

Vrc01-class Precursor Frequency in an African Population is Influenced by Ighv1-2 Allelic Composition and Not Malaria Exposure

Michelle K. Muthui

mmuthui@kemri-wellcome.org

KEMRI-Wellcome Trust Research Programme

Caleb Kibet

International AIDS Vaccine Initiative (IAVI)

Charity Muriuki

KAVI –Institute of Clinical Research, University of Nairobi

Yiakon Sein

KEMRI-Wellcome Trust Research Programme

Sharon Owuor

KEMRI-Wellcome Trust Research Programme

Roselyn Kinoti

KEMRI-Wellcome Trust Research Programme

Wilfrida Ogonda

KEMRI-Wellcome Trust Research Programme

Michael Muteti

KEMRI-Wellcome Trust Research Programme

Oleksandr Kalyuzhniy

IAVI Neutralizing Antibody Center, The Scripps Research Institute

John Kimotho

KEMRI-Wellcome Trust Research Programme

Makobu Kimani

KEMRI-Wellcome Trust Research Programme

Brenda Kamau

KEMRI-Wellcome Trust Research Programme

David C. Owuor

KEMRI-Wellcome Trust Research Programme

Troy Sincomb

IAVI Neutralizing Antibody Center, The Scripps Research Institute

Kelly Ramko

KEMRI-Wellcome Trust Research Programme

Judie Magura

Africa Health Research Institute

Alfred Muia

KEMRI Centre for Clinical Research

Lynn Namwoso

KEMRI Centre for Clinical Research

John Orimbo

Ahero Clinical Trials Unit, CREATES/KEMRI–CCR

Christopher A. Cottrell

IAVI Neutralizing Antibody Center, The Scripps Research Institute

Omu Anzala

KAVI –Institute of Clinical Research, University of Nairobi

Bernhards Ogutu

KEMRI Centre for Clinical Research

Elise Landais

IAVI Neutralizing Antibody Center, The Scripps Research Institute

Linda Murungi

International AIDS Vaccine Initiative (IAVI)

Devin Sok

IAVI Neutralizing Antibody Center, The Scripps Research Institute

William R. Schief

Moderna Inc.

Thumbi Ndungu

Africa Health Research Institute

Daniel Muema

KAVI –Institute of Clinical Research, University of Nairobi

Eunice W. Nduati

KEMRI-Wellcome Trust Research Programme

Article

Keywords:

Posted Date: June 24th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9732671/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: Competing interest reported. W.R.S. is an inventor on patents filed by Scripps and IAVI on the eOD-GT8 monomer and 60mer immunogens (patent number 11248027, “Engineered outer domain (eOD) of HIV gp120 and mutants thereof”). W.R.S. is an employee and shareholder of Moderna, Inc. All other authors declare no financial or non-financial competing interests.

**VRC01-CLASS PRECURSOR FREQUENCY IN AN AFRICAN POPULATION IS
INFLUENCED BY IGHV1-2 ALLELIC COMPOSITION AND NOT MALARIA
EXPOSURE**

Authors and affiliations:

Michelle K. Muthui¹, Caleb Kibet^{2*}, Charity Muriuki^{3*}, Yiakon Sein¹, Sharon Owuor¹, Roselyn Kinoti¹, Wilfrida Ogonda¹, Michael Muteti¹, Oleksandr Kalyuzhniy^{4,5,6}, John Kimotho¹, Makobu Kimani¹, Brenda Kamau¹, David C. Owuor¹, Troy Sincomb^{4,5,6}, Kelly Ramko¹, Judie Magura⁷, Alfred Muia⁸, Lynn Namwoso⁸, John Orimbo⁹, Christopher A. Cottrell^{4,5,6}, Omu Anzala³, Bernhards Ogutu^{8,9}, Elise Landais^{4,5,6}, Linda Murungi^{2§}, Devin Sok^{4,5,6+}, William R. Schief^{4,5,6,10}, Thumbi Ndungu^{7,11,12,13}, Daniel Muema^{3,7*}, Eunice W. Nduati^{1,14*}

¹ Kenya Medical Research Institute (KEMRI)-Wellcome Trust Research Programme, Kilifi, Kenya

² International AIDS Vaccine Initiative (IAVI), Nairobi, Kenya

³ KAVI –Institute of Clinical Research, University of Nairobi, Nairobi, Kenya

⁴ IAVI Neutralizing Antibody Centre, The Scripps Research Institute, La Jolla, California, USA

⁵ Department of Immunology and Microbial Science, The Scripps Research Institute, La Jolla, California, USA

⁶ Centre for HIV/AIDS Vaccine Development, The Scripps Research Institute, La Jolla, California, USA

⁷ Africa Health Research Institute (AHRI), Durban, South Africa

⁸ KEMRI Centre for Clinical Research (KEMRI-CCR), Nairobi, Kenya

⁹ Ahero Clinical Trials Unit, CREATES/KEMRI-CCR, Kisumu, Kenya

¹⁰ Moderna, Cambridge, MA, USA

¹¹ HIV Pathogenesis Programme, The Doris Duke Medical Research Institute, University of KwaZulu-Natal, Durban, South Africa

¹² Ragon Institute of Mass General Brigham, Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA

¹³ Institute of Infection, Immunity and Transplantation, University College London, London, UK

¹⁴ Centre for Tropical Medicine and Global Health, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, UK

USA

*These authors contributed equally to the work. [§]Present address: Science for Africa Foundation, Nairobi, Kenya; ⁺present address: Stripe, 354 Oyster Point Blvd, San Francisco, CA 94080, USA.

Corresponding author: Michelle K. Muthui, mmuthui@kemri-wellcome.org

ABSTRACT

Recent innovations in vaccine design have reignited optimism about the feasibility of developing a safe and effective HIV vaccine. For instance, germline-targeting (GT) immunogens have been designed to selectively expand naïve B cell precursors with broadly neutralizing antibodies (bnAbs)-associated germline-encoded features, thereby priming them for structure-guided maturation. The frequencies and affinities of these precursors in the populations targeted for

vaccination are critical to the success of GT vaccine strategies, but limited information is currently available across different geographies and demographics. We set up a study to characterise CD4 binding site-targeting VRC01-class bnAb precursor frequencies in 60 healthy Kenyan adults with varied levels of prior exposure to malaria. Naïve VRC01-class bnAb precursors were detected using the germline-targeting eOD-GT8 immunogen. The overall frequency of these precursors was similar to that reported in North American populations, with IGHV1-2 allelic composition rather than prior long-term exposure to malaria impacting the precursor frequency estimates. Our findings also revealed four new allelic variants of the IGHV1-2 gene that are not currently included in the immunogenetic reference database. Notably, a novel IGHV1-2 allele (IGHV1-2*06_S5416) demonstrated the ability to engage eOD-GT8, unlike the reference *06 allele. Overall, our results support the development of GT vaccines to induce VRC01-class bnAbs in this population and reiterate the need for participant genotyping in GT vaccine trials, particularly in African populations where genetic diversity is high yet understudied.

INTRODUCTION

A key goal for HIV vaccine development is to elicit broadly neutralizing antibodies (bnAbs) as these would potentially protect individuals from a wide range of infecting HIV clades. Studies in macaques¹⁻⁴, humanised mice⁵ and humans⁶ provide proof-of-principle for bnAb protection against infection with sensitive viruses. However, only a small proportion of HIV-infected individuals naturally generate bnAbs⁷⁻⁹. To date, no HIV vaccine candidate has elicited antibodies with sufficient neutralisation breadth and tires by vaccination to prevent infection^{10,11}. Consequently, novel reverse vaccinology approaches that utilise insights from structural biology and human immunology¹², have improved our ability to elicit antibodies of sufficient breadth.

One of these approaches, the germline targeting (GT) approach, seeks to elicit bnAbs by using priming immunogens engineered to bind to diverse precursors for structurally-defined classes of bnAbs, that can then be shepherded towards breadth using successive boosts with immunogens designed to increasingly resemble the native HIV trimer¹³⁻¹⁸. The priming immunogen first needs to overcome the hurdle of activating rare naïve precursor B cells encoding features within their germline B cell receptors (BCRs) that will allow the recognition of epitopes targeted by bnAbs¹⁹. These precursors will also need to have high affinity for the priming immunogen to outcompete non-precursor B cells²⁰. The presence of high affinity bnAb precursors at sufficient frequencies in most individuals across different populations will therefore be important for the success of this approach, and for ensuring that immunogens are relevant for the target population.

The most advanced series of GT immunogens seek to elicit VRC01-class bnAbs that target the envelope CD4-binding site (CD4bs) and share common genetic and structural features²¹⁻²³. These bnAbs are also among the broadest isolated to date²¹⁻²⁴. VRC01-class bnAbs are defined by their use of immunoglobulin heavy chain variable gene 1-2 (IGHV1-2) of the *02 or *04 alleles^{19,25-28} paired with a light chain (LC) with a short complementarity determining region 3 loop (CDR3) of 5 amino acids (5aa) with a specific motif²⁹. VRC01-class bnAb precursor B cells have been characterised in North American populations,^{24,27,30} and were recently described in one study that included African populations³¹ which suggested higher or equivalent frequencies in African populations compared to the US population³¹.

Allelic variation has been shown to impact VRC01-class bnAb precursor frequencies with higher precursor frequencies and VRC01-class responses post eOD-GT8 priming observed in individuals homozygous for IGHV1-2*02^{27,28,32}. African populations have high genetic heterogeneity^{33,34} which may impact the allelic composition of their naive repertoires and potentially impact the response to GT immunogens. In addition, individuals in Africa are exposed to a myriad of infectious pathogens that differ from those in Western and other populations, which may modify the naïve B cell repertoire composition³⁵, alter immune regulation^{36–38} and/ or modulate vaccine responsiveness as previously suggested^{39–44}. Comprehensive knowledge of precursor frequencies and properties, and of factors that may contribute to differences between populations such as environmental exposures, will inform the potential use of germline targeting immunogens in this target population that also bears the greatest burden of HIV.

Malaria is endemic to multiple regions in Africa and has been shown to cause significant immune dysregulation in affected populations^{45,46}. The heterogeneous levels of exposure to malaria on the continent therefore provide a unique natural experiment to decipher the impact of chronic exposure to a pathogen on the B cell compartment. We thus sought to determine the VRC01-class bnAb precursor frequencies and characteristics in a cohort of individuals living in different geographical regions within Kenya and with varying levels of malaria exposure. In this study we recruited individuals from high, low and intermediate malaria endemic settings. Using the engineered outer domain of gp120 version 8 (eOD-GT8), which targets VRC01-class precursors, as well as other CD4bs bnAb precursors^{30,31}, we sorted epitope-specific naïve B cells and performed high-throughput droplet-based single-cell BCR sequencing analysis to identify VRC01-class bnAb precursors, characterise their genetic and functional properties, and estimate their frequencies. We then incorporated participant germline allele inference analysis using IgDiscover⁴⁷ to determine the participants IGHV1-2 genotypes.

We report similar proportions of VRC01-class bnAb precursors between the high and low malaria endemic regions. Additionally, the overall precursor frequency was similar to previously reported estimates from North American populations. We also confirm the association between participant IGHV1-2 genotype and precursor frequency. Importantly, we identified a novel allele (IGHV1-2*06 S5416) with the ability to engage eOD-GT8, unlike the reference *06 allele. These data confirm the potential use of GT immunogens that aim to elicit VRC01-class of bnAbs in this African population.

RESULTS

Participant demographic and clinical characteristics:

To investigate the potential impact of historical chronic exposure to pathogens on the frequency and characteristics of VRC01-class bnAb precursors, we recruited 60 healthy adults who met the inclusion criteria for collection of a large blood draw, from four regions in Kenya with varied malaria transmission intensities (**Fig. 1a** and **Supplementary Tables 1 and 2**). The median age of the participants was 26 yrs with a majority being male (68%). Anti-*P. falciparum* schizont IgG antibody levels measured by ELISA showed no difference between Junju and Ahero but were significantly lower in Ngerenya and Nairobi ($p < 0.0001$) (**Fig. 1b**) confirming that the selection

criteria correctly captured individuals with differing levels of historic malaria exposure. Based on these results, participants were grouped into high malaria endemicity (Ahero and Junju) and lower malaria endemicity (Nairobi and Ngerenya) ($p < 0.0001$) for assessment of the impact of malaria exposure on VRC01-class precursor frequencies.

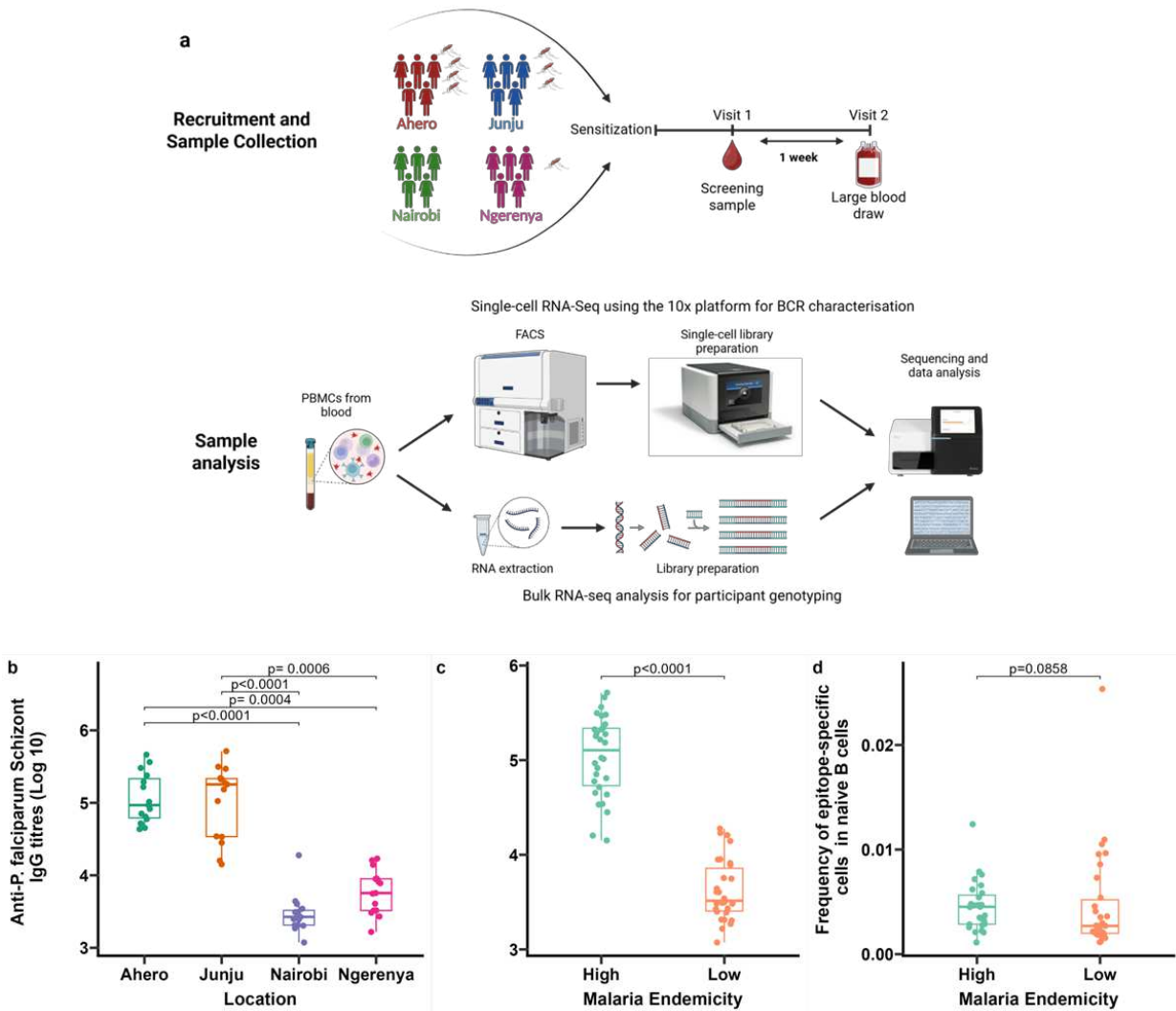


Figure 1: Study design, confirmation of malaria exposure level by *anti-P. falciparum* schizont ELISA and identification of epitope-specific naïve B cells. **a** Schematic showing the regions from which donors were recruited into the study, and the processing of large blood draw samples from samples to identify VRC01-class precursor B cells and genotype participant. **b** Magnitude of *anti P. falciparum* IgG titres per donor stratified by region of recruitment, and **c** by malaria endemicity classification: high – Junju and Ahero, and low – Nairobi and Ngerenya. Statistical significance was determined using the Kruskal-Wallis test with pot hoc analysis done using Dunn’s test with Bonferroni correction applied for multiple comparison adjustment. **d** Frequencies of eOD-GT8-KO11/eOD-GT8⁺⁺ IgD⁺ B cells for each donor stratified by malaria endemicity. Statistical significance was determined using a two-sided Wilcoxon rank sum test. Each dot represents an individual donor, with the thick lines showing the median value, the boxes indicating the 25% and 75% quartiles and the ‘whiskers’ highlighting the minimum and maximum values.

Identification of CD4bs-specific cells

Fifty-seven of the sixty participants had sufficient PBMCs for antigen-specific cell sorting of bnAb precursors. From these, an average of 30.87 million B cells (interquartile range [IQR]: 18.0 – 50.06 million B cells) were obtained after B cell enrichment per participant (**Supplementary Tables 3 and 4**). Using fluorescence-activated cell sorting (FACS) and the gating strategy illustrated in (**Supplementary Fig. 1**) we defined CD4 binding site (CD4bs) epitope-specific naïve B cells as viable CD3⁻CD14⁻CD16⁻CD56⁻/CD19⁺/IgD⁺/IgG⁻/eOD-GT8-KO11⁻/eOD-GT8⁺⁺ cells. Here, eOD-GT8⁺⁺ denotes binding to eOD-GT8 tetramer probes labelled with two different fluorophores. The eOD-GT8 knockout 11 tetramer probe (eOD-GT8-KO11) contains mutations in the CD4bs that abrogate VRC01-class B cell binding^{30,48}. The overall median frequency of epitope-specific cells was 0.0035% (IQR: 0.0022% – 0.0054%) across all donors. There was a trend towards higher frequencies in participants from the high malaria endemicity settings compared to low endemicity settings (0.0045% vs. 0.0027%), although this difference did not reach statistical significance (**Fig. 1d**, $p = 0.086$).

The VRC01-class sequences identified show classical signatures of VRC01-class bnAbs

Using the gating criteria described above, a total of 14,534 epitope-specific cells (eOD-GT8-KO11⁻/eOD-GT8⁺⁺) IgD⁺ B cells were sorted (median: 178 cells per participant, interquartile range (IQR) 92 – 321 cells, **Supplementary Table 3**) and subjected to paired B cell receptor (BCR) characterisation to identify VRC01-class signatures. We obtained sequence data from 54/57 donors and proceeded to assess these sequences for VRC01-class sequence signatures. Consistent with previous reports the IGHV1-2 gene was enriched among CD4bs-specific naïve B cells, comprising 16.97% and 25.23% of sequences from the high and low malaria endemicity regions, respectively (overall frequency: 21.1%) compared to its frequency in frequency in general B cell repertoires (4.26% in this cohort with the reported average in literature being 3 – 4%^{31,49,50}). IGHV1-2 was the dominant gene family observed in both endemicity settings (**Fig. 2a** first panel, and **Supplementary Fig 2**). Similar to previous studies, we observed enrichment of 5aa-LCDR3s in LCs paired with a IGHV1-2 HC among the CD4bs-specific sequences in both low (39.8%) and high (32.4%) malaria endemicity settings. Majority of the light chains were derived from the kappa chain and used various kappa and lambda V genes (**Fig. 2a** middle panel and last panel, and **Supplementary Fig 3a and 3b**).

The VRC01-class signature is minimally characterized by usage of a IGHV1-2*02 or IGHV1-2*04 paired with a 5aa-LCDR3. Using this definition, we detected VRC01-class precursors in approximately 59% of 51 participants with permissive genotypes (30/51) of whom 16 were from high malaria endemicity settings (53%). To minimize unreliable estimates of precursor frequency in samples with very few sequences, we limited the estimation of VRC01-class precursor frequency to 28 participants (52%) that had 10 or more paired HC-LC sequences recovered. Twenty two of the 28 participants had VRC01-class precursors, and for this subset we estimated their VRC01 class precursor frequencies as a proportion of VRC01-class precursors in naïve B cells³¹. Overall, we observed a median precursor frequency of 1 precursor in 345,557 naïve B cells (IQR: 1 in 147,461 – 444,370 naïve B cells). Precursor frequencies were not significantly different between participants from the low compared to the high malaria endemicity settings (median: 1 in 236,200, IQR 1 in 146,092 – 365,896 compared to median: 1 in 402,886, IQR 1 in 231,680 –

542,894 naïve B cells, $p = 0.14$, **Fig. 2b**). When we compared the overall precursor frequencies from our study with estimates previous studies in North American populations, we found no significant differences in the overall precursor frequencies ($p = 0.11$, **Supplementary Fig c**).

In addition to precursor frequencies, we investigated sequence-based metrics associated with VRC01-class BCRs such as light chain V-gene usage, light chain complementarity determining region 3 (LCDR3) motif, and presence of a tryptophan residue at position -5 of the HCDR3 (Trp₁₀₃₋₅). Most of the VRC01-class naïve B cells (~90%) encoded LC genes used by known VRC01-class bnAbs (V κ 3-20, V κ 1-33, V κ 3-15, V κ 1-5, V λ 2-14 and V λ 2-23) or LC genes previously identified in VRC01-class precursors^{27,30} (**Fig. 2c**). Approximately 27.1% of the VRC01 class V κ LCDR3s encoded a glutamate or glutamine (E/Q) at the fourth residue of the 5-aa L-CDR3 (QQYE/QX, **Fig. 2d**), though none possessed the full QQYEF motif found in mature VRC01 class bnAbs. Interestingly, there was 2-fold enrichment of sequences encoding an E/Q in sequences from the low endemicity setting (36.4% vs. 17.8%), however this difference did not reach statistical significance ($p = 0.09$, **Supplementary Fig 3d**). For the HCDR3, we observed enrichment of Trp₁₀₃₋₅ in 40.8% of VRC01-class precursors (**Fig 3d** consistent with a previous study⁵¹, with no differences in enrichment by malaria endemicity ($p = 0.92$, **Supplementary Fig 3e**). Enrichment of J2, that encodes a germline Trp₁₀₃₋₅, was similarly observed (**Supplementary Fig 3f**). The overall HCDR3 length was shorter among the VRC01-class precursors when we compared to control naïve B cells (**Supplementary Fig 3g**).

Noting that many of the key antibody sequence features that mediate the interaction between the CD4bs on gp120 and VRC01-class bnAbs are already encoded in the germline⁵², we generated a ‘bnAb characteristic score’ to assess the number of key features that the isolated VRC01-class precursors share with VRC01-class bnAbs (**Supplementary Data 1**). The bnAb features evaluated were: (i) LCDR3 motif (where each residue was scored based on matching to the QQYEF bnAb motif); (ii) Trp₁₀₃₋₅ in the HCDR3; (iii) having a short LCDR1 (6 – 7 aa); and (iv) having an HCDR3 length within that of VRC01-class bnAbs (11–18 aa). Out of a maximum possible score of eight, the median score was five, with no significant differences observed between the high and low malaria endemicity settings (**Supplementary Fig 3h**)

Taken together, the results highlight that the VRC01-class sequences identified shared several characteristics with VRC01-class bnAbs. Additionally, there was no significant difference in precursor frequencies by malaria endemicity.

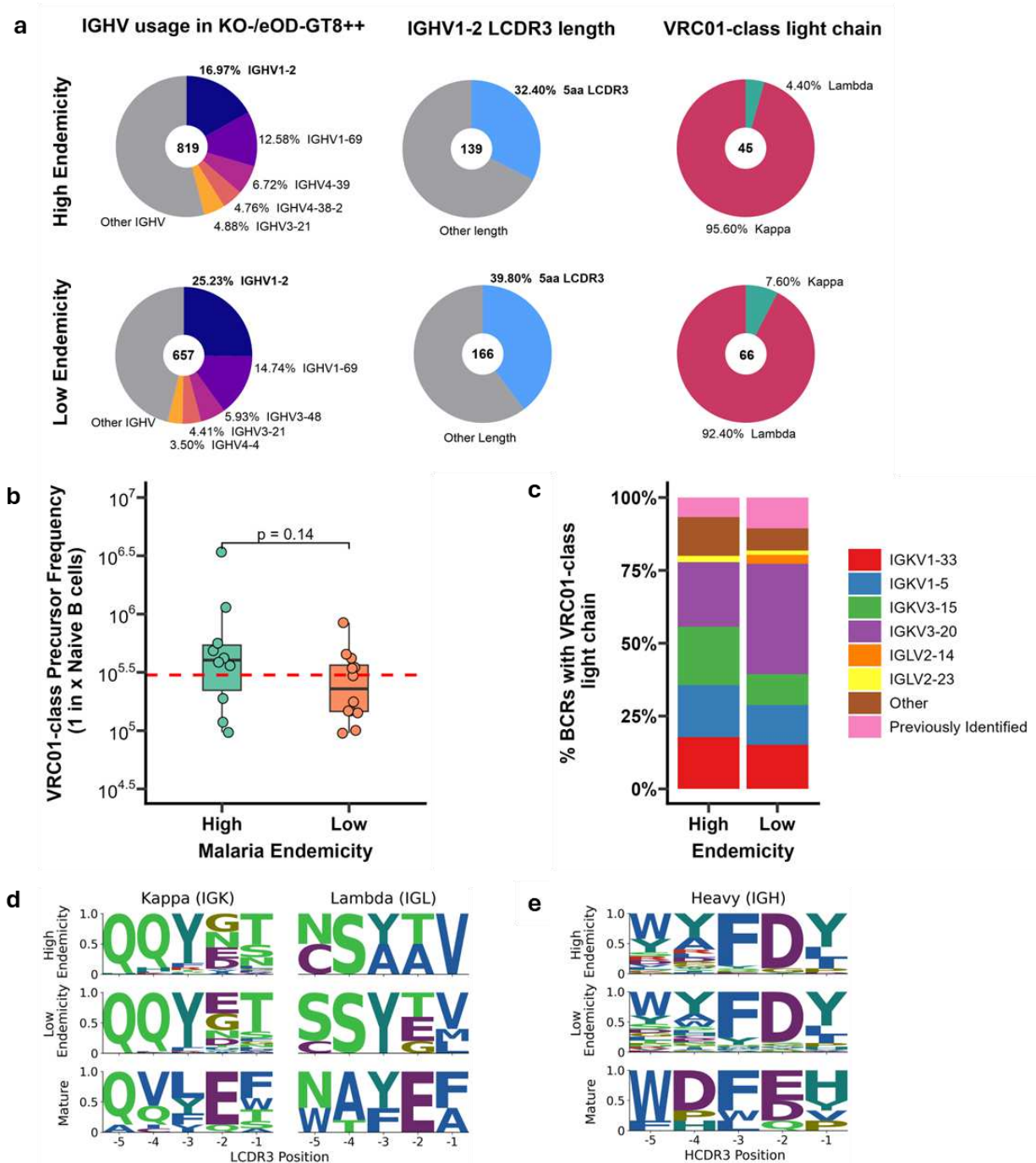


Figure 2: Characteristics of sequences from epitope-specific and VRC01-class naïve B cells. **a** Identification of VRC01-class sequences: first panel – the frequencies of IGHV gene usage among epitope-specific (KO/eOd-GT8⁺⁺) cells. Dark blue shows the frequency of IGHV1-2 gene usage, middle panel – frequency of VRC01-class naïve B cells (IGHV1-2 sequences paired with a 5 aa LCDR3 (light blue)), and last panel – frequency of kappa LC gene use among VRC01-class naïve B cells stratified by malaria endemicity setting (high – top row, low – bottom row). **b** Precursor frequency of VRC01-class naïve B cells among total naïve B cells ($n = 22$). Each dot represents an individual donor, with the thick lines showing the median value, the boxes indicating the 25% and 75% quartiles and the ‘whiskers’ highlighting the minimum and maximum values. The red hashed line indicates a precursor frequency of 1 in 300,000 naïve B cells. **c** Percentage of VRC01 class naïve B cells using bnAb κ or λ LC genes, LC genes previously identified in precursors in other studies (pink) or genes not previously identified in precursors, others (brown). **Supplementary**

Fig. 3b lists the genes identified in the precursors. **d** Sequence logos of the amino acid sequence of the 5aa LCDR3 of VRC01-class naïve B cells from the high malaria endemicity setting (top row), low endemicity (second row), and mature bnAbs (third row). **e** Sequence logos of the amino acid sequence at the C-terminal end of the HCDR3 of VRC01-class naïve B cells highlighting usage at the -5 position (Trp₁₀₃₋₅) stratified by malaria endemicity setting (high – top row, low – bottom row).

The VRC01-class sequences identified bind eOD-GT8 with high affinity

We randomly identified 66 paired heavy and light chain sequences from VRC01 class sequences and 78 paired heavy and light chain sequences from non-VRC01 class sequences and expressed them as monoclonal antibodies (mAbs). We then measured their affinity to eOD-GT8 and eOD-GT8 KO11. All the VRC01-class mAbs showed affinity to eOD-GT8 (median $K_D = 2.01 \mu\text{M}$, **Fig. 3a**), within the range of previously characterized VRC01-class precursors ($3.4 \mu\text{M}^{24}$), with limited affinity to the eOD-GT8-KO11 (median $K_D \geq 100 \mu\text{M}$, **Supplementary Fig 4a**). In contrast, only ~58% of the non-VRC01-class sequences showed detectable binding to eOD-GT8, with an approximately 17-fold lower affinity (median $K_D \geq 33.7 \mu\text{M}$, $p < 0.0001$). Affinity did not differ between VRC01-class mAbs from the high compared to the low malaria endemicity setting (median $K_{Ds} = 2.54 \mu\text{M}$ and $1.53 \mu\text{M}$ respectively, $p = 0.256$, **Fig. 3b**). We observed a trend towards higher affinity in monoclonal antibodies from sequences with an E/Q at position 96 (~1.8-fold, **Fig. 3c**) and those encoding a Trp₁₀₃₋₅ (~2.3-fold, **Fig 3d**), however this was of borderline significance for Trp₁₀₃₋₅ ($p = 0.0564$) and not significant for E/Q at position 96.

Light chain gene usage has previously been used to define VRC01 sub-classes, namely VRC01 (V κ 3-20+), VRC23 (V κ 3-15+), N6 (V κ 1-33+), PCIN63 (V κ 1-5+) and VRC-PG20 (V λ 2-14)³⁰. We did not observe any significant differences in median affinity between the subclasses, however, of the three highest affinity mAbs, two were from the VRC23 subclass and one from the N6 sub-class (median $K_D = 21.3 \text{ nM}$, 16.6 nM for VRC23 and 13.2 nM for N6, **Supplementary Fig. 4b**). The N6-subclass of bnAbs is known for its potency and breadth⁵³, and along with VRC23 does not require LCDR1 deletions to acquire breadth^{30,54}. Finally, we evaluated whether the bnAb score associated with precursor affinity and observed a modest positive correlation between the bnAb characteristic score and affinity to eOD-GT8 ($\rho = -0.35$, $p = 0.0047$, **Supplementary Fig. 4c**).

In summary, the identified VRC01-class precursors showed strong affinity to eOD-GT8, with a non-significant trend toward higher affinity observed in mAbs encoding a Trp₁₀₃₋₅ in the HCDR3. Additionally, the more features the precursors shared with the VRC01-class bnAbs, the better the affinity to eOD-GT8 was.

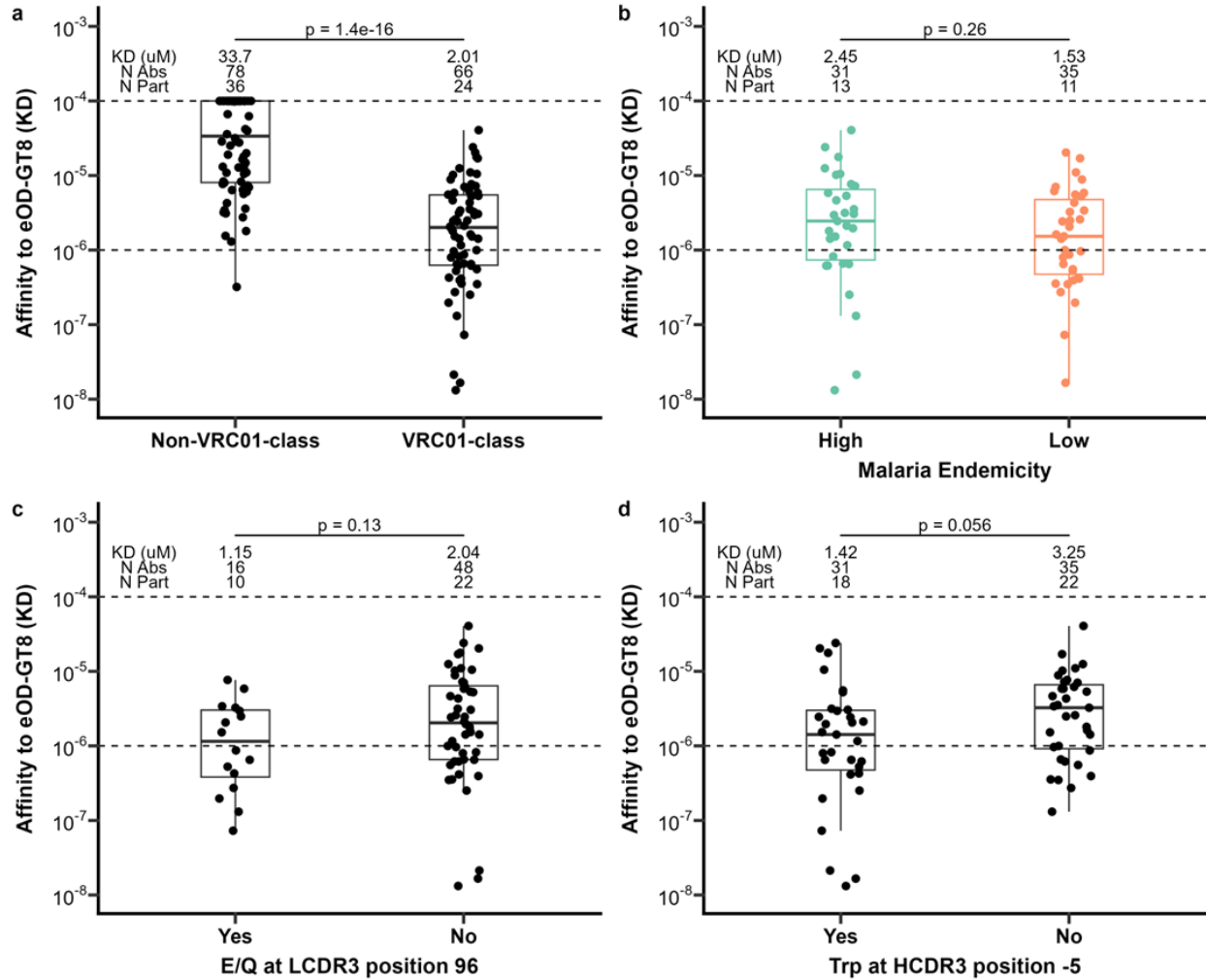


Figure 3: Monovalent affinity of select mAbs produced from VRC01-class and non-VRC01-class sequences derived from epitope-specific cells. a Affinity to eOD-GT8 and eOD-GT8 KO11 of mAbs from VRC01-class and non-VRC01-class BCR sequences. **b** Affinity to eOD-GT8 of VRC01-class mAbs from the low and high malaria endemicity settings. **c** Affinities to eOD-GT8 stratified by possession of an E/Q at position 4 and **d** by possession of Trp₁₀₃₋₅. Each dot represents an individual mAb, with thick lines showing the median value, the boxes indicating the 25% and 75% quartiles and the ‘whiskers’ highlighting the minimum and maximum values. ‘ K_D (uM)’ – median K_D in μM , ‘N Part’ – number of participants, ‘N Abs’ – number of mAbs tested. Statistical significance was determined using a two-sided Wilcoxon rank sum test.

*The IGHV1-2*02 allele was associated with higher VRC01-class precursor frequencies*

Given the reported influence of IGHV1-2 genotype on IGHV1-2 gene usage²⁷, we determined the IGHV1-2 genotypes for all 60 study participants. Twelve different IGHV1-2 genotypes were identified consisting of combinations of *02, *02_S4303, *02_S4953, *04, *05, *06, *06_S5416 and *06_S5931 alleles (**Fig. 4a**). Alleles *02_S4303, *02_S4953, *06_S5416 and *06_S5931 identified in 1, 4, 1 and 1 participants respectively, represent novel alleles. IGHV1-2*02_S4953 allele has previously been identified in Caucasian population^{28,32} and its existence validated in

genomic DNA⁵⁵. Similarly, *06_S5416 and *06_S5931 have been identified in repertoire studies done in other populations^{56,57}.

The *02_S4303 and *02_S4953 alleles encoded the same amino acid sequences as the reference *02 allele and hence we classified them as *02 in subsequent analyses (**Supplementary Fig. 5**). On the other hand, the *06_S5416 and *06_S5931 alleles encoded non-synonymous nucleotide changes hence they were considered distinct from the reference *06 allele. The *06_S5931 allele encodes a valine to methionine mutation at position 11 (V11M), a mutation acquired by some HIV bnAb lineages (e.g., N6 and DH270) during maturation⁵⁸. As expected, we did not detect any precursors from this allele due to the lack of the W50 residue. Overall, *02/*02 was the most dominant genotype identified followed by *02/*04, *02/*06, *04/*06, *06/*06, *04/*04, and *04/*05 in order of decreasing frequencies, with only 5% (3/60) of participants lacking either a *02 or *04 allele. The genotype distribution did not differ between participants from low and high endemicity regions (**Fig. 4b**).

The highest precursor frequencies were associated with the IGHV1-2*02 allele, with participants homozygous for the *02 allele having higher VRC01-class precursor frequencies (~2.7-fold higher than *02/*04 participants and ~4.2-fold higher than *02/*06 participants, **Fig. 4c**), consistent with its association with higher IGHV1-2 gene usage (**Supplementary Fig. 6a**). Homozygous loss of expression of the IGHV7-4-1 gene has also been shown to influence IGHV1-2 gene usage levels, with individuals homozygous for the deletion showing ~2.8-fold higher gene use than those with both copies of the gene present⁵⁵. We inferred IGHV7-4-1 deletion in 27 participants who were heterozygous for the IGHJ6 gene (**Supplementary Table 6**) and observed a 1.8-fold increase in IGHV1-2 gene use in participants with IGHV7-4-1 gene deleted compared to those with both copies of the gene present, though the difference did not reach statistical significance ($p = 0.082$, **Supplementary Fig. 6b**). Higher VRC01 precursor frequencies were also observed in individuals inferred to lack the IGHV7-4-1 gene, although sample sizes in each genotype group were limited (**Fig. 4d**).

In summary, we observed several novel IGHV1-2 alleles among the 60 participants, highlighting the genetic diversity present within this population, as well as a strong influence of IGHV1-2 genotype on precursor frequencies.

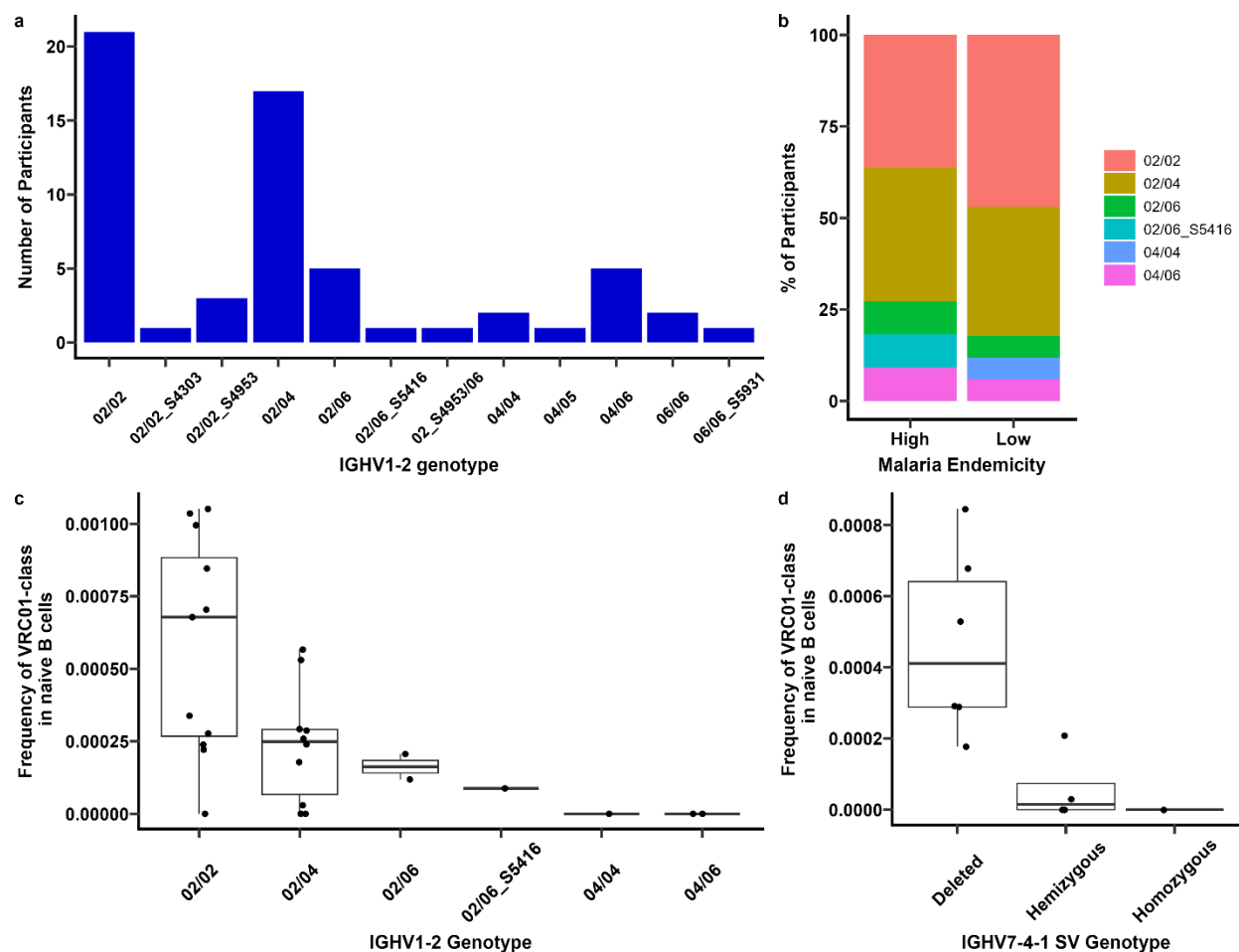


Figure 4: IGHV1-2 genotype distribution and impact on VRC01-class precursor frequency. **a** Distribution of identified genotypes for all donors in the study (N = 60). **b** Distribution of identified genotypes for donors included in the assessment of precursor frequencies (n = 28). Novel alleles that encoded the same amino acid sequence as the reference allele were reassigned as the reference allele. **c** Boxplots showing the frequency of VRC01-class naïve B cells for each stratified by IGHV1-2 genotype (n = 28). **d** Frequency of VRC01-class naïve B cells stratified by IGHV7-4-1 genotype (n = 11). Each dot represents an individual mAb, with thick lines showing the median value, the boxes indicating the 25% and 75% quartiles and the ‘whiskers’ highlighting the minimum and maximum values.

A novel IGHV1-2*06 allele may have ability to engage eOD-GT8

The IGHV1-2*06_S5416 allele encoded a Glutamine (Q) instead of the canonical Arginine (R) residue at aa position 50 (Kabat numbering). This mutation is located at the start position of the conserved ‘WNR’ VRC01-class motif that is important for engaging eOD-GT8^{19,59}. Interestingly, we identified one potential VRC01-class precursor that derived from the 06_S5416 allele suggesting that the R50Q mutation may have conferred some ability to engage with eOD-GT8. We produced a mAb from the BCR sequence of this allele and indeed, the mAb showed detectable binding to eOD-GT8 with a K_D of 2.95 μ M with no binding to eOD-GT8 KO11 (K_D of 100 μ M). This was similar to the median K_D observed for VRC01-class precursors deriving from *02 and *04 alleles (median K_D 2.01 μ M, **Fig. 3a**). Nevertheless, the other eight precursors identified from this individual were derived from the *02 allele potentially implying bias against the use of the

*06_S5416 allele. This finding suggests that there could be other novel IGHV1-2 alleles in existence with important implications for eOD-GT8 binding.

Broad range of CD4bs lineage precursors identified, with the non-VRC01-class VH1-2 naïve B cell repertoire showing some conservation including a dominant shared public light chain clonotype.

In addition to engaging VRC01-class precursors, eOD-GT8 also engages other CD4bs-targeting bnAb precursors^{30,31}. Majority of the non-VRC01-class repertoire also shows specific binding to eOD-GT8 (**Fig. 3a** and **Supplementary Fig. 4a**) though with overall lower affinity compared to VRC01-class bnAbs. Interestingly, we did observe a trend towards a higher frequency of non-VRC01-class naïve B cells in the sequences from the high malaria endemicity setting ($p = 0.053$) (**Fig 5a**), though there were no differences in the non-VRC01-class naïve B cell affinity by malaria endemicity ($p = 0.37$, **Supplementary Fig. 7a**).

Aside from IGHV1-2+ B cells paired with a 5-aa L-CDR3 (VRC01-class), we identified several other eOD-GT8⁺⁺/KO⁻ VH1-2+ naïve B cells. Though there was some enrichment for shorter L-CDR3 lengths (6, 7 and 8 aa) in the V κ light chains when compared to control naïve B cells (**Fig. 5b**), the most frequent L-CDR3 length observed was 9 aa (57.1% and 45.6% in the high and low malaria transmission settings, respectively). Among these epitope-specific 9-aa VH1-2+ B cells, we observed an enriched L-CDR3 motif ‘QQYGSSPWTF’ associated with IG κ 3-20 + J κ 1-01 light chains (**Fig. 5c** and **Supplementary Fig 7b**) that was present in multiple individuals (**Supplementary Fig 7c**). This public light chain was previously identified in 33% of VH1-2+ HCs paired to a 9-aa L-CDR3³⁰, and was present in our study at a frequency of 37.6% with no difference in occurrence by malaria endemicity (**Fig. 5d**). We also observed other public light chains that were present in multiple individuals, however these were less frequent (< 8%, **Supplementary Data 3**).

Similar to V κ light chains, short L-CDR3s were also enriched in the V λ light chains, particularly lengths of 6, 7 and 8 aa (**Fig. 5b**). We identified seven eOD-GT8-specific naïve B cells with an IOMA signature, VH1-2+ paired with an 8-aa λ 2-23 L-CDR3. All the IOMA-class sequences matched positions 2, 3 and 4 of the mature IOMA bnAb L-CDR3 (**Fig. 5e**). Other non-VRC01 class CD4bs bnAbs use diverse HC and LC combinations to engage eOD-GT8, this can occur via H-CDR3 loop-dependent binding (e.g., CH103, VRC16 and VRC13 class bnAbs) or by using their H-CDR2 region to mimic CD4 (VH1-46 derived bnAbs e.g., 8ANC113)⁶⁰. Within the non-VRC01 class CD4bs naïve B cells, we observed several sequences with signatures similar to CH103, VRC13 and VRC33 CD4bs bnAbs as well as several VH1-46+ B cells (**Supplementary Table 5**). The non-VRC01 class CD4bs repertoire thus represents additional B cell lineages that can be engaged for maturation into CD4bs-directed bnAbs.

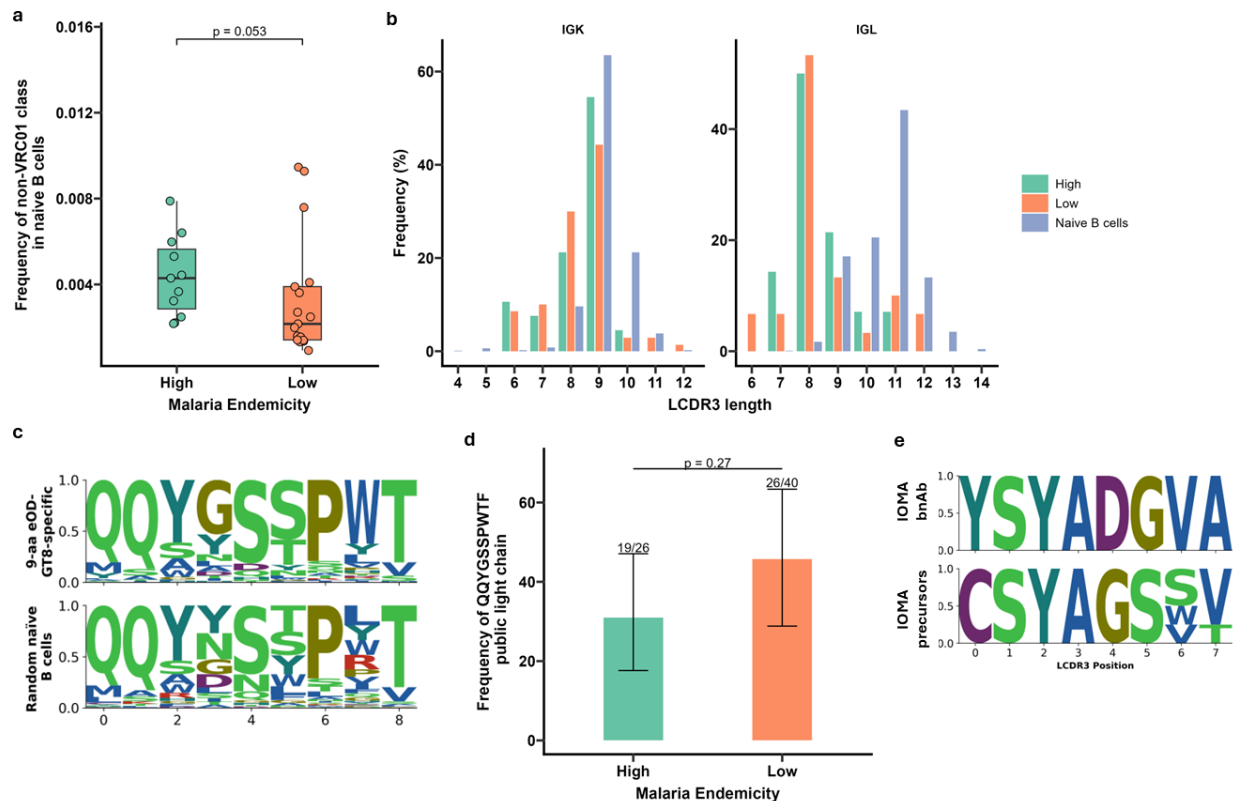


Figure 5: Sequence characteristics of non-VRC01-class eOD-GT8-specific naïve B cells. **a** Frequency of non-VRC01-class epitope-specific cells in naïve B cells (participants with ≥ 10 sequences, $n = 28$). **b** LCDR3 length of non-VRC01-class VH1-2⁺ CD4bs-specific naïve B cells. **b** Amino acid motif of non-VRC01-class VH1-2⁺ CD4bs-specific naïve B cells with a 9-aa LCDR3 and control B cells with a 9aa-LCDR3. **c** Frequency of the public LC in sequences from the low and high malaria endemicity settings. **d** Amino acid motif of IOMA bnAb (top row) and the IOMA-class naïve precursors (bottom row, $n = 7$).

DISCUSSION

Germline-targeting is a promising strategy to induce HIV bnAbs by vaccination. Four proof-of-principle trials conducted so far (IAVI G001²⁸, IAVI G002⁶¹, IAVI G003⁶¹, and IAVI C101²³) have provided evidence that efficient priming of VRC01-class responses using a GT immunogen can be achieved in North American and African populations. The G002 trial additionally provided preliminary evidence that primed VRC01-class bnAb precursors can be shepherded towards bnAb development. These results are encouraging for HIV vaccine development efforts, particularly for African populations where the HIV burden is highest. The study described here sought to isolate and characterise VRC01-class bnAb precursors in healthy Kenyan adults experiencing varied levels of malaria transmission to assess the potential impact of chronic malaria exposure on VRC01-class naïve bnAb precursor frequencies.

We identified VRC01-class naïve B cell precursors in 59% of the participants evaluated, though it is possible that the remaining participants also had VRC01-class precursors, but insufficient cell numbers and/or low precursor frequency may have hindered identification. In the population studied here, precursor frequencies did not differ significantly from previous studies in North American populations, with the overall precursor frequency estimate (1 precursor in 345,557 naïve

B cells) being similar to previous estimates (1 precursor in 300,000 naïve B cells^{24,30}). We postulated that chronic life-long exposure to malaria may drive alterations in the naïve B cell compartment, potentially impacting the HIV CD4bs-specific antibody repertoire. A previous study showed some differences in IGHV gene usage within classical memory and atypical memory B cells subsets driven by malaria exposure⁶². The authors did not observe malaria-related changes in the naïve B cell compartment; however, the sample size of malaria-infected participants included in the analysis was quite limited (three participants).

Infection-related immune impairment has also been observed following measles infection where restructuring of the naïve B cell compartment led to a decrease in naïve B cell diversity in a cohort of unvaccinated children, with alterations in IGHV-J gene usage frequencies observed in 10% of the cohort that persisted post measles resolution³⁵. Contrary to our hypothesis, naïve VRC01-class bnAb precursor frequencies were not influenced by prior chronic exposure to malaria, which is reassuring for future vaccine implementation efforts in endemic populations. There remains a possibility however that chronic exposure to other pathogens may impact precursor frequencies or that chronic malaria exposure may have impacted other repertoire subsets important for responses to other pathogens, therefore this is still an area that requires investigation.

Consistent with other studies, precursor frequencies were influenced by the IGHV1-2 genotype^{27,32,59}. Individuals homozygous for the *02 allele had the highest median precursor frequency estimates compared to all other permissive genotypes identified in our study (*02/*04, *02/*06, *04/*04 and *04/*06). The higher precursor frequency associated with the *02 allele also correlated with the level of VRC01-class response observed post eOD-GT8 vaccination in the IAVI G001 trial³². Encouragingly, ~82% of the study participants had an *02 allele, with only five percent (3/60) lacking an *02 or *04 allele, supporting the development of GT vaccines to induce VRC01-class bnAbs in this population. Notably, IGHV1-2 genotype did not explain all the variation in precursor frequency estimates in our study, nor did it explain all the variation in post-vaccination VRC01-class responses in G001, suggesting a role for other factors such as immune history or additional genetic factors³².

Indeed, other genetic loci are also proposed to influence IGHV1-2 gene use, for example homozygous deletion of the IGHV7-4-1 gene that has been associated with higher gene use⁵⁵. Our observations are in line with this, and we similarly observed a trend towards higher VRC01-class precursor frequencies in individuals lacking expression of the IGHV7-4-1 gene. We do acknowledge, however, that the small sample size of participants with overlapping haplotyping and VRC01-class precursor frequency data limited our ability to perform more comprehensive analyses. We also acknowledge the fact that inference of germline repertoire deletions from RepSeq data may not be efficient for poorly expressed alleles such as IGHV7-4-1*01. Further studies on the possible impact of the IGHV7-4-1 gene deletion on VRC01-class precursor frequencies and vaccine responses and how this compares to/or interacts with the impact of IGHV1-2 genotype are warranted.

In addition to precursor frequencies, precursor affinity is also an important consideration for the success of GT immunogens^{20,63,64}. IGHV1-2 genotype has an impact on precursor affinity, with the *05 and *06 unable to bind eOD-GT8 due to a lack of a tryptophan residue at the start of the

conserved ‘WNR’ motif critical for engaging eOD-GT8^{19,27,32}. Strikingly, we identified a novel *06-like allele (IGHV1-2*06_S5416) encoding a non-synonymous R50Q mutation that we postulate may have conferred some ability to engage with eOD-GT8, however this finding warrants confirmation. The potential functional consequences of novel alleles within the IGHV1-2 genes, and the potential for some of these mutations to be in SHM regions important for bnAb development reiterates the importance of genotyping participants in GT clinical trials.

Aside from IGHV1-2 genotype, the presence of Trp₁₀₃₋₅ showed evidence of an association with increased affinity of VRC01-class mAbs to eOD-GT8. Trp₁₀₃₋₅ forms hydrogen bonds with Loop D of gp120²⁹, and antibodies encoding this key residue may more readily develop into bnAbs as they require fewer maturation steps⁵⁹. In addition to Trp₁₀₃₋₅ other sequence features required for VRC01-class bnAb development, such as use of VH1-2*02 or *04 allele usage and the presence of a 5-aa LCDR3 can readily be encoded in the germline⁵² hence it is likely that the more of these features a naive VRC01-class precursor possesses, the more likely it is to be shepherded towards VRC01-class bnAb development by vaccination. By scoring BCR sequences on the number of bnAb-like characteristics they possessed, we found that a higher score correlated with better affinity to eOD-GT8. In future, as full regimens develop, it will be interesting to evaluate whether the bnAb score of VRC01-class bnAb precursors correlates with accelerated maturation towards broad neutralisation.

In conclusion, we confirm that VRC01-class naïve B cell precursors are present in Kenyan populations, and that the precursor frequency does not seem to differ from that observed in North American populations, nor to be impacted by chronic exposure to malaria. We note that several of these precursors share critical sequence characteristics with VRC01-class bnAbs and hence may be readily shepherded towards breadth by a suitable vaccination regimen. We also identified four novel IGHV1-2 alleles, one that may have functional consequences for binding to eOD-GT8. These data underscore the value of participant IGHV genotyping in GT trials and suggest that rare, function-restoring variants can exist. Moreover, GT immunogens to induce bnAb responses to other target sites on the HIV virus envelope are in development, hence it remains necessary to continue to identify allelic variants that may modulate vaccine responses for other bnAb classes in diverse populations.

METHODS

Study design and participant recruitment

Ethical approval for participation in these cohort studies was given by the Kenya Medical Research Institute Scientific and Ethical Review Unit (KEMRI/SERU/4177). Participants were recruited from four regions in Kenya with varying malaria endemicities. These included Ahero in Western Kenya where there is high malaria transmission, Nairobi (central Kenya) where there is low to no malaria transmission and from two regions in Coastal Kenya, Junju with moderate transmission and Ngerenya with low transmission intensity. Since historical exposure to malaria could vary among Nairobi residents based on previous travels out of the city to malaria-endemic ancestral locations, inclusion criteria in Nairobi included pre-screening to identify individuals who had no to low detectable levels of anti-*P. falciparum* schizont antibodies as a proxy measure of no to low

historical malaria exposure (anti-*P. falciparum* schizont ELISA described elsewhere⁶⁵). Adults between the ages of 18 - 60 years were considered for recruitment if they had lived in the recruitment area for at least the last 10 years. A screening visit was first carried out where adults from each site were invited to the study health facilities for examination. A five ml blood sample was taken for HIV and malaria infection screening, haemoglobin level assessment and complete blood counts. Additionally, medical history was also taken to determine eligibility. Detailed inclusion and exclusion criteria are provided in **Supplementary Table 2**. Following the screening visit, 15 participants from each study site who were deemed fit to provide one blood bag (450 ml blood draw) were recruited into the study. The blood draw was conducted within one week of screening.

Sample collection and PBMC processing

Whole blood from the blood bags was layered onto density centrifugation media (Histopaque, Sigma Aldrich™, catalogue number (#)10771) (15 – 27 mL of blood to 10 mL of Histopaque) and centrifuged at 400 x g for 40 minutes. The PBMCs were then harvested and washed once using Hanks balanced salt solution (HBSS, Sigma Aldrich™, #H8264) and then a second time using RPMI (Sigma Aldrich™, #R0883) 10% HIFCS (Sigma Aldrich™, #F4135) medium (R10) before cell counts were determined on a Vi-Cell XR automated cell counter. The PBMCs were then resuspended in cold freezing media (10% DMSO (Sigma Aldrich™, #D2650) in HIFCS) at a concentration of 1.5×10^7 or 2.0×10^7 cells/mL, and 1 mL of cell suspension transferred to pre-labelled cold cryovials. The cryovials were then transferred to a pre-chilled Mr. Frosty™ Freezing Container (Thermo Scientific™, #5100-0001), and the container placed in a -80°C freezer overnight. The next day the frozen PBMCs were transferred to liquid nitrogen tanks for long term storage.

Sort probe preparation

Sort probes used to isolate epitope-specific cells were prepared as previously described²⁷. Briefly, biotinylated eOD-GT8 wildtype (WT) and knock-out (KO) monomers were combined with fluorescently labelled streptavidin (SA-AlexaFluor647 (Invitrogen™, #S21374) and SA-AlexaFluor488 (Invitrogen™, #S11223) for the WT baits and SA-Brilliant Violet 421 (BioLegend™, #405225) for the KO bait) at a 4:1 molar ratio. The performance of the probes was then tested using beads coated with monoclonal antibodies of known binding characteristics to the probes (germline-VRC01 or KG064-099, for testing eOD-GT8 WT and eOD-GT8-KO probes respectively).

B cell enrichment, flowcytometry and epitope-specific B cell sorting

Cryopreserved PBMC samples were thawed and resuspended in pre-warmed R50 media (RPMI-1640 containing 50% HIFCS). The PBMCs were then pelleted by centrifugation at 500 x g for 7 minutes, supernatants decanted and resuspended in FACS buffer (2% HIFCS, 25 mM HEPES (Gibco™, #H0887) buffer, 2 mM EDTA (Sigma Aldrich™, #03690) and 1 x PBS (Gibco™, #10010-015). The PBMC were counted and the concentrations adjusted to 5×10^7 cells/mL for B cell enrichment using the EasySep™ Human Pan-B Cell Enrichment Kit (STEMCELL™ Technologies, #19054) according to manufacturer's instruction. The enriched B cells were counted

and the quantities of antibody to use for staining determined (for every 10 - 20 million B cells we used a final staining volume of 100 μ L made up of 5 μ L of fluorescently-labelled cell phenotyping markers (**Supplementary Table 6**), 2 μ L of sample multiplexing hashtag (BioLegend TotalSeqC) and 10 μ L of the probes). The enriched B cells were stained using the antibody and KO probe (100 nM final staining volume) cocktail for 15-minutes incubation on ice and in the dark. The WT probes were then added (100 nM) to the B cells and incubated for a further 30 minutes. Zombie AquaTM amine-reactive viability dye (BioLegendTM, #423102) was then prepared at dilution of 1:300 in FACS buffer (1 mL for every 10 – 20 million B cells) and added to the cells, followed by incubation for 30 minutes in the dark. The cells were then washed with 10 mL FACS buffer and resuspended at a concentration of 15,000 cells/ μ L in FACS buffer for acquisition on a FACS Melody, using the gating and sorting strategy illustrated in **Supplementary Fig 1**. Epitope-specific naïve B cells were defined as IgD⁺ B cells that bound both eOD-GT8*AF647 and eOD-GT8*AF488 ('double-positive') but did not bind the eOD-GT8-KO*BV421 probe. Epitope-specific cells were sorted into 20 μ L of 100% FBS. We also sorted between 700 to 1500 random naïve B cells from each donor into the same well as the epitope-specific cells to increase the total cell numbers to above 500 which is the recommended minimum number of cells that can be profiled using the 10x NextGEM reagents. The random naïve B cells were separately hashtagged and to allow removal from the sequence dataset before analysis. After the sort, the sorted epitope-specific cells were resuspended in 200 μ L of cold PBS and pelleted by centrifugation at 2500 rpm for 2 minutes. Excess PBS was then aspirated to leave 38.7 μ L volume required for Gel-beads-in-emulsion (GEMs) preparation using the 10x Genomics[®] platform.

Preparation of single-cell RNAseq libraries using the 10x genomics platform

Single-cell RNAseq libraries of the epitope-specific B cells were prepared using the Chromium Next GEM Single Cell 5' Reagent Kit v2 with Feature Barcoding technology according to the manufacturer's instructions. Immunoglobulin genes were amplified from the fraction of cDNA prepared from cellular mRNA and used to prepare V(D)J libraries, while hashtag libraries were prepared from the cell-hashtag oligonucleotide-derived cDNA. Quantification of the libraries was done by qPCR using the KAPA Library Quantification Kit for Illumina[®] platforms (RocheTM, #KK4824) and library sizes determined using the Agilent Bioanalyzer platform. Libraries were then pooled to achieve a minimum of 5000 reads per cell for both the V(D)J libraries and cell hashtag libraries. The pooled libraries were sequenced on an Illumina NextSeq500 using a 150-cycle high output kit with read lengths set to 26 base pairs (bp) for read 1, 10 bp for index 1, 10 bp for index 2 and 90 bp for read 2.

Next generation sequencing (NGS) library preparation for IGHV1-2 genotype analysis

Bulk RNAseq libraries for genotyping using IgDiscover were prepared using primers from Bernat et al., 2019⁶⁶. Briefly, frozen PBMCs from each study participant (N = 60) were thawed and B cells enriched as described above. The enriched B cells were pelleted by centrifugation at 400 x g for 5 minutes, and the pellets resuspended in TRIzol reagent (InvitrogenTM, #10296010) at a ratio of 1 mL TRIzol to 100 μ L of B cell pellet. Two hundred microlitres of chloroform were added to the TRIzol sample before incubating for three minutes at room temperature after mixing vigorously. The sample was then centrifuged at 1400 x g for 35 minutes at 4°C. The aqueous phase

was then harvested and transferred to a 1.5 mL microcentrifuge tube containing 2 μ L of GlycoBlue™ (Invitrogen™, #AM9515) and 500 μ L of isopropanol (Sigma Aldrich™, #I9516), mixed and incubated overnight at 4°C.

The next day, the sample was spun at 13,000 rpm for 30 minutes at RT, and the pellet then washed gently using 1 ml of ice-cold 75% ethanol and pelleted by centrifugation for 5 min at 7500 x g at 4 degrees two times. The ethanol was decanted and the pellets air-dried for no longer than 5 minutes. Forty microlitres of RNase-free water was then added to each sample and the RNA dissolved by heating the mix at 60°C for 10 minutes. From the extracted RNA, cDNA was prepared using an IgM primer containing a unique molecular identifier (UMI) and the Illumina Read2 sequence. Five prime multiplexed libraries were then prepared using IGHV leader region-specific primers and a universal reverse primer. After this, the 5 ng of each library was then individually indexed. A pool of 15 samples was prepared at equimolar ratios then sequenced on an Illumina MiSeq (version 3 kit, 2 x 300 cycles, #15043762) at 13 pM with 15% PhiX (Illumina, #15017872) spike.

Production of monoclonal antibodies and SPR affinity measurements

MAb production was outsourced to Genscript where the monoclonal antibodies were produced as human IgG1 using the TurboCHO expression service. We measured kinetics and affinity of antibody-antigen interactions on a Catterra LSA using a HC30M sensor chip that has polycarboxylate hydrogel, medium charge density, 30nm coating thickness, moderate ligand capacity and excellent diffusion characteristics for monomeric antigens. We used 1x HBS-EP+ supplemented with BSA at 1mg/ml at pH 7.4 as running buffer. Using manufacturer's instructions about 2800-3300 RU of capture antibody was amine coupled using the N-Hydroxysuccinimide (NHS) and 1-Ethyl-3-(3-dimethylaminopropyl) carbodiimidehydrochloride (EDC) Amine Coupling Kit using 10 times diluted NHS and EDC for final concentrations of NHS and EDC at 1.15 mg/ml and 7.5 mg/ml respectively. The chip surface contact time of EDC/NHS solution varied between 1 and 3 minutes. The Southern Biotech capture antibody was used at a concentration of 25-100 μ g/ml with 10 minutes contact time. We used 1.7% phosphoric Acid as the regeneration solution with 60 seconds contact time and injected three times per each cycle, with ligand concentration maintained at approximately 1 μ g/ml with 5 min contact time. Analyte concentrations were quantified on NanoDrop 2000c Spectrophotometer using absorption signal at 280 nm. Raw sensograms were analyzed using the Catterra Kinetics software. Where a measurement did not result in kinetic-fit K_D from the Catterra Kinetics software analysis, the K_D value was set to ≥ 100 μ M.

BCR Sequence processing for downstream analysis

Analysis of immunoglobulin genes of epitope-specific (KO-/eOD-GT8++) naive B cells

A standardized “*immbase*” analytical pipeline was developed by modifying the “G00X” pipeline previously used for analysis of the IAVI-G00X clinical trials²⁸. The modified pipeline uses CellRanger version 7.1.2 with *cellranger multi* commands, instead of the *count* and *vdj* commands used by the “G00X” pipeline and performs detailed analyses of the sequence features for both VRC01-class and non-VRC01-class BCRs (Suppl. Fig 8). The “*immbase*” pipeline

consists of five main steps: demultiplexing, converting BAM files to FASTQ format, V(D)J processing, Adaptive Immune Receptor Repertoire (AIRR) output generation, and report generation. For AIRR output, the pipeline performs B cell receptor (BCR) V gene assignments, heavy and light chain pairing and somatic hypermutation (SHM) analysis, using the Sequencing Analysis and Data Library for Immunoinformatics Exploration (SADIE, <https://github.com/jwillis0720/sadie>).

A manifest file was created to automate the analysis pipeline and statistical calculations. This manifest captured the GEM number, sample ID, hashtag or cell multiplexing oligo (CMO) ID, number of cells sorted, number of cells in the GEM, Sample index, FASTQ ID, run number, path to FASTQ files, sample type, and endemicity. The sequences were first demultiplexed to separate the different library types (either V(D)J and feature barcode (FBC) libraries) as well as individual samples that were pooled in the same GEM reaction. Next V(D)J annotation was carried out and the pairing of heavy and light chains determined by extracting productive sequences and categorizing them based on their loci (IGH, IGK, IGL). We included only paired HC + LC sequences and filtered the B cell isotype to IgM and/or IgD B cells since we had sorted naïve B cells. To exclude potential IgM memory cells, that as expected, would show additional mutation, we also included an additional filter. A per-sequence mutation cut-off of below 2% was used to exclude sequences with potential SHM while allowing retention of sequences derived from novel alleles absent from the reference datasets used for the annotation.

Next, we identified VRC01-class features based on the described sequence characteristics: usage of the IGHV1-2*02 or IGHV1-2*04 alleles and a light chain CDR3 length of 5 aa. Additionally, the pipeline computed sequence-level statistics, including the number of sequences per sample, the frequency of VRC01-class and IOMA-class sequences. IGHV1-2 gene usage patterns were assessed by calculating the relative frequency of V-gene segment utilization across samples. To classify potential CD4bs antibody lineages, we utilized predefined categories based on their sequence characteristics (**Supplementary Table 5**).

We used the formula below to estimate the VRC01-class precursor frequency in naïve B cells. Estimates of the frequencies of VRC01-class naïve precursors were limited to participants with greater than 10 sequences recovered to minimize the impact of low sampling depth on the frequency estimates.

$$\frac{\text{Number of IgD+epitope-specific cells}}{\text{Number of IgD+B cells}} \times \frac{\text{Number of VRC01-class sequences}}{\text{Number of sequenced epitope-specific cells}} \times 100\%$$

Definition of IGHV1-2 genotypes and VH1-2 gene usage frequency

Sequence reads from each bulk RNAseq library generated for each donor were processed using the IgDiscover analysis pipeline (version 1.0.2.dev4+g7975573) for IGHV germline inference using the default parameters. The number of merged reads for all 60 participants are provided in **Suppl. Data 2** (median reads: 1,198,028, IQR 881,304 – 1,413,149 reads). To confirm the genotyping and identified novel IGHV1-2 novel alleles, the libraries were also genotyped using the corecount genotyping module in IgDiscover⁶⁷, with germline gene inference and genotyping also analysed using TIgGER analysis tool (version 1.1.2)⁶⁸. We observed good concordance

between the two analysis methods (**Suppl. Data 2**). IGHV1-2 gene usage summaries were generated using the ‘countGenes’ function within the Alakazam package (version 1.4.2)⁶⁹.

Haplotype analysis to determine IGHV7-4-1 deletion

We used haplotype-based inference to assess IGHV7-4-1 deletion in a subset of 28 participants (**Supplementary Data 2**) who were heterozygous for IGHJ6 (expressed both IGHJ6*02 and IGHJ6*03 alleles), a commonly used ‘anchor gene’ to define the chromosome from which expressed IGHV genes were derived. As recommended, following novel allele discovery, we reassigned alleles using an individual’s personal genotype then filtered the dataset to retain sequences with no more than three mutations called in the V gene and no mutations in the D gene (mutation analysis was done using SHazaM⁶⁹ analysis tool, version 1.3.1). After this, we collapsed the sequences by clonotype to retain only unique sequences. We then employed RABHIT tool (version 0.3.0),⁷⁰ that adapts a Bayesian framework to infer V, D or J haplotypes, to infer zygosity using the default parameters, however, we set the minimum minor allele fraction of the anchor gene to 0.275 as opposed to 0.3. To detect genes deleted from either or both chromosomes, we first used the ‘deletionsByBinom’ function to detect genes deleted from both chromosomes. We then carried out haplotype inference for the non-deleted genes using the ‘createFullHaplotype’ function to infer single-chromosome deletion events. The priors for Bayesian inference were inferred from the usage frequencies of the anchor gene (IGHJ6). Where IGHV7-4-1 gene deletion was defined as ‘unknown’ for a sample, we excluded that sample from assessment of impact on IGHV1-2 gene usage. We obtained data from 27 of the 28 participants, with only one donor’s IGHV7-4-1 deletion status defined as unknown.

Statistical analysis

Participant demographic characteristics and summaries of B cell sorting and sequencing metrics were stratified by location and malaria endemicity and reported as medians with interquartile ranges (IQR) or summarised as counts and percentages. Comparisons between two independent groups were performed using the Wilcoxon rank-sum test for continuous data and the Chi-Square test for comparisons of frequencies of categorical variables with 95% confidence intervals calculated using the Clopper-Pearson method for binomial proportions. For comparisons of more than two independent groups, we used the Kruskal-Wallis test with post-hoc analysis done using Dunn’s test with the Bonferroni correction applied to adjust for multiple comparisons. We tested correlations between two continuous variables for significance using the Spearman’s test. Statistical and graphical analyses were performed using R (version 4.4.3) using the following packages: tidyverse (version 2.0.0), ggprism (version 1.0.7), ggpubr (version 0.6.3), ggsignif (version 0.6.4), stringr (version 1.6.0), RColorBrewer (version 1.1-3), arrow (version 23.0.1.1), rstatix (version 0.7.3) and DescTools (version 0.99.60). Sequence logos for VRC01-class antibodies were visualized using the Python-based Logomaker⁷¹ and pie charts drawn using Prism (GraphPad, version 10.3.1).

ACKNOWLEDGEMENTS

This research was funded through the IAVI- Accelerate the Development of Vaccines and New Technologies to Combat the AIDS Epidemic (ADVANCE) programme with support of the

American People through the former U.S. President's Emergency Plan for AIDS Relief (PEPFAR) through United States Agency for International Development (USAID). EWN and DMK receive funding from the Gates Foundation (INV-064558). MKM, DMK, TN and EWN receive support from the Sub-Saharan African Network for TB/HIV Research Excellence (SANTHE) which is administered by the Africa Health Research Institute (AHRI) and funded by the Science for Africa Foundation to the Developing Excellence in Leadership, Training and Science in Africa (DELTAS Africa) programme (Del-22-007) with support from Wellcome Trust and the UK Foreign, Commonwealth & Development Office and is part of the EDCPT2 programme supported by the European Union; the Bill & Melinda Gates Foundation (INV-033558); Gilead Sciences Inc., (19275), Aidsfonds (0454) and the Ragon Institute of Mass General Brigham, MIT, and Harvard. DMK is also supported by the Wellcome Trust (220704/Z/20/Z). Funding to W.R.S. was supported by the National Institute of Allergy and Infectious Diseases (NIAID) UM1 AI144462 (Scripps Consortium for HIV/AIDS Vaccine Development), and the Gates Foundation Collaboration for AIDS Vaccine Discovery (CAVD) awards (INV-007522, INV-008813, and INV-034657) to the IAVI NAC centre. The funders played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. We are grateful to Prof. Gunilla Karlsson Hedestam and Dr. Martin Corcoran for their generous donation of the BCR amplification primers used for participant genotyping and to Dr. James Tuju for the *P. falciparum* schizont extract used to define malaria exposure.

COMPETING INTERESTS

W.R.S. is an inventor on patents filed by Scripps and IAVI on the eOD-GT8 monomer and 60mer immunogens (patent number 11248027, "Engineered outer domain (eOD) of HIV gp120 and mutants thereof"). W.R.S. is an employee and shareholder of Moderna, Inc. All other authors declare no financial or non-financial competing interests.

AUTHOR CONTRIBUTIONS

E.W.N., D.M., T.N., E.L., L.M., and D.S. conceived and designed the study. M.K., K.R., J.K., Y.S., C.M., A.M., L.N., J.O., O.A., B.O., D.M., and E.W.N. recruited the study participants, coordinated the sample collection and processing, and organised the data. M.K.M., C.M., Y.S., M.M., S.O., and R.K. performed B cell sorting and sequencing assays and the immunoglobulin genotyping assays. W.O., O.K., and C.A.C. performed the SPR experiments and data analysis. M.K.M., C.K., B.K., B.K., T.S., and D.C.O. analysed the data and created figures and tables. W.R.S. and C.A.C. provided eOD-GT8 probes and control monoclonal antibodies. M.K.M. and E.W.N. wrote the original draft. M.K.M., J.M., D.M., E.W.N., E.L., D.S., T.N., C.A.C., and W.R.S. revised and edited the manuscript. C.K., C.M., D.M., and E.W.N. contributed equally to the work.

DATA AVAILABILITY

The data generated during this study are provided in the main article and the supplementary files. Heavy and light chain sequences shown in Supplementary Data 1 have been deposited at GenBank with the accession codes PZ405537-PZ405758. The underlying primary datasets have been

deposited in the Harvard Dataverse repository <https://doi.org/10.7910/DVN/YEQIEH>. The data are available through a formal request to the KEMRI-Wellcome Trust Research Programme Data Governance Committee if the use of the data is compliant with the consent provided by the participants. Details of the criteria can be found in the KEMRI-Wellcome data sharing guidelines (<https://kemri-wellcome.org/data/>). Requests for the data can be made to the Data Governance Committee (dgc@kemri-wellcome.org) through the corresponding authors.

CODE AVAILABILITY

We have packaged the entire ‘*immbase*’ BCR precursor frequency pipeline with scripts and documentation as well as the scripts used for the novel allele analysis and haplotyping in a private repository that will become public upon manuscript publication https://github.com/ENduatiLab/IMMBASE_immune_repertoire_characterisation. TIgGER, Alakazam, Shazam and RAbHIT are tools from the Immcantation portal available here: <https://immcantation.readthedocs.io/en/stable/index.html>. The IgDiscover code used for participant IGHV genotyping is available from <https://gitlab.com/gkhlab/igdiscover22>.

REFERENCES

1. Moldt, B. *et al.* Highly potent HIV-specific antibody neutralization in vitro translates into effective protection against mucosal SHIV challenge in vivo. *Proc. Natl. Acad. Sci.* **109**, 18921–18925 (2012).
2. Shingai, M. *et al.* Passive transfer of modest titers of potent and broadly neutralizing anti-HIV monoclonal antibodies block SHIV infection in macaques. *J. Exp. Med.* **211**, 2061–2074 (2014).
3. Barouch, D. H. *et al.* Therapeutic Efficacy of Potent Neutralizing HIV-1-Specific Monoclonal Antibodies in SHIV-Infected Rhesus Monkeys. *Nature* **503**, 224–228 (2013).
4. Diskin, R. *et al.* Restricting HIV-1 pathways for escape using rationally designed anti-HIV-1 antibodies. *J. Exp. Med.* **210**, 1235–1249 (2013).
5. Klein, F. *et al.* HIV therapy by a combination of broadly neutralizing antibodies in humanized mice. *Nature* **492**, 118–122 (2012).
6. Caskey, M. *et al.* Viraemia suppressed in HIV-1-infected humans by broadly neutralizing antibody 3BNC117. *Nature* **522**, 487–491 (2015).
7. Moore, P. L., Williamson, C. & Morris, L. Virological features associated with the development of broadly neutralizing antibodies to HIV-1. *Trends Microbiol.* **23**, 204–211 (2015).
8. Hraber, P. *et al.* Prevalence of broadly neutralizing antibody responses during chronic HIV-1 infection. *AIDS* **28**, 163–169 (2014).

- 722 9. Sather, D. N. *et al.* Factors Associated with the Development of Cross-Reactive Neutralizing
723 Antibodies during Human Immunodeficiency Virus Type 1 Infection. *J. Virol.* **83**, 757–769
724 (2008).
- 725 10. Stamatatos, L. HIV vaccine design: the neutralizing antibody conundrum. *Curr. Opin.*
726 *Immunol.* **24**, 316–323 (2012).
- 727 11. Haynes, B. F. & Montefiori, D. C. Aiming to induce broadly reactive neutralizing antibody
728 responses with HIV-1 vaccine candidates. *Expert Rev. Vaccines* **5**, 579–595 (2006).
- 729 12. Rappuoli, R., Bottomley, M. J., D’Oro, U., Finco, O. & Gregorio, E. D. Reverse vaccinology
730 2.0: Human immunology instructs vaccine antigen design. *Journal of Experimental Medicine*
731 **213**, 469–481 (2016).
- 732 13. Briney, B. *et al.* Tailored Immunogens Direct Affinity Maturation toward HIV Neutralizing
733 Antibodies. *Cell* **166**, 1459–1470.e11 (2016).
- 734 14. Chen, X. *et al.* Vaccination induces maturation in a mouse model of diverse unmutated
735 VRC01-class precursors to HIV-neutralizing antibodies with >50% breadth. *Immunity* **54**, 324–
736 339.e8 (2021).
- 737 15. Escolano, A. *et al.* Sequential Immunization Elicits Broadly Neutralizing Anti-HIV-1
738 Antibodies in Ig Knockin Mice. *Cell* **166**, 1445–1458.e12 (2016).
- 739 16. Medina-Ramírez, M. *et al.* Design and crystal structure of a native-like HIV-1 envelope
740 trimer that engages multiple broadly neutralizing antibody precursors in vivo. *J. Exp. Med.* **214**,
741 2573–2590 (2017).
- 742 17. Parks, K. R. *et al.* Overcoming Steric Restrictions of VRC01 HIV-1 Neutralizing Antibodies
743 through Immunization. *Cell Rep.* **29**, 3060–3072.e7 (2019).
- 744 18. Tian, M. *et al.* Induction of HIV Neutralizing Antibody Lineages in Mice with Diverse
745 Precursor Repertoires. *Cell* **166**, 1471–1484.e18 (2016).
- 746 19. Jardine, J. *et al.* Rational HIV immunogen design to target specific germline B cell receptors.
747 *Science* **340**, 711–716 (2013).
- 748 20. Abbott, R. K. *et al.* Precursor Frequency and Affinity Determine B Cell Competitive Fitness
749 in Germinal Centers, Tested with Germline-Targeting HIV Vaccine Immunogens. *Immunity* **48**,
750 133–146.e6 (2018).
- 751 21. Zhou, T. *et al.* Multidonor Analysis Reveals Structural Elements, Genetic Determinants, and
752 Maturation Pathway for HIV-1 Neutralization by VRC01-Class Antibodies. *Immunity* **39**, 245–
753 258 (2013).

- 754 22. Caniels, T. G. *et al.* Germline-targeting HIV-1 Env vaccination induces VRC01-class
755 antibodies with rare insertions. *Cell Rep. Med.* **4**, 101003 (2023).
- 756 23. Caniels, T. G. *et al.* Precise targeting of HIV broadly neutralizing antibody precursors in
757 humans. *Science* eadv5572 (2025) doi:10.1126/science.adv5572.
- 758 24. Jardine, J. G. *et al.* HIV-1 broadly neutralizing antibody precursor B cells revealed by
759 germline-targeting immunogen. *Science* **351**, 1458–1463 (2016).
- 760 25. McGuire, A. T. *et al.* Engineering HIV envelope protein to activate germline B cell receptors
761 of broadly neutralizing anti-CD4 binding site antibodies. *Journal of Experimental Medicine* **210**,
762 655–663 (2013).
- 763 26. West, A. P. *et al.* Structural insights on the role of antibodies in HIV-1 vaccine and therapy.
764 *Cell* **156**, 633–648 (2014).
- 765 27. Lee, J. H. *et al.* Vaccine genetics of IGHV1-2 VRC01-class broadly neutralizing antibody
766 precursor naïve human B cells. *npj Vaccines* **6**, 1–12 (2021).
- 767 28. Leggat, D. J. *et al.* Vaccination induces HIV broadly neutralizing antibody precursors in
768 humans. *Science* **378**, eadd6502 (2022).
- 769 29. West, A. P., Diskin, R., Nussenzweig, M. C. & Bjorkman, P. J. Structural basis for germ-line
770 gene usage of a potent class of antibodies targeting the CD4-binding site of HIV-1 gp120. *Proc.*
771 *Natl. Acad. Sci.* **109**, E2083–E2090 (2012).
- 772 30. Havenar-Daughton, C. *et al.* The human naïve B cell repertoire contains distinct subclasses
773 for a germline-targeting HIV-1 vaccine immunogen. *Science Translational Medicine* **10**, (2018).
- 774 31. Matassoli, F. *et al.* High frequency of HIV precursor-target-specific B cells in sub-Saharan
775 populations. *Cell Rep.* **42**, 113450 (2023).
- 776 32. deCamp, A. C. *et al.* Human immunoglobulin gene allelic variation impacts germline-
777 targeting vaccine priming. *npj Vaccines* **9**, 58 (2024).
- 778 33. Campbell, M. C. & Tishkoff, S. A. African Genetic Diversity: Implications for Human
779 Demographic History, Modern Human Origins, and Complex Disease Mapping. *Annu. Rev.*
780 *Genom. Hum. Genet.* **9**, 403–433 (2008).
- 781 34. Tishkoff, S. A. *et al.* The Genetic Structure and History of Africans and African Americans.
782 *Science* **324**, 1035–1044 (2009).
- 783 35. Petrova, V. N. *et al.* Incomplete genetic reconstitution of B cell pools contributes to
784 prolonged immunosuppression after measles. *Science Immunology* **4**, (2019).

785 36. Stelekati, E. & Wherry, E. J. Chronic Bystander Infections and Immunity to Unrelated
786 Antigens. *Cell Host Microbe* **12**, 458–469 (2012).

787 37. Borkow, G. *et al.* Chronic immune activation associated with intestinal helminth infections
788 results in impaired signal transduction and anergy. *J. Clin. Investig.* **106**, 1053–1060 (2000).

789 38. Clerici, M. *et al.* Immune activation in Africa is environmentally-driven and is associated
790 with upregulation of CCR5. *AIDS* **14**, 2083–2092 (2000).

791 39. Lalor, M. K. *et al.* BCG Vaccination Induces Different Cytokine Profiles Following Infant
792 BCG Vaccination in the UK and Malawi. *J. Infect. Dis.* **204**, 1075–1085 (2011).

793 40. Parker, E. P. K., Kampmann, B., Kang, G. & Grassly, N. C. Influence of Enteric Infections
794 on Response to Oral Poliovirus Vaccine: A Systematic Review and Meta-analysis. *J. Infect. Dis.*
795 **210**, 853–864 (2014).

796 41. Parker, E. P. *et al.* Causes of impaired oral vaccine efficacy in developing countries. *Futur.*
797 *Microbiol.* **13**, 97–118 (2018).

798 42. Muyanja, E. *et al.* Immune activation alters cellular and humoral responses to yellow fever
799 17D vaccine. *J. Clin. Investig.* **124**, 3147–3158 (2014).

800 43. Bowyer, G. *et al.* Reduced Ebola vaccine responses in CMV+ young adults is associated with
801 expansion of CD57+KLRG1+ T cells. *J. Exp. Med.* **217**, e20200004 (2020).

802 44. Safari, M. S. *et al.* A natural experiment in Kenya reveals durable immunosuppressive effects
803 of early childhood malaria: a longitudinal cohort study. (2025) doi:10.7554/elife.107820.1.

804 45. Calle, C. L., Mordmüller, B. & Singh, A. Immunosuppression in Malaria: Do Plasmodium
805 falciparum Parasites Hijack the Host? *Pathogens* **10**, 1277 (2021).

806 46. Bediako, Y. *et al.* Repeated clinical malaria episodes are associated with modification of the
807 immune system in children. *BMC Med.* **17**, 60 (2019).

808 47. Corcoran, M. M. *et al.* Production of individualized v gene databases reveals high levels of
809 immunoglobulin genetic diversity. *Nature Communications* **7**, (2016).

810 48. Sok, D. *et al.* Priming HIV-1 broadly neutralizing antibody precursors in human Ig loci
811 transgenic mice. *Science* **353**, 1557–1560 (2016).

812 49. Briney, B., Inderbitzin, A., Joyce, C. & Burton, D. R. Commonality despite exceptional
813 diversity in the baseline human antibody repertoire. *Nature* **566**, 393–397 (2019).

814 50. DeKosky, B. J. *et al.* In-depth determination and analysis of the human paired heavy- and
815 light-chain antibody repertoire. *Nat. Med.* **21**, 86–91 (2015).

816 51. Havernar-Daughton, C. *et al.* Direct Probing of Germinal Center Responses Reveals
817 Immunological Features and Bottlenecks for Neutralizing Antibody Responses to HIV Env
818 Trimer. *Cell Reports* **17**, 2195–2209 (2016).

819 52. Lin, X. *et al.* Structural insights into VRC01-class bnAb precursors with diverse light chains
820 elicited in the IAVI G001 human vaccine trial. *Proc. Natl. Acad. Sci.* **122**, e2510163122 (2025).

821 53. Huang, J. *et al.* Identification of a CD4-Binding-Site Antibody to HIV that Evolved Near-Pan
822 Neutralization Breadth. *Immunity* **45**, 1108–1121 (2016).

823 54. Georgiev, I. S. *et al.* Delineating Antibody Recognition in Polyclonal Sera from Patterns of
824 HIV-1 Isolate Neutralization. *Science* **340**, 751–756 (2013).

825 55. Rodriguez, O. L. *et al.* Genetic variation in the immunoglobulin heavy chain locus shapes the
826 human antibody repertoire. *Nat. Commun.* **14**, 4419 (2023).

827 56. Omer, A. *et al.* VDJbase: an adaptive immune receptor genotype and haplotype database.
828 *Nucleic Acids Res.* **48**, D1051–D1056 (2019).

829 57. Corcoran, M. *et al.* Human IGH germline gene diversity and allele frequencies in 2486
830 individuals from 25 global populations delineated by ultra-high throughput genotyping. *bioRxiv*
831 2025.08.06.668935 (2025) doi:10.1101/2025.08.06.668935.

832 58. Henderson, R. *et al.* Selection of immunoglobulin elbow region mutations impacts
833 interdomain conformational flexibility in HIV-1 broadly neutralizing antibodies. *Nat. Commun.*
834 **10**, 654 (2019).

835 59. Yacoob, C. *et al.* Differences in Allelic Frequency and CDRH3 Region Limit the
836 Engagement of HIV Env Immunogens by Putative VRC01 Neutralizing Antibody Precursors.
837 *Cell Rep.* **17**, 1560–1570 (2016).

838 60. Zhou, T. *et al.* Structural Basis for Broad and Potent Neutralization of HIV-1 by Antibody
839 VRC01. *Science* **329**, 811–817 (2010).

840 61. Willis, J. R. *et al.* Vaccination with mRNA-encoded nanoparticles drives early maturation of
841 HIV bnAb precursors in humans. *Science* eadr8382 (2025) doi:10.1126/science.adr8382.

842 62. Braddom, A. E. *et al.* B Cell Receptor Repertoire Analysis in Malaria-Naive and Malaria-
843 Experienced Individuals Reveals Unique Characteristics of Atypical Memory B Cells. *mSphere*
844 **6**, 10.1128/msphere.00726-21 (2021).

845 63. Jardine, J. G. *et al.* Priming a broadly neutralizing antibody response to HIV-1 using a
846 germline-targeting immunogen. *Science* **349**, 156–161 (2015).

847 64. Kato, Y. *et al.* Multifaceted Effects of Antigen Valency on B Cell Response Composition
848 and Differentiation In Vivo. *Immunity* **53**, 548–563.e8 (2020).

849 65. Kimotho, J. *et al.* Lifelong malaria exposure drives B cell activation and post-transcriptional
850 regulation while suppressing inflammatory transcriptional programs. (2025)
851 doi:10.21203/rs.3.rs-6974016/v1.

852 66. Bernat, N. V. *et al.* High-quality library preparation for NGS-based immunoglobulin
853 germline gene inference and repertoire expression analysis. *Frontiers in Immunology* **10**, (2019).

854 67. Narang, S., Kaduk, M., Chernyshev, M., Hedestam, G. B. K. & Corcoran, M. M. Adaptive
855 immune receptor genotyping using the corecount program. *Front Immunol* **14**, 1125884 (2023).

856 68. Gadala-Maria, D., Yaari, G., Uduman, M. & Kleinstein, S. H. Automated analysis of high-
857 throughput B-cell sequencing data reveals a high frequency of novel immunoglobulin V gene
858 segment alleles. *Proceedings of the National Academy of Sciences of the United States of*
859 *America* **112**, E862–E870 (2015).

860 69. Gupta, N. T. *et al.* Change-O: a toolkit for analyzing large-scale B cell immunoglobulin
861 repertoire sequencing data. *Bioinformatics* **31**, 3356–3358 (2015).

862 70. Gidoni, M. *et al.* Mosaic deletion patterns of the human antibody heavy chain gene locus
863 shown by Bayesian haplotyping. *Nat. Commun.* **10**, 628 (2019).

864 71. Tareen, A. & Kinney, J. B. Logomaker: beautiful sequence logos in Python. *Bioinformatics*
865 **36**, 2272–2274 (2020).

866

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryData1.xlsx](#)
- [SupplementaryData2.xlsx](#)
- [SupplementaryData3.xlsx](#)
- [Supplementaryfiles.pdf](#)