nature portfolio

Peer Review File

Vaccination elicits HIV broadly neutralizing antibodies in primates

Corresponding Author: Professor William Schief

Any redactions in this file are there to maintain patient confidentiality, the confidentiality of unpublished data, or to remove third-party material.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Referee #1

(Remarks to the Author)
This manuscript titled “Vaccination elicits HIV broadly neutralizing antibodies in nonhuman primates” by Steichen et al. presents an comprehensive nonhuman primate (NHP) study aimed at advancing germline-targeting (GT) vaccination strategies for eliciting HIV broadly neutralizing antibodies (bnAbs). Given the longstanding difficulty of inducing bnAbs by vaccination and the limited success of prior GT approaches in clinical and preclinical settings, the topic is of high significance. The authors, who have been central drivers of the GT paradigm, report iterative refinements to their strategy and provide deep immunogenetic and structural characterization of elicited responses.

The work demonstrates that sequential immunization can expand BG18-class lineages, promote extensive somatic hypermutation (SHM), and yield antibodies that partially recapitulate structural features of mature BG18, including CDR H3 engagement of the V3 GDIR motif and interactions with the N332 glycan. These data indicate that aspects of the lineage can be guided in the intended direction. The structural analyses are a strength of the study and provide convincing evidence of epitope targeting by both type I and type III BG18 modes. The reported overall good serum breadth and especially in one animal with 84% relative to mature BG18 is notable.

However, the central question for a high-impact advance is not whether maturation can be nudged toward a known bnAb template under prolonged and highly engineered conditions, but whether this represents a scalable and generalizable vaccination solution. On this point, the manuscript is less compelling: The regimen requires ~120 weeks, seven boosts, and ~19 distinct immunogens or nanoparticles. Such complexity raises concerns about feasibility, scalability, and clinical translation. While proof-of-concept studies can tolerate complexity, the manuscript does not sufficiently articulate a realistic path toward regimen simplification or define which components are indispensable versus exploratory.

Relatedly, although longitudinal B cell analyses were performed, the manuscript stops short of establishing causal links between specific immunogens (or immunogen combinations) and the acquisition of key SHMs associated with breadth. Without clearer attribution, it is difficult to discern whether breadth development is driven by rational immunogen design or by stochastic maturation under prolonged antigenic exposure. A more rigorous dissection of immunogen-to-mutation relationships would greatly strengthen the mechanistic understanding.

Overall, the study represents a substantial technical effort and provides conceptual progress for GT strategies. However, the translational implications appear limited in the current form, and the manuscript would be strengthened by a clearer demonstration of mechanistic guidance, regimen simplification logic, and generalizability.

Specific comments
1. Figure 2e: The color coding is insufficiently defined and should be clarified.
2. Figure 4: The N332 glycan should be labeled. An additional glycan appears in panel B and should be identified. Quantitative RMSD comparisons between BG18 and elicited type I mAbs would strengthen structural similarity claims.
3. The apparent increase in breadth from post-B3 to post-B4/5 in G2 is striking. Is there direct evidence linking this change to the B46 boost, or could this reflect stochastic maturation?

4. The strong response in NHP RVz20 and presence of non-type I “BG18-like” antibodies warrant deeper analysis, as they may reveal alternative pathways than the BG18 GT.

Referee #2

(Remarks to the Author)
This manuscript reports three important observations. First, broadly HIV-neutralizing antibody responses against this highly mutable virus can be elicited by vaccination in non-human primates, even when precursor B-cell frequencies targeting the desired epitope are extremely low—indeed, lower than those estimated in humans—and without the use of transgenic mice engineered to express “ideal” B-cell receptors. Second, the immunogens used to achieve this outcome were not selected empirically but were rationally designed to present specific, predefined structural features. Finally, the study demonstrates that once naïve B cells are activated by a germline-targeting priming immunogen and enter germinal center reactions, they can be repeatedly boosted to further drive affinity maturation through subsequent immunizations—an outcome that has been questioned by others in the field.
Major queries
1. Following Boost 1, all animals in Groups 2–4 exhibited high frequencies of B cells specific for the boost immunogens (Fig. 1), whereas no animals in Group 1 (immunized only with the prime immunogen, N332-GT5) displayed boost-binding memory B cells. In light of this, can the authors clarify why priming with N332-GT5 is required for the observed responses?
2. BG18 type I and III mAbs isolated after boosts 3, 4, and 5 (i.e., prior to immunization with the AMP cocktail and the nanoparticle trimer cocktail at boosts 6 and 7, respectively) exhibited some degree of neutralization breadth. However, serum cross-neutralizing activity was not detected at these time points (Supplementary Table 5) and only became clearly apparent after the seventh immunization. The authors should therefore either (i) provide comparative data on somatic mutation levels, affinities, and structural features of BG18 mAbs isolated after boosts 6 and 7 (vs mAbs isolated and fully characterized by the previous boost immunizations), or (ii) reconsider emphasizing the contribution of the final two immunizations.
Minor points
• Please clarify what is meant by “BG18-sensitive viruses” in the analysis of monoclonal antibody and serum neutralization. Are these viruses specifically sensitive to BG18 mAb alone, or are they also sensitive to other known anti-N332 broadly neutralizing antibodies?
• On pg 3, line 23-24, the authors discuss that ‘immunological cell biological constraints’ may prevent B cells that become activated by the germline-targeting immunogen to undergo further stimulation/maturation by the necessary subsequent booster immunizations. Can they provide relevant references that discussed this possibility ? It seems to be a critical point for what the authrs try to do here.

Referee #3

(Remarks to the Author)
A. Summary: this is an elegant paper demonstrating the feasibility and improving partial success of germline-targeting HIV Env immunization towards the achievement of induction of a mature sinlge bnAb specificity clonal lineages that become dominant in serum in NHPs

B. Originality/significance: this manuscript represents moving the germline-targeted approach further towards translation with achieving evolution of initiated V3-specific bnAb precursor lineages, defining a new pathway for V3 glycan bnAb evolution (type III), maintenance and expansion of those lineages towards increasing breadth with shaping boosts, dominance of vaccine-elicited and evolved B cell lineages in serum.

C/F. Data & methodolgy and suggestions
1. LN and PBMC B cell analysis should be linked between animals - is there a definition of "responder vs non-responder" that can be proposed based on the flow and/or serum outcome? Are there biomarkers of a responder phenotype?
2. plasma IgG binding strength to vaccine antigens should be reported and related as a potential biomarker of responder status
3. the recognition of evolved/more mature Envs is reported for groups 2-4 after boost 2, but there is not a detailed description of which revertered Env mutations were required for this binding. Can an analysis of plasma and/or mAb binding to reverted Env mutants be included to determine which mutations are the most critical for continuing to engage these lineages?
4. in the bnAb lineage analysis, can the number of improbable mutations that were incorporated be reported? this would be valuable in determining how challenging the AID barriers will be towards eliciting these lineages. Was there one boost immunogen that was better at incorporation of the key improbable mutations?
5. groups 2 vs 3 vs 4 were not compared to help move along the question of which initial shaping immunogen is most successful, this should be included as a separate analysis/figure
6. the achieved geomean ID50 across animals should be related to the IC80 that was calculated from the AMP trial
7.The paper would benefit from inclusion of how the predominate B cell lineages (% antigen-specific B cells, type I/III, number of clones/B cells, etc) relate to serum neuts achieved in each animal - is there a relationship? An early B cell biomarker will benefit the field
9. a definition of "bnAb" for serum and mAbs needs to be included

D. Statistic applied should include more descriptive statistics
1. Animal numbers should be more clearly stated. Every % reported should include absolute numerators and denominators and every median should have an associated range.

2. approach to multiple comparisons should be added

E conclusions
1. there are a few animals highlighted in the conclusion (such as line 20, page 10) for serum neut breadth - but it is unclear how it is defined (how many/which viruses, how is it compared to mAb concentration, etc)
2. plasma cells are mentioned in the discussion (line 40, page 10), but not evaluated here - differentiate the future research needs from data presented

H. Clarity
1. The abstract should include which vaccine platform is being used (adjuvanted Env SOSIP subunit)
2. when reporting % B cells bound to various Envs, this is reported as "higher Env binding" (such as line 35, page 5), but I believe this is referring to larger % B cell (or higher MFI)? the units need to be specific for each of the reported "higher/lower" comparison
3. describe the "sentinel" PSV in more detail (how it was selected/representative) - page 9, line 40

Referee #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Version 1:

Reviewer comments:

Referee #2

(Remarks to the Author)
I remain very positive about this study. The revised manuscript and rebuttal address most of my previous comments, with two important exceptions.
The first outstanding issue concerns the frequency of B cells elicited by the priming immunogen (germline-targeting) that can bind the first heterologous boost immunogens (B46 [Group 2], B54 [Group 3], and B50 [Group 4]). The authors refer to prior work (Extended Data Fig. 1) showing that N332-GT5–immunized macaques developed antibodies capable of binding these booster immunogens, and they also discuss responses at week 12 following the first heterologous boost in the present study. However, what is needed here is direct evidence from the current study on the frequency of B cells, after the prime, that recognize these booster immunogens. For each group, the authors should report: (i) the number of animals with such detectable responses, and (ii) the frequency of B cells recognizing the first boost immunogen in each animal.
The second outstanding issue relates to the absence of data on broadly neutralizing antibodies isolated after the 6th and 7th immunizations, when serum neutralization breadth appears to peak. While the evolution of serum neutralizing activities throughout the immunization regimen is well documented, monoclonal antibodies are only characterized up to the 5th boost. This raises an important question: why do sera collected after the 6th and 7th immunizations exhibit greater breadth? Do antibodies acquire additional beneficial mutations at these later time points, or is the increased breadth primarily due to higher antibody titers?
Notably, even after the 3rd boost, isolated monoclonal antibodies from Groups 2–4 display measurable breadth (particularly in Group 4), whereas those from Group 1 do not (Supplementary Table 2, p. 9).
Minor points:
1. Extended Data Fig. 1c: Please clarify which curve corresponds to min BG18.6 and which to BG18.
2. Page 4: The manuscript states that the immunogen sequence exposes “12 fully glycosylated BG18 epitopes.” I count 10: 1 in Boost 3, 4 in Boosts 4 and 5, and 4 in Boost 6. The immunogens used in Boost 7 appear to be the same as those in Boosts 4–6.

Referee #3

(Remarks to the Author)
Thank you to the reviewers for the extensive work to refine the manuscript, I feel the translatability and strategy for next steps are clearer now in the edited version.

I have a couple of final comments:
- I don't think it is accurate to say that BG18 has few improbable mutations than VRC01 and those in the HCDR3 can't be detected via ARMADILLO, as this has been done. Thus, it would be helpful to assess the probability of the mutations that are in the clonal lineages of vaccine-elicited mAbs characterized in this paper and which are required for the neutralization breadth, moving along the definition of "responder" and how to measure the maturation of responses
See: https://armadillo.dhvi.duke.edu/bnab/BG18/heavy

- The relationship of the elicited neutralization to levels of protection established in NHPs and the VRC01 clinical trial is helpful. However, the serum neutralization analysis seems to have only been performed on BG18-sensitive variants - which should be acknowledged. And only 1 animal has high level serum responses that would be expcted to be protective against these sensitive variants. How can you put this in context with breadth against circulating variants (often measured via the global panel) - what proportion of circulating variants are BG18-sensitive?

- For discussion: since all animals were "responders" based on the B cell sequencing results and elicitation of BG18-similar mAbs, yet the outcome of broad plasma bnAbs against circulating variants is not yet fully achieved, what can you propose as the next definition of "responder"?

- Also for discussion: while a definition of bnAb does not exist in the field (quite surprisingly!), the field will need to get more specific with a definition of biomarkers of successful germline-targeting mutations in order to move forward the best candidates - what is the definition of bnAb precursors and/or bnAbs that you can propose that will relate to eventual efficacy assessments?

thank you!
Sallie Permar

We thank the reviewers for their careful consideration of our manuscript and for their insightful questions and suggestions. We have made changes in response to each comment, and we believe these have improved the manuscript. Below, reviewer comments are in blue, and our responses are in black.

In the course of making our revisions, we have added new data and analyses to the manuscript, including in:

- Fig. 4i (antibody angle of approach in cryoEM structures)
- Fig. 5c (serum neutralization breadth vs. time – data added for 6 weeks post-boost 7)
- Fig. 5d (serum neutralization potency vs. time – data added for 6 weeks post-boost 7)
- Fig. 5e (fold increase in serum neutralization potency for 2 and 6 weeks post-boost 7 vs. 5 weeks post-boost 6)
- Fig. 5g (correlation between post-boost 4/5 mAb and post-boost 7 serum neutralization)
- ED Fig. 3b (negative stain EM images of trimer-NPs)
- ED Fig. 5a (antigen-specific ELISA)
- ED Fig. 7e,f (SPR data on mAbs isolated after early boosters)
- Supplementary Table 10 (neutralization data for 6-weeks post-boost 7)
- Supplementary Table 12 (N332 specificity test for serum)

We also added three Supplementary Data files:

1. HCDR3 sequences from post-boost 5 to illustrate overall diversity and conservation of key residues
2. All SPR K_D values
3. Immunogen amino acid sequences

Referees' comments:

Referee #1 (Remarks to the Author):

This manuscript titled “Vaccination elicits HIV broadly neutralizing antibodies in nonhuman primates” by Steichen et al. presents an comprehensive nonhuman primate (NHP) study aimed at advancing germline-targeting (GT) vaccination strategies for eliciting HIV broadly neutralizing antibodies (bnAbs). Given the longstanding difficulty of inducing bnAbs by vaccination and the limited success of prior GT approaches in clinical and preclinical settings, the topic is of high significance. The authors, who have been central drivers of the GT paradigm, report iterative refinements to their strategy and provide deep immunogenetic and structural characterization of elicited responses.

The work demonstrates that sequential immunization can expand BG18-class lineages, promote extensive somatic hypermutation (SHM), and yield antibodies that partially recapitulate structural features of mature BG18, including CDR H3 engagement of the V3 GDIR motif and interactions with the N332 glycan. These data indicate that aspects of the lineage can be guided in the intended direction. The structural analyses are a strength of the study and provide convincing evidence of epitope targeting by both type I and type III BG18 modes. The reported overall good serum breadth and especially in one animal with 84% relative to mature BG18 is notable.
A: Thank you.

However, the central question for a high-impact advance is not whether maturation can be nudged toward a known bnAb template under prolonged and highly engineered conditions, but whether this represents a scalable and generalizable vaccination solution. On this point, the manuscript is less compelling: The regimen requires ~120 weeks, seven boosts, and ~19 distinct immunogens or nanoparticles. Such complexity raises concerns about feasibility, scalability, and clinical translation. While proof-of-concept studies can tolerate complexity,

the manuscript does not sufficiently articulate a realistic path toward regimen simplification or define which components are indispensable versus exploratory.

A: We thank the reviewer for acknowledging that proof of concept can tolerate complexity, and for requesting that the manuscript address questions about how to reduce regimen complexity and progress toward scalable clinical translation. We are laser focused on trying to develop a real-world implementable HIV vaccine. We do believe that the regimen can be simplified in terms of how many different immunogens are needed in the sequence. Our long-term strategy for development of an HIV sequential GT vaccine that only requires between one to three vaccinations includes using pulsatile release technologies, most probably the atomic layer deposition strategy that is still in early development and that we are actively testing. Our long-term strategy also aims to leverage the mRNA-LNP vaccine platform both for its capacity for economical and rapid production of clinical products and for the ability to deliver membrane-anchored HIV trimers that were indicated to have an advantage over soluble trimers for elicitation of serum neutralizing antibodies in the HVTN302 phase 1 trial.

We have added the following underlined statements to the Discussion to address these points:

> "Our study has limitations. We note opportunities to improve on the regimens tested. The boosting regimens had 7 boosters; future work should focus on developing improved sequential regimens that reproducibly elicit the broadest responses in most vaccine recipients with the fewest number of immunizations. Our data point to potential opportunities for regimen simplification. For example, the relatively minor gains in SHM, affinity, and neutralization capacity at boost 5 suggest that that boost might be dispensable. The favorable impact of trimer-NPs on serum neutralization at boost 7 suggests that such NPs should be tested at earlier stages, and higher valency NPs should be developed and tested[60-62], potentially allowing for fewer vaccinations. The 34-week interval between boosts 1 and 2 was due to logistical reasons, thus regimens should be tested with a shorter interval. Cocktail compositions should be optimized, potentially reducing the number of components. Other technologies could assist translation of improved regimens: pulsatile release technologies might help reduce the number of vaccinations while still allowing for multiple sequential boosters[63], and mRNA vaccine platforms could be tested to facilitate scalable clinical manufacturing and explore potential advantages of membrane-anchored trimer delivery[51,52]. We also note that durability of these bnAb responses has not been explored, which will need to be understood and maximized for an effective vaccine."

We would also like to note that the ongoing HVTN144 phase 1 trial is testing N332-GT5 with SMNP, the same priming immunogen and adjuvant employed in this NHP study. [REDACTED] Given the stringency of the NHP model and the relative consistency of booster-driven maturation observed in the present study, we believe that the ability to boost BG18-class precursors to bnAb development will also likely translate to humans, especially with further optimized regimens.

Relatedly, although longitudinal B cell analyses were performed, the manuscript stops short of establishing causal links between specific immunogens (or immunogen combinations) and the acquisition of key SHMs associated with breadth. Without clearer attribution, it is difficult to discern whether breadth development is driven by rational immunogen design or by stochastic maturation under prolonged antigenic exposure. A more rigorous dissection of immunogen-to-mutation relationships would greatly strengthen the mechanistic

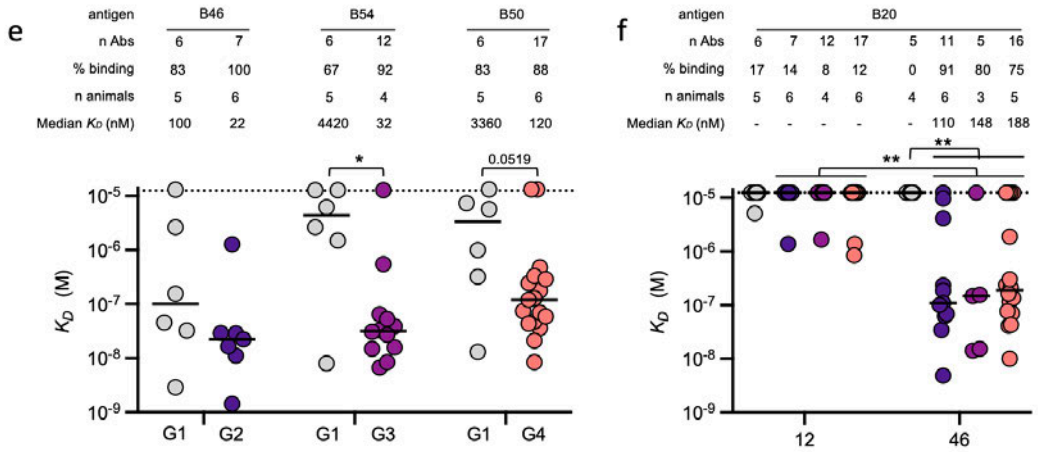
understanding.

A: We thank the review for this question about mechanism.

We think the germline-targeting sequential boosting strategy takes advantage of both rational immunogen design and stochastic maturation trajectories sampled by the immune system. The rational design part includes precise design of the priming immunogen for a broad set of precursors, followed by a series of successively more native-like immunogens designed using the principles described in the first Results section. The stochastic maturation aspect allows for highly diverse responses that traverse distinct maturation trajectories yet share key pre-specified germline-encoded features and converge on similar structural solutions to develop bnAbs. The combination of these aspects can be seen directly in the cryoEM structures of elicited BG18-class antibodies binding to Env trimers.

Regarding the related concept of stochastic maturation under prolonged antigen exposure, a critical component of our study was to include Group 1, which received repeated immunizations with N332-GT5. This provides a direct assessment of the contribution of stochastic maturation under prolonged antigenic exposure, through more than 72 weeks. Notably, this allows for the direct conclusions that N332-GT5 repeated immunization over time is generally insufficient for development of wild-type Env binding, as Group 1 had essentially background levels of ABCG Env cocktail binding at week 72, in contrast to Groups 2-4 ($P < 0.001$, Figure 2b, week 72). These differences were demonstrated at the mAb level, whereby the 6 BG18-class mAbs from Group 1 memory B cells at week 72 had no measurable affinity for WT Env trimers by SPR, in strong contrast to Groups 2-4 for which > 96% of the BG18-class mAbs had affinity for WT Env trimers ($P < 0.01$, Figure 3e, week 72). To underscore this point we have now added statistical comparisons of Group 1 to Groups 2-4 for weeks 46, 56, and 72 in Figure 3e. Note that group 1 received heterologous boosts at the subsequent time points.

To provide additional clarity on the contributions of the early boosters, we have now added SPR data for week 12 (post-boost 1) and week 46 (post-boost 2) to ED Fig 7e,f (shown below). The data in panel e illustrate that heterologous boosting at week 10 was more effective than homologous boosting for generating BG18-class responses with substantial affinity for the heterologous boosters at week 12. The boost 1 trimers are more native-like (have fewer GT mutations) compared to the priming immunogen, hence these affinity improvements represent direction affinity maturation toward bnAb development. The data in panel f show similar findings for the post-boost 2 antibodies.



ED fig 7**e,** SPR binding affinities (K_D values) for BG18 type I antibodies isolated from memory B cells two weeks post-boost 1 binding to boost 1 immunogens B46, B54, and B50. **f,** Binding K_Ds of BG18 type I antibodies isolated from memory B cells two weeks post-boost 1 or boost 2 were measured against the boost 2 immunogen.

In the manuscript we added the following statements to describe these newly added SPR data:

> "Post-boost 1 mAbs isolated after heterologous boosting in Groups 3 and 4 had higher median affinities for the respective fully glycosylated boost 1 trimers (B54 and B50) compared to mAbs isolated from Group 1 (Extended Data Fig. 7e), reflecting favorable directional affinity maturation. By contrast, mAbs from Group 2 and Group 1 had similar affinities for the Group 2 boost 1 trimer (B46). These differences were likely due to B54 and B50 both having substantially fewer germline-targeting mutations than B46 (Fig. 1c). The majority of post-boost 2 mAbs from Groups 2 to 4 had affinity for the boost 2 trimer (B20, possessing only 3 germline-targeting mutations), with median dissociation constant (K_D) values ranging from 110 to 188 nM, illustrating additional favorable affinity maturation (Extended Data Fig. 7f). None of five post-boost 2 mAbs from Group 1 showed affinity for B20 at the matched timepoint (Extended Data Fig. 7f), demonstrating the importance of the heterologous immunogens for the BG18-class mAbs affinity maturation toward binding of authentic HIV Env."

We note that these new data complement the SPR data already present in Fig. 3e that show the capacity for BG18-class BCRs to bind diverse WT trimers started to emerge after boost 2 in the heterologous boost groups and never emerged in Group 1 out to week 72. This data in Fig. 3e indicates that boost 2 was essential to continue the desired directional maturation of BG18-class precursors, as we note in the manuscript where we state:

> "Binding to diverse WT trimers was already detectable after boost 2 (Fig. 3e). Broad affinity for diverse WT trimers increased significantly in Groups 2 to 4 after boost 3 (Fig. 3e)."

We believe that all of the above points provide mechanistic insights on the maturation driven by the early boosters.

Regarding the reviewer's request for us to identify "key SHMs", we appreciate that this question was likely motivated by the identification of key heavy chain residues in VRC01-class responses in humans and mouse models. However, whereas VRC01-class responses all employ the VH1-2 gene, in the case of the HCDR3-dominant responses we are studying here, many different VH genes are observed, as we show in Fig. 3c,d. Thus, to identify key VH gene SHMs, one would need to identify multiple bnAbs for any one subtype of BG18-class type I responses that share the same VH gene, structurally characterize the interactions of these bnAbs with Env trimers, and then identify shared key residues contributing similar structural interactions. (and potentially similarly for the light chain V gene.) Furthermore, an important site of VH-gene SHM for BG18-class maturation occurs in the HCDR1 to facilitate interaction with the conformationally flexible N332 glycan. Germline forms of BG18-class antibodies do not exhibit productive interactions with the N332 glycan. We have previously shown that for the original BG18 VH4-4 gene, maturation from the inferred germline to mature bnAb requires a major conformational change in the HCDR1 accomplished by multiple SHMs (Steichen et al Science 2019). We hypothesize that different VH genes, starting from different germline HCDR1 sequences, will acquire different sets of SHMs and gain the capacity to bind the N332 glycan in different ways. This hypothesis is supported by the interactions of three different BG18-class Fabs with WT trimers shown in the cryoEM structures in Fig. 4. Each of the antibodies interacts with the N332 glycan using distinct VH interactions, including but not limited to distinct HCDR1 interactions.

We also observed high diversity in HCDR3 sequences, but we noted the conservation of key contact residues at nontemplated regions within the HCDR3, and we have added the following text to this effect:

> "BG18-class type I HCDR3s also exhibited high diversity in the V-D and D-J junctions outside the shared D gene, but we noted that 100% (296/296) of such post-boost 5 HCDR3s included key contact residues at nontemplated positions, including a Glu at two residues after the D gene and a Leu or Phe substitution at the last position of the D gene (Supplementary Data 1), as previously reported for BG18-

class type I responses after N332-GT5 priming in RMs[41]. Thus, sequential boosting maintained these key residues."

Overall, the study represents a substantial technical effort and provides conceptual progress for GT strategies. However, the translational implications appear limited in the current form, and the manuscript would be strengthened by a clearer demonstration of mechanistic guidance, regimen simplification logic, and generalizability.

A: We appreciate these critiques, which we have now addressed as described above.

Specific comments

1. Figure 2e: The color coding is insufficiently defined and should be clarified.

A: We thank the reviewers for bringing this to our attention and apologize for any confusion. We have updated the legend for figure 2 to further explain the color coding used in panel e.

2. Figure 4: The N332 glycan should be labeled. An additional glycan appears in panel B and should be identified. Quantitative RMSD comparisons between BG18 and elicited type I mAbs would strengthen structural similarity claims.

A: We have added labels to indicate the N332 glycan as well as the N392 glycan in panel B. We have also now calculated the angle of approach for the elicited antibodies. We have added 1D and 3D plots of the angles of approach for the elicited bnAbs and show they cluster around mature BG18 and are distinct in their approach angles compared to other V3glycan bnAbs (shown in the new Fig. 4i, pasted below for convenience.)

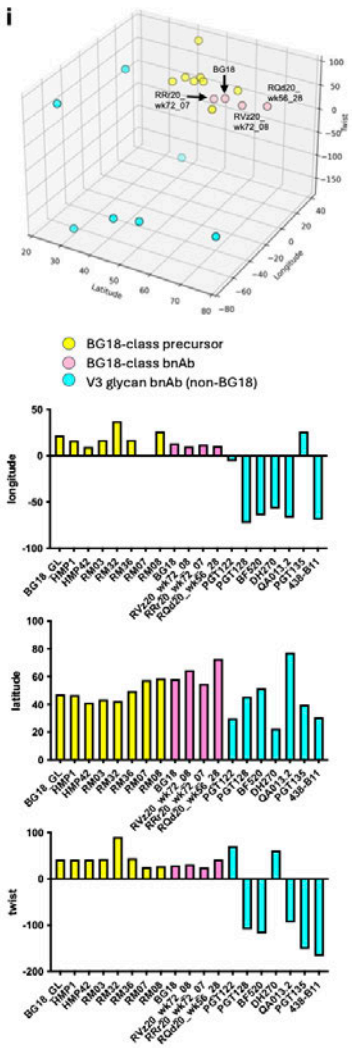


Fig 4 i, Angle of approach analysis for BG18-class antibodies, including the three new antibodies with structures solved in a-h, compared to other V3-glycan bnAbs.

3. The apparent increase in breadth from post-B3 to post-B4/5 in G2 is striking. Is there direct evidence linking this change to the B46 boost, or could this reflect stochastic maturation?

A: The differences in neutralization breadth of mAbs between G2, G3, and G4 post-B4/5 appear to reflect natural variation between animals, as there is not a consistently improved affinity of G2 mAbs for WT Env trimers compared to G3 or G4 by SPR after boost 2, boost 3, or boost 4 (Figure 3e).

4. The strong response in NHP RVz20 and presence of non-type I "BG18-like" antibodies warrant deeper analysis, as they may reveal alternative pathways than the BG18 GT.

A: We thank the reviewer for this query. As we explained in the text, the three BG18-class type III lineages were discovered from an exhaustive analysis of all long HCDR3 expanded lineages at multiple timepoints:

> "The full set of epitope-specific B cells with long HCDR3s (≥20 aa) generated by immunization were screened for expanded lineages present at multiple timepoints, including after the final boost. All but three such lineages were BG18-class type I… The three non-type I lineages were also non-type II. All three lineages used the same V_H gene and had similar HCDR3 features, and two of the three lineages used the same $V_{K/L}$ gene (Extended Data Fig. 9b). As the BCRs had already been characterized as V3-glycan epitope-specific by B cell sorting, we hypothesized these related lineages should each be designated BG18 type III clonal lineages."

We devoted Fig. 4e-i and ED Fig. 9 to an intensive analysis of the type III responses in terms of genetics, lineage expansion, frequencies, affinities, and structural properties, and then we documented the neutralization properties in Fig. 6. We do not believe that the type III responses represent "alternative pathways" to BG18 GT. We previously hypothesized that N332-GT5 may elicit antibodies that share some, but not all features of BG18 (Steichen et al Science 2019). In that study we isolated, from human naïve B cells, BG18 type II antibodies with similar binding approach and LCs with HCDR3s that contact the GDIR. In a 2024 paper (Steichen et al. Science 2024) we identified antibodies elicited in NHPs with the same binding orientation as BG18 but essentially no sequence similarity except for a long HCDR3 – we described those antibodies as BG18 type III. The big question at that time was whether type III antibodies could be matured into bnAbs. In this study we have now identified examples of BG18 type III antibodies that can develop into bnAbs. We believe this is all part of the same germline-targeting strategy. [REDACTED] Therefore, we think these results will be highly relevant to human studies.

Referee #2 (Remarks to the Author):

This manuscript reports three important observations. First, broadly HIV-neutralizing antibody responses against this highly mutable virus can be elicited by vaccination in non-human primates, even when precursor B-cell frequencies targeting the desired epitope are extremely low—indeed, lower than those estimated in humans—and without the use of transgenic mice engineered to express "ideal" B-cell receptors. Second, the immunogens used to achieve this outcome were not selected empirically but were rationally designed to present specific, predefined structural features. Finally, the study demonstrates that once naïve B cells are activated by a germline-targeting priming immunogen and enter germinal center reactions, they can be repeatedly boosted to further drive affinity maturation through subsequent immunizations—an outcome that has been questioned by others in the field.

A: Thank you for your assessment. We agree with all of those points.

Major queries

1. Following Boost 1, all animals in Groups 2–4 exhibited high frequencies of B cells specific for the boost immunogens (Fig. 1), whereas no animals in Group 1 (immunized only with the prime immunogen, N332-GT5) displayed boost-binding memory B cells. In light of this, can the authors clarify why priming with N332-GT5 is required for the observed responses?

A: We thank the reviewer for this question. N332-GT5 is required to prime BG18-class precursor B cells. The naïve BG18-class B cells are exceptionally rare, and the affinity of N332-GT5 for those cells is key to priming. We reported in Steichen et al. Science 2024 that immunization of NHPs with the wt trimer (no germline-targeting mutations) BG505 MD39 failed to induce any detectable BG18-class responses.

We have previously shown that N332-GT5 primed BG18-class B cells in NHPs do gain affinity for B46 prior to boost, via germinal center affinity maturation (Steichen et al., Science 2024, Figure 3F), and we illustrate that again in this manuscript for all three boost 1 candidates (ED Fig. 1b; shows data re-measured from the prior experiment). Most importantly, the SPR data we collected from Group 1 BG18-class BCRs at week 12 (two weeks post-GT5 second immunization; data now included in ED fig. 7e) shows that these antibodies bind with substantial affinity for B46 (median Kd=100nM).

Nevertheless, we did observe by flow cytometry that Group 1 cells had less boost binding compared to Groups 2-4 at week 12, as shown in Fig. 2b. There is very low boost binding observed in the flow-cytometry data for Group 1 at week 12 after receiving a 2nd immunization with N332-GT5, but there are a small number of boost binding epitope-specific cells present. In addition, as shown in Fig. 2d, some of the BG18 type I cells from Group 1 do have detectable binding to B46 using our bioinformatic determination.

We agree with the reviewer that the main text could be clearer, and we have revised the main text as follows, deleted text indicated by strikethrough:

> "Notably, all animals in Groups 2 to 4 possessed high frequencies of boost immunogen¬ binding ("boost-binding") memory B cells at week 12, in contrast to Group 1 ~~where no animals demonstrated appreciable boost binding memory cells~~ (Group 1 vs. Groups 2 to 4 combined, $P < 0.0001$; Fig. 2b). Thus, early after boost 1, for all three heterologous boost 1 immunogens, there was substantial evidence of recall and maturation of epitope-specific B cells capable of accommodating both of the shielding V1 glycans."

As noted in our response to reviewer 1, we have now included SPR data on post-boost 1 mAbs binding to boost 1 candidates (ED Fig. 7e) and post-boost 2 mAbs binding to boost 2 (ED Fig. 7f).

2. BG18 type I and III mAbs isolated after boosts 3, 4, and 5 (i.e., prior to immunization with the AMP cocktail and the nanoparticle trimer cocktail at boosts 6 and 7, respectively) exhibited some degree of neutralization breadth. However, serum cross-neutralizing activity was not detected at these time points (Supplementary Table

5) and only became clearly apparent after the seventh immunization. The authors should therefore either (i) provide comparative data on somatic mutation levels, affinities, and structural features of BG18 mAbs isolated after boosts 6 and 7 (vs mAbs isolated and fully characterized by the previous boost immunizations), or (ii) reconsider emphasizing the contribution of the final two immunizations.

A: As we describe in the manuscript, our B cell sorting analyses were quite exhaustive up to the post-boost 5 timepoint, and we identified BG18-class expanded lineages per animal persisting over time up to that point. Therefore, we hypothesized that the serum neutralization at post-boost 7 was derived from the expanded lineages we documented. We further hypothesized that evidence for this linkage might be found in similar neutralization specificity, in terms of which viruses were neutralized, between the memory mAb neutralization at post-boost 4 or 5 and the serum neutralization at post-boost 7. As we now describe in the text, we carried out a correlation analysis on the neutralization data, now shown in a new figure (Fig. 5g, pasted below) and that analysis did indeed indicate a strong correlation in the specificity, both in each individual animal and in the overall combined data.

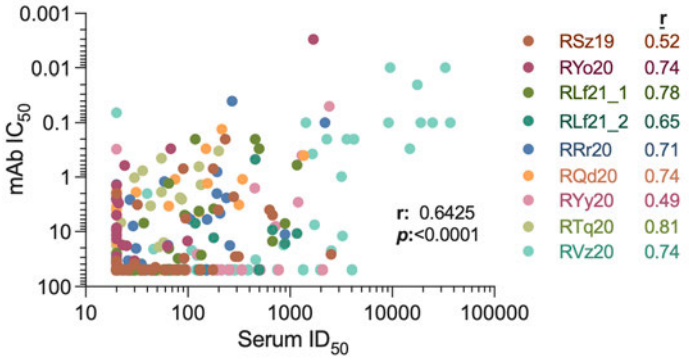


Fig 5g, For each virus of the 63-virus panel the IC_{50} value of the pB4/5 mAb was plotted against the pB7 serum ID_{50} of the animal that mAb was isolated from. The overall r and p-values are listed on the graph and the per animal r values are listed in the legend. The p-values for all per animal correlations are <0.0001.

As we state in the text:

> "Additionally, for the 8 animals with both mAbs and serum tested on the 63-PSV panel (Fig. 5b-f, Supplementary Table 11), all mAb IC_{50} values were plotted against serum ID_{50} values for each individual animal (Fig. 5g). BG18-class mAb and serum neutralization values correlated strongly across all data points ($P < 0.0001$, $r = 0.6425$), and on an individual animal level (Fig. 5g). These correlations indicated that the serum neutralization was mediated by the predominant BG18-class lineages we have characterized in each animal. "

We hope the reviewer will agree that this analysis provides compelling evidence that the serum neutralization post-boost 7 is likely derived from the BG18 lineages described in our post-boost 4/5 B cell sorting and other analyses.

Immediately before those new sentences in the text, we added other new information on the specificity of the post-boost 7 serum neutralization activity. We showed that the serum neut activity is strongly dependent on the presence of the N332 glycan, which illustrates that this activity is directed to the V3-glycan epitope:

> "N332 glycan-variant (±N332 glycan) PSVs were produced for each of 4 viral isolates that were neutralized by the majority of the 8 animals. Serum neutralization from 6 of 8 animals was substantially reduced in the absence of the N332 glycan (Supplementary table 12), indicating that the serum neutralization was predominantly directed to the V3 glycan epitope, per the vaccine design concept."

This finding is consistent with our inference that the post-boost 7 serum neut activity is mediated by BG18-class antibodies.

Minor points

• Please clarify what is meant by "BG18-sensitive viruses" in the analysis of monoclonal antibody and serum neutralization. Are these viruses specifically sensitive to BG18 mAb alone, or are they also sensitive to other known anti-N332 broadly neutralizing antibodies?

A: In the manuscript we use "BG18-sensitive" to indicate the 63 viruses that are neutralized by BG18 in the 118 virus global panel, as first reported by Freund et al. STM 2017 (The first BG18 paper). In the Freund study, the highest concentration of BG18 tested was 30 µg/mL. In the present manuscript, we have included BG18 as a positive control in the neutralization assays (e.g. Supplementary Tables 6 to 11), and our data for which viruses are neutralized by BG18 are generally consistent with the findings from Freund. In general BG18 neutralizes an overlapping subset of viruses with other N332-dependent bnAbs including PGT121 and DH270.

To clarify this definition in the manuscript, in the first sentence of the section "Vaccine-elicited memory B cell bnAbs", where neutralization of vaccine-induced responses is first discussed, we revised the wording as indicated by underlining:

> "BG18 type I mAbs (n=58) isolated post-boost 3, 4, and 5 were tested for neutralization on a panel of 9 to 11 HIV pseudoviruses (PSVs) <u>neutralized by BG18 (BG18-sensitive)</u>."

• On pg 3, line 23-24, the authors discuss that 'immunological cell biological constraints' may prevent B cells that become activated by the germline-targeting immunogen to undergo further stimulation/maturation by the necessary subsequent booster immunizations. Can they provide relevant references that discussed this possibility ? It seems to be a critical point for what the authrs try to do here.

A: Thank you for this query. We have added relevant references with various points of view.

Referee #3 (Remarks to the Author):

A. Summary: this is an elegant paper demonstrating the feasibility and improving partial success of germline-targeting HIV Env immunization towards the achievement of induction of a mature sinlge bnAb specificity clonal lineages that become dominant in serum in NHPs

B. Originality/significance: this manuscript represents moving the germline-targeted approach further towards translation with achieving evolution of initiated V3-specific bnAb precursor lineages, defining a new pathway for V3 glycan bnAb evolution (type III), maintenance and expansion of those lineages towards increasing breadth with shaping boosts, dominance of vaccine-elicited and evolved B cell lineages in serum.
A: Thank you for your assessment, we agree with all of those points.

C/F. Data & methodolgy and suggestions
1. LN and PBMC B cell analysis should be linked between animals - is there a definition of "responder vs non-responder" that can be proposed based on the flow and/or serum outcome? Are there biomarkers of a responder phenotype?
A: Thank you for this comment. PBMC samples are the primary source of B cell data, and all conclusions regarding frequency of BG18-class cells as well as persistent lineages are based on B cell flow cytometry, sorting, and scBCRseq from PBMC samples. LN samples are a valuable additional source of data. For BG18-class lineages with clonal family members detected in any LN sample, the LN B_{GC} sequences are included in the lineage trees and can be seen in Figure 2f as well as in EDF 7a. More than 2,000 sequences from the RQd20 lineage were detected in LN FNAs. The full number of BG18-class sequences found in FNAs will be included in the graphical source data as per Nature policy.

Regarding "*is there a definition of "responder vs non-responder" that can be proposed based on the flow and/or serum outcome,"* in this study 100% of animals were responders, making BG18-class B cell responses (Figure 2e). BG18-class responses were also primed in 100% of animals in our two earlier studies of priming using N332-GT5 plus SMNP in escalating dose immunization (Steichen et al, *Science* 2024, Madden et al, *Immunity* 2025).

A biomarker predicting the animals that develop broadly neutralizing antibodies at the monoclonal and serum level has not been identified, other than serum neutralization itself. Because 100% of animals made BG18-class responses (Figure 2e), and diverse lineages can mature to neutralization (Fig. 3C-D), we believe the immunization regimen can be optimized such that mature bnAbs develop in 100% of animals.

2. plasma IgG binding strength to vaccine antigens should be reported and related as a potential biomarker of responder status
A: We have added ELISA data showing antigen-specific IgG responses to every immunogen in ED Fig 5a. This is referred to in the text at the beginning of the section **Sequential boosting of bnAb-class precursor B cells:** "Serum analysis indicated that all immunogens induced antigen-specific IgG responses (Extended Data Fig. 5a)." One key observation is that of the four trimers used in the trimer-NP immunization; 0860, 1026, and AD8 have stronger antigen-specific responses before administering the trimer-NP at boost 7. The fourth trimer TRO.11 is higher responses post-boost 7 for two groups (G3+G4) but lower in G2 compared to pre-trimer-NP immunization (pB5). Therefore, the ELISA does not appear to predict the serum neutralization as serum breadth and potency was substantially better post-trimer-NP immunization. Because serum binding analyses cannot unambiguously detect the presence or absence of bnAb precursors, we have not attempted to use serum ELISA measurements as a biomarker for bnAb development in the memory compartment. Overall, we prioritized B cell sorting and BCR sequencing to characterize the development of bnAbs triggered by sequential vaccination.

3. the recognition of evolved/more mature Envs is reported for groups 2-4 after boost 2, but there is not a

detailed description of which revertered Env mutations were required for this binding. Can an analysis of plasma and/or mAb binding to reverted Env mutants be included to determine which mutations are the most critical for continuing to engage these lineages?

A: The choice of mutations to revert was rational. BG18-class Abs and epitope-specific competitors induced by the priming immunogen are dependent on the V1 loop but only the BG18-class antibodies can maintain affinity when we restore glycans to the V1 loop. This was a key part of the boost immunogen design to deflect the competitors and is described in the methods "Design of boost immunogens". For boost 2 we left GT mutations that interact directly with the tip of the HCDR3 because analysis of SHM showed that essentially all BG18 type I Abs were acquiring the mature amino acids in the HCDR3 after the prime and boost 1, indicating that the GT mutations were not driving the SHM off-track. This is also described in the methods section with two sentences: "We elected to use a stabilized trimer based on the isolate SF162P3 with three germline-targeting mutations from N332-GT5 designed to interact with the tip of the HCDR3 for BG18 precursors. We hypothesized that these three mutations, which were also present in the prime and boost-1 candidates, would contribute to the boost-2 affinity for post-boost 1 BG18-class BCRs." Then boost 3 was a WT trimer.

4. in the bnAb lineage analysis, can the number of improbable mutations that were incorporated be reported? this would be valuable in determining how challenging the AID barriers will be towards eliciting these lineages. Was there one boost immunogen that was better at incorporation of the key improbable mutations?

A: Improbable mutation calculations are unlikely to be particularly informative for BG18-class bnAb development, for multiple reasons. First, unlike VRC01-class antibodies, BG18 binds in a heavily HCDR3 dependent manner, while VRC01-class bnAb development has a large number of mutations at germline encoded positions in HCDR1 and 2, and in LCDR1. Thus, discreet definable improbable mutations are inherently less numerous in heavily HCDR3-dependent bnAbs like BG18 compared to VRC01-class bnAbs. One representation of BG18-class diversity is the demonstration of diverse vaccine-elicited BG18-class HC V gene usage, and LC V gene usage, among BG18-class type I B cells capable of achieving neutralization (see Fig. 3C-D).

Second, the program that is used for determining improbable mutations, ARMADILLO (Wiehe et al, 2018), relies on counting improbable mutations based on an inferred UCA and because of the inherent uncertainty of HCDR3 UCA inference ARMADILLO masks the HCDR3 sequence and does not include HCDR3 in any improbable mutation reporting. Given that the HCDR3 is the most important region of SHM for BG18-class bnAb development, this renders improbable mutation calculations very limited in value.

5. groups 2 vs 3 vs 4 were not compared to help move along the question of which initial shaping immunogen is most successful, this should be included as a separate analysis/figure

A: Strikingly, all three heterologous Boost 1 candidates successfully achieved boosting and further affinity maturation of BG18-class B cells. Notably, all three boost 1 immunogens reintroduced the V1 glycans, and BG18-class Abs were able to make the jump to all three boost 1 immunogens, even though the boost 1 immunogens had different affinities and surfaces. The differences in neutralization breadth of mAbs between G2, G3, and G4 post-B4/5 appear to reflect natural variation between animals, as there is not a consistently improved affinity of G2 mAbs for WT Env trimers compared to G3 and G4 by SPR after boost 2, boost 3, or boost 4 (Figure 3E). Thus all three candidates are worthy of future exploration.

6. the achieved geomean ID50 across animals should be related to the IC80 that was calculated from the AMP trial

A: We thank the reviewer for this query. We note that Gilbert et al. Nat. Med 2022 ED Fig 5 shows that prevention efficacy in the AMP trials and in NHP passive transfer studies was associated with PT50 (the association in AMP with PT80 was shown in main Gilbert et al. Fig. 3). In the following underlined sentences in the Discussion, both of which cite Gilbert et al. (ref. 5), we have added a citation to Huang et al. *Hum Vaccin Immunother* **18**, 1908030 (2022) (ref. 55) where it is explicitly shown that PT50 is an accurate approximation to ID50. The Gilbert et al. and Huang et al. citations together illustrate that the PT50 protection thresholds stated

in Gilbert et al ED Fig. 5 are the best available estimates for vaccine-induced ID50 requirements for protection. (And the Pauthner et al. citation [ref. 54] shows that an example of vaccine-induced nAbs with an ID50 of 1:500 conferring 90% protection in NHPs.) Based on the data in Gilbert et al. ED Fig. 5b, which associate %protection levels with different PT50 (~ID50) levels for passive mAb studies in NHPs, we have modified our statement in the first underlined sentence to say that 75 to 90% protection would be expected for the best animal, and we have included the second underlined sentence to note that four other animals had titers expected to confer 50% protection.

> "The potency of those serum bnAbs, a geomean ID_{50} neutralization titer of 481, is expected to confer 75 to 90% protection against HIV (or SHIV) for diverse global isolates [5,54,55]. This serum neutralization breadth and potency elicited by vaccination is comparable to that observed in elite neutralizers from natural infection[56], raising the prospect that further optimized vaccines may elicit bnAbs superior to the best bnAbs occurring after HIV infection. Four other animals had geomean neutralization titers expected to confer 50% protection against clinical isolates[5,54,55], supporting further vaccine optimization."

7.The paper would benefit from inclusion of how the predominate B cell lineages (% antigen-specific B cells, type I/III, number of clones/B cells, etc) relate to serum neuts achieved in each animal - is there a relationship? An early B cell biomarker will benefit the field

A: The reviewer raises a good point. As we now describe in the text, we carried out a correlation analysis on the neutralization data, now shown in a new figure (Fig. 5g, pasted below) and that analysis did indeed indicate a strong correlation in the specificity, both in each individual animal and in the overall combined data.

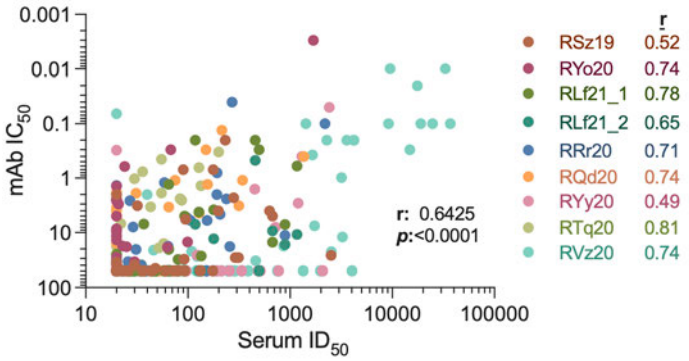


Fig 5g, For each virus of the 63-virus panel the IC_{50} value of the pB4/5 mAb was plotted against the pB7 serum ID_{50} of the animal that mAb was isolated from. The overall r and p-values are listed on the graph and the per animal r values are listed in the legend. The p-values for all per animal correlations are <0.0001.

As we now state in the text:

> "Additionally, for the 8 animals with both mAbs and serum tested on the 63-PSV panel (Fig. 5b-f, Supplementary Table 11), all mAb IC_{50} values were plotted against serum ID_{50} values for each individual animal (Fig. 5g). BG18-class mAb and serum neutralization values correlated strongly across all data points ($P < 0.0001$, r = 0.6425), and on an individual animal level (Fig. 5g). These correlations indicated that the serum neutralization was mediated by the predominant BG18-class lineages we have characterized in each animal. "

We hope the reviewer will agree that this analysis provides compelling evidence that the serum neutralization post-boost 7 is likely derived from the BG18 lineages described in our post-boost 4/5 B cell sorting and other analyses.

Immediately before those new sentences in the text, we added other new information on the specificity of the post-boost 7 serum neutralization activity. We showed that the serum neut activity is strongly dependent on the presence of the N332 glycan, which illustrates that this activity is directed to the V3-glycan epitope:

> "N332 glycan-variant (±N332 glycan) PSVs were produced for each of 4 viral isolates that were neutralized by the majority of the 8 animals. Serum neutralization from 6 of 8 animals was substantially reduced in the absence of the N332 glycan (Supplementary table 12), indicating that the serum neutralization was predominantly directed to the V3 glycan epitope, per the vaccine design concept."

This finding is consistent with our inference that the post-boost 7 serum neut activity is mediated by BG18-class antibodies.

Regarding the second point, about a biomarker, as we discussed above, we have yet to identify an early biomarker other than neutralization itself, but we believe that we can achieve serum neutralization in 100% of animals with regimen optimizations.

9. a definition of "bnAb" for serum and mAbs needs to be included

A: We note that manuscripts in the field rarely if ever indicate precise criteria to define what constitutes a bnAb. We use the term bnAb for this proof of concept study for the following reasons.

The mAbs elicited and characterized in this study show key bnAb features including neutralization of diverse isolates from diverse clades, with breadth reaching up to 46% of the most potent V3-glycan bnAb known on a global PSV panel, and with potencies that are relevant to vaccine protection (≤ 10 μg/mL). They bind to a known conserved bnAb epitope (V3-glycan) using genetic and structural features that mimic the template bnAb. Further, they are dependent on the relatively conserved N332 glycan like other bnAbs that recogenize the V3-glycan epitope.

For the 8 animals with both mAbs and serum neutralization tested on the 63-PSV panel, the serum from 5/8 animals showed greater or similar neutralization breadth compared to the individual BG18-class mAbs from those same animals, with five animals showing geomean ID_{50} values that are expected to provide 50% or more protection against diverse HIV isolates. Our newly added analysis of the serum specificity including correlation of mAb vs serum neutralization of glycan-variant (+/- N332) PSV indicate that serum neutralization is mediated by the BG18-class lineages characterized from memory B cells.

Overall, the vaccine-induced BG18-class mAbs and serum described in this study have breadth and potency in the same range as the known bnAbs CH235 (CD4bs), VRC29.03 (V3-glycan), PCT64-24E (V2-apex), and BF520.1 (V3-glycan) (see Sok and Burton, Nature Immunology 2018, Fig. 1). We think all of these properties together justify using the term "bnAb" to describe the vaccine-elicited BG18-class mAbs and serum in these animals.

D. Statistic applied should include more descriptive statistics

1. Animal numbers should be more clearly stated. Every % reported should include absolute numerators and denominators and every median should have an associated range.

A: We thank the reviewer for raising these points. We have now added a column to Fig. 1b to indicate that all groups are N=6. We will provide all the requested values in the graphical source data in accordance with Nature policy.

2. approach to multiple comparisons should be added

A: The 'Statistical Analysis' section of the methods has been updated to include a statement regarding multiple comparisons.

E conclusions

1. there are a few animals highlighted in the conclusion (such as line 20, page 10) for serum neut breadth - but it is unclear how it is defined (how many/which viruses, how is it compared to mAb concentration, etc)

A: We thank the reviewer for identifying an opportunity for clarification. In the mAb and serum neutralization sections of the manuscript, we have extensively used the 63 virus panel that includes all of the viruses neutralized by BG18 on the commonly used 118 virus global panel. In the course of our revisions we believe the references to this panel and the callouts to relevant figures and supplementary tables have become more clear.

This panel is first described in the section on "Vaccine-elicited memory B cell bnAbs" in the following sentence: "The best 12 mAbs from 12 distinct lineages (9 type I and 3 type III) and 11 different animals were tested on a larger panel of PSVs (n=63) representing all BG18-sensitive PSVs in a widely used global panel of clinical HIV isolates[39]."

In the next section on "Vaccine-elicited serum bnAbs", we again introduce this panel by saying: "For the 8 animals producing the best BG18 bnAbs in the memory compartment post-boost 4 or 5 (Fig. 5b), we tested serum neutralization on the BG18-sensitive 63 PSV panel after boosts 4, 5, 6, and 7 (Fig. 5c-e, and Supplementary Tables 6-11)." That same paragraph later states: "At two weeks post-boost 7, all 8 animals exhibited bnAb activity, with breadth ranging from 16% to a remarkable 84% (Fig. 5c,e, and Supplementary Table 9,11)."

The part of the Discussion section about serum breadth referred to by the reviewer (starting on line 20, page 10 in the original submission) reads as: "The immune responses in the best performing animal provide a remarkable new benchmark. The serum of this animal had 84% neutralization breadth relative to mature BG18, which corresponds to approximately 52% breadth overall on standard HIV global neutralization panels."

We believe this sentence is clear in referring to breadth relative to mature BG18. Also, the prior statements quoted above from the "Vaccine-elicited serum bnAbs" section now refer the reader to Fig. 5c,e, and Supplementary Table 9,11 where the supp tables show the exact viruses neutralized and the summary stats at the bottom of the tables indicate the animal with 84% breadth on the 63-PSV panel.

We also believe that the relevant figure legend was not clear. For clarity we have updated the figure legend for Fig. 5f to state the following:

> "Serum neutralization breadth versus geomean ID_{50} on the ~~same~~ <u>63 PSV</u> BG18-sensitive panel post-boost 7."

Regarding mAb concentration: The BG18-class polyclonal vaccine response is very different from a passively administered bnAb. We are not aware of established methods for measuring concentrations of polyclonal mAbs from specific lineages in serum.

2. plasma cells are mentioned in the discussion (line 40, page 10), but not evaluated here - differentiate the future research needs from data presented

A: The observed increase in serum neutralization is the direct evidence of plasma cell formation that is being referenced here. We have removed the word "fully" from the sentence below to avoid confusion.

> "Lastly, immunization with multivalent nanoparticles demonstrated that extensively affinity-matured vaccine-induced bnAb memory B cells were ~~fully~~ capable of differentiating into plasma cells producing substantial levels of serum neutralizing antibodies."

Additionally, we agree that the biology of plasma cell, and particularly long-lived plasma cell, formation is an important future research need and have highlighted that in the following sentence from the Discussion.

> "We also note that durability of these bnAb responses has not been explored, which will need to be understood and maximized for an effective vaccine."

H. Clarity

1. The abstract should include which vaccine platform is being used (adjuvanted Env SOSIP subunit)

A: We have add the underlined words to the abstract as follows: “Here, we report an adjuvanted protein germline-targeting vaccine tested in outbred nonhuman primates that successfully generated bnAb-class memory B cells and sera capable of neutralizing diverse HIV clinical isolates.”

2. when reporting % B cells bound to various Envs, this is reported as "higher Env binding" (such as line 35, page 5), but I believe this is referring to larger % B cell (or higher MFI)? the units need to be specific for each of the reported "higher/lower" comparison
A: Thank you for bringing this to our attention. We have now updated that line in the text to read “higher frequencies of WT Env trimer-binding” to make it clearer we are referring to frequency of binding cells.

3. describe the "sentinel" PSV in more detail (how it was selected/representative) - page 9, line 40
A: The sentinel PSV 6811 is a tier-2 virus that was potently neutralized by all 12 BG18 type I/III mAbs that were tested on the large 63 PSV panel. Thus, we reasoned that if BG18 type I/III mAbs are present in the serum they should at a minimum be able to neutralize 6811.

We have updated the sentence in question by adding the underlined phrases, as follows:

> “Serum neutralization was evaluated after boosts 4, 5, 6, and 7 on a sentinel heterologous tier 2 PSV (6811.v7.c18) that was neutralized by the best 12 BG18-class mAbs described above (Supplementary Table 4).”

Referee #4 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Response to reviews round 2

We would like to thank the reviewers for their comments and questions, which we believe have helped us improve the manuscript further. Our point-by-point responses are below. Reviewer comments are in blue, and our responses are in black.

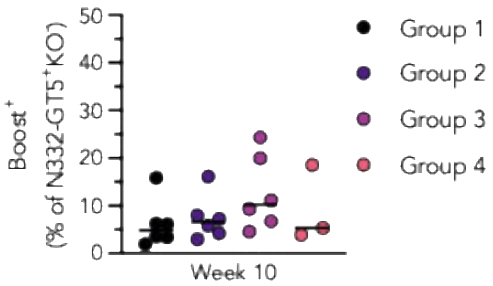
Ref. 2

I remain very positive about this study. The revised manuscript and rebuttal address most of my previous comments, with two important exceptions.

The first outstanding issue concerns the frequency of B cells elicited by the priming immunogen (germline-targeting) that can bind the first heterologous boost immunogens (B46 [Group 2], B54 [Group 3], and B50 [Group 4]). The authors refer to prior work (Extended Data Fig. 1) showing that N332-GT5–immunized macaques developed antibodies capable of binding these booster immunogens, and they also discuss responses at week 12 following the first heterologous boost in the present study. However, what is needed here is direct evidence from the current study on the frequency of B cells, after the prime, that recognize these booster immunogens. For each group, the authors should report: (i) the number of animals with such detectable responses, and (ii) the frequency of B cells recognizing the first boost immunogen in each animal.

A:

We have now generated these data. See the graph below, which is now ED Fig. 6a. All the immunized animals had detectable boost-binding B cells at the pre-boost (week 10) timepoint. The frequency of those cells ranged from 1 to 25% in individual animals, consistent with the observed memory B cell boosting to Boost 1 previously reported in the manuscript (**Fig. 2b**). We call out the new ED Fig 6a in the following sentence on page 4: "Pre-boost, all 24 animals had low levels of boost-binding B cells (Extended Data Fig. 6a)."

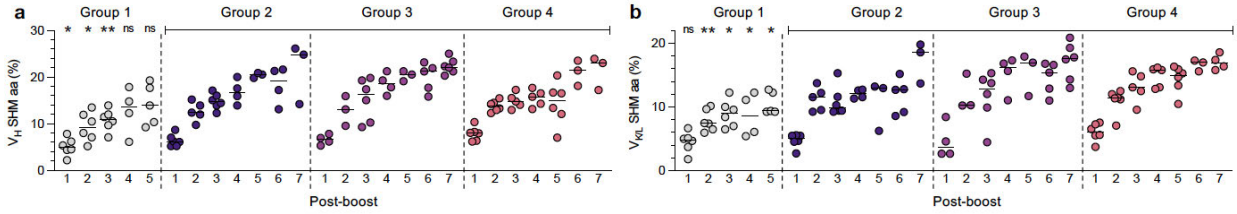


The second outstanding issue relates to the absence of data on broadly neutralizing antibodies isolated after the 6th and 7th immunizations, when serum neutralization breadth appears to peak. While the evolution of serum neutralizing activities throughout the immunization regimen is well documented, monoclonal antibodies are only characterized up to the 5th boost. This raises an important question: why do sera collected after the 6th and 7th immunizations exhibit greater breadth? Do antibodies acquire additional beneficial

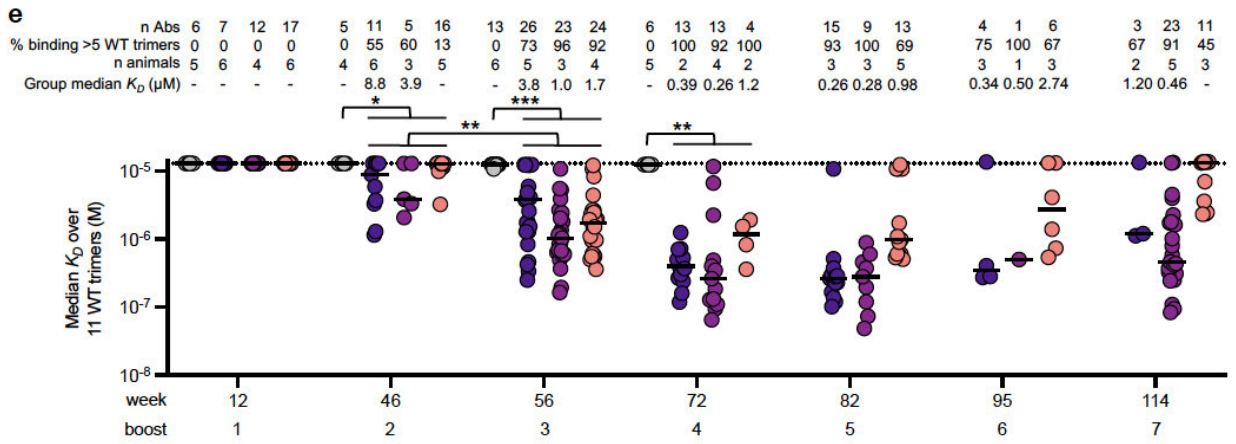
mutations at these later time points, or is the increased breadth primarily due to higher antibody titers?

Notably, even after the 3rd boost, isolated monoclonal antibodies from Groups 2–4 display measurable breadth (particularly in Group 4), whereas those from Group 1 do not (Supplementary Table 2, p. 9).

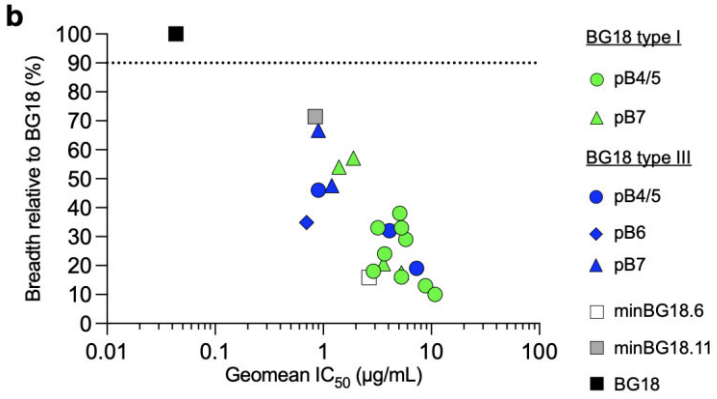
A: We now have these data. We completed memory B cell sorting and BCR sequencing for the post-boost 6 and 7 timepoints, and we have updated Fig. 2 a-g and Fig 3a showing SHM accumulating in BG18-class memory B cells out to post-boost 7. For example, the new Fig 3a-b is below.



We produced 73 new mAbs from memory B cells post-boost 6 and post-boost 7 and measured their affinities to 11 WT trimers. Overall, there is little gain in median affinity across the 11 WT trimers post-boost 6 and post-boost 7 compared to post-boost 4 and post-boost 5 as shown in the SPR figure below, which is the new Fig. 3e.



However, for the highest affinity mAbs, we evaluated neutralization activity on the 63-virus BG18-sensitive fraction of the global panel and found that four of the mAbs have greater neutralizing breadth than any mAb we previously isolated (see the updated mAb neutralization figure below, which is the new Fig. 5b). The post-boost 7 mAbs include the two broadest and most potent BG18-class Type 1 mAbs and the broadest BG18-class Type III mAb in our study. The best post-boost 7 mAb has 67% breadth relative to the original BG18 bnAb. These data provide examples of how boost 6 and 7 increased neutralization breadth and potency in some cases. Therefore, we think the improvement in serum breadth after boost 7 is likely due to a combination of (i) a subset of mAbs having improved neutralization breadth and potency and (ii) higher titers of bnAbs in the serum.

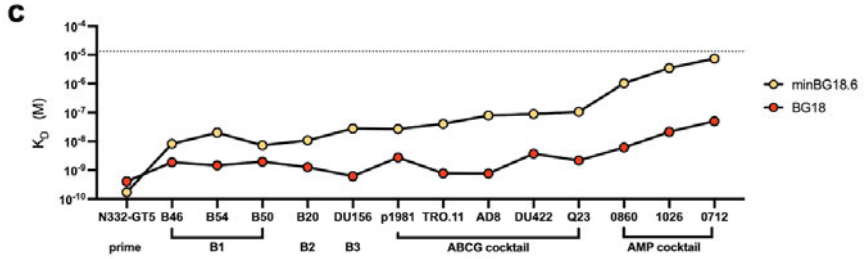


We also note that, with this new post-boost 6/7 data, fully 2/3 of the animals have defined bnAb+ memory B cells, and in group 3, 83% of animals generated bnAb+ memory B cells. These high positivities underscore the reproducible nature of the vaccine bnAb responses.

Minor points:

1. Extended Data Fig. 1c: Please clarify which curve corresponds to min BG18.6 and which to BG18.

A: We have fixed this key, as shown below.



2. Page 4: The manuscript states that the immunogen sequence exposes "12 fully glycosylated BG18 epitopes." I count 10: 1 in Boost 3, 4 in Boosts 4 and 5, and 4 in Boost 6. The immunogens used in Boost 7 appear to be the same as those in Boosts 4–6.

A: Boosts 1 and 2 also have fully glycosylated BG18 epitopes, which explains why 12 is correct. This has been clarified in the text.

Ref. 3

Thank you to the reviewers for the extensive work to refine the manuscript, I feel the translatability and strategy for next steps are clearer now in the edited version.

A: Thank you for this positive assessment. We are also grateful to the reviewers for clarifying many aspects of the manuscript.

I have a couple of final comments:

- I don't think it is accurate to say that BG18 has few improbable mutations than VRC01

and those in the HCDR3 can't be detected via ARMADILLO, as this has been done. Thus, it would be helpful to assess the probability of the mutations that are in the clonal lineages of vaccine-elicited mAbs characterized in this paper and which are required for the neutralization breadth, moving along the definition of "responder" and how to measure the maturation of responses
See: https://armadillo.dhvi.duke.edu/bnab/BG18/heavy

A: We apologize that our original response was unclear. ARMADILLO can visualize HCDR3 if a UCA is provided, but while ARMADILLO visualizes the HCDR3 on the ARMADILLO webpage for BG18 linked above, mutations in the HCDR3 are not counted towards the improbable mutation total. As the authors of ARMADILLO state in their paper "Due to inconsistent uncertainty levels in the inference of the CDRH3 region…only mutations outside the CDR3 are used." (Wiehe et al. CH&M 2018). This is because it is much harder to identify a high confidence HCDR3 than HCDR1/2, particularly for BCRs with long HCDR3s with N-additions, and we know that BG18-class antibodies must have both a long HCDR3 and N-additions.

Equally important, the contrast we were making to VRC01 is that VRC01-class antibodies all use the same HC V gene, VH1-2; but for BG18-class antibodies we see a wide range of HC V-genes being used that can proceed to bnAbs in this vaccine study (**Fig 3c**). This is consistent with the bnAb features of BG18-class antibodies being heavily HCDR3-dependent, and much less dependent on the V gene, consistent with our original proposal for BG18-class germline-targeting vaccine design that included examples of BG18-class bnAbs utilizing different V genes (Steichen et al., Science 2019).

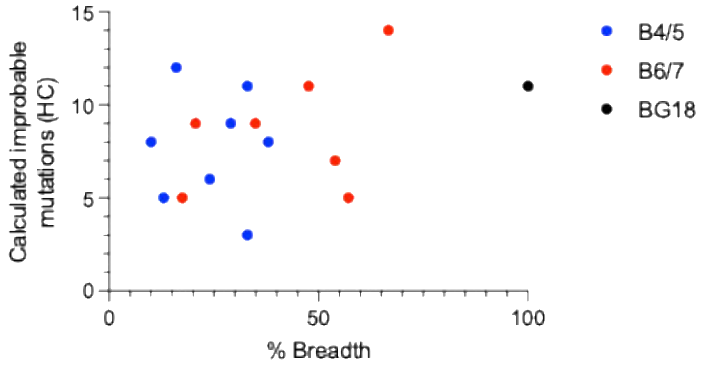
As a result, it is challenging to definitively define discrete improbable mutations in BG18-class antibodies, due to the combination of their diverse V genes and the inherent uncertainties for estimating UCAs for long HCDR3s.

Nevertheless, we have now run multiple of our vaccine-elicited BG18-class bnAbs through ARMADILLO, shown in the table at right. The number of calculated "improbable" mutations in these vaccine-elicited BG18-class bnAbs is in line with other known bnAbs (excluding the HCDR3, as noted above) including multiple that have more improbable mutations than BG18.

	Calculated improbable mutations (HC)
VRC01	11
BG18	11
PGT121	7
RKb21_wk72_46	12
RSz19_wk72_71	9
RYo20_wk82_15	3
RLf21_wk82_30	6
RLf21_wk82_54	5
RRr20_wk72_07	8
REg21_wk82_31	8
RQd20_wk82_59	11
RVz20_w114_42	14
RVz20_w114_23	11
RRr20_w114_47	5
RRr20_w114_34	7
RKm20_w114_40	9
RLf21_w114_08	5
RVj20_wk95_17	9

Regarding "*which are required for the neutralization breadth*", as stated above, it is difficult to define discrete mutations in BG18-class antibodies that are required for breadth due to the diversity of genes used by these lineages, and the inherent nature of HCDR3-dependent binding modalities. However, for completeness, we plotted the number of HC improbable mutations versus the breadth of each mAb on the 63-virus panel and observed no correlation between breadth and number of improbable mutations (see graph below). In fact, we have four vaccine-induced mAbs that have equal or greater numbers of "improbable" mutations compared to fully mature BG18 but are not yet achieving the same breadth. We believe this confirms our emphasis on other features of the vaccine-elicited

BG18-class B cells as being more important for the observed development of vaccine-generated bnAb+ B_{mem} and reinforces our statements above about the importance of the HCDR3.



- The relationship of the elicited neutralization to levels of protection established in NHPs and the VRC01 clinical trial is helpful. However, the serum neutralization analysis seems to have only been performed on BG18-sensitive variants - which should be acknowledged.
A: We agree that this is an important issue. We also believe that we have acknowledged and explained this in the text and figures; however, in agreement with the reviewer, we have found several places where this can be further improved. We provide a summary of places where we think changes are needed or not needed below. We thank the reviewer for raising this issue.

We define viruses that are neutralized by BG18 as "BG18-sensitive" when the term is first used on page 9, lines 2-3: "BG18 type I mAbs (n=58) isolated post-boost 3, 4, and 5 were tested for neutralization on a panel of 9 to 11 HIV pseudoviruses (PSVs) neutralized by BG18 (BG18-sensitive)." We then explain that the 63-virus panel used subsequently in the manuscript represents all viruses in the global panel that are neutralized by BG18 (page 9, lines 6-10): "Based on these results and SPR affinities, the best 19 mAbs from post-boosts 4-7, representing 12 distinct lineages (9 type I and 3 type III) from 12 animals were tested on a larger panel of PSVs (n=63) representing all BG18-sensitive PSVs in a widely used global panel of clinical HIV isolates[37]." We have modified those two sentences as shown below (underline showing new text) to explain that these refer to neutralization at any concentration up to 25 µg/ml, to clarify that we do not use the term "BG18-sensitive" to indicate the "easiest" viruses to neutralize by BG18 but rather every virus that is neutralized by BG18 at a physiologically relevant concentration. We also explicitly note that the global panel contains N=118 viruses.

Revised sentence: "BG18 type I mAbs (n=58) isolated post-boost 3, 4, and 5 were tested for neutralization on a panel of 9 to 11 HIV pseudoviruses (PSVs) neutralized by BG18 <u>at any concentration up to 25 µg/ml</u> (BG18-sensitive)."

Revised sentence: "Based on these results and SPR affinities, the best 19 mAbs from post-boosts 4-7, representing 12 distinct lineages (9 type I and 3 type III) from 12 animals were

tested on a larger panel of PSVs (n=63) representing all BG18-sensitive PSVs in a widely used global panel of n=118 clinical HIV isolates[37]."

In that same paragraph on mAb neutralization, we then refer to different vaccine-induced mAbs achieving breadth relative to BG18, as for example on page 9, lines 12-14 (underline added for emphasis): "The 13 RM BG18 type I mAbs neutralized a median of 15 of the 63 PSVs, with geomean IC_{50} values of 1.4 to 10.8 µg/ml, achieving neutralization breadth up to 57% relative to BG18 (Fig. 5b, Supplementary Table 4 and 11)." And on page 9, lines 12-14: "The six BG18 type III mAbs neutralized a median of 26 PSVs, with geomean IC_{50}s of 0.7 to 7.3 µg/ml, individually reaching breadth up to 67% relative to BG18 (Fig. 5b and Supplementary Table 4)." We close that paragraph by explaining: "In aggregate, mAbs derived from the vaccinated animals exhibited 90% neutralization breadth compared to the mature BG18 bnAb (Fig. 5b and Supplementary Table 4 and 11), demonstrating the potential for germline-targeting vaccines to induce bnAb-class responses with remarkable breadth." We believe these statements are clear.

Furthermore, the y-axis labels for Figs. 5b,c, and f all say "Breadth relative to BG18 (%)", which we believe is clear. Those figure panels refer to both mAb neutralization (5b) and serum neutralization (5c,f).

However, in the section of the text on serum neutralization, we failed to call out in the text itself that the breadth was relative to BG18. To clarify this issue, we have revised these sentences to state (page 9, lines 35-37, underline showing new text): "At two weeks post-boost 7, all 8 animals exhibited bnAb activity, with breadth relative to BG18 ranging from 16% to a remarkable 84% (Fig. 5c,e, and Supplementary Table 9,11). Mean neutralization breadth of the 8 animals was 41% relative to BG18."

We also identified a helpful clarification in Supplementary Tables 4 and 6-11, which report neutralization assays on all 63 PSVs originally reported to be sensitive to BG18 from the 118 PSV standard global panel. One of the summary statistics at the bottom of each table was simply "breadth %". Each of these tables includes a BG18 positive control, for which breadth is shown as 100% or nearly 100% (in some assays BG18 fails to neutralize a few PSVs at the max concentration tested), so we believe the meaning of breadth was pretty clear. Nevertheless, we have changed the label from "breadth %" to "breadth rel. to BG18 (%)", to make the meaning more clear and aligned with Fig. 5b,c,f.

In the Discussion, we return to this issue and provide an explicit example for how breadth relative to BG18 relates to breadth on the full global panel: (page 10, lines 31-33): "The immune responses in the best performing animal provide a remarkable new benchmark. The serum of this animal had 84% neutralization breadth relative to mature BG18, which corresponds to approximately 52% breadth overall on standard HIV global neutralization panels." We believe this statement is clear.

And only 1 animal has high level serum responses that would be expected to be protective against these sensitive variants.

A: We have explained in the Discussion that several animals had neutralization titers expected to confer various degrees of protection. On page 10, lines 34-35, we noted that the best animal had "a geomean ID_{50} neutralization titer of 481, … expected to confer 75 to 90% protection against HIV (or SHIV) for diverse global isolates". Later in the same paragraph (p. 10, lines 38-40) we explain that "Four other animals had geomean neutralization titers expected to confer ~50% protection against clinical isolates[5,54,55], supporting further vaccine optimization."

How can you put this in context with breadth against circulating variants (often measured via the global panel) - what proportion of circulating variants are BG18-sensitive?
A: As noted above, in the Discussion, we provide an explicit example for how breadth relative to BG18 relates to breadth on the full global panel: (page 10, lines 31-33): "The immune responses in the best performing animal provide a remarkable new benchmark. The serum of this animal had 84% neutralization breadth relative to mature BG18, which corresponds to approximately 52% breadth overall on standard HIV global neutralization panels."

- For discussion: since all animals were "responders" based on the B cell sequencing results and elicitation of BG18-similar mAbs, yet the outcome of broad plasma bnAbs against circulating variants is not yet fully achieved, what can you propose as the next definition of "responder"?
A: We agree with the reviewer that this is worth discussion, space permitting. As noted, all animals (100%) were responders based on priming of BG18-class B cells (**Fig 2C**). 83% of animals developed WT HIV cross-binding BG18-class memory B cells at at least one time point. 67% of animals developed bnAb BG18-class memory B cells and thus were bnAb+ memory B cell responders. 44% of animals developed serum bnAbs thereafter (**Fig 5**). One could thus define a series of responder definitions of increasing difficulty or increasing time/boosters : 1) priming responders, 2) bnAb+ memory B cell responders, and 3) serum bnAb responders 4) serum bnAbs with potency needed for protection.

The context is also important. As we note in the manuscript (p.10, lines 27-29), "Achieving these results in NHPs, which have substantially rarer BG18-class bnAb precursors than humans, suggests that similar or better results may be attainable in humans." Thus, the high stringency of the animal model used gives us confidence that in humans the priming responders, bnAb+ memory B cell responders, and serum bnAb responders have a good chance of being higher responder frequencies than we observed in this first-of-its-kind BG18-class bnAb vaccine in NHPs.

- Also for discussion: while a definition of bnAb does not exist in the field (quite surprisingly!), the field will need to get more specific with a definition of biomarkers of successful germline-targeting mutations in order to move forward the best candidates - what is the definition of bnAb precursors and/or bnAbs that you can propose that will relate to eventual efficacy assessments?
A: We agree that this is an important topic. In our view, the correlates of protection identified in the AMP trials establish the clearest benchmarks linking bnAb breadth and potency to protective efficacy. For example, the potency of a bnAb serum response can be described by geomean ID50 or ID80 values that are expected to confer 90%, 75% or 50%

protection, and this number must be coupled with a description of breadth, ideally measured on a standardized global panel. These metrics will provide an objective comparator between vaccine candidates. We do refer to these metrics from the AMP trial analysis in the Discussion.