Title

Development of antibody-dependent cell cytotoxicity function in HIV-1 antibodies

Authors

Laura E. Doepker^{1,2}, Sonja Danon^{1,2}, Elias Harkins³, Duncan Ralph³, Zak Yaffe^{1,5}, Amrit Dhar^{3,4}, Cassia Wagner⁵, Megan M. Stumpf¹, Dana Arenz¹, James A. Williams⁶, Walter Jaoko⁷, Kishor Mandaliya⁸, Kelly K. Lee⁶, Frederick A. Matsen IV³, Julie M. Overbaugh^{1,3*}

Affiliations

¹Human Biology Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA

²These authors contributed equally

³Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA

⁴Department of Statistics, University of Washington, Seattle, WA 98195, USA

⁵Medical Scientist Training Program, University of Washington School of Medicine, Seattle, WA 98195, USA

⁶Department of Medicinal Chemistry, University of Washington, Seattle, WA 98195, USA

⁷Department of Medicinal Microbiology, University of Nairobi, Nairobi, Kenya

⁸Coast Provincial General Hospital, Women's Health Project, Mombasa, Kenya

*Corresponding author and lead contact; joverbau@fredhutch.org

Abstract

A prerequisite for the design of an HIV vaccine that elicits protective antibodies is understanding the developmental pathways that result in desirable antibody features. The development of antibodies that mediate antibody-dependent cellular cytotoxicity (ADCC) is particularly relevant because such antibodies have been associated with HIV protection in humans. We reconstructed the developmental pathways of six human HIV-specific ADCC antibodies using longitudinal antibody sequencing data. Most of the inferred naïve antibodies did not mediate detectable ADCC. Gain of antigen binding and ADCC function typically required mutations in complementarity determining regions of one or both chains. Enhancement of ADCC potency often required additional mutations in framework regions. Antigen binding affinity and ADCC activity were correlated, but affinity alone was not sufficient to predict ADCC potency. Thus, elicitation of human ADCC antibodies may require mutations that first enable

high affinity antigen recognition, followed by mutations that optimize factors contributing to functional ADCC activity.

Introduction

A major concept underlying HIV vaccine research is that we can learn from the processes that lead to potent antibody responses in natural infection to inform immunogen design. For this reason, there have been several detailed studies of the evolutionary pathways of potent and broad neutralizing antibodies (bnAbs) to HIV (Bonsignori et al., 2017; Bonsignori et al., 2016; Doria-Rose et al., 2014; Landais et al., 2017; Liao et al., 2013; MacLeod et al., 2016; Simonich et al., 2019; Wu et al., 2015). These studies highlight the critical role of somatic hypermutation (SHM), particularly in complementary determining regions (CDRs) of antibody heavy chains, in driving the breadth and potency of HIV bnAbs (Kwong & Mascola, 2018). One recent study also highlighted that maturation of the antibody light chain can also be critical for breadth (Simonich et al., 2019). Collectively, these studies have provided important insights into the specific mutations that drive maturation of antibody lineages and specificities of bnAbs.

The keen interest in HIV bnAbs stems in part from a plethora of compelling studies that demonstrate the efficacy of bnAbs in protection of macaques from experimental SHIV infection (Pegu, Hessel, Mascola, & Haigwood, 2017). The data are less convincing for sterilizing protection by HIV-specific non-neutralizing antibodies that mediate antibody-dependent cell cytotoxicity (ADCC) in the same experimental animal models (Fouts et al., 2015), where the functional interactions of antibodies are harder to test due to species differences (Bournazos, Wang, Dahan, Maamary, & Ravetch, 2017). However, they have been implicated in protection from HIV-infection in humans in several settings. In the only partially effective HIV vaccine trial to date, antibodies that mediate ADCC were associated with protection, whereas neutralizing antibodies were not (Haynes et al., 2012). Moreover, in the setting of mother-to-child-transmission, ADCC antibodies have been associated with both decreased transmission risk and disease progression in infants (Mabuka, Nduati, Odem-Davis, Peterson, & Overbaugh, 2012; Milligan, Richardson, John-Stewart, Nduati, & Overbaugh, 2015). For these reasons, ADCC antibodies, which have the potential to eliminate infected cells, have been increasingly recognized as important to consider for vaccine design and prevention efforts (Lewis et al., 2017). Yet, nothing is known about the process of SHM required for HIV specificity and ADCC function of antibodies.

In light of the relevance of ADCC antibodies in human infections, we deep sequenced the antibody repertoire of an individual who developed potent ADCC antibodies to the HIV

envelope (Env) gp120 and gp41 proteins. We used methods that were specifically developed to study antibody evolutionary pathways to infer the naïve precursors and developmental intermediates of six mature ADCC-mediating antibodies that have been previously described (Williams et al., 2015; Williams et al., 2019). Most inferred naïve antibodies did not mediate ADCC. To develop ADCC functionality, inferred naïve antibodies required CDR-localized SHM in one or both chains. In five of six cases, enhancement of ADCC potency to mature levels required mutations in framework regions (FWR), either alone or alongside additional CDR mutations. Interestingly, in one lineage, the developing antibodies first gained the capacity to bind HIV Env and then subsequently acquired ADCC activity due to additional SHM. Overall, binding affinity for Env and facilitation of ADCC were correlated, but we observed cases where binding affinity was similar amongst two antibodies within the same lineage but they differed in their ADCC capacity, suggesting that affinity alone is not sufficient for an antibody to mediate ADCC.

Results

Longitudinal sequencing of six ADCC HIV-1 antibody lineages

We previously reported the isolation of six monoclonal antibodies (mAbs), each with potent antibody-dependent cell cytotoxicity (ADCC) activity, from a sample collected 914 days post-infection (D914) from a clade A-infected Kenyan woman (Williams et al., 2015; Williams et al., 2019). All of these mAbs derived from distinct B cell lineages, with four targeting two distinct gp41 Env epitopes (006, 016, 067, 072), and the other two targeting distinct gp120 Env epitopes (105, 157). One of the gp120 targeting mAbs (105) also had modest capacity to neutralize cell-free virus (Williams et al., 2015), whereas the others were non-neutralizing. To infer the ontogeny of these six mAbs, we used next-generation sequencing (NGS) to sequence antibody variable region genes from five longitudinal blood samples collected from the subject, spanning from pre-infection (D-119) to over 4 years post-infection (D1512). Amongst the different antibody chains and timepoints sequenced, the sequencing libraries ranged widely in their sample coverage, between 6-409% with an average of 100% coverage, based on estimated B cell frequencies within the available peripheral blood mononuclear cell (PBMC) samples (Table 1). It is important to note that, despite the name “deep sequencing,” NGS-sampled datasets are relatively shallow compared to the entire repertoire. This is for two reasons: 1) each 10 mL PBMC sample is only ~0.22% of the subject’s whole-body circulating blood volume (~4.5 L), not counting lymphoid organs or peripheral tissue, and 2) circulating human memory B cell repertoires fluctuate over time (Horns, Vollmers, Dekker, & Quake, 2019; Laserson et al., 2014),

making it difficult to accurately track clonal B cell families. Despite these limitations inherent to studies of this type, our sequencing efforts successfully returned clonal sequences for all six lineages of interest.

QA255 time point	Live PBMC count (excluding non-viable)	PBMC viability	Ab Chain	Replicate	Raw MiSeq reads	Productive replicate-merged deduplicated sequences	Estimated sequencing coverage within sampled blood PBMCs ^b	Estimated sequencing coverage within QA255 d914 whole-body blood repertoire ^c
D-119	1.39E+07	84.22%	IgM	1	486458	200254	21%	0.05%
			IgG	1	604237	199570	48%	0.11%
			IgK	1	705163	292543	21%	0.05%
			IgL	1	406267	178221	13%	0.03%
D462	1.40E+07	86.42%	IgG	1	646809	504029	120%	0.22%
				2	923103			
			IgK	1	57096	506989	36%	0.08%
				2	967693			
			IgL	1	201566	402456	29%	0.06%
				2	586568			
D791	3.90E+06	97.50%	IgG	1	126955	263848	226%	0.22%
				2	776578			
				3 ^a	522308			
			IgK	1	527031	568736	146%	0.22%
				2	934780			
				3 ^a	1057984			
			IgL	1	47368	212922	55%	0.12%
				2	387730			
				3 ^a	537611			
D1174	1.80E+06	100.00%	IgG	1	497094	220740	409%	0.22%
				2	665981			
				3 ^a	678973			
			IgK	1	254890	415601	231%	0.22%
				2	797239			
				3 ^a	929145			
			IgL	1	228992	307415	171%	0.22%
				2	453935			
				3 ^a	510711			
D1512	1.00E+07	79.37%	IgG	1	370198	155262	52%	0.12%
			IgK	1	408498	214651	21%	0.05%
			IgL	1	93848	55785	6%	0.01%
					averages:	293689	100%	0.13%

Table 1. Longitudinal QA255 antibody variable region sequencing statistics.

^a Library replicate performed using unique molecular identifiers during cDNA synthesis step.

^b Sequencing coverage was calculated for PBMCs using QA255 D914 B cell frequency statistics from the cell sort (Williams et al., 2015).

^c Whole-body sequencing coverage was calculated assuming 10 ml of blood was sampled from a total of 4500 ml adult blood volume.

We identified sequences clonally-related to the gamma heavy (VH) and lambda (VL) or kappa (VK) light chains of the six mature mAbs (Table 2; Table 2-figure supplement 1A-B), inferred the most likely naive ancestors of each mAb chain, annotated clonal family V, (D), and J gene usage (Table 2), and calculated the degree of mutation in each lineage over time (Table 2; Table 2-figure supplement 1C). Overall, the level of SHM in the mature D914 mAbs ranged from 3.3-11.5%, similar to our previous reporting of these lineages using different annotation methods (Williams et al., 2015; Williams et al., 2019). Incidentally, consensus among clonal

sequences within families revealed potentially artifactual cloning mutations at the 5' and 3' ends of the variable regions in some of the previously characterized mature mAbs (Williams et al., 2015; Williams et al., 2019) and suggested, overall, that the consensus sequences were the most likely sequences for the antibodies. When directly compared, the sequence-corrected mAbs demonstrated equivalent epitope binding strength and ADCC activity to those previously studied (data not shown). Thus, we corrected the 5' and 3' ends of the mature D914 mAb variable regions to consensus sequences among each clonal family for use in this study (Supplementary File 1).

A

Heavy chain	Lineage	Mature antibody statistics				Clonal family (all timepoints merged, not downsampled)		Clonal family statistics within repertoire (downsampled analyses)															No. ints resolved in lineage path	
		Gene usage (partis)			%SHM mature Ab	# sampled clonal heavy chain seqs	Avg clonal family %SHM	No. unique clonal sequences identified					Size: percent of sampled repertoire					Average %SHM (nt)						
		VH	DH	JH				D-119	D462	D791	D1174	D1512	D-119	D462	D791	D1174	D1512	D-119	D462	D791	D1174	D1512		
	006	3-23*01	3-22*01	4*02	8.7	7	10.8		0	7	0	0	0	0.002%					7.6					1
	016	4-34*01	6-13*01	5*02	9.9	702	9.2		0	581	112	0	9	0.13%	0.05%		0.01%		9.6	10.3			12.6	14
	067	1-69*11#	1-1*01	6*03	8.1	36	11.9		0	35	1	0	0	0.01%					11.6					2
	072	1-69*11#	4-17*01	3*02	11.5	515	12.5		0	11	118	325	61	0.04%	0.05%	0.17%	0.04%		9.5	12.8	13.0	15.1	5	
	105	3-15*01	3-22*01	6*03	9.5	193	4.9		0	193	0	0	0	0.04%					5					2
	157	1-69*11#	1-26*01	6*03	6.5	183	7.4		0	6	165	8	4	0.004%	0.08%	0.04%	0.03%		9.3	7.7	10.3	11.8	4	

B

Light chain	Lineage	Mature antibody statistics				Clonal family (all timepoints merged, not downsampled)		Clonal family statistics within repertoire (downsampled analyses)															No. ints resolved in lineage path
		Gene usage (partis)			%SHM mature Ab	# sampled clonal heavy chain seqs	Avg clonal family %SHM	No. unique clonal sequences identified					Size: percent of repertoire					Average %SHM (nt)					
		VK	VL	JL				D-119	D462	D791	D1174	D1512	D-119	D462	D791	D1174	D1512	D-119	D462	D791	D1174	D1512	
	006		2-11*01	2*01	6.8	673	6.1	12	274	273	52	62	0.14%					7.1					7
	016		1-51*02	2*01	9.9†	1518	7.3	299	339	523	321	36	0.03%	0.02%		0.11%		6.6	7.1			7.7	11
	067		2-11*01	3*02	3.3	20337	6.5	2149	6105	3408	7946	729	2.04%					6.6					2
	072	1-27*01		1*01	7.8	2796	6.2	476	719	702	543	356	0.17%	0.18%	0.18%	0.25%		6.3	6.3	6.6		7.0	5
	105	3-20*01		2*01	7.1	30286	8.0	3058	4884	5607	14271	2466	1.17%					6.3					7
	157	3-11*01		5*01	5.0	18	8.2	0	12	0	1	5	0.24%			0.07%		5.9					7

Table 2. Clonal sequence analyses for ADCC antibody lineages. (A, B) Characteristics and statistics of heavy (A) and light (B) chain clonal families. Percent SHM was calculated as the mutation frequency at the nucleotide level compared to the predicted naïve allele, as determined by the per subject germline inference for QA255. Statistics calculated for individual timepoints within the context of the full subject repertoire were downsampled to 50-150K sequences for computational manageability. Light chain clonal family size statistics are reported but unreliable due to overclustering caveats that are explained in the methods. Per-timepoint clonal family statistics were excluded if ≤ 1 clonal heavy or light chain sequences were identified. #: The VH gene used in 067 and 157 lineages was determined to be an allele not cataloged in IMGT® (www.imgt.org): 1-69*11+A147G+C169T; the 157 lineage VK gene was also a new allele: 3-11*01+T5A+T9A+T36G+G84A. †: Percent SHM for 016 light chain lineage includes a 3-nt insertion in the CDRL3. SHM: somatic hypermutation; ints: intermediates. Longitudinal values for “No. of unique clonal sequences identified” are plotted for heavy and light chains in Table 2-figure supplement 1, along with longitudinal values for “Average %SHM (nt)” for heavy chains. Graphics displaying the most probable routes of antibody lineage maturation corresponding to the “No. of ints resolved in lineage path” are available in Table 2-figure supplement 2.

Table 2-figure supplement 1. Longitudinal characteristics of heavy and light chain clonal families. (A, B) Number of clonal heavy (A) and light (B) chain sequences identified per sequenced timepoint per lineage. Caveats relating to light chain clusters are described in Methods. (C) Average variable region mutation for each clonal heavy chain family per timepoint. Dashed line at D914 indicates the timepoint of isolation of mature antibodies (Williams et al., 2015; Williams et al., 2019). SHM: somatic hypermutation.

Table 2-figure supplement 2. Most probable routes of antibody lineage maturation. (included at end of document, as it is multiple pages)

Supplementary File 1. Sequences of six ADCC antibody lineages. (fasta file) Included are inferred naïve, computationally-inferred lineage intermediate, manually-inferred lineage intermediate, corrected mature, and NGS sequences with high sequence identity to lineage members and/or mature sequences. Sequences are provided as Supplementary File 1 because computationally-inferred sequences cannot be deposited into GenBank.

mAbs constructed from D462 NGS sequences mediated ADCC

Each lineage varied in the time point(s) from which we identified clonal sequences, with most clonal VH sequences existing in the D462 datasets for four of the six lineages (Table 2A). The NGS sequences within each lineage that shared the highest nucleotide (nt) identity with the D914 mature VH variable regions, ranging from 88-99%, were found within the D462 or D791 timepoints (Supplementary File 2). As test cases, we synthesized antibodies with VH sequences derived from the D462 NGS results that were similar to the mature 006 and 016 mAbs (88.5% and 97.5% nt identity, respectively). The 006-NGS_{VH} was paired with 006-NGS_{VL} light chain that

had 95.8% nt identity to the mature 006 VL (Figure 1-figure supplement 1A) and, since this option was not available for 016 VL, the 016-NGS_{VH} was paired with a computationally-inferred light chain (016-2_{VL}) that preceded the mature in development (Figure 1-figure supplement 1B). Both NGS-based mAbs demonstrated full ADCC function (Figure 1), suggesting that ADCC function likely developed within these lineages by D462 post infection.

Supplementary File 2. Select NGS sequences with highest sequence identity to mature D914 mAb sequences. (fasta file)

Figure 1. ADCC activity in antibodies made from NGS-sampled sequences. RFADCC activity of antibodies that use NGS-sampled heavy chains that are most similar in sequence to the mature 006 or 016 antibodies. The 006-NGS_{VH} was paired with an NGS-sampled light chain; the 016-NGS_{VH} was paired with a computationally-inferred mature-like sequence 016-2_{VL}. Data reflect two independent experiments.

Figure 1-figure supplement 1. Graphic summaries of antibodies comprised of paired VH and VL NGS-sampled sequences. Graphic summaries of paired VH and VL antibody sequences for 006 (A) and 016 (B) lineages relative to inferred naïve (0) sequences. Shaded regions demarcate CDRs (gray). Vertical black lines indicate variable region amino acid substitutions. m: mature.

Inferred naïve mAbs vary in antigen binding capability and ADCC function

Inferred naïve ancestor mAbs across lineages varied in their ability to bind HIV Env and facilitate ADCC function. Of the gp41-targeted lineages, only the inferred naïve precursor of the 072 lineage bound gp41 ($K_D = 34.4$ nM) and mediated detectable ADCC, although the activity was very low (Figure 2A,C). Both naïve mAbs from the gp120-targeted lineages bound monomeric gp120 ($K_D = 23.4$ nM for 105-0_{VH0VK} and 41.1 μ M for 157-0_{VH0VK}; Figure 2B). Interestingly, despite the 105 naïve mAb having higher binding affinity compared to 157 naïve mAb, it did not mediate ADCC, while the 157 naïve mAb mediated potent ADCC (Figure 2C).

Typically, the uncertainty in the computational inference of antibody lineage naïve precursors is largely ignored, even though single amino acid differences in the naïve mAb can result in differences in HIV Env binding and functional properties (Yuan, Li, & Zhang, 2011). To account for inference uncertainty, we also tested “alternative naïve” mAbs for each lineage that were inferred not by partis (Ralph & Matsen, 2016b), but by linearham (Dhar, Ralph, Minin, & Matsen, 2020), a second computational method that jointly models V(D)J recombination and evolutionary history. For 7 of 12 antibody chains, this alternative approach generated naïve sequences that differed from the original naïve sequences by one to four amino acids within the CDR3 and/or FWR4 regions (Supplementary File 1). For the remaining five, the alternative method inferred the same naïve sequence as the original method. When tested, the alternative naïve sequences did not alter antigen binding or function in five of six antibody lineages when compared to the original naïve mAbs (Figure 2-figure supplement 1A). However, in the 157 lineage we found that the alternative naïve VK chain (157-A_{VK}) differed from the original naïve (157-0_{VK}) by two adjacent amino acids in the CDRL3 (Figure 2-figure supplement 1D). When 157-A_{VK} was paired with either the original (157-0_{VH}) or alternative (157-A_{VH}) VH chains, it abolished gp120 binding and ADCC functionality (Figure 2-figure supplement 1B,C). Thus, there is some uncertainty about whether the true 157 lineage naïve precursor was specific for HIV and/or capable of mediating ADCC.

Figure 2. Inferred naïve mAbs from six lineages vary in antigen binding capability and ADCC function. (A-B) Binding kinetics of the inferred naïve antibody (ligand) from each lineage to indicated concentrations of monomeric C.ZA.1197MB gp41 ectodomain (A) or BL035.W6M.C1 gp120 (B) (analyte). Best fitting lines (red) to a 1:1 binding model of ligand:analyte binding are shown. Data are representative of two independent experiments. (C) Positive control-normalized RFADCC activity of inferred naïve antibodies compared to their respective mature antibodies. Normalization is described in Methods. Data are represented as mean $\pm$ SEM and reflect at least five independent experiments. 0: naïve; m: mature. See also Figure 2-figure supplement 1.

Figure 2-figure supplement 1. Alternative naïve mAb functionality. (A, C) Functional RFADCC comparisons of each lineage’s “original” (0_{VH} and 0_{VL}) and “alternative” (A_{VH} and A_{VL}) inferred naïve antibodies. In cases where the alternative approach did not infer a naïve sequence that differed from the original naïve sequence identified by partis, alternative chains were paired with original chains, as indicated. For the 157 lineage (C), original and alternative heavy and light chains were also cross-paired. Data are representative of at least two independent experiments. (B) Binding of indicated 157-lineage naïve antibodies (analyte) to monomeric BL035 gp120 (ligand, 250 nM), measured by BLI without the “double reference” approach described in Methods. (D) Graphic summary of paired VH and VL antibody sequences relative to inferred naïve 157-0_{VH}0_{VK}. Shaded regions demarcate CDRs (gray). Vertical black lines indicate variable region amino acid substitutions. 0: original naïve; A: alternative naïve; m: mature

Characterization of lineage intermediates that confer HIV specificity and ADCC function

To reconstruct probable developmental routes that led to the mature D914 mAb heavy and light chains, we used Bayesian phylogenetic lineage inference, which emphasizes only the inferred ancestral intermediates that have high relative statistical confidence (Simonich et al., 2019). Among the twelve heavy and light chain lineages for the six mAbs, we resolved anywhere from 1 to 14 probable intermediate sequences that lay between the naïve and the mature sequences (Table 2, Table 2-figure supplement 2). Of note, 8 of 12 of these

computationally-inferred lineages were validated by the existence of NGS-sampled sequences that were identical either in nucleotide or amino acid sequence to at least one of the inferred intermediates (Table 2-figure supplement 2).

To focus our subsequent lineage studies on determining which mutations enabled HIV Env binding and ADCC gain-of-function, we employed the following strategy to select inferred intermediates of interest (detailed fully in Methods). First, we chose lineage intermediates with high relative confidence that were near the middle of each inferred lineage for preliminary experiments to determine whether antigen binding or ADCC gain-of-function occurred before or after these “middle” intermediate (Table 2-figure supplement 2). If the middle intermediate had function, we focused on pre-middle inferred intermediates, if available within the computationally-inferred lineages. If the middle intermediate lacked function, we focused on post-middle inferred intermediates. mAbs comprised of select intermediate sequences were tested to define the mutations that conferred HIV binding and ADCC activity in each lineage as follows:

gp41 antibody 006: mAb 006 (VH3-23, VL2-11) targets a discontinuous epitope that includes the C-terminal heptad repeat (Williams et al., 2019). Mutations in CDRH2 and CDRH3, totaling 1.4% VH SHM, were required for antigen binding and ADCC activity by 006-1_{VH0VL} in the 006 lineage (Figure 3A). Thereafter, six light chain mutations among CDRL1, CDRL3, and FWRL2 augmented gp41 binding and ADCC potency in 006-1_{VH1VL}.

gp41 antibody 016: mAb 016 (VH4-34, VL1-51) targets a similar gp41 epitope as 006 (Williams et al., 2019). Detectable gp41 binding and ADCC function was demonstrated by the 016-1_{VH1VL} mAb which contains mutations in heavy and light chain CDRs and FWRs (5.6% VH SHM and 3.6% VL SHM) (Figure 3B). A subsequent CDRL3 insertion of residue S94 allowed for augmented ADCC capacity (016-1_{VH2VL}). Additional FWR and CDR mutations in the VH chain (016-2_{VH2VL}) further strengthened gp41 binding and ADCC functionality.

gp41 antibody 067: mAb 067 (VH1-69, VL2-11) targets the C-C' loop (Williams et al., 2019). For the 067 lineage, the 067-0_{VH1VL} mAb incorporated a single CDRL3 mutation (0.3% VL SHM) and displayed detectable antigen binding; we could not determine whether there was meaningful ADCC function or not as levels were just above background (Figure 3C). To achieve more convincing ADCC function, two CDRH2 mutations were also required (1.0% VH SHM). Further mutations in either the heavy or light chain FWRs modestly increased ADCC potency to >50% that of the 067 mature mAb.

gp41 antibody 072: mAb 072 (VH1-69, VK1-27) targets an overlapping but distinct epitope as 067 (Williams et al., 2019). Unlike the other gp41-targeting lineages, the 072 lineage

inferred naïve mAb bound gp41 antigen and mediated weak ADCC, as aforementioned (Figures 2A,C and 3D). To gain ADCC potency, only mutations in the VH chain were necessary; mAbs 072-1_{VH}0_{VK} (0.8% VH SHM) and 072-2_{VH}0_{VK} (0.5% VH SHM) demonstrate that two substitutions in either the CDRH1 or CDRH2 increased activity, but, when combined in 072-3_{VH}0_{VK}, they were not additive or synergistic (Figure 3D). Instead, FWRH3 and FWRH4 mutations in mAb 072-4_{VH}0_{VK} were required for ADCC activity comparable to that of 072 mature mAb.

gp120 antibody 105: mAb 105 (VH3-15, VK3-20) targets a V3 linear epitope (Williams et al., 2015). The inferred 105 lineage naïve and 105-0_{VH}1_{VK} mAbs bound gp120 but displayed only indeterminate levels of ADCC activity, as defined in Methods (Figure 4A). 105-1_{VH}1_{VK} mAb demonstrates that heavy chain CDRH1 and CDRH2 mutations (1.9% VH SHM) were required for ADCC gain of function, even though these mutations did not affect binding affinity (Figure 4A). We note that the FWR4 mutation in the 105-1_{VH} chain was likely a computational artifact and unlikely to affect mAb function, as this mutation is absent in the mature 105 mAb.

gp120 antibody 157: mAb 157 (VH1-69, VK3-11) targets a CD4-induced, C11-like epitope (Williams et al., 2015). As detailed in Figure 2-figure supplement 1 and further shown in Figure 4B, inferred naïves 157-0_{VH}A_{VK} and 157-0_{VH}0_{VK} differed in their binding and ADCC capabilities, leaving us uncertain about whether this lineage was able to bind gp120 and facilitate ADCC since inception or gained these capacities following two CDRL3 mutations (Figure 4B).

Figure 4. HIV specificity and ADCC development in gp120-targeted antibody lineages. RFADCC activity (left), antigen binding to monomeric BL035.W6M.C1 (analyte, 62.5 nM) as measured by BLI (middle), and mAb heavy and light chain pairings (right) for the 105 and 157 lineages all displayed as in Figure 3. Asterisks indicate indeterminate activity, as defined in Methods. RFADCC data are represented as mean $\pm$ SEM and reflect at least four independent experiments, including data presented in Figure 2 for naïve and mature Abs to best account for assay variability.

Themes in ADCC development pathways among six lineages

Overall, the majority of lineages, regardless of epitope specificity, ultimately required substitutions in both the heavy and light chains to develop ADCC function >50% as potent as their respective mature mAb (Figure 5). In five of six lineages, detectable binding accompanied ADCC capacity, with lineage 105 being the exception that bound prior to developing ADCC function. Most notably, substitutions in CDRs were typically required for ADCC gain-of-function, while subsequent FWR substitutions, either alone or alongside additional CDR mutations, augmented this activity (Figure 5). A graphical summary of the key developmental steps for each lineage is presented in Figure 5-figure supplement 1.

Figure 5. Locality of mutations required for gain of function. Summary of the location of mutations critical for acquiring binding and activity functions in all five ADCC Ab lineages, respective to the previous step, i.e. building upon one another to gain functions. Lineages are summarized by row. Detectable ADCC function excludes lineage members with indeterminate ADCC activity, as defined in Methods. Augmented ADCC activity indicates lineage members with activity most comparable to that of their respective mature. *157-0_{VH}A_{VK} required mutations for function while 157-0_{VH}0_{VK} did not; Env binding and detectable ADCC determinants were defined based on this comparison. fnl: functional.

Figure 5-figure supplement 1. Key steps in ADCC development among six lineages. Graphic summaries of development for each antibody lineage. Antibodies (ovals) are arranged from naïve (top) to mature (bottom) with inferred intermediates in between, colored by ADCC functionality: no function or indeterminate (gray), low function (<50% of mature activity; light blue), high function (>50% of mature activity; dark blue). Arcs connecting ovals indicate amino acid substitutions in the heavy chain (left) and light chain (right), colored by their locality: CDR (yellow), FWR (blue), both (green), or no change (gray, thin). The red outline within the 016 lineage denotes that the light chain mutation is an insertion, not a substitution. The blue outline within the 105 lineage denotes the presence of a FR4 mutation that was likely a computational artifact, since this position backmutates within the mature antibody.

Antigen binding affinity and ADCC function correlate

In our detailed analyses of binding and ADCC activity, there were three lineages (067, 072, and 105) where we observed two mAbs from the same lineage having similar binding affinities but different ADCC capabilities. To visualize the relationship between binding affinity and ADCC potency, we plotted mAb binding affinities (K_D) against corresponding RFADCC potencies and found that these functionalities were positively correlated in all lineages (Figure 6). It is notable that the antibodies in the 157 lineage that mediate potent ADCC (naïve 157- $0_{VH}0_{VK}$ and mature 157-m $_{VH}m_{VK}$) bound gp120 antigen with much lower affinity than antibodies with similar ADCC potency in the other five gp41- or gp120-targeted lineages. These data indicate that binding affinity alone does not dictate ADCC potency.

Figure 6. Antigen binding affinity and ADCC function correlate. For antibodies within lineages 006, 016, 067, 072, and 105 (indicated by different colors), the log of reciprocal binding affinity, as determined by representative BLI data of least two independent experiments, is plotted against positive control-normalized RFADCC activity (the data shown in Figures 3 and 4). Antibodies that did not detectably bind antigen by BLI are excluded as none of these mediated ADCC. Dotted lines highlight comparisons between antibodies within the same lineage that have similar binding affinities, but different RFADCC activities.

Discussion

Renewed interest in ADCC-capable antibodies as important to HIV vaccine responses has highlighted the need for a better understanding of their natural development (Forthal & Finzi, 2018). Here we report the evolution of six ADCC antibody lineages within a single HIV-infected individual. The inferred naïve precursors of these ADCC lineages varied in their abilities to bind HIV antigen, a finding that agrees with studies of HIV-neutralizing antibody lineages (Stamatatos, Pancera, & McGuire, 2017). To achieve potent ADCC activity, most lineages required mutations in both their heavy and light chains. Generally, ADCC function was first achieved through mutations in CDRs, while increased ADCC potency required additional mutations in FWR. ADCC activity accompanied antigen binding in all lineages except one; in the exception the V3/gp120-targeting 105 lineage demonstrated binding capacity prior to developing ADCC functionality. Although there was some evidence that binding affinity was not solely responsible for ADCC capability or potency, binding affinity and ADCC activity were largely correlated. In sum, this study presents six examples of developmental pathways taken by ADCC-mediating antibodies, highlighting that there are common themes in the ontogeny of HIV-specific ADCC antibodies, but that each evolutionary pathway has unique features.

Although not widely applied, antibody lineage studies require robust inference methods that account for statistical uncertainty if we are to be confident in their conclusions. This is highlighted in this study by the different properties displayed by two very similar inferred naïve precursors of the 157 lineage. The development of the 157 lineage was ambiguous: computational uncertainty in two CDRL3 residues resulted in multiple probable naïve antibodies with different properties. In these different scenarios, the 157 naïve Ab was either fully capable of both antigen binding and potent ADCC, or it required two CDRL3 substitutions to achieve both functions. In the other five lineages, our use of multiple computational approaches allowed us to report results with high confidence, since the uncertainties present in computational inference of those lineages did not affect functional characteristics.

Three gp41-specific antibody lineages were not capable of binding HIV Env by their inferred naïve precursors, as measured by BLI; they developed this ability through mutation. The lack of detectable antigen binding by inferred naïve precursors is not altogether surprising as this has been observed for a number of HIV bnAbs (Stamatatos et al., 2017). Binding capacity was gained in two of the gp41-directed lineages (006 and 067) through limited mutations localized in the CDRs of single antibody chains (either heavy or light, respectively). The other gp41-directed lineage, 016, which targets a similar epitope as the 006 lineage, required many mutations amongst the CDRs and FWRs of both chains, highlighting that the

gain of activity by different lineages can vary even if they are targeting the same epitope. As is true for all studies of this type, it is possible that our methods were not sensitive enough to detect low level binding or that, despite our best efforts, the inferred naïve sequences were inaccurate. Gp41-directed antibodies have been found to recognize self-antigens (Verkoczy & Diaz, 2014; Verkoczy, Kelsoe, & Haynes, 2014) and this, or its response to another antigen, could explain how the 016 lineage emerged before gaining HIV-specificity.

We chose the RFADCC assay because RFADCC signal has been repeatedly correlated with disease outcome in humans (Dhande et al., 2018; Lewis et al., 2019; Mabuka et al., 2012; Madhavi et al., 2014; Milligan et al., 2015; Ruiz et al., 2017), thus making it relevant to understanding the development of potentially protective responses. There was evidence that the 072 lineage had ADCC function from inception, although the amount of ADCC was low. The remaining lineages required an average of 1.2% SHM to gain detectable ADCC function (ranging 0.1-4.3% SHM). For augmentation of ADCC function to >50% of mature activity, the lineages required an average of 2.2% SHM (range 0.6-4.8% SHM). NGS-sampled sequences that were present at 462 dpi displayed high function with 5.3% SHM (006 lineage). Low mutation levels to achieve potent ADCC are promising for ADCC vaccine efforts, especially since these six characterized lineages have desirable characteristics for combating HIV, such as potent ADCC activity, cross-clade ADCC breadth, highly conserved epitopes, and potential to help clear infected cells (Williams et al., 2015; Williams et al., 2019). Moreover, these levels of mutation are easily attainable by vaccines (Schramm & Douek, 2018).

Importantly, most lineages required FWR mutations in one or both chains to achieve potent ADCC functionality, with or without additional CDR mutations. Though they can cost thermostability (Henderson et al., 2019), FWR mutations are known to contribute to the properties of HIV-specific bnAbs, predominantly by altering Ab conformation in ways that allow conserved epitopes to be bound with high affinity (Henderson et al., 2019; Kepler & Wiehe, 2017; Ovchinnikov, Louveau, Barton, Karplus, & Chakraborty, 2018). FWR residues contribute to the structure of the Ab and, though they do not usually directly contact the antigen, they can affect binding affinity indirectly (Foote & Winter, 1992). Furthermore, since they indeed affect Ab structure, FWR mutations could confer favorable Ab flexibility or an orientation that mediates better Fc receptor dimerization and ADCC potency (Acharya et al., 2014). FWR mutations in Fab “elbow” regions can affect conformational flexibility and paratope plasticity during bnAb development (Henderson et al., 2019). Here, the 006-1_{VH} lineage intermediate caused ADC gain-of-function and it contained, among other mutations, a CDRH3 elbow mutation Y111F.

Future structural studies are warranted to explore how FWR mutations affect Ab structure and function in these lineages.

Binding affinity has been implicated in the regulation of ADCC in several studies on tumor-specific Abs (Mazor et al., 2016; Tang et al., 2007). Here, while binding affinity correlated with ADCC function, it was not solely responsible for ADCC potency. In one lineage, binding capacity preceded detectable ADCC function. Notably, we found that half of the lineages contained clonally-related antibodies with similar binding affinity to one another but markedly different ADCC potencies. Conversely, mAbs within the two gp120-targeted lineages exemplify that there can be dramatic differences in binding affinity among mAbs that mediate equivalent, potent ADCC activity. Thus, ADCC function cannot be predicted based on affinity alone. These data support the notion that factors such as epitope specificity, binding mode, and other antibody-antigen interaction dynamics contribute to ADCC potency, as other studies have emphasized (Acharya et al., 2014; Orlandi et al., 2020; Tolbert et al., 2020)..

Our study offers a granular understanding of how ADCC functionality developed in six human Ab lineages. The results of this study are relevant to rational vaccine design, especially as technologies for guiding Ab development continue to emerge through reverse vaccinology efforts (Burton, 2017; Rappuoli, Bottomley, D'Oro, Finco, & De Gregorio, 2016; Schramm & Douek, 2018). Additionally, the reliance on FWR mutations to achieve high ADCC functionality could inform mAb optimization strategies for therapeutic uses.

Materials and Methods

Human subject

Peripheral blood mononuclear cell (PBMC) samples were obtained between 1997-2002 from a female HIV-1 seroconverted subject, QA255, who was enrolled in a prospective cohort of HIV-1 negative high-risk women in Mombasa, Kenya (Lavreys et al., 2002). QA255 was 40 years old at the time of HIV infection (D0). Study participants were treated according to Kenyan National Guidelines; QA255 did not receive antiretrovirals at any point during the period in which samples were analyzed for this study. Antiretroviral therapy was offered to all participants in the Mombasa Cohort beginning in March 2004, with support from the President's Emergency Plan for AIDS Relief. The infecting virus was clade A based on envelope sequence (Bosch, Rainwater, Jaoko, & Overbaugh, 2010). Approval to conduct this study was provided by the ethical review committees of the University of Nairobi Institutional Review Board, the Fred Hutchinson Cancer Research Center Institutional Review Board, and the University of Washington Institutional Review Board. Study participants provided written informed consent prior to enrollment.

Sample handling and RNA isolation

PBMCs stored in liquid nitrogen were thawed at 37°C, diluted 10-fold in pre-warmed RPMI and centrifuged for 10 min at 300 x g. Cells were washed once in phosphate-buffered saline, counted with trypan blue, centrifuged again, and total RNA was extracted from PBMCs using the AllPrep DNA/RNA Mini Kit (Qiagen, Germantown, MD), according to the manufacturer's recommended protocol. RNA was stored at -80°C.

Sequencing of antibody gene variable regions

Antibody sequencing was performed as previously described (Simonich et al., 2019; Vigdorovich et al., 2016). Library preparation was performed in technical replicate, as indicated in Table 1, using the same RNA isolated from each timepoint: D-119 (119 days prior to HIV infection), D462, D791, D1174, and D1512 post-infection. Briefly, RACE-ready cDNA synthesis was performed using the SMARTer RACE 5'/3' Kit (Takara Bio USA, Inc., Mountain View, CA) using primers with specificity to IgM, IgG, IgK and IgL, as previously reported (Simonich et al., 2019). One replicate each for D791 and D1174 were prepared using template switch adaptor primers that included unique molecular identifiers in the cDNA synthesis step: SmartNNNa 5' AAGCAGUGGTAUCAACGCAGAGUNNNNUNNNNUNNNNUCTTrGrGrGrGrG 3' (Turchaninova et al., 2016), where 'rG' indicates (ribonucleoside) guanosine bases. cDNA was diluted in Tricine-

EDTA according to the manufacturer's recommended protocol. First-round Ig-encoding sequence amplification (20 cycles) was performed using Q5 High-Fidelity Master Mix (New England BioLabs, Ipswich, MA) and nested gene-specific primers. Amplicons were directly used as templates for MiSeq adaption by second-round PCR amplification (10-20 cycles), purified and analyzed by gel electrophoresis, and indexed using Nextera XT P5 and P7 index sequences (Illumina, San Diego, CA) for Illumina sequencing according to the manufacturer's instructions (10 cycles). Gel-purified, indexed libraries were quantitated using the KAPA library quantification kit (Kapa Biosystems, Wilmington, MA) performed on an Applied Biosystems 7500 Fast real-time PCR machine. Libraries were denatured and loaded onto Illumina MiSeq 600-cycle V3 cartridges, according to the manufacturer's suggested workflow.

Sequence analysis and naïve inference

Sequences were preprocessed using FLASH, cutadapt, and FASTX-toolkit as previously described (Simonich et al., 2019; Vigdorovich et al., 2016). The sequences from our NGS replicates were merged either cumulatively or on a per-timepoint basis, as appropriate, to achieve the highest depth possible for each analysis. Sequences were then deduplicated and annotated with partis (<https://github.com/psathyrella/partis>) using default options including per-sample germline inference (Ralph & Matsen, 2016a, 2016b, 2019). Sequences with internal stop codons or CDR3 regions that were out-of-frame or had mutated codons at the start or end were removed. Indel events were identified, tracked, and reversed for alignment purposes. We did not exclude singletons in an attempt to retain very rare or undersampled sequences. Sequencing run statistics are detailed in Table 1. Cumulative and per-timepoint datasets underwent clonal family clustering using both the partis unseeded and seeded clustering methods (Ralph & Matsen, 2016b).

Inference of unmutated common ancestor sequences and simultaneous clonal family clustering was performed using the seed clustering method along with previously-identified QA255 mature ADCC antibody sequences (Williams et al., 2015; Williams et al., 2019) as "seeds". As an additional measure to ensure highest accuracy for inferred naïve sequences, the uncertainty on each inferred naïve sequence was visualized both with the partis --view-alternative-annotations option and by comparing these results to the most likely naïve sequences inferred by linearham software (see next section). The comparison revealed only the minor differences that would be expected based on linearham's method, which uses a more detailed model to infer more accurate naïve sequences for individual clonal families. For the unseeded clustering, each dataset was subsampled to 50-150K sequences for computational

efficiency. For each dataset, three random subsamples were analyzed and compared to ensure that our subsampling was sufficient to minimize statistical uncertainties. Seeded analyses were not subsampled.

Due to the lack of D genes and much shorter non-templated regions in light-chain rearrangements, computational clustering analyses artifactually overestimate clonality in light chain families, i.e. they cluster together sequences that did not originate from the same rearrangement event, but which come from almost identical naïve rearrangements. This caveat, affecting all unpaired antibody chain sequencing studies, ultimately reduces accuracy in inferring true light chain clonal families and their maturation pathways.

Alternative naïve inference

Alternative naïve sequences were inferred by applying a Bayesian phylogenetic Hidden Markov Model approach using the linearham software (<https://github.com/matsengrp/linearham>), using the partis-inferred clonal family clusters for each lineage as input (Dhar et al., 2020). Linearham samples naïve sequences from their posterior distribution rather than providing a single naïve sequence estimate, like other software programs do. The most probable linearham-predicted naïve sequences were compared to the most probable partis-predicted naïve sequences. In families with largely symmetric phylogenetic trees, the two methods return similar results; whereas with highly imbalanced trees, linearham is much more accurate.

Antibody lineage reconstruction

Antibody lineages were inferred as previously described (Simonich et al., 2019). In order to subsample large clonal families according to phylogenetic relatedness to the antibody chain of interest, initial phylogenetic trees of each QA255 clonal family was inferred using FastTree 2 (Price, Dehal, & Arkin, 2010). This allowed clonal families to be reduced to the 75-100 sequences most relevant to inferring the lineage history of the antibody chain of interest (<https://github.com/matsengrp/cft/blob/master/bin/prune.py>), which was then analyzed with RevBayes (Höhna et al., 2016) using an unrooted tree model with the general time-reversible (GTR) substitution model. All settings for RevBayes runs were customized to ensure likelihood and estimated sample sizes were >100 for each lineage. MCMC iterations ranged from 10,000-200,000, with thinning frequencies between 10-200 iterations and number of burn-in samples between 10-190. All RevBayes runs were done in technical duplicate (i.e. specifying different starting trees) and duplicates were all confirmed to agree on lineage chronology. RevBayes

output was summarized for internal node sequences (<https://github.com/matsengrp/ecgtheow>, Table 2-figure supplement 2), resulting in summary graphics where relative confidence in unique inferred sequences and amino acid substitutions are represented by color intensity. For each lineage, inferred intermediate sequences found on the most probable lineage paths were selected for study (Supplementary File 1, Table 2-figure supplement 2).

In the 016-VL lineage, we determined that a 3-nt insertion event occurred prior to the inferred_7_712 intermediate within this antibody's most likely developmental pathway (Table 2-figure supplement 2). Since most insertion-containing clonal NGS sequences (136 of 142) encoded serine at the insertion site (amino acid position 94), we inserted S94 into 016-2_{VL} instead of N94 that would correspond to the mature 016-VL sequence. For thoroughness, the inferred_5_789 016-VL intermediate (which precedes inferred_7_712) was synthesized both without (016-1_{VL}) and with (016-2_{VL}) the S94 insertion.

Identification of lineage-like NGS sequences

To determine if the computationally-inferred naïve and ancestor sequences were observed in the NGS data, we performed the following procedure for each lineage (implemented in a script found here: <https://git.io/Je7Zp>). A local BLAST database was created for each seeded clonal family and queried for sampled sequences that had high nucleotide sequence identity to lineage members using the 'blastn' command (Biopython package). E-value of 0.001 was used; other settings were default. Blastn matches for each lineage member were sorted according to their percent nt identity and alignment length. Sampled sequences with 100% nt or 100% aa identity in common with lineage members were noted (Supplementary File 1, Table 2-figure supplement 2). To identify the closest sampled sequences to each VH mature sequence, blastn results per VH mature query were viewed and the highest percent aa identity match was selected.

ADCC-focused lineage determination

As outlined in our Results, we first selected "middle" lineage intermediates for studying ADCC gain-of-function based on their moderate inclusion of amino acid substitutions (between naïve and mature sequences; Supplementary File 1) and high statistical confidence (Table 2-figure supplement 2), along with, in some cases, their validation by sampled NGS sequences (Table 2-figure supplement 2), and/or concentration of substitutions in complementarity-determining regions (CDRs) (Supplementary File 1). Middle intermediate chains were paired with partner naïve, middle, and mature chains and tested for antigen binding and ADCC

function. Based on preliminary results using middle intermediates, ADCC-focused lineages were chosen from remaining pre-middle inferred intermediates or post-middle intermediates depending on whether the middle intermediate showed activity. If multiple relevant intermediate choices were available within the pertinent (early or late) portion of an inferred lineage (Table 2-figure supplement 2), we selected sequences that had amino acid substitutions that were concentrated in CDRs (Supplementary File 1) and, whenever possible, we selected intermediates that were validated by NGS-sampled sequences, as aforementioned (Table 2-figure supplement 2). For 067-VH, 067-VK, 072-VH, and 105-VH lineages, early mutations were implicated in ADCC gain-of-function, but we lacked early lineage resolution and therefore could not select early inferred intermediates to study. Instead, we selectively incorporated CDR substitutions from the middle intermediate sequence into the inferred naïve sequence. Non-conservative FWR3 substitutions were also incorporated into the 067-VH early lineage. For each lineage, intermediate sequences were numbered consecutively based on their position in the developmental pathway between naïve (0) and mature (m). Heavy and light chain pairings for lineage intermediates were based on chronology, which reflected levels of mutation. Each chain was paired with several partner chains for functional assessment.

Ultimately, developmental lineages featuring the minimal number of mAbs to study ADCC development were defined based on levels of mAb mutation and ADCC activity. For each lineage, the minimal antibodies included 1) the inferred naïve, 2) the most mutated intermediate(s) that lacked ADCC activity, if available, 3) the least mutated intermediate(s) that gained detectable ADCC function, and 4) the least mutated intermediate(s) that demonstrated ADCC activity comparable to the mature antibodies (>50%). Mature antibodies were always included as benchmark positive controls.

Cell lines

For antibody production: HEK 293-F cells (RRID:CVCL_D603; originally derived from female human embryonic kidney cells) were obtained from Invitrogen (Thermo Fisher Scientific, Waltham, MA, catalog #R790-07) and grown at 37°C in Freestyle™ 293 Expression Medium (Thermo Fisher Scientific, catalog #12338002) in baffle-bottomed flasks orbiting at 135 rpm. These cells were not further authenticated in our hands.

For RFADCC: CEM.NKR cells (RRID:CVCL_X622; originally derived from female human T-lymphoblastoid cells) were obtained from NIH AIDS Reagent Program (Germantown, MD, catalog #458) and grown at 37°C in RPMI 1640 media with added penicillin (100 U/mL), streptomycin (100 µg/mL), Amphotericin B (250 ng/mL), L-glutamine (2mM), and fetal bovine

serum (10%) (all from Thermo Fisher Scientific). These cells were not further authenticated in our hands.

Monoclonal antibody production

Antibody heavy and light chain variable regions were synthesized as FragmentGENES (GENEWIZ®, South Plainfield, NJ) and subsequently cloned into corresponding Igy1, Igk or Igλ expression vectors (Tiller et al., 2008). Equal ratios of heavy and light chain plasmids were co-transfected into HEK 293F cells using FreeStyle MAX (Thermo Fisher Scientific, catalog #16447100) according to the manufacturer's instructions. Column-based Pierce™ protein G (Thermo Fisher Scientific, catalog #20397) purification of IgG was done according to the manufacturer's instructions.

Rapid and fluorometric ADCC (RFADCC) assay

The RFADCC assay was performed as described (Gomez-Roman et al., 2006; Williams et al., 2019). In short, CEM-NK cells (NIH AIDS Reagent Program, catalog #458) were double-labeled with PKH-26-cell membrane dye (Sigma-Aldrich, St. Louis, MO) and a cytoplasmic-staining dye (Vybrant CFDA SE Cell Tracer Kit, Thermo Fisher Scientific). The double-labeled cells were coated with clade A gp120 (BL035.W6M.Env.C1, (Wu et al., 2006)) or clade C gp41 ectodomain (C.ZA.1197MB, Immune Technology Corp, New York, NY) (Rousseau et al., 2006) for 1 hr at room temperature at a ratio of 1.5 µg protein : 1 x 10⁵ double-stained target cells. Coated targets were washed once with complete RPMI media (RPMI supplemented with 10% FBS, 4.0 mM Glutamax, and 1% antibiotic-antimycotic, all from Thermo Fisher Scientific). Monoclonal antibodies were diluted in complete RPMI media to a concentration of 100-500 ng/mL and mixed with 5 x 10³ coated target cells for 10 min at room temperature. PBMCs (peripheral blood mononuclear cells; Bloodworks Northwest, Seattle, WA) from an HIV-negative donor were added at a ratio of 50 effector cells : 1 target cell. The coated target cells, antibodies, and effector cells were co-cultured for 4 hr at 37°C then fixed in 1% paraformaldehyde (Affymetrix, Santa Clara, CA). Cells were analyzed by flow cytometry (Symphony I/II, BD Biosciences, San Jose, CA) and ADCC activity was defined as the percent of PKH-26+ CFDA- cells after background subtraction, where background (antibody-mediated killing of uncoated cells) was standardized to be 3–5% as analyzed using FlowJo software (FlowJo LLC, Ashland, OR). To mitigate differences in activity observed with different PBMC donors and between experiments, ADCC activity for each sample was normalized to monoclonal positive control mAbs: 167-D (NIH AIDS Reagent Program, Cat #11681) for gp41-

targeted Abs and C11 (NIH AIDS Reagent Program, Cat #7374) for gp120-targeted Abs. The activity of an unrelated antibody, FI6v3, that recognizes influenza hemagglutinin protein was used to define the limit of detection. For each replicate experiment, samples were categorized in the following manner: positive (sample > 2*FI6v3), indeterminate (1*FI6v3 ≤ sample ≤ 2*FI6v3), negative (sample < 1*FI6v3). Experiments were excluded if FI6v3 signal was >10% of monoclonal positive control signal. Background (uncoated cells) and negative-control (1*FI6v3) signal was subtracted from each sample's activity and the resultant values were averaged across experimental replicates, normalized to the respective lineage's mature mAb activity, and plotted in Prism v8.0c (GraphPad, San Diego, CA). A designation of indeterminate was also assigned in cases where samples were negative in the majority of experimental replicates, but indeterminate or positive in any replicate(s). In such cases, average activity was reported as usual.

Biolayer Interferometry

QA255 monoclonal antibody binding to monomeric gp120 or gp41 was measured using biolayer interferometry on an Octet RED instrument (ForteBio, Fremont, CA). Antibodies diluted to 8 μg mL⁻¹ in a filtered buffer solution of 1X PBS containing 0.1% BSA, 0.005% Tween-20, and 0.02% sodium azide were immobilized onto anti-human IgG Fc capture biosensors (ForteBio). C.ZA.1197MB or BL035.W6M.C1 gp41 was diluted to 250 nM, or as indicated, in the same buffer (above) and a series of up to six, two-fold dilutions were tested as analytes in solution at 30°C. The kinetics of mAb binding were measured as follows: association was monitored for 180 seconds, dissociation was monitored for 180 seconds, and regeneration was performed in 10mM Glycine HCl (pH 1.5). For experiments in which dilution series were run, binding-affinity constants (K_D ; on-rate, K_{on} ; off-rate, K_{dis}) were calculated using ForteBio's Data Analysis Software 7.0. Responses (nanometer shift) were calculated and background-subtracted using double referencing against the buffer reference signal and non-specific binding of biosensor to analyte. Data were processed by Savitzky-Golay filtering prior to fitting using a 1:1 model of binding kinetics.

Data and Code Availability

The QA255 longitudinal antibody deep sequencing datasets (Table 1) generated during this study are publicly available at BioProject SRA, accession PRJNA639297 [<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA639297/>]. The inferred antibody variable region sequences generated in this study have not been deposited in GenBank because

computationally-inferred sequences are not accepted, but they are available in Supplementary File 1. Mature antibody GenBank accession numbers are MT791224-MT791235. GenBank accession numbers for HIV Env variant are ABA61538.1 (BL035.W6M.Env.C1 gp120) and AAR27754.1 (C.ZA.1197MB gp41). The custom code generated or used in this study is publicly available on GitHub: prune.py (<https://github.com/matsengrp/cft/blob/master/bin/prune.py>), ecgtheow (<https://github.com/matsengrp/ecgtheow>), CFT (<https://github.com/matsengrp/cft>), and Blast validation (<https://git.io/Je7Zp>).

Quantification and Statistical Analysis

Where applicable, raw data were normalized and/or averaged across replicates using Microsoft Excel. Plots were generated using GraphPad Prism version 8.0c. Relevant experimental details, such as use of biological and technical replicates, can be found in figure legends. For RFADCC plots, bars with error bars represent mean and SD. RFADCC experimental exclusion criteria are detailed within the RFADCC method section above.

Acknowledgements

We thank the participants, staff, Scott McClelland, and Ludo Lavreys for continued efforts to oversee the Mombasa cohort. We acknowledge Brian Oliver, Vladimir Vigdorovich, and D. Noah Sather for technical assistance with NGS library preparation and sequence processing. We thank Vrasha Chohan, Haidyn Weight, and Tucker Price for their help with antibody production and testing.

Competing Interests

The authors declare no competing interests.

References

- Acharya, P., Tolbert, W. D., Gohain, N., Wu, X., Yu, L., Liu, T., . . . Pazgier, M. (2014). Structural definition of an antibody-dependent cellular cytotoxicity response implicated in reduced risk for HIV-1 infection. *J Virol*, 88(21), 12895-12906. doi:10.1128/JVI.02194-14
- Bonsignori, M., Kreider, E. F., Fera, D., Meyerhoff, R. R., Bradley, T., Wiehe, K., . . . Haynes, B. F. (2017). Staged induction of HIV-1 glycan-dependent broadly neutralizing antibodies. *Sci Transl Med*, 9(381). doi:10.1126/scitranslmed.aai7514
- Bonsignori, M., Zhou, T., Sheng, Z., Chen, L., Gao, F., Joyce, M. G., . . . Haynes, B. F. (2016). Maturation Pathway from Germline to Broad HIV-1 Neutralizer of a CD4-Mimic Antibody. *Cell*, 165(2), 449-463. doi:10.1016/j.cell.2016.02.022
- Bosch, K. A., Rainwater, S., Jaoko, W., & Overbaugh, J. (2010). Temporal analysis of HIV envelope sequence evolution and antibody escape in a subtype A-infected individual with a broad neutralizing antibody response. *Virology*, 398(1), 115-124. doi:10.1016/j.virol.2009.11.032
- Bournazos, S., Wang, T. T., Dahan, R., Maamary, J., & Ravetch, J. V. (2017). Signaling by Antibodies: Recent Progress. *Annu Rev Immunol*, 35, 285-311. doi:10.1146/annurev-immunol-051116-052433
- Burton, D. R. (2017). What Are the Most Powerful Immunogen Design Vaccine Strategies?: Reverse Vaccinology 2.0 Shows Great Promise. In *Cold Spring Harb Perspect Biol* (Vol. 9).
- Dhande, J., Angadi, M., Murugavel, K. G., Poongulali, S., Nandagopal, P., Vignesh, R., . . . Thakar, M. (2018). Brief Report: The Anti-HIV-1 ADCC-Mediating Antibodies From Cervicovaginal Secretions of HIV-Infected Women Have an Ability to Mediate Lysing of Autologous CD4+ HIV-Infected Cells. *J Acquir Immune Defic Syndr*, 79(2), 277-282. doi:10.1097/qai.0000000000001788
- Dhar, A., Ralph, D. K., Minin, V. N., & Matsen, F. A. t. (2020). A Bayesian phylogenetic hidden Markov model for B cell receptor sequence analysis. *PLoS Comput Biol*, 16(8), e1008030. doi:10.1371/journal.pcbi.1008030
- Doria-Rose, N. A., Schramm, C. A., Gorman, J., Moore, P. L., Bhiman, J. N., DeKosky, B. J., . . . Mascola, J. R. (2014). Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies. *Nature*, 509(7498), 55-62. doi:10.1038/nature13036
- Foote, J., & Winter, G. (1992). Antibody framework residues affecting the conformation of the hypervariable loops. *J Mol Biol*, 224(2), 487-499. doi:10.1016/0022-2836(92)91010-m
- Forthal, D. N., & Finzi, A. (2018). Antibody-dependent cellular cytotoxicity in HIV infection. *AIDS*, 32(17), 2439-2451. doi:10.1097/QAD.0000000000002011
- Fouts, T. R., Bagley, K., Prado, I. J., Bobb, K. L., Schwartz, J. A., Xu, R., . . . Gallo, R. C. (2015). Balance of cellular and humoral immunity determines the level of protection by HIV vaccines in rhesus macaque models of HIV infection. *Proc Natl Acad Sci U S A*, 112(9), E992-999. doi:10.1073/pnas.1423669112

771 Gomez-Roman, V. R., Florese, R. H., Patterson, L. J., Peng, B., Venzon, D., Aldrich, K., &
772 Robert-Guroff, M. (2006). A simplified method for the rapid fluorometric assessment of
773 antibody-dependent cell-mediated cytotoxicity. *J Immunol Methods*, 308(1-2), 53-67.
774 doi:10.1016/j.jim.2005.09.018

775 Haynes, B. F., Gilbert, P. B., McElrath, M. J., Zolla-Pazner, S., Tomaras, G. D., Alam, S. M., . . .
776 Kim, J. H. (2012). Immune-correlates analysis of an HIV-1 vaccine efficacy trial. *N Engl J*
777 *Med*, 366(14), 1275-1286. doi:10.1056/NEJMoa1113425

778 Henderson, R., Watts, B. E., Ergin, H. N., Anasti, K., Parks, R., Xia, S. M., . . . Alam, S. M.
779 (2019). Selection of immunoglobulin elbow region mutations impacts interdomain
780 conformational flexibility in HIV-1 broadly neutralizing antibodies. In *Nat Commun* (Vol.
781 10).

782 Höhna, S., Landis, M. J., Heath, T. A., Boussau, B., Lartillot, N., Moore, B. R., . . . Ronquist, F.
783 (2016). RevBayes: Bayesian Phylogenetic Inference Using Graphical Models and an
784 Interactive Model-Specification Language. *Syst Biol*, 65(4), 726-736.
785 doi:10.1093/sysbio/syw021

786 Horns, F., Vollmers, C., Dekker, C. L., & Quake, S. R. (2019). Signatures of selection in the
787 human antibody repertoire: Selective sweeps, competing subclones, and neutral drift.
788 *Proc Natl Acad Sci U S A*, 116(4), 1261-1266. doi:10.1073/pnas.1814213116

789 Kepler, T. B., & Wiehe, K. (2017). Genetic and Structural Analyses of Affinity Maturation in the
790 Humoral Response to HIV-1. *Immunol Rev*, 275(1), 129-144. doi:10.1111/imr.12513

791 Kwong, P. D., & Mascola, J. R. (2018). HIV-1 Vaccines Based on Antibody Identification, B Cell
792 Ontogeny, and Epitope Structure. *Immunity*, 48(5), 855-871.
793 doi:<https://doi.org/10.1016/j.immuni.2018.04.029>

794 Landais, E., Murrell, B., Briney, B., Murrell, S., Rantalainen, K., Berndsen, Z. T., . . . Poignard,
795 P. (2017). HIV Envelope Glycoform Heterogeneity and Localized Diversity Govern the
796 Initiation and Maturation of a V2 Apex Broadly Neutralizing Antibody Lineage. *Immunity*,
797 47(5), 990-1003 e1009. doi:10.1016/j.immuni.2017.11.002

798 Laserson, U., Vigneault, F., Gadala-Maria, D., Yaari, G., Uduman, M., Vander Heiden, J. A., . . .
799 Church, G. M. (2014). High-resolution antibody dynamics of vaccine-induced immune
800 responses. *Proc Natl Acad Sci U S A*, 111(13), 4928-4933.
801 doi:10.1073/pnas.1323862111

802 Lavreys, L., Baeten, J. M., Overbaugh, J., Panteleeff, D. D., Chohan, B. H., Richardson, B. A., .
803 . . Kreiss, J. K. (2002). Virus load during primary Human Immunodeficiency Virus (HIV)
804 type 1 infection is related to the severity of acute HIV illness in Kenyan women. *Clin*
805 *Infect Dis*, 35(1), 77-81. doi:10.1086/340862

806 Lewis, G. K., Ackerman, M. E., Scarlatti, G., Moog, C., Robert-Guroff, M., Kent, S. J., . . .
807 Thyagarajan, B. (2019). Knowns and Unknowns of Assaying Antibody-Dependent Cell-
808 Mediated Cytotoxicity Against HIV-1. *Front Immunol*, 10, 1025.
809 doi:10.3389/fimmu.2019.01025

810 Lewis, G. K., Pazgier, M., Evans, D. T., Ferrari, G., Bournazos, S., Parsons, M. S., . . . Finzi, A.
811 (2017). Beyond Viral Neutralization. *AIDS Res Hum Retroviruses*, 33(8), 760-764.
812 doi:10.1089/AID.2016.0299

813 Liao, H. X., Lynch, R., Zhou, T., Gao, F., Alam, S. M., Boyd, S. D., . . . Haynes, B. F. (2013).
814 Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature*,
815 496(7446), 469-476. doi:10.1038/nature12053

816 Mabuka, J., Nduati, R., Odem-Davis, K., Peterson, D., & Overbaugh, J. (2012). HIV-specific
817 antibodies capable of ADCC are common in breastmilk and are associated with reduced
818 risk of transmission in women with high viral loads. *PLoS Pathog*, 8(6), e1002739.
819 doi:10.1371/journal.ppat.1002739

820 MacLeod, D. T., Choi, N. M., Briney, B., Garces, F., Ver, L. S., Landais, E., . . . Poignard, P.
821 (2016). Early Antibody Lineage Diversification and Independent Limb Maturation Lead to
822 Broad HIV-1 Neutralization Targeting the Env High-Mannose Patch. *Immunity*, 44(5),
823 1215-1226. doi:10.1016/j.immuni.2016.04.016

824 Madhavi, V., Wren, L. H., Center, R. J., Gonelli, C., Winnall, W. R., Parsons, M. S., . . . Stratov,
825 I. (2014). Breadth of HIV-1 Env-specific antibody-dependent cellular cytotoxicity:
826 relevance to global HIV vaccine design. *AIDS*, 28(13), 1859-1870.
827 doi:10.1097/qad.0000000000000310

828 Mazor, Y., Yang, C., Borrok, M. J., Ayriss, J., Aherne, K., Wu, H., & Dall'Acqua, W. F. (2016).
829 Enhancement of Immune Effector Functions by Modulating IgG's Intrinsic Affinity for
830 Target Antigen. In *PLoS One* (Vol. 11).

831 Milligan, C., Richardson, B. A., John-Stewart, G., Nduati, R., & Overbaugh, J. (2015). Passively
832 acquired antibody-dependent cellular cytotoxicity (ADCC) activity in HIV-infected infants
833 is associated with reduced mortality. *Cell Host Microbe*, 17(4), 500-506.
834 doi:10.1016/j.chom.2015.03.002

835 Orlandi, C., Deredge, D., Ray, K., Gohain, N., Tolbert, W., DeVico, A. L., . . . Lewis, G. K.
836 (2020). Antigen-Induced Allosteric Changes in a Human IgG1 Fc Increase Low-Affinity
837 Fcγγ Receptor Binding. *Structure*, 28(5), 516-527 e515.
838 doi:10.1016/j.str.2020.03.001

839 Ovchinnikov, V., Louveau, J. E., Barton, J. P., Karplus, M., & Chakraborty, A. K. (2018). Role of
840 framework mutations and antibody flexibility in the evolution of broadly neutralizing
841 antibodies. *Elife*, 7. doi:10.7554/eLife.33038

842 Pegu, A., Hessel, A. J., Mascola, J. R., & Haigwood, N. L. (2017). Use of broadly neutralizing
843 antibodies for HIV-1 prevention. *Immunol Rev*, 275(1), 296-312. doi:10.1111/imir.12511

844 Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2--approximately maximum-likelihood
845 trees for large alignments. *PLoS One*, 5(3), e9490. doi:10.1371/journal.pone.0009490

846 Ralph, D. K., & Matsen, F. A. t. (2016a). Consistency of VDJ Rearrangement and Substitution
847 Parameters Enables Accurate B Cell Receptor Sequence Annotation. *PLoS Comput*
848 *Biol*, 12(1), e1004409. doi:10.1371/journal.pcbi.1004409

849 Ralph, D. K., & Matsen, F. A. t. (2016b). Likelihood-Based Inference of B Cell Clonal Families.
850 *PLoS Comput Biol*, 12(10), e1005086. doi:10.1371/journal.pcbi.1005086

851 Ralph, D. K., & Matsen, F. A. t. (2019). Per-sample immunoglobulin germline inference from B
852 cell receptor deep sequencing data. *PLoS Comput Biol*, 15(7), e1007133.
853 doi:10.1371/journal.pcbi.1007133

854 Rappuoli, R., Bottomley, M. J., D'Oro, U., Finco, O., & De Gregorio, E. (2016). Reverse
855 vaccinology 2.0: Human immunology instructs vaccine antigen design. *J Exp Med*,
856 213(4), 469-481. doi:10.1084/jem.20151960

857 Rousseau, C. M., Birditt, B. A., McKay, A. R., Stoddard, J. N., Lee, T. C., McLaughlin, S., . . .
858 Mullins, J. I. (2006). Large-scale amplification, cloning and sequencing of near full-length
859 HIV-1 subtype C genomes. *J Virol Methods*, 136(1-2), 118-125.
860 doi:10.1016/j.jviromet.2006.04.009

861 Ruiz, M. J., Salido, J., Abusamra, L., Ghiglione, Y., Cevallos, C., Damilano, G., . . . Turk, G.
862 (2017). Evaluation of Different Parameters of Humoral and Cellular Immune Responses
863 in HIV Serodiscordant Heterosexual Couples: Humoral Response Potentially Implicated
864 in Modulating Transmission Rates. *EBioMedicine*, 26, 25-37.
865 doi:10.1016/j.ebiom.2017.11.001

866 Schramm, C. A., & Douek, D. C. (2018). Beyond Hot Spots: Biases in Antibody Somatic
867 Hypermutation and Implications for Vaccine Design. *Front Immunol*, 9.
868 doi:10.3389/fimmu.2018.01876

869 Simonich, C. A., Doepker, L., Ralph, D., Williams, J. A., Dhar, A., Yaffe, Z., . . . Overbaugh, J.
870 (2019). Kappa chain maturation helps drive rapid development of an infant HIV-1 broadly
871 neutralizing antibody lineage. *Nat Commun*, 10(1), 2190. doi:10.1038/s41467-019-
872 09481-7

873 Stamatatos, L., Pancera, M., & McGuire, A. T. (2017). Germline-targeting immunogens.
874 *Immunol Rev*, 275(1), 203-216. doi:10.1111/imr.12483

875 Tang, Y., Lou, J., Alpaugh, R. K., Robinson, M. K., Marks, J. D., & Weiner, L. M. (2007).
876 Regulation of antibody-dependent cellular cytotoxicity by IgG intrinsic and apparent
877 affinity for target antigen. *J Immunol*, 179(5), 2815-2823.
878 doi:10.4049/jimmunol.179.5.2815

879 Tiller, T., Meffre, E., Yurasov, S., Tsuiji, M., Nussenzweig, M. C., & Wardemann, H. (2008).
880 Efficient generation of monoclonal antibodies from single human B cells by single cell
881 RT-PCR and expression vector cloning. *J Immunol Methods*, 329(1-2), 112-124.
882 doi:10.1016/j.jim.2007.09.017

883 Tolbert, W. D., Sherburn, R., Gohain, N., Ding, S., Flinko, R., Orlandi, C., . . . Pazgier, M.
884 (2020). Defining rules governing recognition and Fc-mediated effector functions to the
885 HIV-1 co-receptor binding site. *BMC Biol*, 18(1), 91. doi:10.1186/s12915-020-00819-y

886 Turchaninova, M. A., Davydov, A., Britanova, O. V., Shugay, M., Bikos, V., Egorov, E. S., . . .
887 Chudakov, D. M. (2016). High-quality full-length immunoglobulin profiling with unique
888 molecular barcoding. *Nat Protoc*, 11(9), 1599-1616. doi:10.1038/nprot.2016.093

889 Verkoczy, L., & Diaz, M. (2014). Autoreactivity in HIV-1 broadly neutralizing antibodies:
890 implications for their function and induction by vaccination. *Curr Opin HIV AIDS*, 9(3),
891 224-234. doi:10.1097/COH.0000000000000049

892 Verkoczy, L., Kelsoe, G., & Haynes, B. F. (2014). HIV-1 envelope gp41 broadly neutralizing
893 antibodies: hurdles for vaccine development. *PLoS Pathog*, 10(5), e1004073.
894 doi:10.1371/journal.ppat.1004073

895 Vigdorovich, V., Oliver, B. G., Carbonetti, S., Dambrauskas, N., Lange, M. D., Yacoob, C., . . .
896 Sather, D. N. (2016). Repertoire comparison of the B-cell receptor-encoding loci in
897 humans and rhesus macaques by next-generation sequencing. *Clin Transl Immunology*,
898 5(7), e93. doi:10.1038/cti.2016.42

899 Williams, K. L., Cortez, V., Dingens, A. S., Gach, J. S., Rainwater, S., Weis, J. F., . . .
900 Overbaugh, J. (2015). HIV-specific CD4-induced Antibodies Mediate Broad and Potent
901 Antibody-dependent Cellular Cytotoxicity Activity and Are Commonly Detected in Plasma
902 From HIV-infected humans. *EBioMedicine*, 2(10), 1464-1477.
903 doi:10.1016/j.ebiom.2015.09.001

904 Williams, K. L., Stumpf, M., Naiman, N. E., Ding, S., Garrett, M., Gobillot, T., . . . Overbaugh, J.
905 (2019). Identification of HIV gp41-specific antibodies that mediate killing of infected cells.
906 *PLoS Pathog*, 15(2), e1007572. doi:10.1371/journal.ppat.1007572

907 Wu, X., Parast, A. B., Richardson, B. A., Nduati, R., John-Stewart, G., Mbori-Ngacha, D., . . .
908 Overbaugh, J. (2006). Neutralization escape variants of human immunodeficiency virus
909 type 1 are transmitted from mother to infant. *J Virol*, 80(2), 835-844.
910 doi:10.1128/JVI.80.2.835-844.2006

911 Wu, X., Zhang, Z., Schramm, C. A., Joyce, M. G., Kwon, Y. D., Zhou, T., . . . Shapiro, L. (2015).
912 Maturation and Diversity of the VRC01-Antibody Lineage over 15 Years of Chronic HIV-
913 1 Infection. *Cell*, 161(3), 470-485. doi:10.1016/j.cell.2015.03.004

914 Yuan, T., Li, J., & Zhang, M. Y. (2011). A single mutation turns a non-binding germline-like
915 predecessor of broadly neutralizing antibody into a binding antibody to HIV-1 envelope
916 glycoproteins. In *MAbs* (Vol. 3, pp. 402-407).
917
918

006 VH

006 VL

016 VH

016 VL

Key

- ★ 100% nt identity with NGS
- ★ 100% aa identity with NGS
- MID "middle intermediate"
- Name used in study

Table 2-figure supplement 2: Most probable routes of antibody lineage maturation, cont'd

Table 2-figure supplement 2: Most probable routes of antibody lineage maturation, cont'd

Table 2-figure supplement 2. Most probable routes of antibody lineage maturation. Bayesian clonal family phylogenies for each antibody chain were sampled from associated posterior distributions and summarized to display relative confidences for internal node sequences. Resulting graphics display multiple possible lineages of amino acid transitions and their relative confidences for heavy- and light-chain development. Amino acid substitutions (arrows) connect the inferred naïve sequence (white node, top) to the mature sequence (white node, bottom) via reconstructed ancestral intermediate sequences (red-shaded nodes). The red shading of nodes is proportional to the posterior probability that this ancestral sequence was present in the lineage. For a given node, the blue shading across arrows arising from that node is proportional to the corresponding transition probability. Low probability nodes were filtered out, resulting in some incomplete pathways within the graphics. Transient mutations are labeled in gray instead of black. Sequences used in functional studies are labeled with their manuscript aliases. Stars denote inferred sequences that were identical in nucleotide (green) or amino acid (yellow) sequence to sampled NGS sequences. “MID” (red) denotes middle intermediate sequence. Labels with cyan highlighting denote inferred sequences that were used in lineage studies. These are incomplete because some lineage members were manually designed (see Methods). If inferred “alternative” naïve sequences matched a lineage member, we included these labels as well for convenience.