and B degenerate codons and the exclusion of adenine at every third position. (E) Amino acid frequency at each of the ten randomized positions of the 10NNB insert, prior to selection. The rightmost column (“NNB”) shows the overall amino acid frequency across the entire pre-sort 10NNB library, for comparison with the per-position distributions. The pre-sort panel and the two post-sort panels (EP1–EP3 and EP4–EP6) in (A) were acquired as separate gels and are shown as separate panels, each with its own size markers. In (A), lanes #1–#6 correspond to conditions EP1–EP6. In (E), the asterisk denotes a stop codon.

Table 2. NGS-based sequence composition of the EP1 condition across sorted populations.

	(top)			
	Pre-sort	hIgG-Fc-positive	hIgG-Fc-middle	hIgG-Fc-negative
10NNB (non-stop)	26,325	43,074	2197	2290
10NNB (stop codon)	6825	26	18,892	20,869
Indel	6078	1142	11,772	12,699
Unedited (WT)	485	510	127	291
Total	39,713	44,752	32,988	36,149
	(bottom)			
	Pre-sort	hIgG-Fc-positive	hIgG-Fc-middle	hIgG-Fc-negative
10NNB (non-stop)	66.3%	96.3%	6.7%	6.3%
10NNB (stop codon)	17.2%	0.1%	57.3%	57.7%
Indel	15.3%	2.6%	35.7%	35.1%
Unedited (WT)	1.2%	1.1%	0.4%	0.8%

Read counts (top) and the corresponding percentage of total reads (bottom) for each sequence category (10NNB [non-stop], 10NNB [stop codon], indel, and unedited [WT]) in the pre-sort population and in the hIgG-Fc-positive, -middle, and -negative populations sorted from the EP1 condition.

The degenerate oligonucleotide donor was synthesized using defined N and B base compositions (N, A:T:G:C = 20:30:30:20; B, T:G:C = 40:35:25). Accordingly, the nucleotide frequencies showed the expected three-nucleotide periodicity across the 30 positions of the 10NNB insert (Figure 3D). Although amino acid frequencies varied among positions, all 20 amino acids were represented in the pre-selection library (Figure 3E).

3.5. Long-Amplicon Sequencing Identified Editing Products Not Resolved by Short-Amplicon Analysis

To determine whether sequence alterations outside the locally analyzed CDR-H3 region contributed to the discrepancy between short-amplicon sequencing and surface antibody expression, we performed long-amplicon NGS spanning the antibody-coding sequence. In contrast to the short-amplicon analysis, which classified reads primarily according to the sequence surrounding the insertion site, the long-amplicon analysis required the amino acid sequence outside the inserted region to match the parental Δ DH antibody sequence.

Among 45,075 analyzed reads, only 12.70% contained a complete 10NNB insertion without detectable amino acid changes outside the randomized region and without an internal stop codon, whereas reads with no mutation or with substitutions alone were rare (<1% combined; Table 3 and Data S4). Most reads instead carried more complex products: partial or other-length insertions accounted for 58.45% of reads and deletions for 19.57%, and a further 8.73% could not be assigned a defined start or end position within the antibody-coding sequence.

Collectively, these findings indicate that long-amplicon sequencing revealed a substantial frequency of complex editing products that were not resolved by analysis restricted to the immediate CDR-H3 insertion site, providing a possible explanation for the discrepancy between the high apparent insertion frequency detected by short-amplicon sequencing and the lower frequency of surface antibody expression.

Long-amplicon sequencing identified 2408 unique complete 10NNB sequences, of which 494 contained a stop codon, leaving 1914 unique non-stop sequences that met the sequence criteria described above. Several of these sequences were also detected preferentially in the hIgG-Fc-middle or -negative fractions in the short-amplicon analysis (Table 4 and Data S5). For example, GYCTLVHAWF and AGAAAWILIK were recovered

predominantly from the middle or negative populations despite meeting the long-amplicon sequence criteria.

Table 3. Long-amplicon NGS classification of full-length antibody-coding sequences.

Category	Reads	%	Notes
Deletion	8820	19.57%	200–2339 bp Includes reads with a differing nucleotide sequence (n = 19)
Substitution	165	0.37%	
No mutation	85	0.19%	2370 bp; excludes complete 10NNB insertions Sequence outside CDR-H3 fully matches the ΔDH antibody; no stop codon except possibly at the C-terminus
Partial 10NNB insertion	7133	15.82%	
Complete 10NNB insertion (sequence-intact)	5726	12.70%	Insertions outside the 10NNB length range (2341–2369 bp, 2371–3470 bp) Lacking a definite start or stop codon
Insertion of other length	19,212	42.62%	
Other	3934	8.73%	
Total	45,075	100%	

Reads from long-amplicon PCR products spanning the entire antibody-coding sequence, classified by mutation type relative to the pre-editing sequence (see Notes column for category definitions), expressed as read counts and percentage of total reads. Unlike short-amplicon analysis, which classified reads based on a match within the 5-bp regions flanking the insertion site, long-amplicon reads were classified based on whether the entire amino acid sequence outside the inserted region was identical to the pre-editing sequence.

Table 4. Representative CDR-H3 sequences with discordant short- and long-amplicon classification.

CDR-H3 Sequence	Genome Editing (Pre-Sort)	hIgG-Fc-Positive	hIgG-Fc-Middle	hIgG-Fc-Negative	Long-Amplicon (Pre-Sort and in-Frame Sequence)
LSLCRCLNLH	–	–	202	–	10
GYCTLVHAWF	30	–	100	119	9
AGAAAWILIK	–	–	94	555	1
MCLAWNGPSA	20	–	–	81	12
LPNDYIVSDV	25	3	–	12	1
FYSALSDSDV	–	–	–	6	3
NRMRICPLFI	–	–	–	5	8

Read counts for representative 10NNB CDR-H3 sequences in the pre-sort genome-editing population and in the hIgG-Fc-positive, -middle, and -negative populations sorted by short-amplicon analysis, alongside the corresponding read counts among in-frame, stop-codon-free sequences identified in the pre-sort population by long-amplicon analysis. Sequences shown are representative sequences preferentially recovered in the hIgG-Fc-middle or -negative populations (short-amplicon analysis); the complete sequence set for both short- and long-amplicon analyses is provided in Data S5.

These observations indicate that an intact reading frame and absence of a stop codon are not by themselves sufficient to predict detectable surface IgG expression for all CDR-H3 variants.

3.6. Large-Scale 6NNB Screening Generated an XEC RBD-Reactive Population

Using the optimized editing conditions, we next generated a large-scale 6NNB library. A total of 9.6×10^7 cells were subjected to genome editing. Following recovery and expansion, 1.42×10^8 cells were subjected to MACS-based enrichment for surface hIgG-Fc expression, yielding approximately 1.42×10^7 cells (~10% of the input; Figure 4A). At the sequencing depth analyzed, 3158 unique CDR-H3 amino acid sequences were directly observed in the pre-selection sample (Data S6). Additional unique sequences were observed in independently sampled later populations, but these observed counts were not used

to extrapolate total functional library complexity. Biotin- and His-tagged SARS-CoV-2 Omicron XEC RBD was used as the target antigen. In the first selection round, cells were enriched by MACS using the biotinylated antigen, and antigen binding was subsequently assessed by flow cytometry using a fluorescent anti-His antibody. The binding-positive proportion was 3.13% in the presence of XEC RBD and 1.58% in the antigen-free control (Figure 4B).

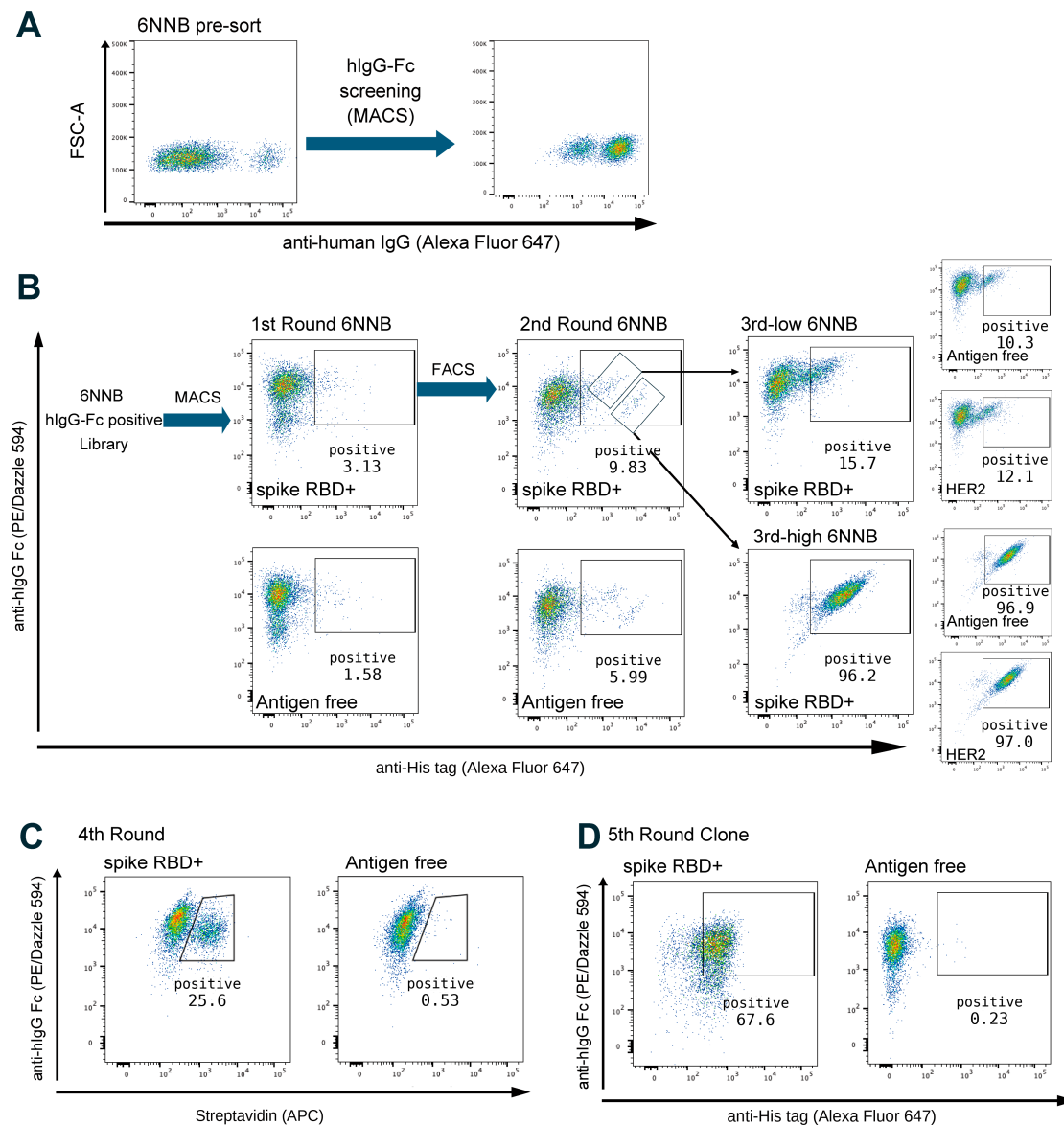


Figure 4. Screening for antigen-binding antibodies from a large-scale (10^8 -cell) 6NNB library. (A) MACS-based enrichment of hIgG-Fc-positive cells from the pre-sort 6NNB library, based on surface staining with an anti-human IgG antibody. (B) Sequential enrichment of the 6NNB hIgG-Fc-positive library against the SARS-CoV-2 Omicron (XEC) RBD protein (spike RBD+), with antigen-free conditions run in parallel at each round, across the 1st, 2nd, and 3rd rounds of MACS/FACS enrichment. Following the 2nd round, two populations were resolved based on anti-His tag signal and were separately analyzed in the 3rd round, designated 3rd-high and 3rd-low. (C) Flow-cytometric analysis of the 3rd-low population after further enrichment (4th round), using APC-streptavidin to detect the biotinylated antigen, for spike RBD+ and the antigen-free control. (D) Flow cytometry results for the spike RBD+ population from the 5th round. Spike RBD+ and antigen-free controls.

Following a second enrichment step, two populations differing in anti-His signal intensity were isolated separately and designated 3rd-high and 3rd-low. The 3rd-high

population showed a strong signal but displayed comparable staining in the antigen-free and HER2 control conditions, indicating substantial antigen-independent or reagent-dependent binding. The 3rd-low population also showed background staining, but the binding-positive proportion was larger in the presence of XEC RBD than with HER2 (15.7% versus 12.1%).

The 3rd-low population was therefore subjected to an additional selection round using an alternative detection strategy. When biotinylated XEC RBD was detected using APC-streptavidin, 25.6% of cells fell within the binding-positive gate, compared with 0.53% in the antigen-free control (Figure 4C). Subsequent enrichment generated a population in which 67.6% of cells were positive in the presence of XEC RBD and 0.23% in the antigen-free control (Figure 4D).

A parallel screening experiment was performed using the 10NNB library. Although an anti-His-positive population emerged during early selection, evaluation with APC-streptavidin showed only a small difference between XEC RBD and antigen-free conditions (3.06% versus 0.64%) in the higher-signal fraction, whereas the lower-signal fraction showed no antigen-dependent increase (Figure S4).

3.7. NGS Tracking Showed Progressive Convergence of the 6NNB Library on AFFVVF

To track sequence composition during XEC RBD selection, we performed short-amplicon NGS of the CDR-H3 region in the genome-edited population, the surface hIgG-Fc-enriched population, and each subsequent selection population. In the pre-sort 6NNB population, 19,895 of 30,451 total reads contained a correctly sized 6NNB sequence, corresponding to 3158 unique sequences. The most abundant sequence accounted for only approximately 0.25% of total reads (Data S6).

Across the 6NNB selection series, the number of observed unique sequences increased to 9131 after the first selection round and then declined during subsequent enrichment, reaching fewer than 100 observed sequences in the final populations (Figure 5A and Data S6). These results show progressive convergence of the library during selection.

Sequence-class analysis further showed that indel-containing reads decreased after selection for surface hIgG-Fc expression and remained low during subsequent antigen-selection rounds (Figure 5B and Table S2).

Clone-level analysis revealed marked differences between the two populations isolated after the second selection round. The 3rd-high population was dominated by PFTTVS (50.6%) and LFFVVS (44.8%). LFFVVS differed by only two amino acid positions from LFVSVS, the Alexa Fluor 647-reactive clone identified in the pilot screen, consistent with the reagent-dependent binding phenotype observed for the 3rd-high population.

In contrast, selection of the 3rd-low population generated a different sequence trajectory (Figure 5C). PFWMVA and CRYYYVS were among the most abundant sequences after the 4th round but were not retained in the 5th-round population following the change in detection reagent. AFFVVF, which was present in the earlier selected populations, became dominant in the 5th round, accounting for 94.7% of reads. Because AFFVVF remained enriched after switching from Alexa Fluor 647-based detection to APC-streptavidin-based detection, it was selected for subsequent characterization as a candidate XEC RBD-binding clone.

We next examined the physicochemical composition of sequences recovered during selection. The read-weighted Kyte–Doolittle hydrophobicity [23] distribution was similar among the pre-sort, hIgG-Fc-positive, and 1st-round populations. The distribution shifted toward higher hydrophobicity in later selection rounds, with the read-weighted mean reaching 2.99 in the 5th-round population as the highly hydrophobic AFFVVF sequence (Kyte–Doolittle score, 3.10) became dominant (Figure 5D).

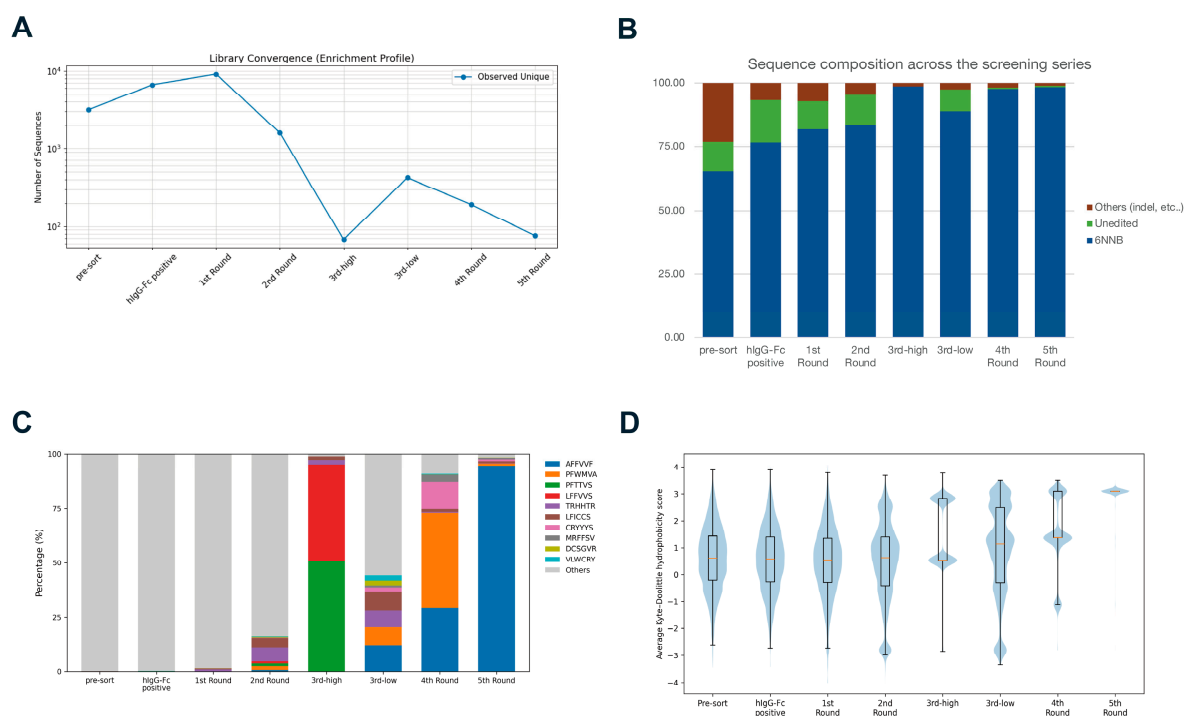


Figure 5. NGS-based tracking of library diversity and clonal enrichment throughout the spike RBD screening series. **(A)** Number of observed unique CDR-H3 sequences at each stage of the screening series (log scale). These values represent sample-dependent observed richness and are not estimates of total functional library complexity. **(B)** Sequence composition of the library at each stage of the screening series, classified as in-frame 6NNB sequences, unedited (wild-type) sequences, or other sequences (including indels and out-of-frame insertions), expressed as a percentage of total reads. **(C)** Relative abundance of the ten most frequently observed CDR-H3 clones across successive stages of the screening series (pre-sort, hlgG-Fc-positive, 1st–5th rounds), expressed as a percentage of total reads at each stage. All other sequences are grouped as “Others.” Raw read counts are provided as Supplementary Data. **(D)** Distribution of the average Kyte–Doolittle hydrophobicity score of CDR-H3 sequences at each stage of the screening series, weighted by read count. Violin plots show the full distribution; boxes indicate the interquartile range, with the median marked in orange. The analysis script used to generate panels (A–D) is provided as Code S1.

3.8. AFFVVF Recognized Multiple SARS-CoV-2 RBD Variants and Showed Assay-Dependent Binding to Spike Trimer

The AFFVVF clone recovered after the 5th selection round was next evaluated for antigen binding. By flow cytometry, AFFVVF bound RBD proteins from all seven SARS-CoV-2 variants tested: Alpha, Delta, Omicron, BA.5, EG.5.1, JN.1, and XEC (Figure 6A,B). The binding-positive proportion ranged from 61.1% for Delta RBD to 86.6% for JN.1 RBD, with 71.5% of cells scoring positive for the XEC RBD used during selection.

In contrast, no comparable binding was detected by flow cytometry using Alpha or XEC spike trimer proteins. AFFVVF also showed no detectable binding above the background level to the unrelated antigens EGFR, GPA33, CD19, HER2, or influenza HA under the same assay conditions (Figure 6D).

Among the RBD variants tested, JN.1 showed the strongest binding signal by flow cytometry (86.6% positive; Figure 6A,B). The binding properties of recombinant AFFVVF were subsequently examined by indirect ELISA against the Delta, BA.5, and JN.1 RBDs and the JN.1 spike trimer. In ELISA, binding to Delta, BA.5, and JN.1 RBDs was weak and an EC_{50} could not be determined within the tested concentration range ($\leq 50,000$ ng/mL). In contrast, dose-dependent binding to JN.1 spike trimer was observed, with an estimated EC_{50} of 389 ng/mL (Figure 6C).

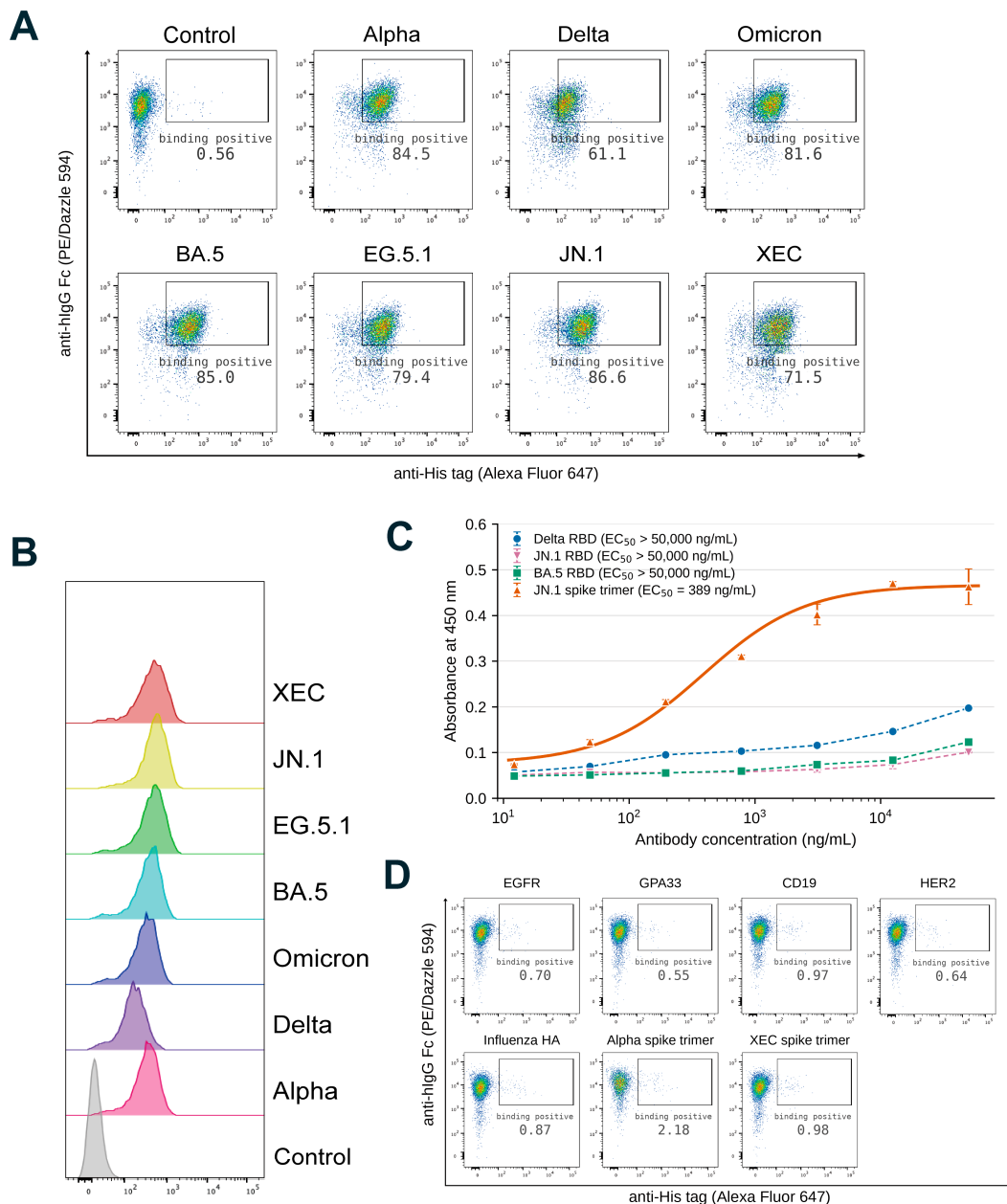


Figure 6. Binding profile of the isolated antibody clone (AFFVVF) against SARS-CoV-2 spike RBD variants. **(A)** Flow-cytometric binding of AFFVVF to the receptor-binding domain (RBD) of seven SARS-CoV-2 variants and a control (Control, Alpha, Delta, Omicron, BA.5, EG.5.1, JN.1, XEC), detected via anti-His tag (Alexa Fluor 647) staining. The percentage of binding-positive cells is indicated for each variant. **(B)** Overlaid histograms of anti-His tag (Alexa Fluor 647) signal for AFFVVF binding to each RBD variant, arranged in chronological order of variant emergence (Control, Alpha, Delta, Omicron, BA.5, EG.5.1, JN.1, XEC). **(C)** Indirect ELISA of recombinant AFFVVF antibody protein against Delta RBD, JN.1 RBD, BA.5 RBD, and JN.1 spike trimer. EC₅₀ could not be defined for Delta, JN.1, and BA.5 RBD within the tested concentration range ($\leq 50,000$ ng/mL); EC₅₀ against JN.1 spike trimer was 389 ng/mL. Data are mean $\pm$ SD of technical triplicates, except where outlier replicates were excluded: Delta RBD at 48.8 ng/mL (one replicate excluded, $n = 2$), BA.5 RBD at 12,500 ng/mL (two replicates excluded, $n = 1$), and JN.1 spike trimer at 3125 and 12,500 ng/mL (one replicate each excluded, $n = 2$). Solid curves show four-parameter logistic fits; dashed lines connect the means of conditions for which a satisfactory fit could not be obtained. All replicate values, including those excluded, are provided in Data S7. **(D)** Flow-cytometric analysis of AFFVVF binding to five unrelated antigen proteins (EGFR, GPA33, CD19, HER2, and influenza HA) and to spike trimer proteins (Alpha, XEC).

To exclude dependence on a single detection reagent, binding to XEC RBD was re-examined using biotinylated antigen detected with APC-streptavidin. Overall, 47.8% of AFFVVF-expressing cells were positive for binding to XEC RBD, compared with 0.61% in the absence of antigen and 0.39% with biotinylated CD19, whereas 0.95% of cells expressing the parental Δ DH antibody were positive for binding to XEC RBD (Figure S5). Binding to XEC RBD was thus also observed with a detection system that does not rely on Alexa Fluor 647. Because the two detection systems differ in fluorophore and detection mechanism, the percentages obtained with them are not directly comparable, and the two data sets are interpreted qualitatively.

To obtain biochemical confirmation of binding to the antigen used for selection, recombinant AFFVVF was examined by indirect ELISA against the XEC spike trimer and the XEC RBD, in parallel with the JN.1 spike trimer, with EGFR as an unrelated antigen and sotrovimab as a reference antibody (Figure S6). Sotrovimab bound the JN.1 and XEC spike trimers with comparable maximal signals (EC₅₀ 1661 and 1641 ng/mL, respectively), confirming that both antigens were immobilized in a functional form (Figure S6A). AFFVVF bound the XEC spike trimer in a concentration-dependent manner on both plates processed in parallel (EC₅₀ values of 1223 and 816 ng/mL; Figure S6A,B). Binding to the JN.1 spike trimer was also observed on both plates (EC₅₀ values of 790 and 927 ng/mL), approximately 2.0- to 2.4-fold higher than the estimate obtained in the earlier experiment (389 ng/mL; Figure 6C). The maximal absorbance obtained with AFFVVF was approximately one-third of that obtained with sotrovimab on the same antigens; because the two antibodies reached different maxima, their EC₅₀ values are not directly comparable. In contrast, no concentration-dependent binding was detected against the XEC RBD or against EGFR, and no EC₅₀ could be defined for either within the tested range (Figure S6B). At the highest antibody concentration, the absorbance obtained with the XEC RBD (0.084) was lower than that obtained with the unrelated antigen EGFR (0.120), compared with 0.481 for the XEC spike trimer.

In summary, AFFVVF recognized isolated SARS-CoV-2 RBD proteins by cell-based binding analysis, and this recognition was confirmed for the XEC RBD using two independent detection systems. As a recombinant protein, AFFVVF bound both the JN.1 and the XEC spike trimer by ELISA, the latter corresponding to the same SARS-CoV-2 variant as the XEC RBD used for selection, but gave no detectable signal against monomeric RBD proteins in this format despite their ready detection on cells. Binding of the recombinant antibody was therefore antigen-dependent, but the antigen formats recognized differed between the cell-based and plate-based assays.

3.9. Screening Against EGFR and GPA33 Produced Antigen-Dependent Enriched Populations

To explore whether the screening workflow yields antigen-dependent enrichment with antigens other than SARS-CoV-2 RBD, we performed exploratory selections against EGFR and GPA33 using 6NNB, 10NNB, and CX4C library designs. The CX4C design places a randomized four-residue segment between two fixed cysteines to form a disulfide-constrained loop.

After successive selection rounds, flow-cytometric analysis showed antigen-concentration-dependent shifts in selected populations for both targets (Figure 7). For GPA33, the largest concentration-dependent shift was observed in the 10NNB population, with smaller shifts in the 6NNB and CX4C populations. For EGFR, antigen-dependent shifts were observed in the 6NNB and CX4C populations, whereas the 10NNB selection against EGFR did not yield an enriched population and was therefore not analyzed further.

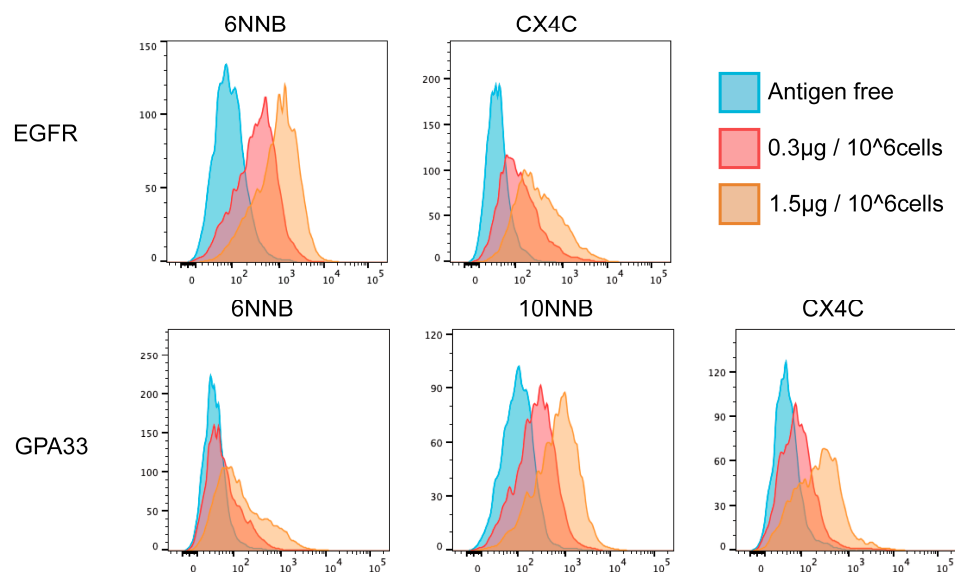


Figure 7. Screening against EGFR and GPA33. Flow-cytometric analysis of the final enriched population for each library against EGFR (6NNB, CX4C) and GPA33 (6NNB, 10NNB, CX4C), showing overlaid histograms for the antigen-free condition (Antigen free) and for two antigen concentrations, 0.3 $\mu\text{g}/10^6$ cells and 1.5 $\mu\text{g}/10^6$ cells. The 10NNB library was not included for EGFR because for EGFR because no enriched population was obtained from this selection.

To reduce persistent enrichment of clones reacting with a single detection reagent, different antigen-detection schemes were used during successive selection rounds, including biotin/APC-streptavidin and anti-His antibodies conjugated to Alexa Fluor 647 or PE/Dazzle 594.

Sequences recovered after selection were analyzed by bulk Sanger sequencing or by blunt-end cloning followed by Sanger sequencing of individual clones, depending on the library (Table S3). Following EGFR selection, CDFFS was predominant in the 6NNB population. Two CX4C sequences, QCFMNFCL and VCATDGCE, were recovered among individually cloned sequences.

GPA33 selection produced different sequence patterns among the three library designs. In the 6NNB population, 7 of 16 analyzed clones carried the sequence CALLPKYGYFDY, which arose from a deletion and insertion in the ΔDH bridging sequence rather than a clean six-residue insert (Table S3), whereas additional sequences included SVSCAT, AYFMMS, and VCWWVC. In the 10NNB population, bulk sequencing identified a predominant sequence containing a single-codon deletion in the ΔDH bridging sequence, CARSYFDY. In the CX4C population, FCTPARCL was the predominant recovered sequence, with LCLVATCS detected as a minor variant.

Thus, selection against EGFR and GPA33 produced antigen-dependent population shifts accompanied by enrichment of discrete sequence variants. Because the binding properties of the individual recovered clones have not yet been independently confirmed to reflect genuine recombinant antibody activity, these sequences were considered candidate binders rather than validated antigen-specific antibodies.

4. Discussion

Mammalian display enables antibodies to be screened directly as full-length IgG molecules in a mammalian cellular environment, but construction of genetically well-defined libraries remains challenging. Variation in integration site, copy number, and expression can weaken genotype–phenotype linkage and complicate binder selection [3,6,7]. In this study, we established a defined mammalian display system by first integrating an

antibody-expression cassette at a single HAC acceptor site and subsequently introducing diversity specifically into CDR-H3 by CRISPR-Cas9-mediated HDR. This strategy separates establishment of the display cassette from library diversification and provides a common genetic background for selection of antibody variants.

Although targeted integration at the HAC acceptor site was supported by junction PCR and by the relatively uniform surface IgG expression of the selected clonal population, these assays do not themselves establish the copy number of the integrated cassette or formally exclude unintended additional genomic integrations. Direct assessment by methods such as quantitative PCR or fluorescence in situ hybridization would strengthen this characterization and represents an important direction for future work. In this study, the term “genetically defined” therefore refers to site-specific integration at a defined acceptor site and to the resulting common genetic background shared by all library members, rather than to a verified single-copy configuration.

The ΔDH scaffold was designed as a defined starting framework without pre-existing target specificity using the human germline elements IGHV3-23/IGHJ4 and IGKV3-20/IGKJ1. Its minimized CDR-H3, CARSGYFDY, provided a defined site for insertion of predetermined or randomized sequences. Introduction of the P3E6 CDR-H3 restored antigen binding, demonstrating that an appropriate CDR-H3 can generate binding activity within this fixed scaffold, whereas three other CDR-H3 sequences from reported SARS-CoV-2 antibodies failed to restore detectable binding (Figure 1E). These results indicate that CDR-H3-only diversification offers a simple and genetically tractable strategy [15–17], but the accessible binding landscape is constrained by the need for compatibility with the fixed framework, the remaining CDRs, and the light chain.

We used NNB codons to generate randomized CDR-H3 sequences because this scheme encodes all 20 amino acids while reducing the frequency of stop codons relative to NNK or NNS [27]. Although a six-position NNB design theoretically encodes 6.4×10^7 amino acid variants, our results highlight the distinction between theoretical sequence diversity and the repertoire that is actually generated and displayed in mammalian cells. Short-amplicon sequencing suggested efficient insertion at the targeted CDR-H3 site, yet only approximately 5–10% of edited cells retained strong surface IgG expression (Figure 3B). Long-amplicon sequencing revealed that partial insertions, insertions of other lengths, and deletions were common (Table 3), indicating that local analysis around the HDR site can overestimate the proportion of structurally intact antibody alleles.

The CDR-H3 sequence itself also affected antibody display. Several in-frame, non-stop sequences that met the long-amplicon sequence criteria were preferentially recovered from the hIgG-Fc-middle or -negative population (Table 4). Thus, the experimentally available library represents not simply the diversity encoded by the donor oligonucleotide but the subset of edited sequences compatible with antibody production and surface display. This provides an intrinsic expression-related filter while at the same time reducing the effective repertoire available for antigen selection.

The present sequencing data do not permit a reliable estimate of total functional library complexity. In the large-scale 6NNB experiment, 3158 unique CDR-H3 amino acid sequences were directly observed in the pre-selection sample, and additional unique sequences were observed in separately sampled hIgG-Fc-positive and first-round populations. Because sequencing depth, clone abundance, PCR/sequencing error, and the composition of each sampled population can all influence the number of observed unique sequences, cumulative observed counts should not be interpreted as either a lower-bound estimate or an extrapolated total functional library complexity. As a separate reference point, the number of electroporated cells expected to survive and express a functional antibody provides a theoretical ceiling on the number of distinct variants that could in principle be present.

Assuming post-electroporation viability of 86.2% (Figure S2B) and an hIgG-Fc-positive proportion of approximately 10% among the 9.6×10^7 electroporated cells, this ceiling is approximately 8.3×10^6 . This calculation assumes one variant per surviving cell and does not account for redundancy; because the NGS data already show substantial duplication even at limited sampling depth, this theoretical ceiling likely substantially overestimates the number of distinct functional variants actually present. The observed sequence count and this theoretical ceiling should be understood as independent reference points reflecting different sources of information—sequencing depth and cell input, respectively—rather than as defensible lower and upper bounds on the true functional library complexity. We therefore report observed sequence richness and the number of cells entering selection separately. Quantitative determination of functional library complexity will require a dedicated sampling and error-correction strategy in future studies.

The pilot screening identified another important limitation: Strong enrichment does not necessarily indicate target-specific binding. LFVSVS rapidly became dominant but was subsequently shown to recognize an Alexa Fluor 647-associated determinant rather than the intended antigen (Figure 2). Similar reagent-dependent populations were encountered during the larger screening campaign. We therefore alternated detection formats and incorporated antigen-free and unrelated-antigen controls. Several sequences enriched under the initial detection conditions disappeared after the labeling strategy was changed, whereas AFFVVF became dominant (Figure 5C). These findings emphasize that orthogonal detection systems are essential for distinguishing antigen-directed selection from enrichment of reagent-reactive clones.

AFFVVF provides proof of principle that a newly selected antigen-binding antibody variant can be recovered directly from the randomized Δ DH mammalian display system. The antibody recognized RBDs from all seven SARS-CoV-2 variants tested by flow cytometry (Figure 6A,B) and showed no detectable binding to several unrelated antigens under the same conditions (Figure 6D). Binding to the XEC RBD used for selection was confirmed using two detection systems that differ in both fluorophore and detection chemistry (Figures 6A and S5), indicating that the signal is not attributable to reactivity with a single detection reagent. The affinity of AFFVVF was not determined, and quantitative kinetic characterization remains to be performed.

The isolation of AFFVVF reported here derives from a single library-construction and selection trajectory, and the present study does not address whether independent library construction and selection experiments would reproducibly yield antigen-binding clones, or whether such clones would share sequence or binding characteristics with AFFVVF. Addressing these questions will require performing the entire workflow, from genome editing through selection, as an independent biological replicate.

Binding to the XEC antigen used for selection was also examined biochemically. The soluble recombinant antibody bound the XEC spike trimer by indirect ELISA on two plates processed in parallel, under conditions in which a reference antibody confirmed that both the XEC and the JN.1 spike trimer had been immobilized in a functional form (Figure S6). Recognition of the monomeric XEC RBD, by contrast, was observed only with the cell-displayed antibody and was not reproduced with the soluble antibody in this plate-based format. Orthogonal confirmation of monomeric RBD binding by a label-free method such as surface plasmon resonance or biolayer interferometry, together with quantitative determination of affinity, therefore remains an important validation step for future work.

The enrichment of AFFVVF was accompanied by a shift toward higher CDR-H3 hydrophobicity during later selection rounds (Figure 5D). Because highly hydrophobic sequences may be favored by both specific and nonspecific interactions [28], this shift should not itself be interpreted as evidence of antigen specificity. The absence of detectable

AFFVVF binding to the unrelated antigens tested is encouraging, but additional assessment of polyspecificity and developability will be important, particularly given the strongly hydrophobic CDR-H3 sequence.

The binding profile of AFFVVF differed between assay formats, and a uniform binding phenotype should not be inferred. On cells, the displayed antibody bound monomeric RBD but gave no comparable signal with the Alpha or XEC spike trimer; as a soluble recombinant protein in indirect ELISA, it showed the converse pattern, binding the JN.1 and XEC spike trimers but giving no detectable signal with monomeric RBD, including the XEC RBD used for selection (Figures 6C and S6). These formats differ in how antigen and antibody are presented. In flow cytometry, the antibody is displayed as a bivalent IgG on the cell surface and encounters soluble antigen, so avidity from multivalent engagement may contribute to the signal. In ELISA, the antibody is a soluble recombinant protein and the antigen is adsorbed to the plate, which may alter epitope accessibility or orientation; adsorption may be particularly unfavorable in this respect for a small monomeric domain, whereas a trimer presents three RBD copies at higher local density that could favor bivalent engagement by IgG. We did not test these explanations directly, and they remain possibilities rather than conclusions. Within the ELISA format, the response was nevertheless clearly antigen-dependent: the signal obtained with the XEC spike trimer was approximately fourfold that obtained with EGFR, an unrelated antigen tested in the same experiment, whereas the signal obtained with the XEC RBD did not exceed that obtained with EGFR (Figure S6B). Thus, the present data support recognition of isolated RBD by AFFVVF displayed on the cell surface and recognition of trimeric spike by the soluble recombinant antibody, with the form of antigen recognized differing between the two formats. Quantitative characterization of affinity, together with orthogonal confirmation of monomeric RBD binding by a label-free method, would strengthen the interpretation of AFFVVF as a *de novo* RBD-binding proof-of-principle clone. Selection against EGFR and GPA33 further produced antigen-concentration-dependent population shifts and enrichment of discrete sequence variants across the 6NNB, 10NNB, and CX4C library designs (Figure 7, Table S3). Because the individual recovered clones have not yet been validated as recombinant antibodies, these experiments show antigen-dependent enrichment but do not demonstrate the isolation of antigen-specific antibodies for either target; the workflow may be applicable to additional antigens, but this remains to be established by independent recombinant expression and binding validation of representative clones. Notably, some of the recovered GPA33-selected sequences (CALLLPKYGYFDY in the 6NNB population and CARSYFDY in the 10NNB population) did not correspond to an actually designed insertion but instead arose from a deletion or insertion within the Δ DH bridging sequence itself. Because such scaffold-derived variants are not constrained to the intended randomized-region length or reading frame, their enrichment may reflect selection for altered scaffold conformations or expression-related artifacts rather than genuine antigen-specific binding, underscoring the importance of independent recombinant validation for any candidate recovered from exploratory screens.

Because candidate antibodies are screened and characterized in the same full-length IgG format throughout selection, without requiring conversion from an initially screened fragment format (e.g., scFv or Fab), the platform obviates the reformatting-related uncertainty inherent to fragment-based display systems [3,14]. The experiments also identify important practical constraints, including complex genome-editing outcomes, sequence-dependent antibody expression, limited experimentally sampled diversity, and enrichment of detection-reagent binders. Despite these limitations, isolation of AFFVVF demonstrates that the platform can generate an antigen-binding antibody without requiring a pre-existing binder or an immune-derived repertoire [11–13]. Further improvements in editing fidelity,

library architecture, and selection design should extend the utility of mammalian display from optimization of known antibodies toward de novo discovery of full-length IgG binders.

5. Conclusions

This study establishes a genetically defined mammalian display platform in which CDR-H3 diversity is introduced directly into a pre-established, site-specifically integrated full-length IgG scaffold and screened through iterative selection. Using this system, we isolated AFFVVF, a previously undefined antigen-binding antibody recognizing multiple SARS-CoV-2 RBD variants without requiring a pre-existing antigen-specific binder. We also observed antigen-dependent enrichment in exploratory selections against two additional targets, EGFR and GPA33, for which the recovered sequences remain candidates pending recombinant validation. These results provide a proof of principle for de novo binder generation through CDR-H3 diversification within a predefined mammalian IgG scaffold, rather than a general demonstration of de novo antibody discovery. Complex genome-editing outcomes, sequence-dependent antibody expression, and limited experimentally sampled diversity remain important constraints to be addressed in future work.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/antib15060099/s1>. Figure S1: Establishment of a site-specifically integrated antibody-display cell line (HAC/293T-ΔDH); Figure S2: Effect of HDR Enhancer V2 on genome editing efficiency and cell viability; Figure S3: Antibody surface expression across genome editing conditions (EP1–EP6); Figure S4: Screening of the 10NNB library against SARS-CoV-2 XEC spike RBD; Figure S5: Binding of AFFVVF to XEC RBD detected with biotinylated antigen and APC-streptavidin; Figure S6: Binding of recombinant AFFVVF to SARS-CoV-2 spike trimers, monomeric RBD, and an unrelated antigen by indirect ELISA; Table S1: Reagent ratios and pulse conditions tested during optimization of homology-directed repair efficiency (EP1–EP6); Table S2: Sequence composition of the large-scale 6NNB library across the screening series; Table S3: CDR-H3 sequences identified in post-selection libraries screened against EGFR and GPA33; Data S1: Raw NGS read counts for individual CDR-H3 sequences across the pilot 6NNB screening series (pre-sort, hIgG-Fc-positive, 1st and 2nd rounds; Figure 2); Data S2: Raw ELISA absorbance data and 4PL curve-fitting parameters for the LFVSVS clone and unmodified ΔDH antibody (Figure 2D); Data S3: Raw short-amplicon NGS read counts for individual CDR-H3 sequences across the 10NNB screening series (pre-sort, hIgG-Fc-positive, -middle, and -negative populations; Figure 3); Data S4: Comparison of short-amplicon and long-amplicon CDR-H3 read counts for the 10NNB library (Figure 3); Data S5: Raw long-amplicon sequencing reads, including nucleotide and amino acid sequences, mutation classification, and read counts, for the 10NNB library (Figure 3); Data S6: Raw NGS read counts for individual CDR-H3 sequences and total/classified read counts at each stage of the large-scale 6NNB screening series (Figures 4 and 5); Data S7: Raw ELISA absorbance data and 4PL curve-fitting parameters for AFFVVF binding to variant spike RBD proteins (Figure 6C); Data S8: Raw ELISA absorbance data and 4PL curve-fitting parameters for AFFVVF binding to SARS-CoV-2 spike trimers, monomeric XEC RBD, and EGFR (Figure S6); Code S1: Python script for NGS-based diversity, clone enrichment, and hydrophobicity tracking (Figure 5A–D), using Data S6 as input; Source Data: Reagents, antigens, primers, and instrument details used in this study.

Author Contributions: Conceptualization, K.T., N.U., A.H. and Y.K.; methodology, R.S. and N.U.; validation, R.S.; formal analysis, R.S.; investigation, R.S. and N.H.; resources, A.H. and Y.K.; data curation, R.S.; writing—original draft preparation, R.S.; writing—review and editing, R.S., N.U., A.H., Y.K. and K.T.; visualization, R.S.; supervision, N.U., A.H., Y.K. and K.T.; project administration, N.U., A.H., Y.K. and K.T.; funding acquisition, R.S., A.H., Y.K. and K.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by JSPS KAKENHI, grant number 23K27306 (to K.T., A.H., and Y.K.); by the Japan Agency for Medical Research and Development (AMED) under Grant Number JP26ama121046 (to Y.K.); by the Collaborative Research Program of Exploratory Research Center on Life and Living Systems (ExCELLS) (Grant No. 21-101) (to Y.K.); and by JST SPRING, grant number JPMJSP2134 (to R.S.).

Institutional Review Board Statement: Not applicable. This study involved established cultured cell lines, including HEK293T and CHO-derived cells, and did not involve human participants or experimental animals.

Informed Consent Statement: Not applicable.

Data Availability Statement: The NGS raw sequencing data generated in this study are available in the DDBJ Sequence Read Archive (DRA) under BioProject accession number PRJDB45658 (Run accessions DRR1092909–DRR1092925). Raw ELISA data are provided as Supplementary Data (Data S2, S7 and S8). The flow cytometry data (FCS files) generated and analyzed during this study are available from the corresponding author upon reasonable request. Plasmids generated in this study are subject to a material transfer agreement (MTA) with collaborating institutions; requests for these materials will be considered by the corresponding author on a case-by-case basis. The human artificial chromosome-carrying cell lines used in this study are available from the corresponding author upon reasonable request, subject to the agreement of the institutions holding the relevant materials.

Acknowledgments: The authors thank Daisuke Kuroda (Nihon University) for valuable technical discussions, and Hiromu Maruko, Hidetsugu Honda, Kana Uno, and Yuki Ishizu (all Tokyo University of Pharmacy and Life Sciences) for their technical assistance. The authors also thank all members of the Laboratory of Bioengineering for their support. During the preparation of this manuscript, the authors used Claude (Anthropic), accessed between June and October 2026, for English-language refinement of a Japanese draft and for aspects of NGS and ELISA data analysis, as described in Section 2.15. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

CDR	Complementarity-Determining Region
IgG	Immunoglobulin G
hIgG-Fc	Human Immunoglobulin G Fc domain
HDR	Homology-Directed Repair
ssODN	Single-Stranded Oligodeoxynucleotide
PAM	Protospacer Adjacent Motif
PCR	Polymerase Chain Reaction
MACS	Magnetic-Activated Cell Sorting
FACS	Fluorescence-Activated Cell Sorting
HAC	Human Artificial Chromosome
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
RBD	Receptor-Binding Domain
DNA	Deoxyribonucleic Acid
ELISA	Enzyme-Linked Immunosorbent Assay
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
NGS	Next-Generation Sequencing
PBS	Phosphate-Buffered Saline
HPRT	Hypoxanthine Phosphoribosyltransferase
GPA33	Glycoprotein A33

EGFR	Epidermal Growth Factor Receptor
HAT	Hypoxanthine-Aminopterin-Thymidine
FBS	Fetal Bovine Serum
EDTA	Ethylenediaminetetraacetic Acid
TBS	Tris-Buffered Saline
TMB	3,3',5,5'-Tetramethylbenzidine
RNP	Ribonucleoprotein
FSC	Forward Scatter
SSC	Side Scatter
APC	Allophycocyanin
DRA	DDBJ Sequence Read Archive
DDBJ	DNA Data Bank of Japan
PDGFR	Platelet-Derived Growth Factor Receptor
RRID	Research Resource Identifier

References

1. Lu, R.-M.; Chiang, H.-L.; Yuan, J.P.-Y.; Wang, H.-H.; Chen, C.-Y.; Panda, S.S.; Liang, K.-H.; Peng, H.-P.; Ko, S.-H.; Hsu, H.-J.; et al. Technological advancements in antibody-based therapeutics for treatment of diseases. *J. Biomed. Sci.* **2025**, *32*, 98. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Bauer, J.; Rajagopal, N.; Gupta, P.; Gupta, P.; Nixon, A.E.; Kumar, S. How can we discover developable antibody-based biotherapeutics? *Front. Mol. Biosci.* **2023**, *10*, 1221626. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Valldorf, B.; Hinz, S.C.; Russo, G.; Pekar, L.; Mohr, L.; Klemm, J.; Doerner, A.; Krah, S.; Hust, M.; Zielonka, S. Antibody display technologies: Selecting the cream of the crop. *Biol. Chem.* **2022**, *403*, 455–477. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Hoogenboom, H.R. Selecting and screening recombinant antibody libraries. *Nat. Biotechnol.* **2005**, *23*, 1105–1116. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Zhou, H.; Wang, Y.; Wang, W.; Jia, J.; Li, Y.; Wang, Q.; Wu, Y.; Tang, J. Generation of monoclonal antibodies against highly conserved antigens. *PLoS ONE* **2009**, *4*, e6087. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Parthiban, K.; Perera, R.L.; Sattar, M.; Huang, Y.; Mayle, S.; Masters, E.; Griffiths, D.; Surade, S.; Leah, R.; Dyson, M.R.; et al. A comprehensive search of functional sequence space using large mammalian display libraries created by gene editing. *mAbs* **2019**, *11*, 884–898. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Mason, D.M.; Weber, C.R.; Parola, C.; Meng, S.M.; Greiff, V.; Kelton, W.J.; Reddy, S.T. High-throughput antibody engineering in mammalian cells by CRISPR/Cas9-mediated homology-directed mutagenesis. *Nucleic Acids Res.* **2018**, *46*, 7436–7449. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Huhtinen, O.; Salbo, R.; Lamminmäki, U.; Prince, S. Selection of biophysically favorable antibody variants using a modified Flp-In CHO mammalian display platform. *Front. Bioeng. Biotechnol.* **2023**, *11*, 1170081. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Parola, C.; Neumeier, D.; Friedensohn, S.; Csepregi, L.; Di Tacchio, M.; Mason, D.M.; Reddy, S.T. Antibody discovery and engineering by enhanced CRISPR-Cas9 integration of variable gene cassette libraries in mammalian cells. *mAbs* **2019**, *11*, 1367–1380. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Bowers, P.M.; Horlick, R.A.; Neben, T.Y.; Toobian, R.M.; Tomlinson, G.L.; Dalton, J.L.; Jones, H.A.; Chen, A.; Altobelli, L.; Zhang, X.; et al. Coupling mammalian cell surface display with somatic hypermutation for the discovery and maturation of human antibodies. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 20455–20460. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Hellmann, I.; Waldmeier, L.; Bannwarth-Escher, M.-C.; Maslova, K.; Wolter, F.I.; Grawunder, U.; Beerli, R.R. Novel antibody drug conjugates targeting tumor-associated receptor tyrosine kinase ROR2 by functional screening of fully human antibody libraries using Transpo-mAb display on progenitor B cells. *Front. Immunol.* **2018**, *9*, 2490. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Hamamichi, S.; Uno, N.; Shimoya, K.; Fukushima, T.; Abe, M.; Watanabe, A.; Ito, M.; Wakasa, Y.; Yajima, Y.; Suzuki, R.; et al. Rapid immunization and antibody redesign platform discovers broadly neutralizing antibodies against non-immunized SARS-CoV-2 variant. *Mol. Ther.* **2026**, *34*, 620–638. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Waldmeier, L.; Hellmann, I.; Gutknecht, C.K.; Wolter, F.I.; Cook, S.C.; Reddy, S.T.; Grawunder, U.; Beerli, R.R. Transpo-mAb display: Transposition-mediated B cell display and functional screening of full-length IgG antibody libraries. *mAbs* **2016**, *8*, 726–740. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Robertson, N.; Lopez-Anton, N.; Gurjar, S.A.; Khaliq, H.; Khalaf, Z.; Clerkin, S.; Leydon, V.R.; Parker-Manuel, R.; Raeside, A.; Payne, T.; et al. Development of a novel mammalian display system for selection of antibodies against membrane proteins. *J. Biol. Chem.* **2020**, *295*, 18436–18448. [\[CrossRef\]](#) [\[PubMed\]](#)

15. Weitzner, B.D.; Dunbrack, R.L.; Gray, J.J. The origin of CDR H3 structural diversity. *Structure* **2015**, *23*, 302–311. [[CrossRef](#)] [[PubMed](#)]
16. Kuroda, D.; Shirai, H.; Kobori, M.; Nakamura, H. Structural classification of CDR-H3 revisited: A lesson in antibody modeling. *Proteins* **2008**, *73*, 608–620. [[CrossRef](#)] [[PubMed](#)]
17. Xu, J.L.; Davis, M.M. Diversity in the CDR3 region of VH is sufficient for most antibody specificities. *Immunity* **2000**, *13*, 37–45. [[CrossRef](#)] [[PubMed](#)]
18. Tiller, T.; Schuster, I.; Deppe, D.; Siegers, K.; Strohnner, R.; Herrmann, T.; Berenguer, M.; Poujol, D.; Stehle, J.; Stark, Y.; et al. A fully synthetic human Fab antibody library based on fixed VH/VL framework pairings with favorable biophysical properties. *mAbs* **2013**, *5*, 445–470. [[CrossRef](#)] [[PubMed](#)]
19. Persson, H.; Ye, W.; Wernimont, A.; Adams, J.J.; Koide, A.; Koide, S.; Lam, R.; Sidhu, S.S. CDR-H3 diversity is not required for antigen recognition by synthetic antibodies. *J. Mol. Biol.* **2013**, *425*, 803–811. [[CrossRef](#)] [[PubMed](#)]
20. Kagita, A.; Lung, M.S.Y.; Xu, H.; Kita, Y.; Sasakawa, N.; Iguchi, T.; Ono, M.; Wang, X.H.; Gee, P.; Hotta, A. Efficient ssODN-mediated targeting by avoiding cellular inhibitory RNAs through precomplexed CRISPR-Cas9/sgRNA ribonucleoprotein. *Stem Cell Rep.* **2021**, *16*, 985–996. [[CrossRef](#)] [[PubMed](#)]
21. Kazuki, Y.; Hoshiya, H.; Takiguchi, M.; Abe, S.; Iida, Y.; Osaki, M.; Katoh, M.; Hiratsuka, M.; Shirayoshi, Y.; Hiramatsu, K.; et al. Refined human artificial chromosome vectors for gene therapy and animal transgenesis. *Gene Ther.* **2011**, *18*, 384–393. [[CrossRef](#)] [[PubMed](#)]
22. Uno, N.; Takata, S.; Komoto, S.; Miyamoto, H.; Nakayama, Y.; Osaki, M.; Mayuzumi, R.; Miyazaki, N.; Hando, C.; Abe, S.; et al. Panel of human cell lines with human/mouse artificial chromosomes. *Sci. Rep.* **2022**, *12*, 3009. [[CrossRef](#)] [[PubMed](#)]
23. Kyte, J.; Doolittle, R.F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, *157*, 105–132. [[CrossRef](#)] [[PubMed](#)]
24. Almagro, J.C.; Mellado-Sánchez, G.; Pedraza-Escalona, M.; Pérez-Tapia, S.M. Evolution of anti-SARS-CoV-2 therapeutic antibodies. *Int. J. Mol. Sci.* **2022**, *23*, 9763. [[CrossRef](#)] [[PubMed](#)]
25. Valadon, P.; Pérez-Tapia, S.M.; Nelson, R.S.; Guzmán-Bringas, O.U.; Arrieta-Oliva, H.I.; Gómez-Castellano, K.M.; Pohl, M.A.; Almagro, J.C. ALTHEA Gold Libraries™: Antibody libraries for therapeutic antibody discovery. *mAbs* **2019**, *11*, 516–531. [[CrossRef](#)] [[PubMed](#)]
26. Onodera, T.; Kita, S.; Adachi, Y.; Moriyama, S.; Sato, A.; Nomura, T.; Sakakibara, S.; Inoue, T.; Tadokoro, T.; Anraku, Y.; et al. A SARS-CoV-2 antibody broadly neutralizes SARS-related coronaviruses and variants by coordinated recognition of a virus-vulnerable site. *Immunity* **2021**, *54*, 2385–2398. [[CrossRef](#)] [[PubMed](#)]
27. Nov, Y. When second best is good enough: Another probabilistic look at saturation mutagenesis. *Appl. Environ. Microbiol.* **2012**, *78*, 258–262. [[CrossRef](#)] [[PubMed](#)]
28. Jain, T.; Sun, T.; Durand, S.; Hall, A.; Houston, N.R.; Nett, J.H.; Sharkey, B.; Bobrowicz, B.; Caffry, I.; Yu, Y.; et al. Biophysical properties of the clinical-stage antibody landscape. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 944–949. [[CrossRef](#)] [[PubMed](#)]

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