Selection of HIV Vaccine Candidates for Concurrent Testing in an Efficacy Trial

Ying Huang^{1*}, Carlos DiazGranados², Holly Janes¹, Yunda Huang¹, Allan C. deCamp¹, Barbara Metch¹, Shannon Grant¹, Brittany Sanchez¹, Sanjay Phogat², Marguerite Koutsoukos³, Niranjana Kanasa-Thanan³, Patricia Bourguignon³, Alix Collard³, Susan Buchbinder⁴, Georgia D. Tomaras⁵, Julie McElrath¹, Glenda Gray⁶, James G. Kublin¹, Lawrence Corey¹, Peter B. Gilbert^{1*}

¹Vaccine & Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA

²Sanofi Pasteur, Swiftwater, PA

³GlaxoSmithKline Vaccines, Belgium

⁴Department of Medicine and Epidemiology/Biostatistics, UCSF, San Francisco, CA

⁵Perinatal HIV Research Unit, University of the Witwatersrand, Johannesburg, South Africa

⁶Duke Human Vaccine Institute and Department of Surgery, Duke University Medical Center, Durham, NC

*To whom correspondence should be addressed to: yhuang@fhcrc.org, pgilbert@fhcrc.org

Conflict of interest statement:

Carlos DiazGranados and Sanjay Phogat are fulltime employees of Sanofi Pasteur.

Word counts (Introduction to Acknowledgement): 2000

Abstract

Phase IIb or III HIV-1 vaccine efficacy trials are generally large and operationally challenging. To mitigate this challenge, the HIV Vaccine Trials Network is designing a Phase IIb efficacy trial accommodating the evaluation of multiple vaccine regimens concurrently. As this efficacy trial would evaluate a limited number of vaccine regimens, there is a need to develop a framework for optimizing the strategic selection of regimens from the large number of vaccine candidates tested in Phase I/IIa trials. In this paper we describe the approaches for the selection process, including the choice of immune response endpoints and the statistical criteria and algorithms. We illustrate the selection approaches using data from HIV-1 vaccine trials.

Introduction

The World Health Organization [1], and the Joint United Nations Programme on HIV/AIDS [2], estimate that there were 36.9 million people living with HIV globally by the end of 2014, with Sub-Saharan Africa accounting for 70% of the global HIV burden. Development of a preventive HIV vaccine remains a global health priority. Unfortunately, despite decades of research, only a single study, RV144, has demonstrated modest protection against HIV. This study was a phase III randomized controlled study conducted in Thailand. The vaccine regimen (two doses of prime with ALVAC-HIV (vCP1521), followed by two boosts with ALVAC-HIV + AIDSVAX clades B/E gp120 protein) demonstrated 31.2% efficacy compared to placebo ($p = 0.04$) at 3.5 years [3] and 60.5% efficacy through 12 months [4]. These results have reinvigorated the scientific community by suggesting that developing a preventive HIV vaccine may be possible.

To better understand how the RV144 vaccine regimen reduced the risk of HIV infection, a large consortium of independent laboratories worked together systematically to perform a case control study within the RV144 trial to identify correlates of risk (CoR) for HIV infection [5], i.e. vaccine-induced immune response biomarkers that are associated with subsequent HIV infection [6, 7]. Two CoRs were identified: immunoglobulin G (IgG) Antibody that binds to a scaffolded gp70 V1V2 recombinant protein (inversely correlated with risk); and plasma Env-specific binding immunoglobulin A (IgA) (directly correlated with risk). Four additional variables correlated inversely with infection risk when the level of IgA binding was low. Recently, several studies have further enhanced our understanding of the efficacy seen in RV144 [8, 9, 10, 11, 12, 13, 14, 15] and the potential relevance of Env V1V2-specific IgG3 [16, 17]. These studies lay the groundwork for immunogenicity analyses in several of the ongoing and future HIV vaccine trials.

Since its inception in the late nineties, the HIV Vaccine Trials Network (HVTN) has been conducting multiple clinical trials with a large number of candidate HIV vaccine regimens in different regions of the world. The vast majority of these studies are Phase I/IIa safety and immunogenicity studies, with only a few Phase IIb trials evaluating the impact of vaccine candidates on HIV-1 infection [18, 19, 20]. Phase IIb efficacy trials are generally large and operationally challenging. To mitigate this challenge, the HVTN is considering the inclusion of several vaccine regimens concurrently in one Phase IIb efficacy study, to augment study design and operational efficiencies [21, 22]. As this efficacy trial would only allow evaluation of a limited number of candidate vaccine regimens, we need a framework to select the most

promising vaccine regimens from the reagents tested in Phase I/IIa trials. We describe approaches for the selection process and illustrate our rationale using data from completed HIV-1 vaccine trials.

Methods

A number of phase I/IIa studies testing multiple vaccine regimens comprising combinations of non-protein components (viral vectors or DNA), Env proteins and adjuvants, and administered at different schedules (3 injections at 0, 1 and 6 months or 4 injections at 0, 1, 3 and 6 months) will provide the data for selecting vaccine regimens for efficacy testing.

Three-Step Down-Selection Scheme

We plan a three-step scheme for down-selection (Figure 1). In Step 1, all regimens entering down-selection are screened based on safety and peak immunogenicity from a core set of immune assays measured on month 6.5 (two weeks after the last prime-boost vaccination) samples. In Steps 2 and 3, regimens passing Step 1 will be evaluated based on immunogenicity data from a full set of assays measured on month 6.5 samples, and a subset of months 3.5 (two weeks after the month 3 vaccination) and month 12 (6 months after the last vaccination) samples. The latter two time points are included for differentiating onset of immune responses among vaccine recipients and for assessing the durability of induced immune responses [23]. Step 2 involves a comparison of each putative regimen with the reference ALVAC-gp120 C/C in MF59 regimen, a similar regimen as the one used in RV144 but adapted for the region of southern Africa (HIV clade C based). ALVAC-gp120 C/C in MF59 is currently being tested in HVTN 100, a phase I trial in South Africa that may lead to a subsequent pivotal efficacy trial. Candidate regimens that are not superior to ALVAC-gp120 C/C for at least one immunological endpoint will be filtered out. Step 3 conducts head-to-head comparisons of the remaining regimens for final down-selection.

All steps rely on the identification of CoRs and possible immune correlates of protection (CoP). The latter is an immune response biomarker that predicts vaccine efficacy [6, 7]. Although several potential CoRs were identified in RV144, no CoP has been validated in the HIV-1 vaccine field [5, 15]. One approach to assessing a potential CoP requires first demonstrating overall VE > 0, augmenting the efficacy trial design, and making additional assumptions in order to

estimate the immune responses of placebo participants if they had received vaccine [24, 25, 26]. In order to maximize the possibility that selected vaccine regimens would confer protection in a future efficacy trial, the down-selection will incorporate existing knowledge on HIV-1 vaccines as well as emerging evidence and knowledge of mechanistic CoPs for other licensed vaccines.

Take/Potency Criteria for Step 1 An immunogenicity criterion called take/potency is considered in Step 1, which requires response rates above pre-specified thresholds for designated core immunological endpoints considered essential for a vaccine regimen to potentially confer adequate protection based on current knowledge. Specifically, the majority of vaccine recipients must generate IgG binding antibodies to the vaccine gp120s and a minimum frequency of vaccine recipients must generate two of three types of responses: V2 antibodies, neutralizing antibodies (NAb), and CD4+ or CD8+ T cell responses.

Down-selection based on Additional Immunological Endpoints for Steps 2 and 3 A potentially large number of immunological endpoints are considered in Steps 2 and 3, including all immune classes that are part of a putative CoP (Figure 2a). For each immune class, one or more endpoint scores will be used to summarize a vaccine recipient's immune response in that class, which includes either a continuous score indicating magnitude or a binary score indicating positive/negative response. Down-selection will also be performed using a subset of immunological endpoints that were demonstrated to be statistically significant CoRs in RV144 (Figure 2b).

Head-to-Head Comparison between Regimens in Step 3

Selection Criteria To down-select vaccine regimens based on comparisons of their multivariate immune response profiles in Step 3, we developed two criteria (Figure 3). The first is “**superiority**,” where selected vaccine regimens should be immunologically superior to un-selected regimens in a statistically defined way. Assuming a larger immune endpoint score is associated with a presumed better protective effect of vaccine, we define superiority for each endpoint score as a larger mean and consider a regimen to be superior to another with respect to its immune profile if it is superior with respect to at least one endpoint score and is not inferior with respect to any endpoint score. The second criterion is “**non-redundancy**.” The protective mechanisms of HIV-1 infection via the generation of different immune

responses are not yet well understood. It is desirable to select vaccine regimens with non-redundant immune profiles [27], so that diverse mechanisms of vaccine protective effects can be investigated in the efficacy trial. We define two regimens A and B to be non-redundant if A is superior to B with respect to at least one endpoint score and B is superior to A with respect to at least one endpoint score.

Selection Algorithms Based on these criteria, we developed two statistical algorithms for down-selection that integrates hypothesis testing, ranking, and clustering (Y Huang et al., unpublished). The first, the “ranking, filtering, and selection” (RFS) algorithm, 1) ranks all regimens according to an overall summary score that weights each endpoint score according to its putative predictive clinical importance to vaccine efficacy, 2) selects the top-ranked regimen, and 3) sequentially evaluates regimens ranked next, selecting regimens non-redundant with the existing set and excluding regimens inferior to the newly selected one, based on hypothesis tests of each individual endpoint score between regimens. To minimize errors due to multiple testing, we implement multiplicity correction to individual tests such that the probability of violating the non-redundancy criterion during down-selection can be controlled at a pre-specified level.

The second algorithm applies a “clustering and ranking” (CR) step before the RFS, which groups the candidate vaccine regimens into different clusters based on similarities in their immune response profiles and selects from each cluster the top-ranked regimen based on the aforementioned summary score to enter the RFS. We name this algorithm CR+RFS.

Data Example

As an example, we applied the down-selection algorithms to evaluate five regimens studied previously in HIV-1 vaccine trials: the RV144 vaccine regimen (RV144.T) and four regimens studied in the phase 1 trial HVTN096, including NYVAC prime plus NYVAC + AIDSVAX B/E boosts (096.T1), NYVAC + AIDSVAX B/E prime plus NYVAC + AIDSVAX B/E boosts (096.T2), DNA prime plus NYVAC + AIDSVAX B/E boosts (096.T3), and DNA + AIDSVAX B/E prime plus NYVAC + AIDSVAX B/E boosts (096.T4). Peak immune assay data at month 6.5 were available for 205, 19, 18, 17, and 19 HIV-1 uninfected vaccinees, respectively, for eight immunological endpoints common across trials including IgG binding antibody responses to six different gp120 antigens, mean NAb responses to six HIV-1 isolates and CD4+ T-cell response. We

generated a dataset that included 43 individuals per regimen by sampling with replacement from individuals with complete endpoint scores within each regimen.

For the demonstration we equally weighted IgG, CD4+, and NAb and subdivided the weight for IgG equally between the six different antigens. The ranking of the five regimens based on the weighted average score from highest to lowest was 096.T3, 096.T1, 096.T4, 096.T2, and RV144.T (Figure 4a). We use a principal component biplot to graphically display the structure of the data (Figure 4b). It showed that the IgG and NAb scores were correlated and were along the first principal component direction capturing the maximum variation in responses, whereas the CD4+ score captured additional variation more in the second principal component direction. Hierarchical clustering with weighted Manhattan distance [28] grouped the five vaccine regimens into four clusters: {RV144.T}, {096.T1, 096.T3}, {096.T2}, {096.T4} (Figure 4c-d). Both RFS and CR+RFS algorithms selected 096.T3 only; the other regimens were either inferior or redundant and not superior (Figure 4d).

Discussion

The development of the down-selection plan is a multi-disciplinary collaboration demanding sophisticated statistical and scientific considerations. The framework is a new advancement for the field of HIV vaccine research: the selection of more than one regimen would lead to the first multi-regimen HIV vaccine efficacy trial.

The down-selection framework is broadly applicable. It can accommodate different assumptions about immune response profiles that are consistent with protection using different ranking strategies, e.g., by favoring regimens with best immune response on average, or best immune responses in a few classes, etc. This flexibility is essential given the lack of full knowledge about the nature of immune responses needed for protection. The choice of weights of immune response endpoints affects selection outcome, the finalization of which requires continuing discussions among researchers and updates to accommodate emerging research in the field. Lastly, the framework was developed with focus on candidate vaccines in a relatively homogenous class of regimens (pox-protein), and additional research would be needed to optimize the framework for regimens generating totally novel responses (e.g., broadly neutralizing Tier 2 antibodies).

Performance of the down-selection algorithms depends critically on the sample sizes and the differences between regimens. Simulation studies of the phase I/IIa trials demonstrated reasonable selection performance if regimens have immunological differences consistent with a 20% difference in vaccine efficacy, based on CoR and CoP analyses of RV144 (Supplementary Figure 1).

ACKNOWLEDGEMENTS

The authors thank the participants, investigators, and sponsors of the HVTN096 and RV144 trials.

The authors thank the HVTN lab program for generating the immune response data for the HVTN096 trial and Dr. Gepi Pantaleo for chairing the HVTN096 trial. This research is supported by NIH NIAID award UM1AI068635.

REFERENCES

- [1] WHO HIV/AIDS Fact sheet. Updated July 2015. Available at: <http://www.who.int/mediacentre/factsheets/fs360/en>.
- [2] UNAIDS. MDG 6: 15 years, 15 lessons of hope from the AIDS response fact sheet. Available at http://www.unaids.org/sites/default/filesmedia_asset/20150714_FS_MDG6_Report_en.pdf.
- [3] Rerks-Ngarm S, Pitisuttithum P, Nitayaphan S, Kaewkungwal J, Chiu J, Paris R, Premisri N, Namwat C, de Souza M, Adams E, et al. Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. *N Engl J Med* 2009, 361(23):2209-20.
- [4] Robb ML, Rerks-Ngarm S, Nitayaphan S, Pitisuttithum P, Kaewkungwal J, Kunasol P, Khamboonruang C, Thongcharoen P, Morgan P, Benenson M, et al. Risk behaviour and time as covariates for efficacy of the HIV vaccine regimen ALVAC-HIV (vCP1521) and AIDSVAX B/E: a post-hoc analysis of the Thai phase 3 efficacy trial RV144. *The Lancet Infectious Diseases* 2012, 12(7), 531-537.
- [5] Haynes BF, Gilbert PB, McElrath MJ, Zolla-Pazner S, Tomaras GD, Alam SM, Evans DT, Montefiori DC, Karnasuta C, Sutthent R, et al. Immune-correlates analysis of an HIV-1 vaccine efficacy trial. *N Engl J Med* 2012, 366(14):1275-86.
- [6] Qin L, Gilbert PB, Corey L, McElrath MJ, and Self SG. A framework for assessing immunological correlates of protection in vaccine trials. *J Infect Dis* 2007, 196(9):1304-12.
- [7] Plotkin SA, and Gilbert PB. Nomenclature for immune correlates of protection after vaccination. *Clin Infect Dis* 2012, 54(11):1615-7.
- [8] Rolland M, Edlefsen PT, Larsen BB, Tovanabutra S, Sanders-Buell E, Hertz T, deCamp AC, Carrico C, Menis S, Magaret CA, et al. Increased HIV-1 vaccine efficacy against viruses with genetic signatures in Env V2. *Nature* 2012, 490(7420), 417-420.

- [9] Liao HX, Bonsignori M, Alam SM, McLellan JS, Tomaras GD, Moody MA, Kozink DM, Hwang KK, Chen X, Tsao CY et al. Vaccine induction of antibodies against a structurally heterogeneous site of immune pressure within HIV-1 envelope protein variable regions 1 and 2. *Immunity* 2013, 38(1), 176-186.
- [10] Gottardo R, Bailer RT, Korber BT, Gnanakaran S, Phillips J, Shen X, Tomaras GD, Turk E, Imholte G, Eckler L et al. Plasma IgG to linear epitopes in the V2 and V3 regions of HIV-1 gp120 correlate with a reduced risk of infection in the RV144 vaccine efficacy trial. *PLoS One* 2013, 8(9), e75665.
- [11] Tomaras GD, Ferrari G, Shen X, Alam SM, Liao HX, Pollara J, Bonsignori M, Moody MA, Fong Y, Chen X. Vaccine-induced plasma IgA specific for the C1 region of the HIV-1 envelope blocks binding and effector function of IgG. *Proceedings of the National Academy of Sciences* 2013, 110(22), 9019-9024.(*)

This paper investigated the hypothesis that IgA could attenuate the protective effect of IgG responses through competition for the same Env binding sites. It showed that Env-specific plasma IgA/IgG ratios are higher in infected than in uninfected vaccine recipients in RV144. This suggests including relative IgA/IgG responses as an endpoint for characterizing pox-protein HIV vaccine candidates.

- [12] Zolla-Pazner S, deCamp A, Gilbert PB, Williams C, Yates NL, Williams WT, Howington R, Fong Y, Morris DE, Soderberg KA et al. Vaccine-induced IgG antibodies to V1V2 regions of multiple HIV-1 subtypes correlate with decreased risk of HIV-1 infection. *PLoS One* 2014, 9(2), e87572.(*)

This paper reported that the IgG V2 antibody correlate of low HIV-1 infection risk observed in RV144 (Haynes et al., 2012) was reproduced in a follow-up study with multiple V2 antigens and two assays in two labs. This study also demonstrated that the V2 IgG correlate targeted a breadth of cross-clade V2 sequences. This replication provides some statistical robustness to the original V2 correlate of risk finding of Haynes et al. (2012).

- [13] Tomaras GD, Haynes BF. Advancing toward HIV-1 vaccine efficacy through the intersections of immune correlates. *Vaccines* 2014, 2(1), 15-35.

- [14] Lin L, Finak G, Ushey K, Seshadri C, Hawn TR, Frahm N, Scriba TJ, Mahomed H, Hanekom W, Bart PA, et al. COMPASS identifies T-cell subsets correlated with clinical outcomes. *Nature Biotechnology* 2015, 33(6): 610-6.(*)

This paper developed a new statistical approach for measuring T cell polyfunctionality with a polyfunctionality score, and showed that this score for CD4 T cells in RV144 was a better predictor of low HIV-1 infection risk than magnitude of CD4 T cell response. This paper supports further study of the quality of CD4 T cells as a correlate of protection for vaccines against viral diseases.

- [15] Corey L, Gilbert PB, Tomaras G, Haynes BF, Pantaleo G, and Fauci AS. Immune correlates of vaccine protection against HIV-1 acquisition: A review. *Sci Transl Med* 2015;In Press.(**)

This review paper summarizes evidence from follow-up research about whether the original IgG V2 correlate of risk finding from Haynes et al. (2012) could also be a correlate of protection. The research has been supportive, with outcomes: (1) replication of the correlate result with cross-clade V2 antigens with two assays and two labs; (2) a viral sieve analysis showed differential vaccine efficacy at amino acid sites of IgG V2 loop reactivity; and (3) crystal structures and alanine scanning of V2 loop monoclonal antibodies from RV144 infected participants showed direct immune pressure on V2 mediated by ADCC. This synthesis motivates future efficacy testing of vaccines with putatively improved IgG V2 antibody responses.

- [16] Yates NL, Liao HX, Fong Y, deCamp A, Vandergrift NA, Williams WT, Alam SM, Ferrari G, Yang Z, Seaton KE et al. Vaccine-induced Env V1-V2 IgG3 correlates with lower HIV-1 infection risk and declines soon after vaccination. *Science Translational Medicine* 2014, 6(228), 228ra39-228ra39.(*)

This paper noted that Env V1V2-specific IgG3 was the immunoglobulin subclass showing the strongest inverse correlation low HIV-1 infection risk in RV144.

- [17] Chung AW, Ghebremichael M, Robinson H, Brown E, Choi I, Lane S, Dugast A, Shoen MK, Rolland M, Suscovich TJ, et al. Polyfunctional Fc-effector profiles mediated by IgG subclass selection distinguish RV144 and VAX003 vaccines. *Science Translational Medicine* 2014, 6(228), 228ra38-228ra38.(*)

This paper demonstrated that the IgG3 subclass engaged Fc-mediated effector responses more effectively than other IgG subclasses, thereby providing a possible mechanism explaining the association of Env V1V2 IgG3 with a lower rate of HIV acquisition.

- [18] Buchbinder SP, Mehrotra DV, Duerr A, Fitzgerald DW, Mogg R, Li D, Gilbert PB, Lama JR, Marmor M, Del Rio C, et al. Efficacy assessment of a cell-mediated immunity HIV-1 vaccine (the Step Study): a double-blind, randomised, placebo-controlled, test-of-concept trial. *The Lancet* 2008, 372(9653), 1881-1893.
- [19] Gray GE, Allen M, Moodie Z, Churchyard G, Bekker LG, Nchabeleng M, Mlisana K, Metch B, de Bruyn G, Latka MH, et al. Safety and efficacy of the HVTN 503/Phambili study of a clade-B-based HIV-1 vaccine in South Africa: a double-blind, randomised, placebo-controlled test-of-concept phase 2b study. *The Lancet Infectious Diseases* 2011, 11(7), 507-515.
- [20] Hammer SM, Sobieszczyk ME, Janes H, Karuna ST, Mulligan MJ, Grove D, Koblin BA, Buchbinder SP, Keefer MC, Tomaras GD, et al. Efficacy trial of a DNA/rAd5 HIV-1 preventive vaccine. *N Engl J Med* 2013, 369(22), 2083-2092.
- [21] Corey L, Nabel GJ, Dieffenbach C, Gilbert PB, Haynes BF, Johnston M, Kublin J, Lane HC, Pantaleo G, Picker LJ, et al. HIV-1 vaccines and adaptive trial designs. *Sci Transl Med*, 2011;3(79):79ps13.
- [22] Gilbert PB, Grove D, Gabriel E, Huang Y, Gray G, Hammer SM, Buchbinder SP, Kublin J, Corey L, Self SG. A sequential phase 2b trial design for evaluating vaccine efficacy and immune correlates for multiple HIV vaccine regimens. *Stat Commun Infect Dis* 2011;3(1).

- [23] Lewis GK, DeVico AL, Gallo RC. Antibody persistence and T-cell balance: two key factors confronting HIV vaccine development. *Proc Natl Acad Sci USA* 2014, 111(44):15614-21.(*)

Authors in this paper discussed the evidence illustrating the poor persistence of antibody responses to Env and the related problem of CD4+ T-cell responses that compromise vaccine efficacy by creating excess cellular targets of HIV-1 infection, and proposed solutions to both problems applicable to Env-based AIDS vaccines.

- [24] Follmann D. Augmented designs to assess immune response in vaccine trials. *Biometrics* 2006, 62(4):1161-9.
- [25] Gilbert PB, Hudgens MG. Evaluating candidate principal surrogate endpoints. *Biometrics* 2008, 64(4):1146-54.
- [26] Huang Y, Gilbert PB, and Wolfson J. Design and estimation for evaluating principal surrogate markers in vaccine trials. *Biometrics* 2013, 69(2):301-9.
- [27] Manrique A, Adams E, Barouch DH, Fast P, Graham BS, Kim JH, Kublin JG, McCluskey M, Pantaleo G, Robinson HL, Russell N, Snow W, Johnston MI. The immune space: a concept and template for rationalizing vaccine development. *AIDS Res Hum Retroviruses* 2014, 30 (11): 1017 – 1022.(*)

This paper proposed an "immune space template" that provides a standardized approach by which the quality, level, and durability of immune responses elicited in early human trials by a candidate vaccine can be described.

- [28] Black P (2006) "Manhattan distance", in Dictionary of algorithms and data structures [online], Vreda Pieterse and Paul E. Black, eds.
- [29] Tibshirani R, Walther G, and Hastie T. Estimating the number of clusters in a data set via the gap statistic. *J. R. Stat. Soc., B* 2001, 63(2): 411-423.

Figure 1. Down-selection scheme

Figure 2. Lists of immune classes for the second step of down-selection.

- (a) Inclusive set of immune classes, which are putatively part of a CoP based on one of the following conditions: i) a known CoP for one or more licensed vaccines; ii) evidence as a CoP based on CoR analysis of the RV144 trial and assumptions for assessing CoRs as CoPs; or iii) evidence as a CoP in non-human primates challenge trials for HIV-1 vaccines.

(b) Subset of significant inverse CoRs of HIV-1 infection in RV144 for Step 2 [weighting by strength of correlation]

	Class	Assay	Weight*
T Cell	CD4 Polyfunctionality	1. CD4 Polyfunctionality Score	1.12
		2. PBMC Cytokine Score MBA	$\frac{1}{4} \times 0.57$
	T Cell Cytokines	3. IL-5	$\frac{1}{4} \times 0.89$
		4. IL-10	$\frac{1}{4} \times 1.24$
		5. IL-13	$\frac{1}{4} \times 1.34$
Humoral	V2 Binding Antibody	6. IgG V1V2	$\frac{1}{2} \times 1.12$
		7. IgG3 V1V2	$\frac{1}{2} \times 0.87$
	V3 Binding Antibody / IgA	8. V3 Binding / IgA score**	0.83
	IgA Binding Antibody	9. -IgA score	$\frac{1}{2} \times 0.91$
		10. -IgA C1 Peptide	$\frac{1}{2} \times 1.15$
	Avidity / IgA	11. Avidity / IgA score	0.99
	Neutralization / IgA	12. Nab / IgA score	0.77
T Cell + Humoral	ADCC / IgA	13. ADCC / IgA score	0.50
	CD4 ICS / IgA	14. CD4 ICS / IgA score	0.80

*Weight = Absolute estimated log OR per 2 SD increase, from bottom to top tertile, or from negative to positive in the identified CoR from Haynes et al. 2012 (5), Yates et al. 2014 (16), Gottardo et al. 2013 (10), Tomaras et al. 2013 (11), Zolla-Pazner et al. 2014 (12), Lin et al. 2015 (14)

**IgA score: Primary IgA variable from Haynes et al. 2012 (5)

Figure 3. Demonstration of the superiority and non-redundancy criteria. **(a)** Suppose six regimens with mean immune endpoint scores enter the down-selection. The pairwise relationship is shown in **(b)**, where $\rightarrow$ indicates that one regimen is superior to the other with respect to their immune profiles, with the arrow pointing to the inferior regimen; and $\leftrightarrow$ indicates equivalence between two regimens with respect to their immune profiles. Panels **(c)-(g)** display some possible outcomes based on the down-selection. The non-redundancy criterion is satisfied if there is not a $\rightarrow$ or $\leftrightarrow$ among the selected set. Here the non-redundancy criterion is satisfied in **(f)** and **(g)** but not in **(c)-(e)**. Among the two panels where the non-redundancy criterion is satisfied, the superiority criterion is also satisfied in **(g)** but not in **(f)**. For the latter, regimen B is selected but there is a regimen A superior to B in the original set before the down-selection.

Figure 4. Results of the data example. **(a)** Assay-specific mean immune endpoint scores for five HIV-1 vaccine regimens. The eight endpoint scores are: IgG binding antibody responses to six different gp120 antigens measured using the binding antibody multiplex assay (BAMA) (log-transformed blank-subtracted MFI readout at a 1:50 dilution), mean NAb response (log-transformed ID50 titer) to six HIV-1 isolates measured using the TZM-bL assay and the A3R5 assay, and CD4+ T-cell response measured using the intracellular cytokine staining (ICS) assay (log-transformed % of T cells expressing IFN- γ and /or IL-2) . The scores have been scaled by the standard deviations of the corresponding endpoint measures in RV144.T. **(b)** Principal component (PC) biplot by regimen. The x-axis is the value from the first principal component and the y-axis is the second principal component, where each axis label includes the percentage of variation in the total set of readouts captured by the principal component. Points on the plot represent the values of the principal components of each observation. Points that are close together correspond to observations that have similar values on the components displayed in the plot. The top axis is the first principal component loadings and the right axis is the second principal component loadings, where loadings are the weights by which each original immunogenicity endpoint score should be multiplied to get the value of the first or second principal component. An arrow (vector) is drawn for each immunogenicity endpoint from the origin to the point defined by its first two principal component loadings. Vectors that point in the same direction correspond to endpoints that have similar response profiles based on the first two PCs. The observations whose points project furthest in the direction in which the vector points are the observations that have the most weight of whatever the endpoint measures. Those points that project at the other end have the least weight of whatever the endpoint measures. The angle between two arrows conveys information about the correlation of the two endpoint scores, with a zero degree angle denoting perfect correlation and a 90 degree angle denoting no correlation. **(c)** Heatmap of the mean immune endpoint scores for the vaccine regimens. The number of clusters is estimated to be 4 based on the GAP statistics proposed in Tibshirani et al. (2001) [29]. The value for each endpoint score has been centered by the average of mean scores for this particular endpoint across regimens. Hierarchical clustering analysis is performed based on the complete linkage method; the distance matrix is generated as weighted Manhattan distance of mean endpoint scores. The blue vertical line cuts the hierarchical tree into four clusters. **(d)** Process of down-selection by RFS and CR+RFS.

(a)

	RV144.T	096.T1	096.T2	096.T3	096.T4
BAMA AE.A244 V1V2 Tags/293F (B1)	5.99	5.53	5.39	5.81	4.86
BAMA gp70_B.CaseA2 V1/V2/169K (B2)	4.37	4.54	3.83	4.52	3.77
BAMA gp70_B.CaseA_V1_V2 (B3)	3.76	3.83	3.81	4.00	3.60
BAMA A244 gp 120 gDneg/293F/mon (B4)	3.52	4.52	3.72	4.50	3.83
BAMA Vaccine Insert (B5)	3.85	4.16	3.68	4.31	3.61
BAMA Con S gp140 CFI (B6)	3.81	5.37	4.85	5.37	4.73
ICS	-3.23	-3.42	-3.84	-2.41	-2.79
NAb	6.13	7.50	6.78	7.20	7.20
Weighted AS*	2.37	2.91	2.39	3.18	2.83

*: weight=1/6 for each BAMA endpoint, and weight=1 for ICS and NAb

Boldface in each row indicates the best ranked regimen according to an individual endpoint score or the weighted score

(b)

(c)

(d)

RFS

Sorted Regimens by average score:

096.T3, 096.T1, 096.T4, 096.T2, RV144.T

Iteration 1. Regimens selected: 096.T3

096.T1 vs. 096.T3:

- superior (none), inferior (ICS)

Iteration 2. Regimens selected: 096.T3

096.T4 vs. 096.T3:

- superior (none), inferior (none)

Iteration 3. Regimens selected: 096.T3

096.T2 vs. 096.T3:

- superior (none), inferior (B4, B5, ICS)

Iteration 4. Regimens selected: 096.T3

RV144.T vs. 096.T3:

- superior (none), inferior (B4, B6, ICS, NAb)

Final regimens selected: 096.T3

CR+RFS

(I) CR step:

Four clusters generated

{RV144.T}, {096.T1, 096.T3}, {096.T2}, {096.T4}

Regimens entering RFS:

RV144.T, 096.T3, 096.T2, 096.T4

(II) RFS step

Iteration 1. Regimens selected: 096.T3

096.T4 vs. 096.T3:

- superior (none), inferior (none)

Iteration 2. Regimens selected: 096.T3

096.T2 vs. 096.T3:

- superior (none), inferior (B2, B4, B5, B6, ICS)

Iteration 3. Regimens selected: 096.T3

RV144.T vs. 096.T3:

- superior (none), inferior (B4, B5, B6, ICS, NAb)

Final regimens selected: 096.T3