Supplementary Information for
Two complementary features of humoral immune memory confer protection against the same or variant antigens
Matthew Van Beek1, Michel C. Nussenzweig2, 3, and Arup K. Chakraborty1, 4-7,*

Arup K. Chakraborty
Email: arupc@mit.edu

This PDF file includes:

Supplementary text
Figures S1 to S16
SI References
Supplementary Information Text

Supplement 1- Width and Mean of Naïve B Cell’s Affinity Distribution

Fig. S1 shows the median binding free energies seeding and exiting the prime GC when initialized with different naïve B cell affinity distributions. The initial binding free energy distribution was a uniform distribution with four different medians: -1, -0.5, 0, and 1. To make the graph easier to read, medians of -1 and 0 are shown on the left and medians of -0.5 and 1 are shown on the right. Different widths of the uniform distribution are color-coded, with darker colors representing larger widths of the initial distribution. The distribution with median = -1 only has widths of 0, 1, and 2, because larger widths would put the lowest affinity B cells underneath the affinity floor of -2. The distributions with median = -0.5 or median = 1 have widths of 0, 1, 2, and 3. The distribution with median = 0 has all of the widths.

In most regimes, there is a small increase or decrease in median binding affinity, including for results shown in the main text, where the median is 0 and the width is 2. There are three exceptions. First, a very low initial affinity (median = -1) leads to a larger increase in median affinity, because selection is stronger at low affinities. Second, a very high initial affinity (median = 1) leads to a decrease in median affinity because selection is not strong in this affinity range and mutations are more likely to be deleterious. This range of initial affinities is not relevant for GC processes. Third, high widths lead to a larger increase in binding free energy, because selection acts upon variance; this only occurs when the width is large enough that mutation does not significantly add variance (Supplement 3).

Fig. S1. Median binding free energies seeding and exiting the prime GC at different binding free energy distributions of naïve B cells. The median of the uniform distribution occurs at -1, -0.5, 0, and 1. The width represents the total width of the uniform distribution, and the color-coding indicates a larger width. The left and right half of the graph are separated only to make the graph clearer; it does not represent any change of variables.
### Supplement 2- Selection Strength

\[
\beta_i = \beta_0 \frac{C + A_i}{C + (A)}
\]  \hspace{1cm} \text{(Eq. S1)}

A given B cell’s birth rate, \( \beta_i \), is defined above where, \( \beta_0 \) is the basal birth rate, \( C \) represents the amount of T cell help, and \( A_i \) is the amount of antigen captured. A smaller value of \( C \) means stricter T-cell selection.

Fig. S2 shows how selection strength affects the binding free energy distribution exiting the prime GC. Dashed lines indicate medians. Fig. S2 shows results for when the value of \( C \) in Eq. S1 is changed.

In all cases, stronger selection increases the median affinity and increases the density of the high-affinity tails. Selection strength in linear regimes does not qualitatively change the GC results. The GC processes do not result in a large median affinity increase, but the high-affinity right-tail becomes more populated.

**Fig. S2.** Binding free energy distributions exiting the prime GC at different values of \( C \) in Eq. S1. Dashed lines indicate medians. A darker color means stronger selection.
**Supplement 3- Temporal Dynamics of Binding Free Energy Variance**

Fig. S3 shows the variance of binding free energies in the prime GC over time when initialized with different naïve B cell affinity distributions. The initial affinity distribution was a uniform distribution with means of -0.5 (Fig. S3A) or 0 (Fig. S3B). Different widths of the uniform distribution are color-coded, with darker colors representing larger widths of the initial distribution. The GC rapidly creates variance in most parameter regimes, including the main text results, where the median is 0 and the width is 2. Higher widths, such as 3 or 4 are unlikely, because such a high initial variance would make a large amount of mutation obsolete and disagrees with experiments showing an increase in variance(1). For smaller widths, the median binding free energy exiting the prime GC is only weakly dependent upon the width (Supplement 1) because the initial variance rapidly increases to similar levels.

**Fig. S3.** The variance of binding free energies in the prime GC over time when initialized with different affinity distributions of naïve B cells. The mean of the uniform distribution is A) -0.5 and B) 0. The width represents the total width of the uniform distribution, and darker colors have more initial width.
Supplement 4- Correlation and Binding Free Energy

Fig. S4 shows how the affinity distributions exiting the recall GC and EGC change based upon the correlation factor between the prime and boost antigen. Fig. S4A shows the affinity distributions exiting the GC and Fig. S4B shows the EGC’s products. When the correlation factor is 1 (darker lines), the boost GC and EGC perform the best. As the correlation becomes more negative (lighter colors), the boost response performs worse. Only memory B cells seed the secondary GCs for the results shown in this figure.

**Fig. S4.** The probability distributions of the binding free energies of B cells after boosting with antigens with different correlation factors. The vertical dashed lines represent median affinities. A) The blue curves show secondary GC products when only memory B cells seed the GC. B) The red curves show EGC products.
Supplement 5-More generations in EGC

Fig. S5 shows the effect of increasing the number of generations of replication in the EGC. This is performed by increasing the exponential decay of memory B cells, so there are fewer initial B cells. The mean number of generations is defined below, where $\text{EGC}_{\text{out}}$ is the number of B cells exiting the EGC and $\text{EGC}_{\text{in}}$ is the number of B cells seeding the EGCs.

$$\text{Generations} = \log_2(\frac{\text{EGC}_{\text{out}}}{\text{EGC}_{\text{in}}})$$  \hspace{1cm} (Eq. S2)

The correlation factor between boost and prime antigens is 1 (Fig. S5A) or -1 (Figs. S5B). In all cases, more generations (births) means more selection events, pushing the distribution further to the right and increasing EGC’s performance. More generations occur when the correlation is negative, because the weak, variable-targeting B cells capture less antigen.

Fig. S5. The probability distributions of the binding free energies of EGC products after boosting with antigens at a correlation factor of A) 1 and B) -1. The vertical dashed lines represent median affinities.
Supplement 6- Alternative probability distribution of affinity changes upon mutation

To test the robustness of our results to different probabilities of affinity changes upon mutation, we also carried out studies using the mutational scheme from Zhang and Shaknovich’s model based off the PINT database(2, 3). This mutational scheme is similar, because most mutations are affinity-decreasing. However, this mutational scheme is more punitive than the one used by us in that mutations are more deleterious. We think the real distribution of mutations is less punitive, because AID is biased toward mutating the Complementarity-determining regions, and not the framework regions(4).

Fig. S6 shows the results of a prime-boost injection with the mutational scheme in reference 2, when the correlation factor is 1 (Fig. S6A) or -1 (Fig. S6B). In Fig. S6 the EGCs perform similarly well, and expand the high-affinity clones. However, the GCs perform much worse in this mutational scheme. There is a large density at the affinity floor of -2, and the Boost GC is outperformed by the boost EGC. The GCs are largely unable to make high affinity clones. However, the absolute highest-affinity B cells and antibodies are derived from the secondary GC, although these are rare and in the range of 7-11 kcal/mol. So, using a much more punitive mutational scheme does not alter significant qualitative results – the GC products have lower affinities in general and there are fewer high affinity clones emerging from secondary GCs.

Fig. S6. A and B) The probability distributions of binding free energies of different populations of B cells under Zhang and Shaknovich’s mutational scheme(2). Dashed lines show median affinities. The light blue curves show the memory B cells available after the first antigen exposure, the red curves show EGC products after re-exposure to the antigen, and the dark blue curves show secondary GC products. Equal proportions (50/50 mixture) of naive and memory B cells seed secondary GCs. The correlation factor between prime and boost is A) 1 or B) -1.
Supplement 7- Methods
Our model of affinity maturation is very similar to that described in Amitai et al(5). A short description of the method is included below, but more details can be found in reference 5.

MD simulations and viral geometry.
For estimating how on-rates to different epitopes depend upon the geometry of surface antigens, we used the spike of influenza as a model. However, we note that as long as there is a variation in the access to various epitopes, our qualitative results should hold. To model the influenza virus, a viral spike was generated from pdb files. The spike had 184 solvent-accessible residues, 13 of which represent a conserved region on the stem. A smaller virus (for computation speed) with a similar spike density was generated. The antibody was approximated using different-sized beads model an antibody’s geometry. MD simulations were performed using Lammps(6) with the antibody starting at different initial positions. The time it takes for the antibody to bind to an individual epitope was averaged over many MD simulations to determine the initial on-rate for the first arm, $k_{on}$. This is a mathematical representation of the steric/geometric constraints of the spike. A conditional on-rate for the second arm, $k_{on,2^nd,Ep}$, was calculated by determining how long it takes for the second arm to bind after the first arm is bound.

Affinity Maturation Simulations
Before the GC, B cells are initialized. B cells have a given precursor frequency to each epitope. $N_{variable}$ is the number of naïve precursors to the variable portion of the antigen, while $N_{Conserved}$ is the number of naïve precursors to the 13 conserved epitopes. Each B cell has a random binding free energy generated from $U[-1, 1]$. In the paper, $N_{Conserved}$ is 40 and $N_{Variable}$ is 8.

1. **GC formation**: A B cell needs to capture a threshold amount of antigen, $A_{seed} = 100$, and have captured more antigen than the median B cell in order to enter the GC. Antigen is discretized in units of 100. B cells that capture antigen are randomly selected to seed the GC, with a probability $N_{seed}/N_{eligible}$, where $N_{seed}$ is the number of B cells seeding the GC (200) and $N_{eligible}$ is the number of B cells that have captured enough antigen and are eligible to seed the GC.

2. **Growth Phase**. B cells expand according to a stochastic birth-death model for 1 day. At each time step $\Delta t$ (0.1 days), a B cell gives birth with probability $\beta_0 \Delta t$ and dies with probability $\mu_0 \Delta t$. $\beta_0$ is $\frac{1.5}{day}$ and $\mu_0$ is $\frac{0.6}{day}$.

3. **Competitive Phase** B cells then enter a competitive phase where the birth rate is dependent upon the amount of antigen captured. The competitive phase lasts for 16 days. At each timestep four different events happen.
   a. **Antigen Capture**-A B cell attempts to capture antigen, which is dependent upon the affinity of the B cell receptor and the amount of antigen left. Antigen capture is detailed below in the “Antigen Capture” section. Antigen that is captured is removed from the immune complexes causing the antigen to be depleted over time.
   b. **T Cell Selection**- B cells that capture antigen have a birth rate proportional to the amount of antigen, detailed in Supplemental 2.
   c. **Death**- B cells die according to a stochastic version of the logistic model(7). The death rate, $\mu_c$, is a function of the basal death rate, $\mu_0$, the average birth rate, $\langle \beta \rangle$, the number of B cells in the GC, $N_{cells}$, and the carrying capacity, $N_{capacity}$, which is 5000 B cells.

$$\mu_c = \mu_0 + \left( \langle \beta \rangle - \mu_0 \right) \frac{N_{cells}}{N_{capacity}} \quad (eq. \ S3)$$

d. **Mutation** Progeny will mutate and change its affinity according to the given log-normal distribution (LN).

$$\Delta E = c^* LN(\mu, \sigma, a) \quad (eq. \ S4)$$

With mean $\mu=1$, standard deviation $\sigma=0.5$, offset $a=4$, and scaling factor $c = \sqrt{0.1}$. This is shown visually below.
Fig. S7. A) The empirical distribution function of the change in binding free energy upon mutation. B) The cumulative distribution function of the change in binding free energy upon mutation. The dashed blue and orange lines show $x_1$ and $x_2$ described in the methods. The red dots show the change in binding free energy $\Delta E_1$ and $\Delta E_2$.

e. Exit B cells exit the GC as memory B cells with probability $P_{\text{Exit}} = 0.1$. These B cells will join the naïve pool and any existing memory B cells and will be eligible to seed GCs and EGCs upon re-exposure to antigen.

4. Memory B Cell Expansion Memory B cells that do not seed the GC will expand outside of the germinal center in parallel EGC processes. This expansion can potentially continue over the course of 17 days, which is the same duration as the GC processes.

a. Antigen Capture Antigen capture happens in the same way as in the GC. Antigen that is captured is removed from the immune complexes in the EGC, depleting it over time. Note: These immune complexes represent a different store of antigen from the GC.

b. T cell selection- B cells that capture antigen have a birth rate proportional to the amount of antigen.

c. Death- Memory B cells are selected to die with probability $\mu$. This death rate is low at 0.01. Higher death rates would result in a slower consumption of antigen and more cycles of EGC processes.

d. Exiting- When all of the antigen in the EGC has been consumed or 17 days pass, the memory B cells exit the EGC.

5. Memory B cells decay To keep simulations more computationally tractable, memory B cells decay according to some death rate, $\mu_{\text{decay}}$ over time $T_{\text{decay}}$. $\mu_{\text{decay}} T_{\text{decay}} = 1.5$.

6. Repeat Steps 1 – 5 are repeated for the primary exposure and re-exposure to antigens.

Antigen Capture Each of the B cell’s 20 BCRs independently search for antigen. It is assumed that a BCR arm has a characteristic time, $t$, to find an antigen molecule. The probability it finds an antigen before a time, $\tau$

$$P_{\text{find antigen}} = 1 - e^{-k_{\text{on}} \tau} \quad (\text{Eq. S5})$$

where $k_{\text{on}}$ is the individual on-rate for the first arm of a B cell, which is specific to the epitope. The antigen capture is then modeled as each BCR binding to a pair of antigens in a Markov process to estimate the probability of antigen capture. The probability of different transitions is detailed here(5) and re-shown in Figure S8. Note: Figure S8 shows only half of the transitions from the state where both arms are bound; symmetry is ignored. This symmetry is the source of the factor of 2 in $A_1$ and $A_2$. The accompanying table details all of the transition probabilities.
**Fig. S8.** A schematic showing the blue BCR attempting to capture the red antigen from the yellow immune complex. This image shows the probability of different transitions from state to state in the Markov Chain. The table details these transition probabilities.

Here, $\xi$ and $\lambda$ is the capture rate of the Antigen from the IC when two arms or one arm is bound, respectively. Here, $x_b F$ (1 nm * 1 pN) relates to the work done by the B cell pulling upon the antigen. $E_{Ag-mem}$ is the interaction energy between the Antigen and the membrane (Immune Complex). $E_{Ag-Mem}$ is 2 $k_b T$.

\[
\xi = e^{-\beta E_{Ag-mem} - e^{\beta x_b F}/2} \quad \text{(Eq. S6)}
\]

\[
\lambda = e^{-\beta E_{Ag-Mem}} e^{\beta x_b F} \quad \text{(Eq. S7)}
\]

$k$ is the failure rate of antigen capture (rupture rate of the arm) when two arms are bound; $r$ is the rupture rate when one arm is bound. $E_{Ag-Ab}$ is the binding free energy of a BCR to the antigen, which is variable.

\[
k = e^{-\beta E_{Ag-Ab} - e^{\beta x_b F}/2} \quad \text{(Eq. S8)}
\]

\[
r = e^{-\beta E_{Ag-Ab}} e^{\beta x_b F} \quad \text{(Eq. S9)}
\]
Supplement 8- EGC Mutation

Transcriptomic evidence suggests that AID is expressed substantially less in the SPF/EGCs than in the GC(8). Simulations were run with an EGC mutational rate varying between 0 to 100% of GC mutation. Fig. S9 shows that an increase in mutation rate acts like an interpolation between the Secondary GC-MBC Only and recall EGC behavior shown in Figure 2A and 3A. As the mutation rate increases, the median EGC binding free energy becomes worse, but the tails increase in thickness. Fig. S9A shows when the correlation factor is 1, and Fig. S9B shows when the correlation factor is -1. Given that AID has dramatically decreased expression in the SPF/EGCs(8), there is little difference between a model with zero mutation rate and the true small mutation rate.

**Fig. S9.** The empirical probability distributions of binding free energies of B cells exiting the recall EGC at different mutational rates when the correlation factor is A) 1 or B) -1.
Supplement 9- Varying Precursor Frequency of B Cells Targeting Conserved Site

Fig. S10 shows the affinity distributions of EGC products (orange) and GC products (blue) when changing the number of B cells targeting the conserved fraction to a variable antigen with a correlation of -1. The lighter-colored lines show that when fewer B cells target the conserved portion (only 7% instead of 27.5%), the secondary GC and EGC response fares worse because there are fewer high-affinity B cells targeting the conserved portion. The dark lines show that when more B cells target the conserved portion (65.5% of B cells), the GC and EGC fare better. The qualitative results do not differ from Figs 3A.

Fig. S10. A) The probability distribution of binding free energies of B cells exiting the recall GC and EGC when the correlation factor is -1. The light blue and light orange lines show when the number of B cells targeting the conserved site is 20% of that in Fig. 3A. The dark blue and dark orange lines show when the number of B cells targeting the conserved site is 5x that in Fig. 3A. Only memory B cells seed secondary GCs.
Supplement 10- Naïve cells seeding GCs can enhance secondary GC output against variant antigens in some parameter regimes.

Fig. S11 shows regimes in which the secondary GC performs worse with more memory B cells seeding GCs. Fig. S11A shows that when there are no conserved epitopes (or precursors to conserved epitopes), seeding the recall GC with only naïve B cells (light blue) perform better than memory B cells (dark blue) in terms of median affinity. Figure S11B shows that when T-cell selection is weaker (C=800), then memory B cells cause the median affinity to be lower. This is because selection is needed to filter out the weak pool of prime memory B cells (light blue line in Figure 3). However, the main effect that we find is that the distribution of affinities becomes narrower as more naïve B cells seed secondary GCs; the differences in median affinities are small. However, Figure S11C shows when a more punitive mutational scheme(2) is used (see supplement 6), naïve B cells seeding secondary GCs results in a higher median affinity. In all graphs, the right-tails are thicker for memory B cells, because mutation introduces variance that allows for the possibility of higher-affinity B cells seeding the GC.

The graphs shown here are just a few examples of regimes in which naïve B cells seeding GCs can outperform the situation where only memory B cells do. Other examples include when the naïve B cell affinity distribution has a higher affinity or when the priming antigen is more distantly related to the variant virus than a correlation factor of -1, which have been considered in studies focused on the evolution of an optimal immune response(9).

Fig. S11. The probability distribution of binding free energies of B cells exiting the secondary GC when the correlation factor is -1. A) There are no B cells targeting the conserved sites. The light blue lines show when only naïve B cells seed the recall GC. The dark blue lines show when only memory B cells seed the recall GC. B) T cell-selection is weaker (C=800). The light blue line shows when the recall GC is seeded by 12% memory B cells, and the dark blue line shows only Memory B cells seed the recall GC. C) A more punitive mutational scheme(2) is used.
**Supplemental 11- Evolutionary Model: Mutation rate = 0.25 and 0.5**

Fig. S12 shows the results corresponding to Fig. 6 when the mutation rate, $\mu$, is 0.25 and 0.5, instead of 0.05. S12A and S12B show the response to a variant virus when $\mu = 0.25$ (S12A) or $\mu = 0.5$ (S12B). The qualitative behavior remains the same against a variant virus when $\mu = 0.25$ (S12A) but changes when $\mu = 0.5$ (S12B). When $\mu$ is high (S12B), the viral load decreases as $r_{GC}$ increases because the prime GC produces many diverse memory B cells at high mutation rates; for example, if $\mu = 1$, the B cells would have a uniform distribution. At large $\mu$, more boost GC antibodies (increasing $r_{GC}$) is more beneficial than a more diverse memory pool (decreasing $r_{GC}$).

In our model, this qualitative change happens around $\mu = 0.35$. Figs. S12C and S12D show the distribution of B cells leaving the recall GC and EGC when $\mu = 0.5$. S12C shows the distribution against the variant antigen of $[0.8, 0.8, 0.8]$, and S12D against a nearby variant of $[1, 1, 1]$. When $\mu$ is higher, the prime GCs generate more B cells that have higher affinities to the variant (compare Recall EGC to 6C and 6D).

**Fig. S12.** Results corresponding to Fig. 6B-D when the mutation rate is increased. A and B) The maximum viral load during a recall infection at different values of $\beta$ where $\alpha = 1$. The red lines represent a distant virus at the point $[0.8, 0.8, 0.8]$ in shape space, while the green lines represent the same virus at the origin. The dashed lines represent simulations where only earlier B cells leave the GC as MBCs. A) $\mu = 0.25$ B) $\mu = 0.5$. C and D) The role of diversity created in the GC. The fitness distributions for GC and EGC products after the recall response ($t=32\tau$) with a variant antigen at $[0.8, 0.8, 0.8]$ when $\alpha=\beta=1$ and $\mu = 0.5$. C) The fitness distribution to the antigen at $[0.8, 0.8, 0.8]$ or D) The fitness distribution to the antigen at $[1, 1, 1]$ (nearby variant).
Supplement 12- Evolutionary Model: Plasma cells leave EGC

Fig. S13 shows results for the case when plasma cells leaving the GC at a rate of $PC_{EGC}$ is not ignored in Eq. 3. The results are roughly the same as in Fig. 6 if $r_{ECC}$ is shifted by the value of $PC_{EGC}$ (0.6). This is because the average effective population growth rate of the EGC is equal to $r_{EGC} - PC_{EGC}$, and by not including $PC_{EGC}$ we effectively have a higher population growth rate in the results shown in the main text. The average effective growth rate of the EGC and GC is what is observed in experiments(8). The only difference is that EGC selection is slightly stronger than for the results shown in the main text. This only affects the response to a variant pathogen (compare to Fig. 6B), as the fitness to the same pathogen is already very high. This stronger selection is also shown by comparing the orange peaks in 6C and S13C.

Fig. S13. Results for the case when plasma cells leaving the GC at a rate of $PC_{EGC}$ is not ignored in Eq. 3. A) The effect of changing $\alpha$ ($r_{EGC}/r_v$) and $\beta$ ($r_{GC}/r_v$) on the log_{10} (maximum viral load) during a recall infection with the same virus. The abcissa is the growth rate of the GC compared to EGC, ($\beta/\alpha$). The lines show different values of $\alpha$. B) The maximum viral load during a recall infection at different values of $\beta$ where $\alpha = 1$. The red lines represent a distant virus at the point [0.8, 0.8, 0.8] in shape space, while the green lines represent the same virus at the origin. The dashed lines represent simulations where only earlier B cells leave the GC as MBCs. C and D) The role of diversity created in the GC. The fitness distributions for GC and EGC products after the recall response ($t=32\tau$) with a variant antigen at [0.8, 0.8, 0.8] when $\alpha=\beta=1$. C) The fitness distribution to the antigen at [0.8, 0.8, 0.8] or D) The fitness distribution to the antigen at [1, 1, 1] (nearby variant).
Supplement 13- Evolutionary Model: Continuous Model

Fig. S14 shows the results of a continuous model without B cell numbers below 1 being set to zero. Thus, there is no extinction of memory B cell clones. The results are approximately the same as the discrete model (Figs. 6A), except against a variant antigen. In this scenario, the immune system responds better to the variant antigen (Fig. S14B), because memory B cells near [0.8, 0.8, 0.8] in shape space do not go extinct after the prime, allowing the EGC to expand these high-affinity clones to the variant pathogen (compare the orange bars in Fig. S14C and 4C). Just as in Supplement 11, more boost GC antibodies (increasing \( r_{GC} \)) is more beneficial than a more diverse memory pool (decreasing \( r_{GC} \)) because there is no extinction of clones that target the variant antigen.

Fig. S14. Results of a continuous model, where there is no extinction of fractional MBCs. A) The effect of changing \( \alpha \) (\( r_{EGC}/r_{v} \)) and \( \beta \) (\( r_{GC}/r_{v} \)) on the log_{10} (maximum viral load) during a recall infection with the same virus. The abscissa is the growth rate of the GC compared to EGC, \( (\beta/\alpha) \). The lines show different values of \( \alpha \), the scaled EGC growth rate. B) The maximum viral load during a recall infection at different values of \( \beta \) where \( \alpha = 1 \). The red lines represent a distant antigen at the point [0.8, 0.8, 0.8] in shape space, while the green lines represent the same antigen at the origin. The dashed lines represent simulations where only earlier B cells leave the GC as MBCs. C and D) The role of diversity created in the GC. The fitness distributions for GC and EGC products after the recall response \( t=32\tau \) with a variant antigen at [0.8, 0.8, 0.8] when \( \alpha=\beta=1 \). C) The fitness distribution to the antigen at [0.8, 0.8, 0.8] or D) The fitness distribution to the antigen at [1, 1, 1] (nearby variant).
**Supplement 14- Phylogenetic Trees**

Fig. S15 shows the phylogenetic trees used to sort clones into "likely EGC" or not. Phylogenetic trees show the mutational distance on the abcissa, with each blue dot being a common ancestor. Identical sequences are connected vertically. The labels show the patient ID, when the sample was taken, the isotype, and the B cell type. Fig. S15A shows an identical sequence after prime and boost, leading us to think these come from the EGC. Fig. S15B shows two identical boost sequences. We assume these come from some undetected clone in the prime, and so we deem them likely EGC clones. Fig. S15C shows identical prime sequences, which could not have come from the EGC, so they are not deemed likely EGC clones.

**Fig. S15.** Phylogenetic trees of clonal families detailing the process of deeming a "likely EGC" clone or not. A) A tree where with identical prime-boost clones deemed "likely EGC" clones. B) Identical boost clones deemed "likely EGC clones". C) Identical prime clones not deemed likely EGC clones.
Supplement 15- 5 Month after Boost Data

Five different EGC clones were identified after 5 months(10). Of those five, one had an affinity measured at 5 months, three had identical clones with affinities measured at 1.3 months, and one had no affinity measurement. Identical clones with affinities measured at separate time points were not included in Fig. 4, however, Fig. S16 includes these. Fig. S16 shows the distribution of clones sampled 5 months after the boost. The behavior is similar to the behavior shown in Fig. 4B. The EGC clones are concentrated at high-affinities, meanwhile the other clones and singlets have a higher variance and contain the highest-affinity clones. The low affinity tail belonging to 1.3 month clones is not present in the 5 month clones.

Fig. S16. The cumulative distributions of binding free energies of B cells sampled 5 months after the second dose of the vaccine (boost). Data for those identified as expanded in EGCs (red), part of other clonal families (blue), or singlets (light blue). Dashed lines indicate means.
SI References

1. M. Kang, T. J. Eisen, E. A. Eisen, A. K. Chakraborty, H. N. Eisen, Affinity Inequality among Serum Antibodies That Originate in Lymphoid Germinal Centers. *PLoS One* 10, e0139222 (2015).

2. J. Zhang, E. I. Shakhnovich, Optimality of Mutation and Selection in Germinal Centers. *PLOS Comput. Biol.* 6, e1000800 (2010).

3. M. D. S. Kumar, M. M. Gromiha, PINT: Protein–protein Interactions Thermodynamic Database. *Nucleic Acids Res.* 34, D195–D198 (2006).

4. T. T. Wu, E. A. Kabat, An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *J. Exp. Med.* 132, 211–250 (1970).

5. A. Amitai, *et al.*, Defining and Manipulating B Cell Immunodominance Hierarchies to Elicit Broadly Neutralizing Antibody Responses against Influenza Virus. *Cell Syst.* (2020) https://doi.org/10.1016/j.cels.2020.09.005.

6. A. P. Thompson, *et al.*, LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput. Phys. Commun.* 271, 108171 (2022).

7. I. NÄSELL, Extinction and Quasi-stationarity in the Verhulst Logistic Model. *J. Theor. Biol.* 211, 11–27 (2001).

8. I. Moran, *et al.*, Memory B cells are reactivated in subcapsular proliferative foci of lymph nodes. *Nat. Commun.* 9, 3372 (2018).

9. C. Victor, V. Massimo, W. A. M., M. Thierry, Affinity maturation for an optimal balance between long-term immune coverage and short-term resource constraints. *Proc. Natl. Acad. Sci.* 119, e2113512119 (2022).

10. A. Cho, *et al.*, Anti-SARS-CoV-2 receptor-binding domain antibody evolution after mRNA vaccination. *Nature* 600, 517–522 (2021).