

Immunity, Volume 57

Supplemental information

**An explainable language model for antibody
specificity prediction using curated influenza
hemagglutinin antibodies**

Yiquan Wang, Huibin Lv, Qi Wen Teo, Ruipeng Lei, Akshita B. Gopal, Wenhao O. Ouyang, Yuen-Hei Yeung, Timothy J.C. Tan, Danbi Choi, Ivana R. Shen, Xin Chen, Claire S. Graham, and Nicholas C. Wu

Table S4. Cryo-EM data collection statistics, Related to Figure 7.

	SI06HA-18A5 complex (EMD-41849)
Data collection and processing	
Magnification	130,000
Voltage (kV)	200
Electron exposure (e ⁻ /Å ²)	52.76
Defocus range (μm)	-0.6 to -3
Pixel size (Å)	0.96
Symmetry imposed	C3
Initial particle images (no.)	41,774
Final particle images (no.)	39,446
Map resolution (Å)	4.81
FSC threshold	0.143
Map resolution range (Å)	N/A

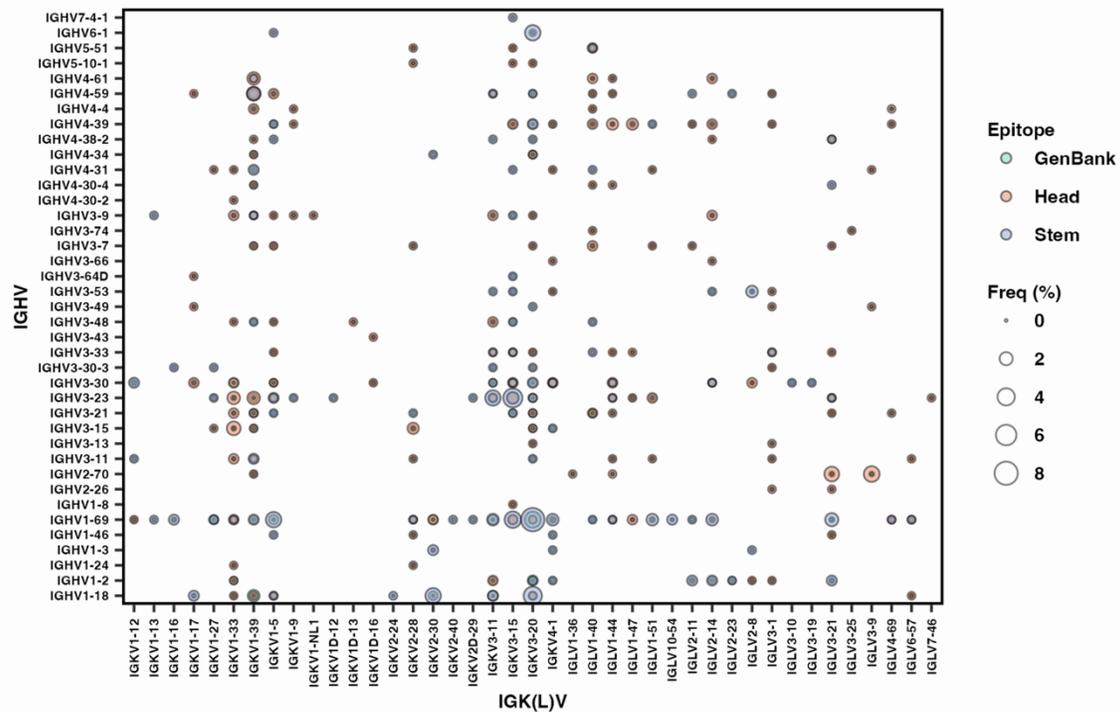


Figure S1. Preference of V gene pairings in influenza HA antibodies, Related to Figure 1.

The frequencies of different V gene pairings between heavy and light chains are shown for influenza HA antibodies to the head and stem domains. Antibodies from GenBank were also included as a reference. The size of each data point represents the frequency of the corresponding IGHV/IGK(L)V pair within its specificity category. Only those antibodies with both IGHV and IGK(L)V gene information available were included in this analysis.

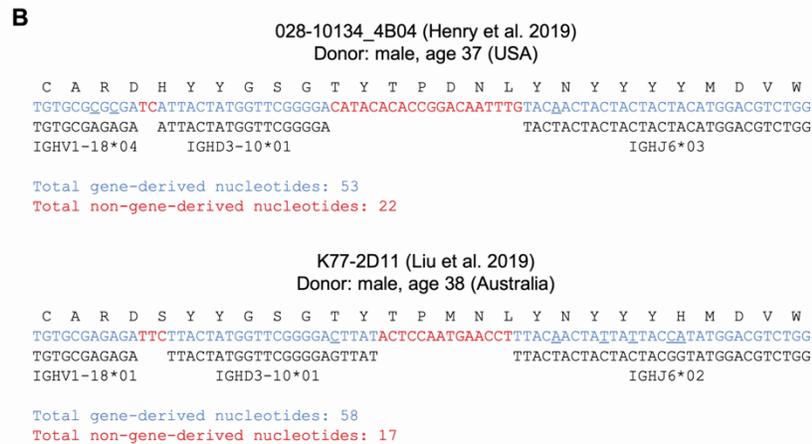
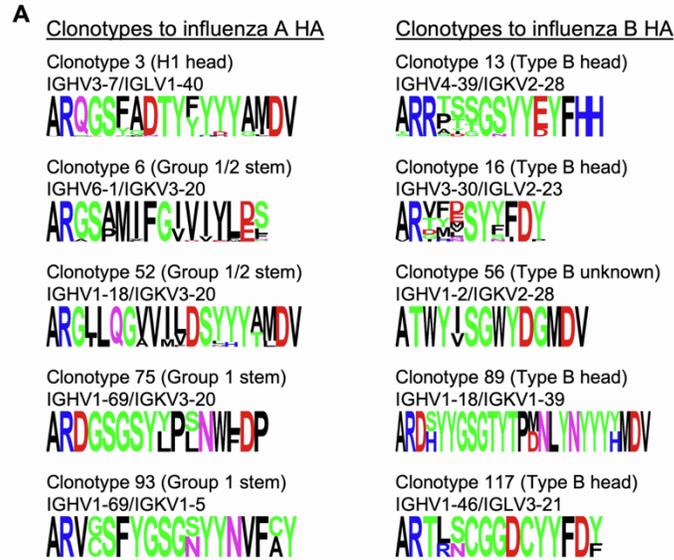


Figure S2. Public clonotypes of influenza HA antibodies, Related to Figure 1. (A) Antibodies with the same IGHV/IGK(L)V genes and at least 80% sequence identity in the CDR H3 were defined as a clonotype. A clonotype with antibodies from at least two donors was defined as a public clonotype. A total of 10 public clonotypes were identified. The V gene usage and CDR H3 sequence are shown for each of these 10 public clonotypes. The CDR H3 sequences are shown as a sequence logo, where the height of each letter represents the frequency of the corresponding amino-acid variant (single-letter amino acid code) at the indicated position. This analysis captured many known recurring sequence features in HA stem antibodies, including IGHV6-1 with an [I/V]FG[I/L/V] motif (clonotype 6) [S1,2], VH1-18 with a QxxV motif in CDR H3 (clonotype 52) [S3], and IGHV1-69 with a Tyr in the CDR H3 (clonotypes 75 and 93) [S4]. The recurring usage of

IGHV3-7/IGLV1-40 among HA head antibodies (clonotype 3) was also noted previously [S5,6].

(B) Among the five public clonotypes to influenza type B HA, clonotypes 13, 16, 56, and 117 consisted of antibodies from the same study (**Table S1**) [S7,8]. In contrast, clonotype 89 consisted of antibodies from two different studies [S8,9]. Our dataset contained two antibodies within clonotype 89, namely 028-10134_4B04 and K77-2D11, which were isolated from donors in the US and Australia, respectively [S8,9]. Amino acid and nucleotide sequences of the V-D-J junction are shown for 028-10134_4B04 and K77-2D11. Antibody 028-10134_4B04 was isolated from a 37-year-old male in the US [S8], whereas K77-2D11 was isolated from a 38-year-old male in Australia [S7]. Putative germline sequences and segments were identified by IgBlast [S10] and are indicated. Somatic mutations are underlined. Intervening spaces at the V-D and D-J junctions are N-nucleotide additions. Both 028-10134_4B04 and K77-2D11 have a long CDR H3 with 25 amino acids (IMGT numbering), including a YYGSGTY that is largely encoded by IGHD3-10 and a TPxNL motif that is encoded by N-nucleotide addition. While previous studies of recurring sequence features among HA antibodies have mainly focused on influenza type A HA [S1-4,11-15], our results suggest that recurring sequence features among antibodies to influenza type B HA may also be quite common.

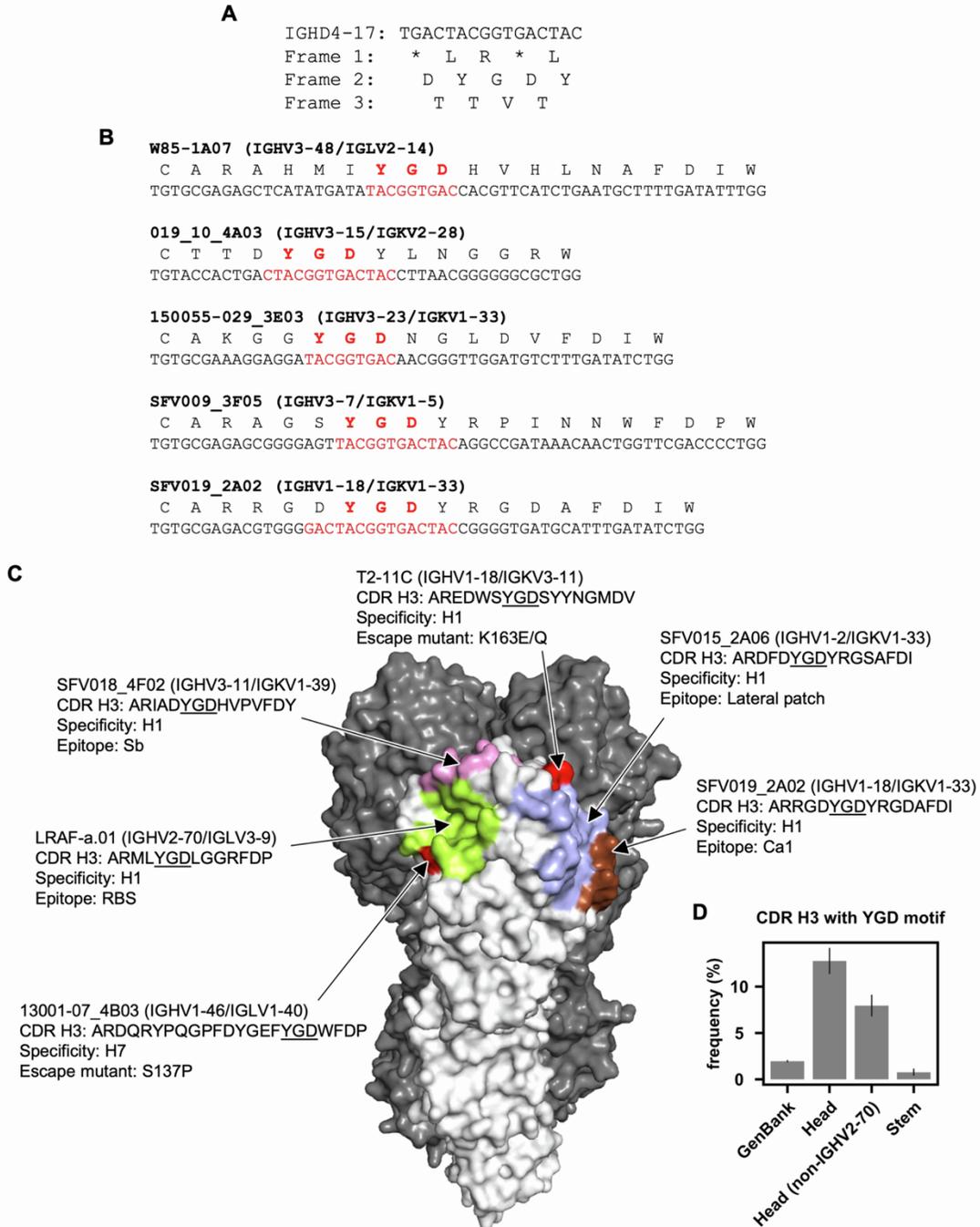


Figure S3. YGD motif in IGHV4-17 HA head antibodies, Related to Figure 1. (A) The nucleotide sequence of IGHV4-17 and its amino acid sequences in all three translation frames are shown. **(B)** CDR H3 sequences of representative IGHV4-17 HA head antibodies with different V gene usages. The YGD motif and the IGHV4-17-encoded region are highlighted in red. **(C)** Different epitopes, including the receptor-binding site (RBS, lime), Sb (pink), Ca1 (brown), and

lateral patch (blue), are shown on the HA structure (PDB: 3LZG) [S16]. Information of representative IGHD4-17 HA head antibodies that target these epitopes is shown. The locations of K163E/Q and S137P, which escape antibodies T2-11C [S6] and 13001-07_4B03 [S17], respectively, are colored in red. Of note, S137P (H3 numbering) was named as S152P in the original paper [S17]. **(D)** Frequency of antibodies with a YGD motif in the CDR H3 among all antibodies from GenBank, HA head antibodies, non-IGHV2-70-encoded HA head antibodies, and HA stem antibodies.

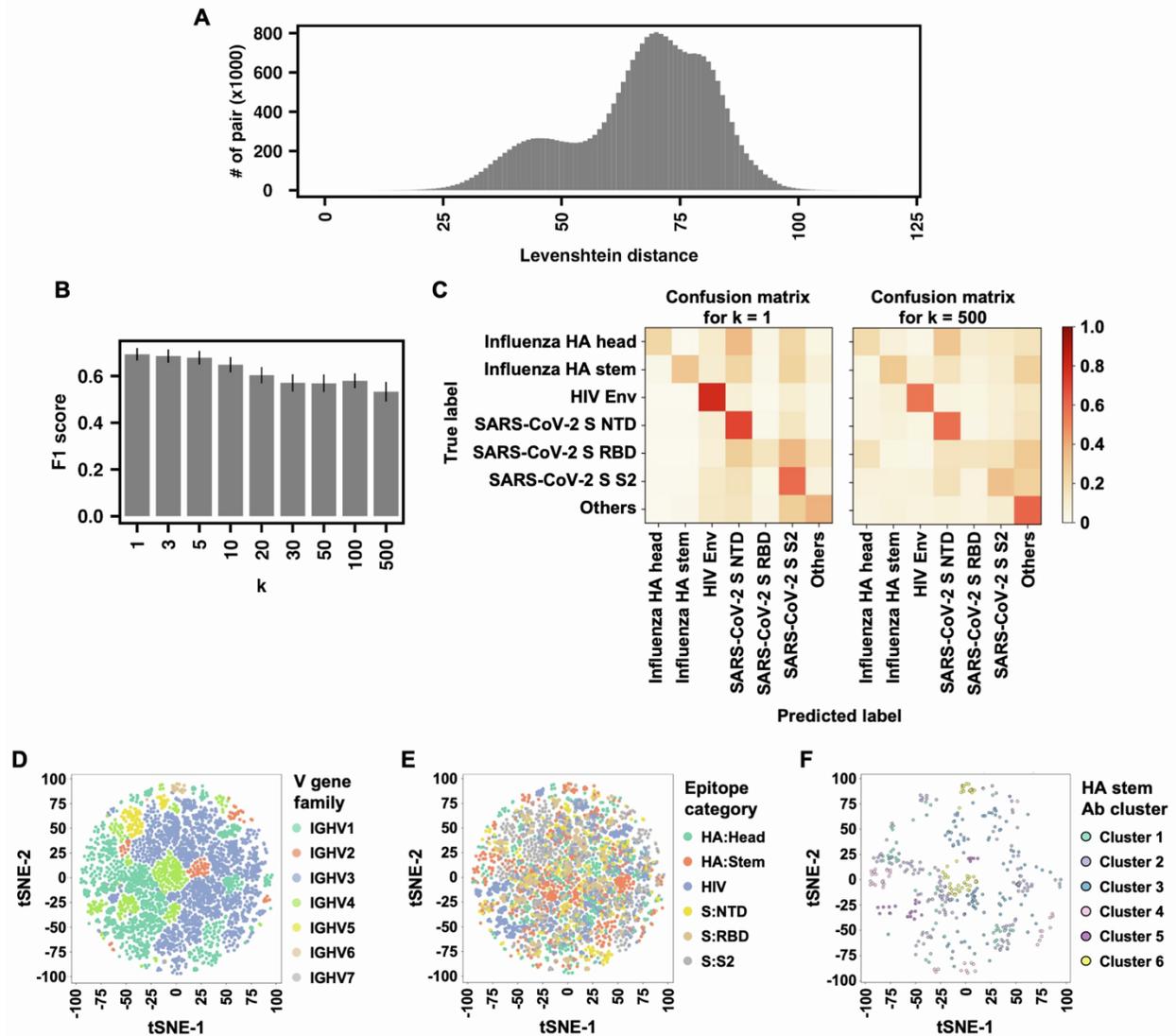


Figure S4. t-SNE analysis of the final-layer embeddings of the pre-trained mBLM, Related to Figure 3 and Figure 4. (A) Distribution of pairwise Levenshtein distance between the training and test sets is shown. (B-C) Model performance of kNN classifier on the test set was evaluated by (B) F1 score and (C) confusion matrix. We have tested a range of k values as indicated. (D-E) The final-layer embeddings of the pre-trained mBLM model (i.e. prior to fine-tuning for specificity prediction) was analyzed by t-SNE (t-distributed Stochastic Neighbor Embedding). Heavy chain sequences in the training set for fine-tuning were used in this analysis. Each datapoint represents one heavy chain sequence. Datapoints are colored by (D) V gene families, (E) specificity categories, or (F) clusters of HA stem antibodies.

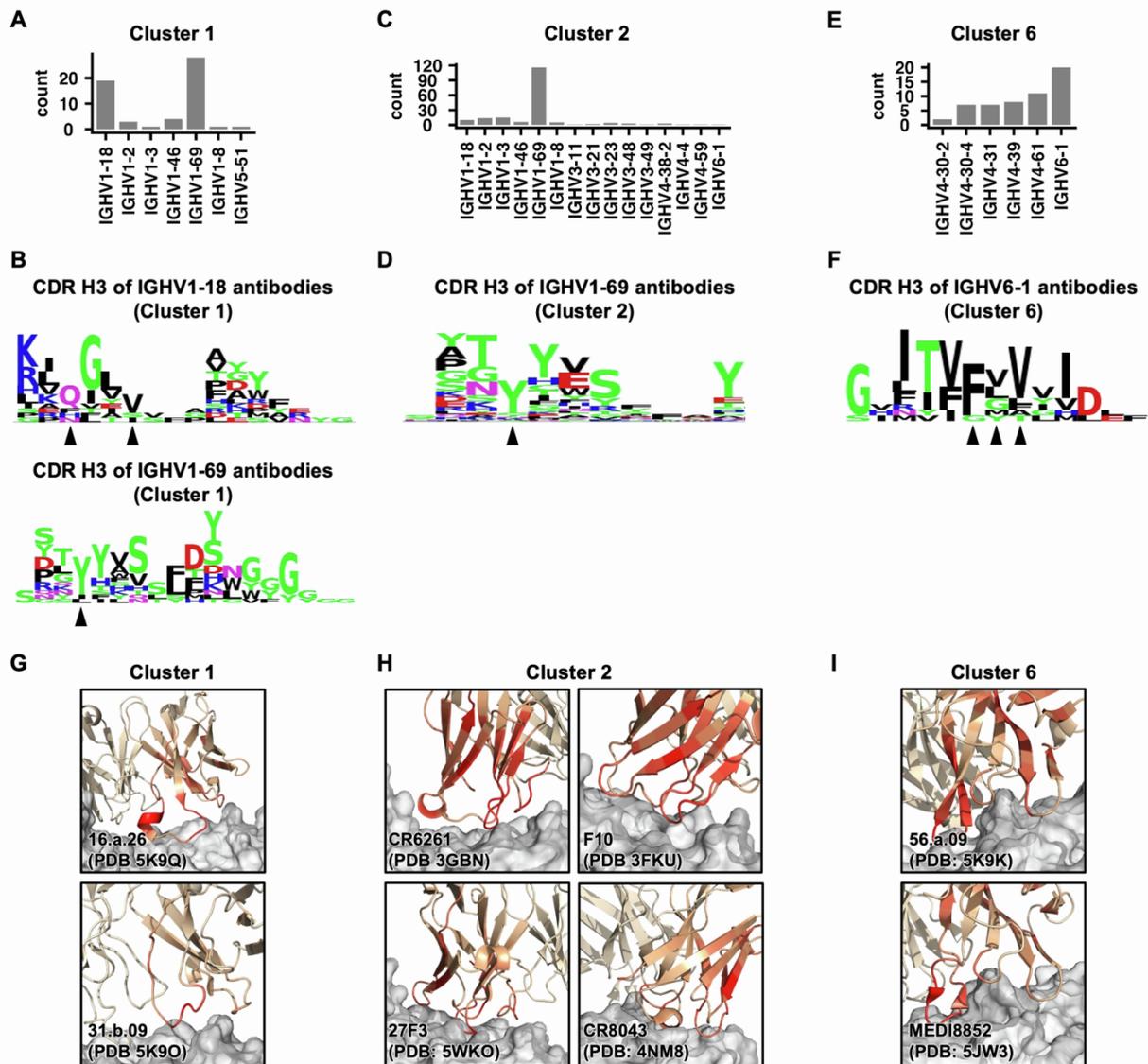


Figure S5. Sequence features of clusters 1, 2, and 6 of HA stem antibodies, Related to Figure 4. (A, C, E) IGHV gene usages among antibodies in (A) cluster 1, (C) cluster 2, and (E) cluster 6 are shown (B, D, F) The saliency score of each CDR H3 residue in (B) IGHV1-18 antibodies (top) and IGHV1-69 antibodies (bottom) within cluster 1, (D) IGHV1-69 antibodies within cluster 2, and (F) IGHV6-1 antibodies within cluster 6 was analyzed. The frequency of each amino acid for residues with a saliency score >0.5 is shown as a sequence logo. Arrows at the bottom indicate the residues of interest, including (B) a QxxV motif (top), Y98 (bottom), (D) Y98, and (F) an FGV motif (G-I) Saliency scores are projected onto the structures of (G) two antibodies

in cluster 1, namely 16.a.26 (PDB 5K9Q) and 31.b.09 (PDB 5K9O) [S3], **(H)** four antibodies in cluster 2, namely CR6261 (PDB 3GBN) [S18], F10 (PDB 3FKU) [S19], 27F3 (PDB 5WKO) [S11], and CR8043 (PDB 4NM8) [S20], as well as **(I)** two antibodies in cluster 6, namely 56.a.09 (PDB 5K9K) [S3] and MEDI8852 (PDB 5JW3) [S21]. Color scheme is the same as that in **Figure 4A**.

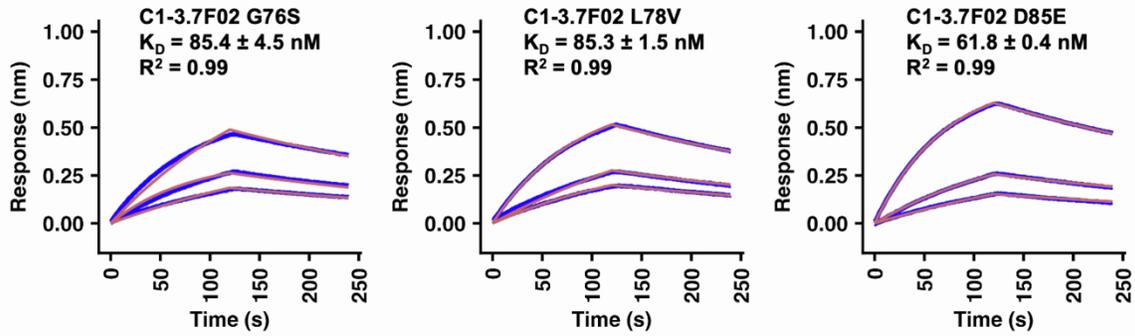


Figure S6. SHMs in the DE loop that are important for HA-stem binding affinity, Related to Figure 5. Binding kinetics of different C1-3.7F02 mutant Fabs against recombinant H3 mini-HA[S22] were measured by biolayer interferometry (BLI). The y-axis represents the response. Blue lines represent the response curves, and red lines represent a 1:1 binding model. Binding kinetics were measured for four concentrations of Fab at 3-fold dilution ranging from 300 nM to 33 nM. Dissociation constant (K_D) and the goodness of model fitting (R^2) are indicated.

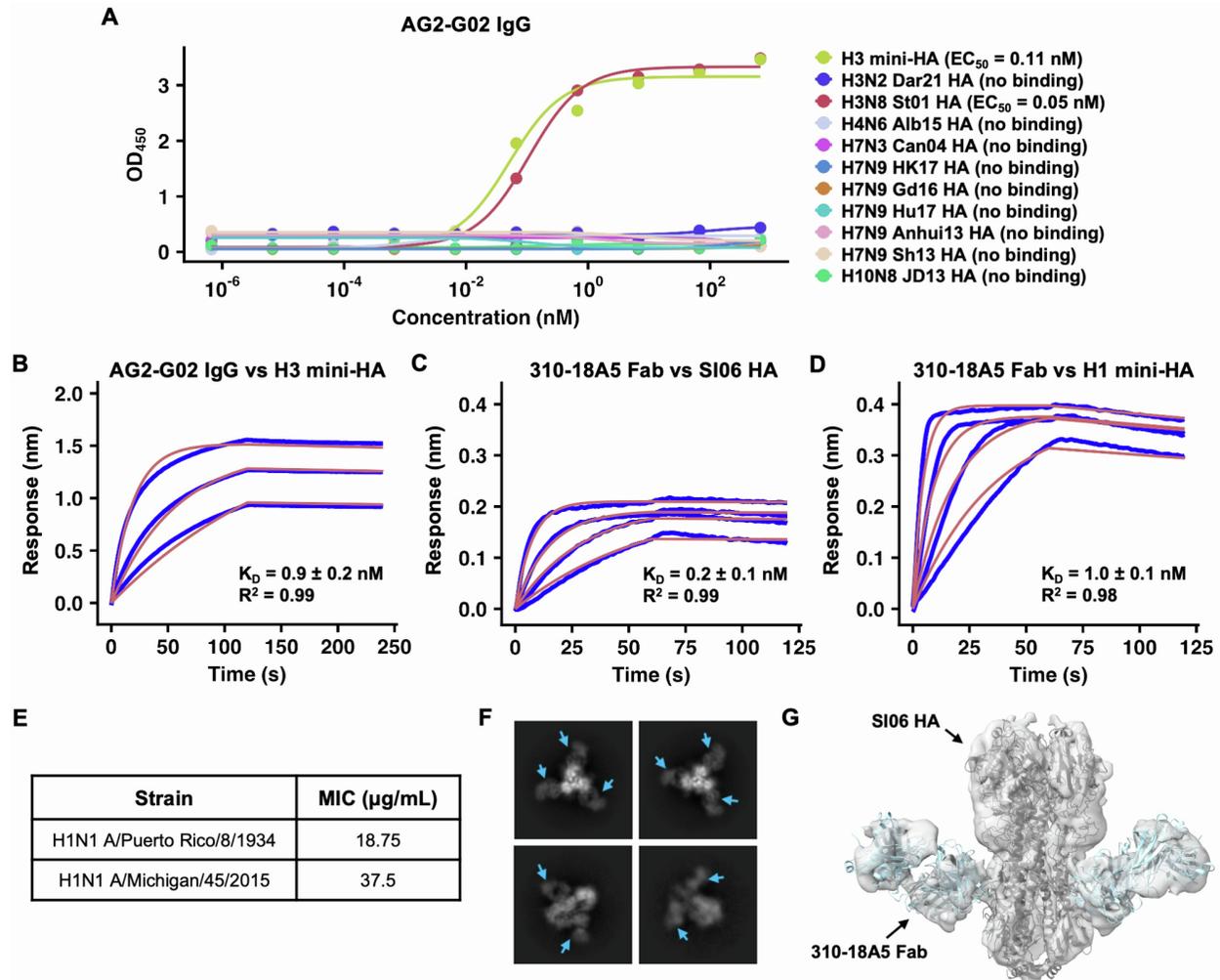


Figure S7. Characterization of antibodies that were predicted to target HA stem, Related to Figure 7. (A) The binding affinity of AG2-G02 IgG [S8] against H3 mini-HA, H3N2 A/Darwin/9/2021 (Dar21) HA, H3N8 A/duck/Shantou/1283/2001 (St01) HA, H4N6 A/mallard/Alberta/455/2015 (Alb15) HA, H7N3 A/Canada/rv444/2004 (Can04) HA, H7N9 A/Hong Kong/125/2017 (HK17) HA, H7N9 A/Guangdong/17SF003/2016 (Gd16) HA, H7N9 A/Hunan/02285/2017 (Hu17) HA, H7N9 A/Anhui/1/2013 (Anhui13) HA, H7N9 A/Shanghai/1/2013 (Sh13) HA, and H10N8 A/Jiangxi-Donghu/346/2013 (JD13) HA, was measured by ELISA. The EC_{50} values are indicated. **(B-D)** Binding kinetics of **(B)** AG2-G02 IgG [S8] against H3 mini-HA, **(C)** 310-18A5 Fab [S23] against H1N1 A/Solomon Islands/3/2006 (SI06) HA, and **(D)** 310-18A5 Fab [S23] against H1 mini-HA were measured by bi-layer interferometry (BLI). Y-axis represents

the response. Blue lines represent the response curve and red lines represent the 1:1 binding model. Binding kinetics were measured for four concentrations of Fab at 2-fold dilution ranging from 200 nM to 25 nM. Dissociation constant (K_D) and the goodness of model fitting (R^2) are indicated. **(E)** Neutralization activity of 310-18A5 was tested against two H1N1 strains, namely A/Puerto Rico/8/1934 and A/Michigan/45/2015. Minimal inhibitory concentration (MIC) is indicated. **(F)** Representative 2D classes from cryo-EM analysis of 310-18A5 Fab in complex with H1N1 A/Solomon Islands/3/2006 (SI06) HA are shown. Cyan arrows point to the 310-18A5 Fabs. **(G)** Cryo-EM 3D reconstruction of 310-18A5 Fab in complex with SI06 HA. Structural models of SI06 HA (PDB 6XSK) [S24] and CR9114 (PDB 4FQH) [S25] were docked into the 3D reconstruction.

SUPPLEMENTAL REFERENCES

- S1. Chuang, G.Y., Shen, C.H., Cheung, C.S., Gorman, J., Creanga, A., Joyce, M.G., Leung, K., Rawi, R., Wang, L., Yang, E.S., et al. (2021). Sequence-signature optimization enables improved identification of human HV6-1-derived class antibodies that neutralize diverse influenza A viruses. *Front Immunol* 12, 662909. 10.3389/fimmu.2021.662909.
- S2. Wu, N.C., Andrews, S.F., Raab, J.E., O'Connell, S., Schramm, C.A., Ding, X., Chambers, M.J., Leung, K., Wang, L., Zhang, Y., et al. (2020). Convergent evolution in breadth of two V_H6-1-encoded influenza antibody clonotypes from a single donor. *Cell Host Microbe* 28, 434-444. 10.1016/j.chom.2020.06.003.
- S3. Joyce, M.G., Wheatley, A.K., Thomas, P.V., Chuang, G.Y., Soto, C., Bailer, R.T., Druz, A., Georgiev, I.S., Gillespie, R.A., Kanekiyo, M., et al. (2016). Vaccine-induced antibodies that neutralize group 1 and group 2 influenza A viruses. *Cell* 166, 609-623. 10.1016/j.cell.2016.06.043.
- S4. Avnir, Y., Tallarico, A.S., Zhu, Q., Bennett, A.S., Connelly, G., Sheehan, J., Sui, J., Fahmy, A., Huang, C.Y., Cadwell, G., et al. (2014). Molecular signatures of hemagglutinin stem-directed heterosubtypic human neutralizing antibodies against influenza A viruses. *PLoS Pathog* 10, e1004103. 10.1371/journal.ppat.1004103.
- S5. Krause, J.C., Tsibane, T., Tumpey, T.M., Huffman, C.J., Briney, B.S., Smith, S.A., Basler, C.F., and Crowe, J.E., Jr. (2011). Epitope-specific human influenza antibody repertoires diversify by B cell intracлонаl sequence divergence and interclonal convergence. *J Immunol* 187, 3704-3711. 10.4049/jimmunol.1101823.
- S6. Huang, K.Y., Rijal, P., Schimanski, L., Powell, T.J., Lin, T.Y., McCauley, J.W., Daniels, R.S., and Townsend, A.R. (2015). Focused antibody response to influenza linked to antigenic drift. *J Clin Invest* 125, 2631-2645. 10.1172/JCI81104.
- S7. Liu, Y., Tan, H.X., Koutsakos, M., Jegaskanda, S., Esterbauer, R., Tilmanis, D., Aban, M., Kedzierska, K., Hurt, A.C., Kent, S.J., and Wheatley, A.K. (2019). Cross-lineage protection by human antibodies binding the influenza B hemagglutinin. *Nat Commun* 10, 324. 10.1038/s41467-018-08165-y.
- S8. Henry, C., Zheng, N.Y., Huang, M., Cabanov, A., Rojas, K.T., Kaur, K., Andrews, S.F., Palm, A.E., Chen, Y.Q., Li, Y., et al. (2019). Influenza virus vaccination elicits poorly adapted B cell responses in elderly individuals. *Cell Host Microbe* 25, 357-366.e6. 10.1016/j.chom.2019.01.002.
- S9. Thomson, C.A., Wang, Y., Jackson, L.M., Olson, M., Wang, W., Liavonchanka, A., Keleta, L., Silva, V., Diederich, S., Jones, R.B., et al. (2012). Pandemic H1N1 influenza infection and vaccination in humans induces cross-protective antibodies that target the hemagglutinin stem. *Front Immunol* 3, 87. 10.3389/fimmu.2012.00087.
- S10. Ye, J., Ma, N., Madden, T.L., and Ostell, J.M. (2013). IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res* 41, W34-W40. 10.1093/nar/gkt382.

- S11. Lang, S., Xie, J., Zhu, X., Wu, N.C., Lerner, R.A., and Wilson, I.A. (2017). Antibody 27F3 broadly targets influenza A group 1 and 2 hemagglutinins through a further variation in V_H1-69 antibody orientation on the HA stem. *Cell Rep* 20, 2935-2943. 10.1016/j.celrep.2017.08.084.
- S12. Wu, N.C., Yamayoshi, S., Ito, M., Uraki, R., Kawaoka, Y., and Wilson, I.A. (2018). Recurring and adaptable binding motifs in broadly neutralizing antibodies to influenza virus are encoded on the D3-9 segment of the Ig gene. *Cell Host Microbe* 24, 569-578.e4. 10.1016/j.chom.2018.09.010.
- S13. Cheung, C.S., Fruehwirth, A., Paparoditis, P.C.G., Shen, C.H., Foglierini, M., Joyce, M.G., Leung, K., Piccoli, L., Rawi, R., Silacci-Fregni, C., et al. (2020). Identification and structure of a multidonor class of head-directed influenza-neutralizing antibodies reveal the mechanism for its recurrent elicitation. *Cell Rep* 32, 108088. 10.1016/j.celrep.2020.108088.
- S14. Crowe, J.E., Jr. (2019). Influenza virus-specific human antibody repertoire studies. *J Immunol* 202, 368-373. 10.4049/jimmunol.1801459.
- S15. Andrews, S.F., Graham, B.S., Mascola, J.R., and McDermott, A.B. (2018). Is it possible to develop a "universal" influenza virus vaccine? immunogenetic considerations underlying B-cell biology in the development of a pan-subtype influenza A vaccine targeting the hemagglutinin stem. *Cold Spring Harb Perspect Biol* 10, a029413. 10.1101/cshperspect.a029413.
- S16. Xu, R., Ekiert, D.C., Krause, J.C., Hai, R., Crowe, J.E., Jr., and Wilson, I.A. (2010). Structural basis of preexisting immunity to the 2009 H1N1 pandemic influenza virus. *Science* 328, 357-360. 10.1126/science.1186430.
- S17. Henry Dunand, C.J., Leon, P.E., Huang, M., Choi, A., Chromikova, V., Ho, I.Y., Tan, G.S., Cruz, J., Hirsh, A., Zheng, N.Y., et al. (2016). Both neutralizing and non-neutralizing human H7N9 influenza vaccine-induced monoclonal antibodies confer protection. *Cell Host Microbe* 19, 800-813. 10.1016/j.chom.2016.05.014.
- S18. Ekiert, D.C., Bhabha, G., Elsliger, M.A., Friesen, R.H., Jongeneelen, M., Throsby, M., Goudsmit, J., and Wilson, I.A. (2009). Antibody recognition of a highly conserved influenza virus epitope. *Science* 324, 246-251. 10.1126/science.1171491.
- S19. Sui, J., Hwang, W.C., Perez, S., Wei, G., Aird, D., Chen, L.M., Santelli, E., Stec, B., Cadwell, G., Ali, M., et al. (2009). Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. *Nat Struct Mol Biol* 16, 265-273. 10.1038/nsmb.1566.
- S20. Friesen, R.H., Lee, P.S., Stoop, E.J., Hoffman, R.M., Ekiert, D.C., Bhabha, G., Yu, W., Juraszek, J., Koudstaal, W., Jongeneelen, M., et al. (2014). A common solution to group 2 influenza virus neutralization. *Proc Natl Acad Sci U S A* 111, 445-450. 10.1073/pnas.1319058110.
- S21. Kallewaard, N.L., Corti, D., Collins, P.J., Neu, U., McAuliffe, J.M., Benjamin, E., Wachter-Rosati, L., Palmer-Hill, F.J., Yuan, A.Q., Walker, P.A., et al. (2016). Structure

- and function analysis of an antibody recognizing all influenza A subtypes. *Cell* 166, 596-608. 10.1016/j.cell.2016.05.073.
- S22. Corbett, K.S., Moin, S.M., Yassine, H.M., Cagigi, A., Kanekiyo, M., Boyoglu-Barnum, S., Myers, S.I., Tsybovsky, Y., Wheatley, A.K., Schramm, C.A., et al. (2019). Design of nanoparticulate group 2 influenza virus hemagglutinin stem antigens that activate unmutated ancestor B cell receptors of broadly neutralizing antibody lineages. *mBio* 10, e02810-18. 10.1128/mBio.02810-18.
- S23. Whittle, J.R., Wheatley, A.K., Wu, L., Lingwood, D., Kanekiyo, M., Ma, S.S., Narpala, S.R., Yassine, H.M., Frank, G.M., Yewdell, J.W., et al. (2014). Flow cytometry reveals that H5N1 vaccination elicits cross-reactive stem-directed antibodies from multiple Ig heavy-chain lineages. *J Virol* 88, 4047-4057. 10.1128/JVI.03422-13.
- S24. Moin, S.M., Boyington, J.C., Boyoglu-Barnum, S., Gillespie, R.A., Cerutti, G., Cheung, C.S., Cagigi, A., Gallagher, J.R., Brand, J., Prabhakaran, M., et al. (2022). Co-immunization with hemagglutinin stem immunogens elicits cross-group neutralizing antibodies and broad protection against influenza A viruses. *Immunity* 55, 2405-2418.e7. 10.1016/j.immuni.2022.10.015.
- S25. Dreyfus, C., Laursen, N.S., Kwaks, T., Zuijdgeest, D., Khayat, R., Ekiert, D.C., Lee, J.H., Metlagel, Z., Bujny, M.V., Jongeneelen, M., et al. (2012). Highly conserved protective epitopes on influenza B viruses. *Science* 337, 1343-1348. 10.1126/science.1222908.