bioRxiv preprint doi: https://doi.org/10.1101/2021.05.12.443737; this version posted May 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Monocyte-derived transcriptome signature indicates antibody-dependent cellular phagocytosis as the primary mechanism of vaccine-induced protection against HIV-1

Shida Shangguan[1,2], Philip K. Ehrenberg[1], Aviva Geretz[1,2], Lauren Yum[1,2], Gautam Kundu[1,2], Kelly May[1,2], Slim Fourati[3], Krystelle Nganou-Makamdop[4], LaTonya D. Williams[5], Sheetal Sawant[5], Eric Lewitus[1,2], Punnee Pitisuttithum[6], Sorachai Nitayaphan[7], Suwat Chariyalertsak[8], Supachai Rerks-Ngarm[9], Morgane Rolland[1,2], Daniel Douek[4], Peter Gilbert[10], Georgia D. Tomaras[5], Nelson Michael[1], Sandhya Vasan[1,2], Rasmi Thomas[1,*]

[1]US Military HIV Research Program (MHRP), Walter Reed Army Institute of Research, Silver Spring, Maryland, United States, [2]Henry M. Jackson Foundation for the Advancement of Military Medicine, Bethesda, Maryland, United States, [3]Department of Pathology and Laboratory Medicine, Emory University, Atlanta, Georgia, United States, [4]Vaccine Research Center, NIH, Bethesda, Maryland, United States, [5]Departments of Surgery, Immunology and Molecular Genetics and Microbiology, Duke University School of Medicine, Durham, North Carolina, United States, [6]Vaccine Trial Centre, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand, [7]Armed Forces Research Institute of Medical Sciences, Bangkok, Thailand, [8]Research Institute for Health Sciences and Faculty of Public Health, Chiang Mai University, Chiang Mai, Thailand, [9]Department of Disease Control, Ministry of Public Health, Nonthaburi, Thailand, [10]Fred Hutchinson Cancer Research Center, Seattle, Washington, United States.

*To whom correspondence should be addressed: Rasmi Thomas, PhD, U. S. Military HIV Research Program (MHRP), Walter Reed Army Institute of Research, 503 Robert Grant Avenue,

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.12.443737; this version posted May 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Rm 2W07, Silver Spring, MD, USA 20910. Email: rthomas@hivresearch.org; phone: 301-319-9960; fax: 301-319-7512

Conflict of interest statement

The authors have declared that no conflict of interest exists.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.12.443737; this version posted May 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Abstract

A gene signature previously correlated with mosaic adenovirus 26 vaccine protection in simian immunodeficiency virus (SIV) and SHIV challenge models in non-human primates (NHP). In this report we investigated presence of this signature as a correlate of reduced risk in human clinical trials and potential mechanism for protection. The absence of this gene signature in the DNA/rAd5 human vaccine trial which did not show efficacy, strengthens our hypothesis that this signature is only enriched in studies that demonstrated protection. This gene signature was enriched in the partially effective RV144 human trial that administered the ALVAC/protein vaccine, and we find that the signature associates with both decreased risk of HIV-1 acquisition and increased vaccine efficacy. Total RNA-seq in a clinical trial that used the same vaccine regimen as the RV144 HIV vaccine implicated antibody-dependent cellular phagocytosis (ADCP) as a potential mechanism of vaccine protection. CITE-seq profiling of 53 surface markers and transcriptomes of 53,777 single cells from the same trial, showed that genes in this signature were primarily expressed in cells belonging to the myeloid lineage including monocytes, which are major effector cells for ADCP. The consistent association of this transcriptome signature with vaccine efficacy represents a tool both to identify potential mechanisms, as with ADCP here, and to screen novel approaches to accelerate development of new vaccine candidates.

Keywords

HIV vaccine, vaccine efficacy, single cell, CITE-seq, transcriptomics, RNA-seq, ADCP, RV144

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.12.443737; this version posted May 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Introduction

HIV vaccines are being tested in ongoing HIV human efficacy trials including the HVTN 705 and 706 studies (Hsu & O'Connell, 2017; Mega, 2019). Vaccines from these trials were previously tested in non-human primate (NHP) and showed partial protection from infection (Barouch et al., 2015; Barouch et al., 2018). These vaccines are based on the adenovirus serotype 26 (Ad26)-based regimen and have been tested in human immunogenicity trials (Baden et al., 2020; Barouch et al., 2018). To date, the pivotal RV144 phase 3 human efficacy trial conducted in Thailand is the only vaccine to show any protection against HIV (Rerks-Ngarm et al., 2009). The vaccine used a canary-pox ALVAC-based vector with a bivalent gp120 protein boost. Although neither of the vaccine regimens using different viral vectors were fully efficacious, there is some consensus that current preventive and treatment methods along with a moderately effective vaccine could potentially reduce the HIV pandemic (Anderson, Swinton, & Garnett, 1995; Andersson et al., 2007; Fauci, 2017; Medlock et al., 2017). While a number of correlates of vaccine protection have been described for these studies, protection mediated by the humoral immune system including HIV-1 specific IgG antibody titers, antibody Fc polyfunctionality, antibody interactions with HLA class II gene products and antibody effector functions have been key features of these partially effective vaccines (Barouch et al., 2015; Barouch et al., 2018; Haynes et al., 2012; Prentice et al., 2015).

Recently, we showed that a vaccine-induced gene signature identified in B cells by an unbiased transcriptome-wide RNA-seq approach associated with decreased risk against SIV/SHIV infection in NHP studies evaluating the Ad26 vaccine (Ehrenberg et al., 2019). This geneset was also enriched in NHP and the human RV144 trial that employed a vaccine containing the ALVAC viral vector (Ehrenberg et al., 2019). This gene signature is not merely a

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.12.443737; this version posted May 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

general response elicited by vaccination as it was not enriched in the Ad26-MVA arm of the SHIV challenge in NHP that showed some protection previously (Barouch et al., 2018). This gene signature was initially defined when comparing differentially expressed genes between B cells and monocytes from vaccinated individuals in an Influenza immunogenicity trial (Nakaya et al., 2011). The geneset that was submitted to the molecular signature database (MSigdb) comprised the top 200 genes that were upregulated in monocytes compared to B cells. In our previous study, specific genes in this geneset were upregulated in the uninfected compared to infected group in multiple SIV/HIV trials (Ehrenberg et al., 2019). Genes that were previously correlated with immunogenicity in human vaccine trials of influenza and yellow fever, including TNFSF13 (APRIL), were also enriched in uninfected rhesus monkeys in the NHP studies (Li et al., 2014; Nakaya et al., 2011). Although we first identified this geneset in sorted B cells (Ehrenberg et al., 2019), we were able to identify the same signature associating with reduced infection in published microarray datasets from both bulk unstimulated and *in vitro* antigen-stimulated PBMCs from three independent preclinical and clinical studies (Fourati et al., 2019; Vaccari et al., 2018; Vaccari et al., 2016). To determine if specific immune responses might be driving protection in conjunction with the gene signature, we examined whether this geneset associated with these responses measured in the NHP studies. We observed that this gene signature was also enriched in animals with increased magnitude of ADCP (Ehrenberg et al., 2019). We propose that this gene signature is a correlate of reduced risk of infection in efficacy studies and further investigation of the enriched genes in the geneset could potentially help uncover the mechanism of vaccine protection. Here we investigate this gene signature further as a proxy of vaccine-induced protection in human clinical trials, to identify the cellular origin, as well as to investigate potential mechanisms for the decreased risk of infection.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.12.443737; this version posted May 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Results

Gene signature is absent in a human HIV vaccine trial that did not show efficacy

Since the gene signature associated with vaccine protection in multiple studies using different regimens, we wanted to further confirm that this signature was truly associated with protection by looking for its presence (or absence) in a human vaccine trial that failed to show efficacy. We screened for this gene signature in whole-transcriptome data from participants vaccinated with the DNA/rAd5 HIV-1 preventive vaccine in the HVTN 505 human efficacy trial. Immunizations in this trial were halted prior to reaching the clinical endpoint due to lack of efficacy (Hammer et al., 2013). When comparing infection status, enrichment of this gene signature, as defined by the normalized enrichment score (NES), was not significant in transcriptome data from sorted B cells or monocytes one month after the final immunization in the vaccinated arm of the study (NES=-1.18, P=0.09 and NES=1.12, P=0.18 respectively). This finding further supports our hypothesis that this gene signature is associated with protection as summarized in Table 1.

GES is the strongest correlate of protection in RV144

In previous analyses of NHP preclinical studies, we utilized a composite gene expression score (GES) consisting of an average of standardized expression of the specific number of enriched genes in one study, to predict infection status in an independent study using the overlapping expressed genes from the first study (Ehrenberg et al., 2019). While this method is successful in evaluating gene signatures in studies using similar vaccine strategies, we wanted to explore this approach across different studies using diverse platforms. For each independent study, we computed a GES derived from genes within the geneset that were enriched in uninfected donors by averaging standardized expression and showed that it associates with

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.12.443737; this version posted May 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

120 decreased HIV-1 infection (Fig. 1). The magnitude of the GES and total number of enriched
121 genes present in the gene signature is specific to each study and is higher in the uninfected
122 animals in the two NHP preclinical trials evaluating the mosaic Ad26 vaccine (09-11 and 13-19),
123 including the different arms of the 13-19 study (13-19a-b) (Fig. 1A-C). Further, in the RV144
124 trial, the GES of 63 enriched genes in the gene signature also was higher in the vaccinated
125 individuals that remained uninfected (Fig. 1D). The number of enriched genes in each study
126 might vary due to global differences in the vaccine strategies, but we consistently observed that
127 higher GES associated with protection from HIV acquisition. We took advantage of the
128 composite GES measurement to compare it with the other known primary correlates of HIV-1
129 infection risk in the human RV144 trial. IgG antibodies binding to the variable regions 1-2
130 (V1V2) of the HIV-1 Envelope (Env) have been shown to correlate with decreased risk of
131 infection, while IgA binding to Env associated with increased risk of infection (Haynes et al.,
132 2012). We show that the association of the GES in RV144 is a stronger correlate of reduced risk
133 of infection than the previously described V1V2-specific IgG antibodies (Fig. 2A). Cumulative
134 incidence curves of HIV-1 infection showed decreased rates of infection among vaccine
135 recipients with high GES (Fig. 2B) Estimated vaccine efficacy was higher among vaccine
136 recipients with higher GES (Fig. 2C). The distribution of area under the receiver operating
137 characteristic curve (AUC) and accuracy suggested that GES was also able to predict HIV-1
138 infection (Fig. 2D). The effect of GES was also tested in RV144 vaccine and placebo
139 participants who became infected during the trial (Rolland et al., 2012). If the GES was
140 associated with vaccine efficacy, we would expect that vaccinees with a high GES did not get
141 infected, hence vaccinees who became infected should have lower GES than placebo participants
142 (who reflect the entire distribution of GES). This was observed across 43 breakthrough

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.12.443737; this version posted May 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

participants, with a significant difference among participants infected with single HIV-1 founder variants (N=29) (Supplementary Fig. 1). These findings strengthen the hypothesis that the GES is associated with vaccine efficacy.

Gene signature associates with an antibody effector function in a human vaccine trial

Immune responses correlating with this signature can provide additional insights into mechanisms that could be harnessed to improve vaccine design. We previously showed in the NHP studies that the protective gene signature that was enriched in uninfected monkeys after Ad26/gp140 vaccination also associated with higher magnitude of ADCP (Ehrenberg et al., 2019). In the RV144 human trial a number of immunological parameters were previously measured as part of the immune-correlates analysis, but not ADCP. The RV306 immunogenicity trial that employed a similar prime boost RV144 vaccine regimen with additional late boosts provided us with a unique opportunity to test if the gene signature was associated with ADCP (Pitisuttithum et al., 2020). We generated transcriptome-wide gene expression data from peripheral blood two weeks after the RV144 vaccine regimen (prior to the additional boosts) and assessed for enrichment of the gene signature with the magnitude of ADCP measured at the same timepoint in 24 participants. The gene signature with 118 enriched genes was significantly associated with higher magnitude of ADCP (NES=3.0, $P<0.001$)(Fig. 3A). Using the same geneset, 93 genes were found to be enriched in a subset of overlapping participants (N=21), where samples were collected 3 days after the RV144 immunizations (NES=2.5, $P<0.001$)(Fig. 3B). A GES from the list of enriched genes associating with ADCP from both timepoints was computed in the RV144 dataset. ADCP GES from both timepoints correlated strongly with the protective RV144 GES (rho=0.74, P= 2.2e-16, rho=0.75, P= 2.2e-16)(Fig. 3C).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.12.443737; this version posted May 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

166

167 **Pathways are shared between ADCP and vaccine protection phenotypes**

168 These findings demonstrated a strong link of the geneset with both vaccine protection and ADCP
169 in NHP and human studies. We sought to broaden our understanding of the relationship between
170 the different enriched genes in the geneset and establish some of the top pathways with gene
171 membership from the different studies. Genes that were significantly enriched with either the
172 ADCP or infection phenotypes from the 09-11, 13-19, RV144, and RV306 studies
173 (Supplementary table 1) were classified into statistically significant terms from different
174 biological pathways and represented as networks. The top significant pathways were leukocyte
175 activation involved in immune responses, lysosome function and signaling by interleukins (Fig.
176 4A-B). Gene ontology of the 63 genes in the RV144 signature revealed that the top non-
177 redundant enriched clusters with gene membership were myeloid leukocyte activation, lysosome
178 and cellular response to oxidative stress genes (Fig. 4C).

179

180 **Cellular origin of the protective genes by single cell transcriptomics**

181 To dissect the cellular origin of these genes we performed simultaneous detection of mRNA and
182 cell surface expression from single cells using the cellular indexing of transcriptomes and
183 epitopes by sequencing (CITE-seq) technology in a subset of the vaccinated RV306 participants
184 (Fig. 5A). This technology allows simultaneous detection of cell surface markers and gene
185 expression from the same single cells. Our analysis revealed that a majority of the genes in the
186 RV144 signature were expressed in cells of the myeloid lineage, with monocyte subsets having
187 the highest average gene expression (Fig. 5B). A subset of genes were also significantly
188 associated with decreased risk of acquisition in a univariate analysis (Odds ratio<1.0, P<0.05,

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.12.443737; this version posted May 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

189 q<0.1) (Fig. 5C). A stepwise logistic regression analysis identified specific genes (SEMA4A,
190 SLC36A1, SERINC5, IL17RA, CTSD, CD68, GAA) to have independent associations with
191 reduced risk of acquisition and were mainly expressed in the monocyte compartment (Fig. 5D).
192 CD14 monocytes also had the greatest number of differentially expressed genes that were
193 associated with ADCP (Fig. 5E).
194
195 **Discussion**
196 Though an effective vaccine has been a challenge for the HIV field, we see a glimpse of
197 optimism in partially protective NHP and human studies (Barouch et al., 2015; Barouch et al.,
198 2013; Barouch et al., 2018; Rerks-Ngarm et al., 2009; Vaccari et al., 2018; Vaccari et al., 2016).
199 These studies provide a unique opportunity to identify correlates of reduced risk that could help
200 inform protective signals and enable design of enhanced vaccine strategies. Targeted and
201 unbiased approaches have implicated non-neutralizing antibodies as the major correlate of
202 reduced risk of HIV infection (Barouch et al., 2015; Barouch et al., 2013; Barouch et al., 2018;
203 Haynes et al., 2012; Vaccari et al., 2018; Vaccari et al., 2016). We previously showed that a
204 transcriptomic signature first identified in sorted B cells at timepoints prior to challenge was a
205 correlate of protection in two NHP studies after administration of the Ad26/gp140 vaccine. This
206 signature also associated with increased magnitude of ADCP in the vaccinated monkeys
207 (Ehrenberg et al., 2019). We also identified this signature in bulk PBMCs from other studies that
208 used the ALVAC/protein regimen, suggesting that this vaccine might be an indicator of effective
209 vaccination. In this report we further investigated this gene signature to answer the following
210 questions including: 1) can the gene signature's association with protection be substantiated in

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.12.443737; this version posted May 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

211 additional human efficacy trials, 2) does it associate with ADCP in human trials, and 3) what is
212 the cellular origin of the signature at the single cell level?

213 The gene signature previously associated with HIV vaccine protection in a number of
214 studies with partial protection (Ehrenberg et al., 2019). We hypothesized that if this gene set was
215 a true marker of HIV vaccine protection, it would not be enriched in a failed vaccine trial. HVTN
216 505 is a DNA based vaccine which despite not showing overall efficacy in a Phase 2b trial,
217 demonstrated both cellular and antibody effector mediated protection in specific subgroups of
218 individuals in follow-up studies (Fong et al., 2018; Janes et al., 2017; Neidich et al., 2019). We
219 performed transcriptomics on sorted cell subsets from HVTN 505 vaccinated individuals and did
220 not observe enrichment of the gene signature, further strengthening our notion that the geneset
221 could be a proxy for vaccine protection.
222 Next, we developed a method to assess this gene signature compared to other correlates of risk in
223 the human RV144 study. This method employs an analytical method using a GES which is
224 computational score generated from the average expression of all genes enriched in the signature
225 and associating with a phenotype. This method was tested across different NHP studies and
226 RV144 and showed consistent association with reduced risk of infection based on the study
227 specific GES. The composite GES computed from RV144 consisting of the standardized
228 expression of 63 genes had the strongest association with decreased risk of infection and
229 increased efficacy. The RV144 GES was also able to accurately predict infection status in the
230 study. This study shows that the GES composite score provides a robust analytical measurement
231 to explore the effect of genes as a continuous variable in immune-correlates analyses, and that it
232 could be applied to other ongoing efficacy studies. Further, the analysis of RV144 breakthrough
233 infections were consistent with GES being lower in the vaccinated infected participants

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.12.443737; this version posted May 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

234 compared to placebos and hence protective. These observations albeit only significant in the
235 group infected with single founder viruses, strengthen the premise of the RV144 GES being a
236 correlate of reduced risk of infection.

237 We previously showed in NHP challenge studies that the gene signature correlated with
238 increased magnitude of functional antibody responses (Ehrenberg et al., 2019). Although this
239 geneset associated with ADCP in in NHP, the same analyses were previously not possible in the
240 human RV144 study since this immune response was not reported (Haynes et al., 2012). ADCP
241 has since been implicated with vaccine protection in a number of NHP challenge studies
242 (Ackerman et al., 2018; Barouch et al., 2015; Barouch et al., 2013; Barouch et al., 2018; Bradley
243 et al., 2017; Neidich et al., 2019). It is reported that ADCP could be involved in most studies that
244 previously showed antibody-dependent correlates of protection against viruses (Tay, Wiehe, &
245 Pollara, 2019).To investigate the effect of the gene signature on the magnitude of ADCP, we
246 performed transcriptomics in samples from a human trial (RV306) that employed the same
247 RV144 regimen. At both day 3 and two weeks after the 4^{th} vaccine corresponding to the last
248 RV144 vaccine dose, this signature correlated with increased magnitude of ADCP responses. A
249 strong correlation was also observed between GES from the ADCP enriched genes and the
250 vaccine protection genes in RV144. These findings provide evidence supporting the antibody-
251 mediated effector function as the mechanistic basis of this signature.

252 The enriched genes from ADCP and infection risk in multiple studies of both NHP and
253 human were involved in overlapping functions related to leukocyte activation, lysosomal
254 degradation and immune stimulation by cytokines. The 63 genes from the RV144 signature were
255 mostly clustered in the myeloid leukocyte activation pathway, alluding to the cellular origin of
256 this signature. The specific genes in the gene set that associated with the greatest odds of reduced

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.12.443737; this version posted May 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

risk of infection including SEMA4A, CTSD, CD68 and GAA were all members of this pathway, but not TNFSF13 (APRIL) which was the most protective gene in the NHP studies. Although the geneset of interest was first seen in sorted B cells from vaccinated NHP, it was subsequently identified in transcriptomic data from PBMCs in the RV144 study (Ehrenberg et al., 2019). While samples were exhausted from the RV144 primary dataset, the RV306 clinical trial that employed the same ALVAC-protein vaccine regimen gave us a unique chance to explore the cellular origin of the RV144 signature using single cell transcriptomics. Single cell surface expression data revealed that the majority of genes were expressed in monocytes, which was not surprising given the fact that this geneset was originally defined as genes downregulated in B cells compared to monocytes after influenza vaccination (Nakaya et al., 2011). While our initial study found this signature in sorted B cells from the NHP challenge studies, single cell data provides further insight that monocytes could be the cellular origin of these genes in the RV144 study. Although monocytes were classified as mononuclear phagocytes almost fifty years ago, assays designed to specifically measure monocyte ADCP were not widely used in the context of vaccination until a few years ago (van Furth et al., 1972). Although monocytes have recently been implicated in vaccine-induced protection in preclinical vaccine trials of SIV challenge, our findings in human trials at the single cell level provides greater impetus to explore the role of other non-lymphoid cell populations on HIV-1 vaccine efficacy. (Gorini et al., 2020; Vaccari et al., 2018). Though we think that monocytes are important in the vaccine responses observed in RV144, it would be remiss not to mention that the effect of granulocytes including neutrophils in response to vaccination is missed when transcriptomics is performed in PBMC compared to blood. Other than the phagocytic cell, both antibody and Fc receptor diversity can influence ADCP mediated immune responses to viral pathogens and are also elements that warrant further

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.12.443737; this version posted May 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

study and may potentially be manipulated to improve vaccine efficacy (Chung & Alter, 2017; Geraghty, Thorball, Fellay, & Thomas, 2019; Tay, Wiehe, et al., 2019).

Our data demonstrate the potential to discover novel protective correlates using an approach that mines transcriptomic data in multiple preclinical and clinical trials. Unbiased transcriptome-wide analyses are able to identify biological perturbations that associate with vaccine protection even when differences are small, but credibility can only be strengthened by replicating findings across multiple studies. Gene signatures that associate not only with vaccine protection, but specific immune responses can be a prospective tool to evaluate vaccine effectiveness even prior to challenge or infection. Developing analytical tools that can interface with phenotypes such as vaccine protection across human and preclinical studies can allow for more systematic meta-analyses of data emerging from the ongoing HIV vaccine clinical trials, as well as the recently halted HVTN 702 trial (NIH, 2020). We propose that assessment of such gene signatures with immune responses in human immunogenicity trials could provide orthogonal insight for down-selection of vaccine candidates. Identifying overlapping correlates of protection in these studies could be pivotal to making discoveries that may allow for licensure and subsequent bridging studies of an effective HIV vaccine.

Methods

Study design

The aim of the study was post-hoc analyses of a protective gene expression signature identified previously in five SIV/HIV vaccine studies with efficacy and immune response data (Ehrenberg et al., 2019). To enable interpretation of this gene signature, bulk RNA-seq, scRNA-seq and functional data was generated in clinical samples from the RV306 and HVTN 505 human trials.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.12.443737; this version posted May 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

The RV306 vaccine trial was conducted in Thailand and all participants received the primary RV144 ALVAC/gp120 vaccine series, with additional late boosts assigned to specific groups (Pitisuttithum et al., 2020). Bulk RNA-seq was performed in 24 participants two weeks after the RV144 vaccine regimen (week 26). Additionally, RNA-seq was also performed 3 days after the same primary endpoint. The HVTN 505 trial used a DNA/rAd5 vaccine regimen to test safety and efficacy in a US population (Hammer et al., 2013). PBMC collected one month after the final immunization (month 7) was available from 47 vaccinees in the HVTN 505 study for RNA-seq (Hammer et al., 2013). The infection status of the vaccinees (22 cases and 25 controls) was categorized based on infection status between months 7-24. Microarray transcriptome data from PBMCs and immune response data for 170 vaccinated individuals from the RV144 study at timepoint two weeks post last vaccination was used for correlates analyses (Fourati et al., 2019; Haynes et al., 2012). All studies were approved by the participating local and international institution review boards. Informed consent was obtained from all participants in the different trials included in this study (Hammer et al., 2013; Pitisuttithum et al., 2020).

Bulk transcriptomics

RNA was extracted from sorted B cells (Aqua live/dead$^-$CD20$^+$CD3$^-$) and monocytes (Aqua live/dead$^-$CD20$^-$CD3$^-$CD56$^-$HLA-DR$^+$CD14$^+$ from PBMC of HVTN 505 vaccinees using RNAzolRT (MRC Inc.) as per recommendations from the manufacturer. For the preparation of mRNA libraries, polyadenylated transcripts were purified on oligo-dT magnetic beads, fragmented, reverse transcribed using random hexamers and incorporated into barcoded cDNA libraries based on the Illumina TruSeq platform. Next, libraries were validated by electrophoresis, quantified, pooled and clustered on Illumina TruSeq v2 flow cells. Clustered

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.12.443737; this version posted May 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

flow cells were sequenced on an Illumina HiSeq (2000/4000) using 2 x 75 base paired-end runs. Total RNA from RV306 participants was extracted from whole blood collected in PAXgene Blood RNA tubes, using the PAXgene Blood RNA kits (both Qiagen; Germantown, MD) and subsequently subjected to the GlobinClear kit (ThermoFisher Scientific; Waltham, MA) per manufacturer's suggestions. NGS RNA-seq was performed using the SMART-Seq technology (Picelli et al., 2014; Ramskold et al., 2012). Briefly, cDNA was generated from 10 ng of RNA using the SMART-Seq v4 UltraLow Input RNA Prep kit (Takara Bio Inc) per manufacturer's suggestions, with control RNA spiked-in (ThermoFisher Scientific). Sequencing libraries were generated using the Nextera XT DNA Sample Prep kit (Illumina, San Diego, CA). Concentration of each sample in the pooled libraries were determined using the paired-end 300-cycle MiSeq Reagent Nano Kit v2 (2 x 150 bp) on a MiSeq instrument (both Illumina). NGS was performed on a final adjusted library pool using the paired-end 300-cycle NovaSeq 6000 S2 XP Reagent Kit (2 x 150 bp) on a NovaSeq instrument (both Illumina) per the manufacturer's instructions. Fastp v0.19.7 and Trimmomatic v0.33 with default parameters was used to trim low-quality bases from both ends of each read. Trimmed reads were aligned to the human genome (GRCh38 build 88-92) using HISAT2 v2.1.0 or the STAR aligner (v2.4.2a) and HTSeq (v0.6.1-0.9.1) was used for counting. Trimmed mean of M-values normalization method, as implemented in the R package edgeR, was used for normalization.

Single cell transcriptomics

Simultaneous evaluation of mRNA and cell surface expression from single cells was performed using feature barcoding technology from 10x Genomics, based on the CITE-seq technology (Stoeckius et al., 2017) . Cell hashing (HTO) was used in conjunction with the 10x Genomics

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.12.443737; this version posted May 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

5'V(D)J Feature Barcoding kit to generate single cell mRNA (GEX) and antibody-derived tag (ADT) libraries (Stoeckius et al., 2017; Stoeckius et al., 2018). Briefly, PBMCs from 12 samples were hashed using TotalSeq-C anti-human Hashtag antibodies and combined into two batches. In each batch, surface proteins were stained with a cocktail of 53 TotalSeq-C antibodies (Biolegend). Antibody concentrations were either predetermined by titration (Kotliarov et al., 2020) or used at a default concentration. 50,000 cells from each batch were loaded onto each of 4 wells of a Chromium Chip, and GEX and ADT (HTO and Feature Barcode (FB)) libraries were constructed following the manufacturer's protocol. Libraries were pooled and quantitated using a MiSeq Nano v2 reagent cartridge. Final libraries were sequenced on the NovaSeq 6000, S4 reagent cartridge (2x100 bp) (Illumina).

CITE-seq data analyses

FASTQ files were demultiplexed with bcl2fastq v2.20 (Illumina). Alignment and counting were performed using Cell Ranger v3.1.0 (10x Genomics) and the human reference files provided by 10x Genomics (human genome GRCh38 and Ensembl annotation v93). The average number of genes per cell was 1453 and the average number of unique molecular identifiers (UMI) was 4248. The mean read depth per cell was approximately 65,000–84,000. The minimum fraction of reads mapped to the genome was 88% and sequencing saturation was above 85% for all lanes with an average of 88%. The computational analysis of ADT data was performed using the Seurat v3.1 package (Stuart et al., 2019). HTO expression matrices were CLR (Centered Log-Ratio) normalized and demultiplexed using MULTIseqDemux. The FB matrices from the Seurat objects were split into cell-positive and negative droplet matrices using the HTO demultiplexing results, and were used for denoising by DSB (Denoised and Scaled by Background)

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.12.443737; this version posted May 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

372 normalization (Kotliarov et al., 2020)(https://cran.r-project.org/web/packages/dsb/index.html).
373 Only cells with <10% mitochondrial genes were retained, and cells were assigned to specific
374 donors using the HTO demultiplexing results. A total of 53,777 single cells remained after the
375 quality control process. The gene expression matrices for all samples were normalized and
376 integrated into a single object in Seurat (Stuart et al., 2019). Based on the workflow described in
377 Kotliarov et al., a distance matrix was generated from cell surface protein features (Kotliarov et
378 al., 2020). This matrix was used for shared-nearest-neighbor finding and clustering at
379 resolution=0.5. Neighbor finding and clustering were performed on the integrated gene
380 expression data at a resolution=0.75 and dimensions=1:30. A tSNE (t-distributed stochastic
381 neighbour embedding) was generated from the protein data PCA. Seurat was used to generate a
382 heatmap, dotplot and featureplots. Differential gene expression testing was performed within
383 each cluster between the high and low ADCP groups using Seurat's FindMarkers function.
384 ADCP DEG were filtered to genes with >10% expression in either group, a log fold change
385 value > 0.25, and a Bonferroni p value < 0.05.
386
387 **ADCP assay**
388 The antibody effector function ADCP was measured as previously described (Ackerman et al.,
389 2011; Tay, Kunz, et al., 2019; Tay et al., 2016). Briefly, A244 gp120 Env-coated fluorescent
390 beads were incubated at 37°C for 2 hours with diluted plasma (1:50) collected at week 26, two
391 weeks after administration of the RV144 vaccination series. Anti-CD4 monoclonal antibody-
392 treated THP-1 cells (human monocytic cell line; ATCC TIB-201) (treated for 15 minutes at 4°C)
393 were added to immune complexes and spinoculated for 1 hour at 4°C to allow phagocytosis to
394 occur. Supernatant was removed, cells were washed and fixed in

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.12.443737; this version posted May 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

paraformaldehyde. Phagocytosis was measured by flow cytometry and a phagocytosis score was calculated as follows: phagocytosis score = (% pos * MFI of Sample) / (% pos * MFI of no-antibody PBS control). The CD4 binding-site broadly neutralizing antibody (bnAb), CH31, was used as a positive control, and the influenza receptor binding site-specific bnAb, CH65, was used as a negative control. Results are representative of two independent experiments.

Pathway analyses

Association of the protective gene signature with infection (HVTN 505) or magnitude of median ADCP (RV306) responses were analyzed using the Gene Set Enrichment Analysis (GSEA) method as described previously (Ehrenberg et al., 2019; Subramanian et al., 2005). GSEA was performed on vaccinated HVTN 505 participants at the visit 7 timepoint, one month after the last immunization. RNA-seq was performed on samples prior to infection, but participants were categorized based on their infection status. GSEA was performed on 45 RV306 RNA-seq samples that also had ADCP scores obtained at the week 26 (week 2 after the 4^{th} vaccination) time-point. Participants were categorized into high and low ADCP groups based on the median values of ADCP measured in a total 79 vaccinated participants. The RNA-seq gene expression values at the day 3 and week 2 time-points were then analyzed for gene enrichment using a gene set of 200 genes, obtained from the Broad Institute (GSE29618_BCELL_VS_MONOCYTE_DAY7_FLU_VACCINE_DN), between the two groups of samples based on the classification, in each time-point. The gene signature of interest was considered significantly enriched using a threshold of NES$\geq$1.4 and P<0.001 as described previously (Ehrenberg et al., 2019). Gene ontology and network analyses of genes were performed using Metascape with default parameters (Zhou et al., 2019).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.12.443737; this version posted May 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Correlates of protection

Composite gene expression score (GES) was computed as the average of standardized expression of normalized enriched genes in the gene signature in different vaccine studies. The samples in each vaccine study were grouped into outcomes after challenge or infection status after immunization (Barouch et al., 2015; Barouch et al., 2018; Rerks-Ngarm et al., 2009). Logistic regression was used for evaluating the association between GES and HIV-1 infection in the RV144 study. The fitting methods accommodate the 2-phase sampling design via maximum likelihood estimation (Breslow & Holubkov, 1997). Cumulative HIV-1 incidence curves were plotted for the three subgroups of vaccine recipients defined by tertiles into the lower, middle, and upper third of the GES (Low, Medium, High subgroups), as well as for the entire placebo group HIV negative at week 24 (n=6267 subjects) for reference. These curves were estimated using the Kaplan-Meier method with inverse probability weighting that accounted for the sampling design. Next, vaccine efficacy (VE) for the GES subgroups versus the entire placebo group was estimated as one minus the odds of infection in vaccine recipients with Low/Medium/High response divided by the odds of infection in the entire placebo group HIV-1 negative at week 24 of enrollment in the study. The RV144 prediction analysis was implemented by logistic regression. The dataset was randomly split into training and testing sets in a 7:3 ratio, while retaining class distributions within the groups. The training dataset consisted of 119 individuals while the test dataset consisted of 51 individuals. A logistic regression of GES was fit on to the training dataset (Prentice et al., 2015). The model's discriminative ability was evaluated by generating a receiver operator characteristic curve (ROC) and the corresponding AUC on the test dataset. The prediction accuracy of the model was also assessed on the test dataset. The probability that gives minimum mis-classification error was chosen as the cut-off. This process

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.12.443737; this version posted May 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

was repeated 1000 times and the distribution of the resulting AUC and accuracy were demonstrated by a histogram with a density curve.

Among 121 RV144 participants who became infected during the trial and had their HIV-1 genome sequenced at diagnosis, 43 had GES measurements computed from microarray data (Fourati et al., 2019; Rolland et al., 2012). Vaccine and placebo groups were compared overall and after stratifying infections with single HIV-1 founders.

Other statistical analyses

Logistic regression that accounted for the sampling design was used for the univariate analyses of the 63 enriched genes. A radar plot of the significant genes was generated to illustrate odds ratios (OR) and 95% confidence intervals. All ORs were reported per 1-SD increase. Significant genes resulted from univariate logistic regressions of the 63 enriched genes were further analyzed with a multivariate stepwise logistic regression to identify genes that independently associated with HIV protection. Akaike information criterion (AIC) was used to identify the optimal set of genes. The expressed enriched genes associated with higher magnitude of ADCP in RV306 at day 3 and 2 weeks post the RV144 vaccine regimen were used to compute the ADCP GES in RV144. Spearman correlation was calculated between the ADCP GES from the two timepoints and the infection GES respectively.

All descriptive and inferential statistical analyses were performed using GraphPad Prism 8 (GraphPad Software) and R 3.6.1 (or later) software packages. Comparison of groups was performed using Mann-Whitney tests or t-tests when assumptions were met. All logistic regression models were adjusted for gender and baseline risk behavior and one significant principal component axis (Haynes et al., 2012; Prentice et al., 2015). A two-sided P value of less

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.12.443737; this version posted May 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

464 than 0.05 was considered significant. The Benjamini and Hochberg method was used to calculate
465 false discovery rate (FDR)-adjusted p values for multiple testing corrections.

466
467 **Acknowledgements:** We would like to thank the volunteers and staff of the RV306, RV144 and
468 HVTN 505 clinical trials. We also acknowledge DeAnna Tenney and Derrick Goodman, Duke
469 University for expert technical assistance. The views expressed are those of the authors and
470 should not be construed to represent the positions of the U.S. Army or the U.S. Department of
471 Defense (DOD). This work was supported by a cooperative agreement (W81XWH-07-2-0067)
472 between the Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., and
473 the DOD. This research was funded in part by the U.S. National Institute of Allergy and
474 Infectious Disease.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.12.443737; this version posted May 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

References

Ackerman, M. E., Das, J., Pittala, S., Broge, T., Linde, C., Suscovich, T. J., . . . Alter, G. (2018). Route of immunization defines multiple mechanisms of vaccine-mediated protection against SIV. *Nat Med, 24*(10), 1590-1598. doi:10.1038/s41591-018-0161-0

Ackerman, M. E., Moldt, B., Wyatt, R. T., Dugast, A. S., McAndrew, E., Tsoukas, S., . . . Alter, G. (2011). A robust, high-throughput assay to determine the phagocytic activity of clinical antibody samples. *J Immunol Methods, 366*(1-2), 8-19. doi:10.1016/j.jim.2010.12.016

Anderson, R. M., Swinton, J., & Garnett, G. P. (1995). Potential impact of low efficacy HIV-1 vaccines in populations with high rates of infection. *Proc Biol Sci, 261*(1361), 147-151. doi:10.1098/rspb.1995.0129

Andersson, K. M., Owens, D. K., Vardas, E., Gray, G. E., McIntyre, J. A., & Paltiel, A. D. (2007). Predicting the impact of a partially effective HIV vaccine and subsequent risk behavior change on the heterosexual HIV epidemic in low- and middle-income countries: A South African example. *J Acquir Immune Defic Syndr, 46*(1), 78-90. doi:10.1097/QAI.0b013e31812506fd

Baden, L. R., Stieh, D. J., Sarnecki, M., Walsh, S. R., Tomaras, G. D., Kublin, J. G., . . . Traverse, H. H. P. X. S. T. (2020). Safety and immunogenicity of two heterologous HIV vaccine regimens in healthy, HIV-uninfected adults (TRAVERSE): a randomised, parallel-group, placebo-controlled, double-blind, phase 1/2a study. *Lancet HIV, 7*(10), e688-e698. doi:10.1016/S2352-3018(20)30229-0

Barouch, D. H., Alter, G., Broge, T., Linde, C., Ackerman, M. E., Brown, E. P., . . . Schuitemaker, H. (2015). Protective efficacy of adenovirus/protein vaccines against SIV

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.12.443737; this version posted May 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

challenges in rhesus monkeys. *Science, 349*(6245), 320-324. doi:10.1126/science.aab3886

Barouch, D. H., Stephenson, K. E., Borducchi, E. N., Smith, K., Stanley, K., McNally, A. G., . . . Michael, N. L. (2013). Protective efficacy of a global HIV-1 mosaic vaccine against heterologous SHIV challenges in rhesus monkeys. *Cell, 155*(3), 531-539. doi:10.1016/j.cell.2013.09.061

Barouch, D. H., Tomaka, F. L., Wegmann, F., Stieh, D. J., Alter, G., Robb, M. L., . . . Schuitemaker, H. (2018). Evaluation of a mosaic HIV-1 vaccine in a multicentre, randomised, double-blind, placebo-controlled, phase 1/2a clinical trial (APPROACH) and in rhesus monkeys (NHP 13-19). *Lancet, 392*(10143), 232-243. doi:10.1016/S0140-6736(18)31364-3

Bradley, T., Pollara, J., Santra, S., Vandergrift, N., Pittala, S., Bailey-Kellogg, C., . . . Haynes, B. F. (2017). Pentavalent HIV-1 vaccine protects against simian-human immunodeficiency virus challenge. *Nat Commun, 8*, 15711. doi:10.1038/ncomms15711

Breslow, N. E., & Holubkov, R. (1997). Weighted likelihood, pseudo-likelihood and maximum likelihood methods for logistic regression analysis of two-stage data. *Stat Med, 16*(1-3), 103-116. doi:10.1002/(sici)1097-0258(19970115)16:1<103::aid-sim474>3.0.co;2-p

Chung, A. W., & Alter, G. (2017). Systems serology: profiling vaccine induced humoral immunity against HIV. *Retrovirology, 14*(1), 57. doi:10.1186/s12977-017-0380-3

Ehrenberg, P. K., Shangguan, S., Issac, B., Alter, G., Geretz, A., Izumi, T., . . . Thomas, R. (2019). A vaccine-induced gene expression signature correlates with protection against SIV and HIV in multiple trials. *Sci Transl Med, 11*(507). doi:10.1126/scitranslmed.aaw4236

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.12.443737; this version posted May 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

521 Fauci, A. S. (2017). An HIV Vaccine Is Essential for Ending the HIV/AIDS Pandemic. *JAMA, 318*(16), 1535-1536. doi:10.1001/jama.2017.13505

523 Fong, Y., Shen, X., Ashley, V. C., Deal, A., Seaton, K. E., Yu, C., . . . Tomaras, G. D. (2018). Modification of the Association Between T-Cell Immune Responses and Human Immunodeficiency Virus Type 1 Infection Risk by Vaccine-Induced Antibody Responses in the HVTN 505 Trial. *J Infect Dis, 217*(8), 1280-1288. doi:10.1093/infdis/jiy008

527 Fourati, S., Ribeiro, S. P., Blasco Tavares Pereira Lopes, F., Talla, A., Lefebvre, F., Cameron, M., . . . Sekaly, R. P. (2019). Integrated systems approach defines the antiviral pathways conferring protection by the RV144 HIV vaccine. *Nat Commun, 10*(1), 863. doi:10.1038/s41467-019-08854-2

531 Geraghty, D. E., Thorball, C. W., Fellay, J., & Thomas, R. (2019). Effect of Fc Receptor Genetic Diversity on HIV-1 Disease Pathogenesis. *Front Immunol, 10*, 970. doi:10.3389/fimmu.2019.00970

534 Gorini, G., Fourati, S., Vaccari, M., Rahman, M. A., Gordon, S. N., Brown, D. R., . . . Franchini, G. (2020). Engagement of monocytes, NK cells, and CD4+ Th1 cells by ALVAC-SIV vaccination results in a decreased risk of SIVmac251 vaginal acquisition. *PLoS Pathog, 16*(3), e1008377. doi:10.1371/journal.ppat.1008377

538 Hammer, S. M., Sobieszczyk, M. E., Janes, H., Karuna, S. T., Mulligan, M. J., Grove, D., . . . Gilbert, P. B. (2013). Efficacy trial of a DNA/rAd5 HIV-1 preventive vaccine. *The New England journal of medicine, 369*(22), 2083-2092. doi:10.1056/NEJMoa1310566

541 Haynes, B. F., Gilbert, P. B., McElrath, M. J., Zolla-Pazner, S., Tomaras, G. D., Alam, S. M., . . . Kim, J. H. (2012). Immune-correlates analysis of an HIV-1 vaccine efficacy trial. *The New England journal of medicine, 366*(14), 1275-1286. doi:10.1056/NEJMoa1113425

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.12.443737; this version posted May 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

544 Hsu, D. C., & O'Connell, R. J. (2017). Progress in HIV vaccine development. *Hum Vaccin*
545 *Immunother, 13*(5), 1018-1030. doi:10.1080/21645515.2016.1276138
546 Janes, H. E., Cohen, K. W., Frahm, N., De Rosa, S. C., Sanchez, B., Hural, J., . . . McElrath, M.
547 J. (2017). Higher T-Cell Responses Induced by DNA/rAd5 HIV-1 Preventive Vaccine
548 Are Associated With Lower HIV-1 Infection Risk in an Efficacy Trial. *J Infect Dis,*
549 *215*(9), 1376-1385. doi:10.1093/infdis/jix086
550 Kotliarov, Y., Sparks, R., Martins, A. J., Mule, M. P., Lu, Y., Goswami, M., . . . Tsang, J. S.
551 (2020). Broad immune activation underlies shared set point signatures for vaccine
552 responsiveness in healthy individuals and disease activity in patients with lupus. *Nat*
553 *Med*. doi:10.1038/s41591-020-0769-8
554 Li, S., Rouphael, N., Duraisingham, S., Romero-Steiner, S., Presnell, S., Davis, C., . . .
555 Pulendran, B. (2014). Molecular signatures of antibody responses derived from a systems
556 biology study of five human vaccines. *Nat Immunol, 15*(2), 195-204. doi:10.1038/ni.2789
557 Medlock, J., Pandey, A., Parpia, A. S., Tang, A., Skrip, L. A., & Galvani, A. P. (2017).
558 Effectiveness of UNAIDS targets and HIV vaccination across 127 countries. *Proc Natl*
559 *Acad Sci U S A, 114*(15), 4017-4022. doi:10.1073/pnas.1620788114
560 Mega, E. R. (2019). 'Mosaic' HIV vaccine to be tested in thousands of people across the world.
561 *Nature, 572*(7768), 165-166. doi:10.1038/d41586-019-02319-8
562 Nakaya, H. I., Wrammert, J., Lee, E. K., Racioppi, L., Marie-Kunze, S., Haining, W. N., . . .
563 Pulendran, B. (2011). Systems biology of vaccination for seasonal influenza in humans.
564 *Nat Immunol, 12*(8), 786-795. doi:10.1038/ni.2067

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.12.443737; this version posted May 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

565 Neidich, S. D., Fong, Y., Li, S. S., Geraghty, D. E., Williamson, B. D., Young, W. C., . . .
566 Tomaras, G. D. (2019). Antibody Fc effector functions and IgG3 associate with
567 decreased HIV-1 risk. *J Clin Invest, 129*(11), 4838-4849. doi:10.1172/JCI126391
568 NIH. (2020). NIH news release. Experimental HIV Vaccine Regimen Ineffective in Preventing
569 HIV. National Institutes of Health Web Site (2020). https://www.nih.gov/news-
570 events/news-releases/experimental-hiv-vaccine-regimen-ineffective-preventing-hiv
571 [Press release]
572 Picelli, S., Faridani, O. R., Bjorklund, A. K., Winberg, G., Sagasser, S., & Sandberg, R. (2014).
573 Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc, 9*(1), 171-181.
574 doi:10.1038/nprot.2014.006
575 Pitisuttithum, P., Nitayaphan, S., Chariyalertsak, S., Kaewkungwal, J., Dawson, P., Dhitavat, J., .
576 . . group, R. V. s. (2020). Late boosting of the RV144 regimen with AIDSVAX B/E and
577 ALVAC-HIV in HIV-uninfected Thai volunteers: a double-blind, randomised controlled
578 trial. *Lancet HIV, 7*(4), e238-e248. doi:10.1016/S2352-3018(19)30406-0
579 Prentice, H. A., Tomaras, G. D., Geraghty, D. E., Apps, R., Fong, Y., Ehrenberg, P. K., . . .
580 Thomas, R. (2015). HLA class II genes modulate vaccine-induced antibody responses to
581 affect HIV-1 acquisition. *Sci Transl Med, 7*(296), 296ra112.
582 doi:10.1126/scitranslmed.aab4005
583 Ramskold, D., Luo, S., Wang, Y. C., Li, R., Deng, Q., Faridani, O. R., . . . Sandberg, R. (2012).
584 Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor
585 cells. *Nat Biotechnol, 30*(8), 777-782. doi:10.1038/nbt.2282
586 Rerks-Ngarm, S., Pitisuttithum, P., Nitayaphan, S., Kaewkungwal, J., Chiu, J., Paris, R., . . .
587 Kim, J. H. (2009). Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.12.443737; this version posted May 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

588 in Thailand. *The New England journal of medicine, 361*(23), 2209-2220.
589 doi:10.1056/NEJMoa0908492
590 Rolland, M., Edlefsen, P. T., Larsen, B. B., Tovanabutra, S., Sanders-Buell, E., Hertz, T., . . .
591 Kim, J. H. (2012). Increased HIV-1 vaccine efficacy against viruses with genetic
592 signatures in Env V2. *Nature, 490*(7420), 417-420. doi:10.1038/nature11519
593 Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P. K.,
594 Swerdlow, H., . . . Smibert, P. (2017). Simultaneous epitope and transcriptome
595 measurement in single cells. *Nat Methods, 14*(9), 865-868. doi:10.1038/nmeth.4380
596 Stoeckius, M., Zheng, S., Houck-Loomis, B., Hao, S., Yeung, B. Z., Mauck, W. M., 3rd, . . .
597 Satija, R. (2018). Cell Hashing with barcoded antibodies enables multiplexing and
598 doublet detection for single cell genomics. *Genome Biol, 19*(1), 224. doi:10.1186/s13059-
599 018-1603-1
600 Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., 3rd, . . . Satija,
601 R. (2019). Comprehensive Integration of Single-Cell Data. *Cell, 177*(7), 1888-1902
602 e1821. doi:10.1016/j.cell.2019.05.031
603 Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., . . .
604 Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for
605 interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A, 102*(43), 15545-
606 15550. doi:10.1073/pnas.0506580102
607 Tay, M. Z., Kunz, E. L., Deal, A., Zhang, L., Seaton, K. E., Rountree, W., . . . Permar, S. R.
608 (2019). Rare Detection of Antiviral Functions of Polyclonal IgA Isolated from Plasma
609 and Breast Milk Compartments in Women Chronically Infected with HIV-1. *J Virol,*
610 *93*(7). doi:10.1128/JVI.02084-18

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.12.443737; this version posted May 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

611 Tay, M. Z., Liu, P., Williams, L. D., McRaven, M. D., Sawant, S., Gurley, T. C., . . . Tomaras,
612 G. D. (2016). Antibody-Mediated Internalization of Infectious HIV-1 Virions Differs
613 among Antibody Isotypes and Subclasses. *PLoS Pathog, 12*(8), e1005817.
614 doi:10.1371/journal.ppat.1005817
615 Tay, M. Z., Wiehe, K., & Pollara, J. (2019). Antibody-Dependent Cellular Phagocytosis in
616 Antiviral Immune Responses. *Front Immunol, 10*, 332. doi:10.3389/fimmu.2019.00332
617 Vaccari, M., Fourati, S., Gordon, S. N., Brown, D. R., Bissa, M., Schifanella, L., . . . Franchini,
618 G. (2018). HIV vaccine candidate activation of hypoxia and the inflammasome in
619 CD14(+) monocytes is associated with a decreased risk of SIVmac251 acquisition. *Nat*
620 *Med, 24*(6), 847-856. doi:10.1038/s41591-018-0025-7
621 Vaccari, M., Gordon, S. N., Fourati, S., Schifanella, L., Liyanage, N. P., Cameron, M., . . .
622 Franchini, G. (2016). Adjuvant-dependent innate and adaptive immune signatures of risk
623 of SIVmac251 acquisition. *Nat Med, 22*(7), 762-770. doi:10.1038/nm.4105
624 van Furth, R., Cohn, Z. A., Hirsch, J. G., Humphrey, J. H., Spector, W. G., & Langevoort, H. L.
625 (1972). The mononuclear phagocyte system: a new classification of macrophages,
626 monocytes, and their precursor cells. *Bull World Health Organ, 46*(6), 845-852.
627 Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/4538544
628 Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., . . . Chanda, S.
629 K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-
630 level datasets. *Nat Commun, 10*(1), 1523. doi:10.1038/s41467-019-09234-6
631
632

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.12.443737; this version posted May 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

633 **Figure Legends**

634 **Fig. 1. Composite gene expression scores (GES) are higher in the uninfected compared to**
635 **infected groups.** GES computed from enriched genes in the geneset is higher in the uninfected
636 compared to infected vaccinated NHP and humans. (A) Ad26/gp140 (09-11 NHP SIV challenge
637 study, 58 enriched genes, N = 10), (B) Ad26/gp140 (13-19 NHP SHIV challenge study, 58
638 enriched genes, N = 11), (C) Ad26/Ad26+gp140 (13-19 NHP SHIV challenge study, 68 enriched
639 genes, N = 12), and (D) ALVAC/gp 120 (RV144 human efficacy trial, 63 enriched genes, N =
640 170). The statistical significance was calculated by either Mann-Whitney or unpaired t-test.

641

642 **Fig. 2. GES is a stronger correlate of reduced risk of infection in RV144.** A GES of the 63
643 enriched genes in the RV144 study was examined as a continuous variable (N = 170). (A) GES is
644 associated with lower odds of HIV acquisition compared to the other two primary correlates of
645 risk. Variables were measured at week 26, two weeks post last vaccination. For each variable, the
646 odds ratio is reported per 1-SD increase. Transcriptome data was available only in a subset of the
647 246 donors. (B) Probability of acquiring HIV-1 is lower in individuals with higher GES. (C)
648 Vaccine efficacy is increased significantly in individuals with high GES. (D) Distribution of AUC
649 and accuracy plotted after repeating the process 1000 times showed that GES could predict HIV-
650 1 infection with AUC of 0.67 ± 0.08 and with accuracy of 0.83 ± 0.02.

651

652 **Fig. 3. Strong relationship between functional ADCP responses in a human vaccine trial and**
653 **the protective RV144 signature.** The geneset that associated with protection in an efficacy study
654 was also enriched with higher magnitude of ADCP measured two weeks after vaccination in an
655 immunogenicity trial that employed the RV144 vaccine regimen. NES from RNA-seq data at

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.12.443737; this version posted May 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

timepoints (A) 2 weeks (118 enriched genes) (N = 24) and (B) 3 days (93 enriched genes) (N = 21) post the RV144 vaccine regimen in the RV306 trial are indicated. (C) GES computed from the enriched genes associating with ADCP correlated strongly with the protective GES in the RV144 study (N =170).

Fig. 4. Pathway analyses of the enriched genes in the different vaccine studies. A meta-analysis of pathways including enriched genes with reduced infection or higher ADCP was performed. (A) The top related pathways and their (B) corresponding significance are indicated for the RV144 infection (N = 172), 09-11 (infection & ADCP) (N = 10), two arms of 13-19 (infection & ADCP) (N = 23), and RV306 ADCP (2 timepoints) (N = 45) enriched genes. Each term is represented by a circle node, where its size is proportional to the number of input genes that fall into that term, and its color represents its cluster identity. Terms with a similarity score > 0.3 are linked by an edge (the thickness of the edge represents the similarity score). The network is visualized with Cytoscape (v3.1.2) with "force-directed" layout and with edge bundled for clarity. (C) Clustering of the 63 enriched genes that associated with reduced infection in the RV144 study.

Fig. 5. Cellular origin of the RV144 signature. Single cell CITE-seq in vaccinated participants (N = 12) that employed the RV144 vaccine regimen (day 3 after last vaccination) identified expression of the genes in the signature in cells from the myeloid lineage. (A) Clustering based on cell surface expression of CITE-seq data (B) Heat map of the mRNA expression of the 63 genes from the RV144 signature from single cells. Columns represent single cells from different protein cell subsets and rows the mRNA gene expression (C) Radar plot showing significant genes in the signature that associated with decreased risk of infection in RV144 ($P<0.05$, $q<0.1$) (N =170). (D)

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.12.443737; this version posted May 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

679 Feature plots of the expression of the most protective genes show that SEMA4A, IL17RA, CTSD,
680 CD68 and GAA were mainly expressed in monocytes. (E) CD14+ monocytes had the highest
681 number of differentially expressed genes (DEG) when comparing high versus low ADCP (2 weeks
682 after vaccination) from single cell CITE-seq vaccinated participants that employed the RV144
683 vaccine regimen (day 3 after last vaccination).
684
685 **Supplementary Material**
686 **Supplementary Fig. 1. Association of the GES with HIV-1 breakthrough infections in a**
687 **human vaccine trial.** GES was significantly higher in the placebo (blue) compared to vaccine
688 recipients (red) among the RV144 participants with single-founder breakthrough infection
689 ($P<0.05$). Numbers indicate the number of participants plotted. Asterisks indicate significant
690 pairwise differences by Mann-Whitney test.
691
692 **Supplementary table 1.** Number of enriched genes from the geneset in different comparisons
693 from multiple studies.

was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 1
and is also made available for use under a CC0 license.

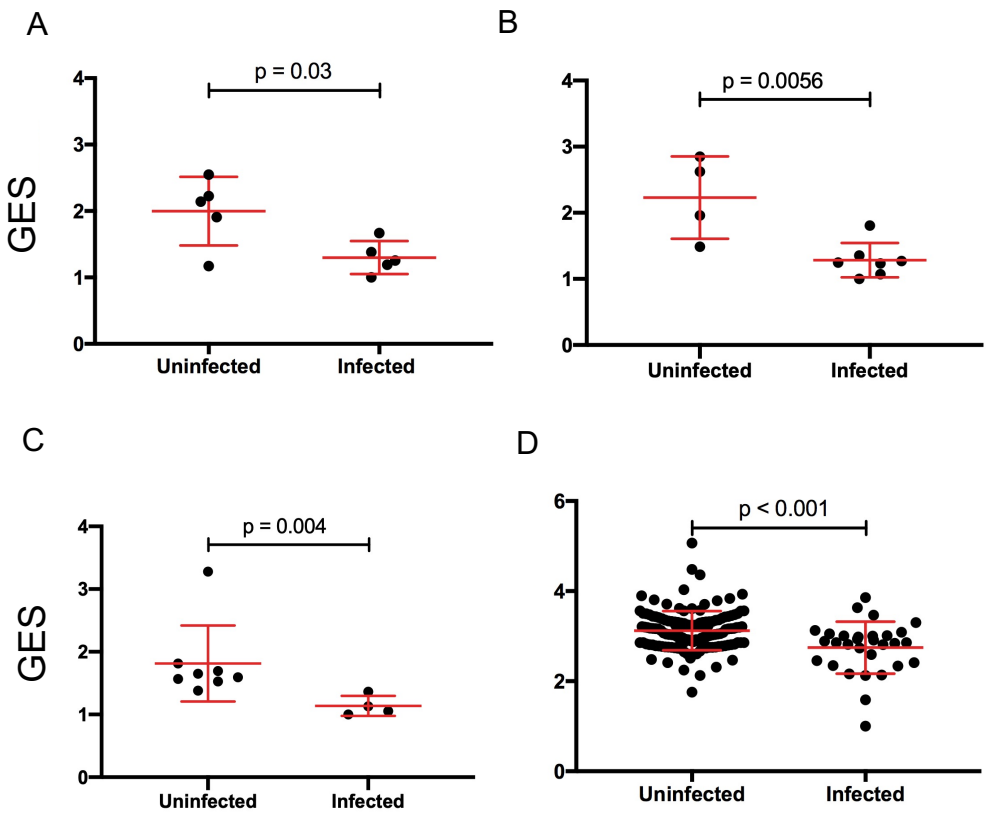


Fig. 1. Composite gene expression scores (GES) are higher in the uninfected compared to infected groups. GES computed from enriched genes in the geneset is higher in the uninfected compared to infected vaccinated NHP and humans. (A) Ad26/gp140 (09-11 NHP SIV challenge study, 58 enriched genes, N = 10), (B) Ad26/gp140 (13-19 NHP SHIV challenge study, 58 enriched genes, N = 11), (C) Ad26/Ad26+gp140 (13-19 NHP SHIV challenge study, 68 enriched genes, N = 12), and (D) ALVAC/gp 120 (RV144 human efficacy trial, 63 enriched genes, N = 170). The statistical significance was calculated by either Mann-Whitney or unpaired t-test.

was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 1
and is also made available for use under a CC0 license.

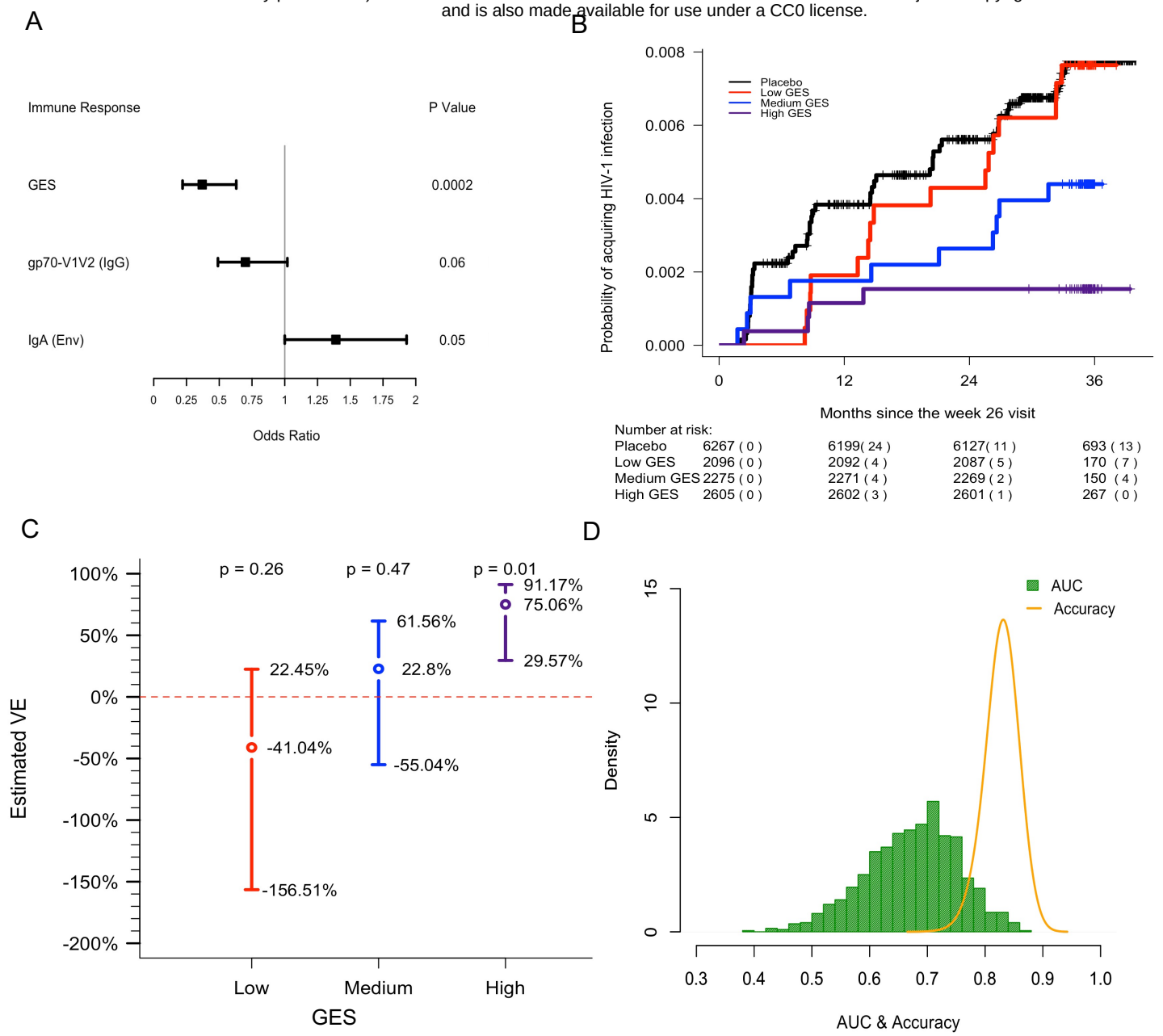


Fig. 2. GES is a stronger correlate of reduced risk of infection in RV144. A GES of the 63 enriched genes in the RV144 study was examined as a continuous variable (N = 170). (A) GES is associated with lower odds of HIV acquisition compared to the other two primary correlates of risk. Variables were measured at week 26, two weeks post last vaccination. For each variable, the odds ratio is reported per 1-SD increase. Transcriptome data was available only in a subset of the 246 donors. (B) Probability of acquiring HIV-1 is lower in individuals with higher GES. (C) Vaccine efficacy is increased significantly in individuals with high GES. (D) Distribution of AUC and accuracy plotted after repeating the process 1000 times showed that GES could predict HIV-1 infection with AUC of 0.67 ± 0.08 and with accuracy of 0.83 ± 0.02.

was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 1
and is also made available for use under a CC0 license.

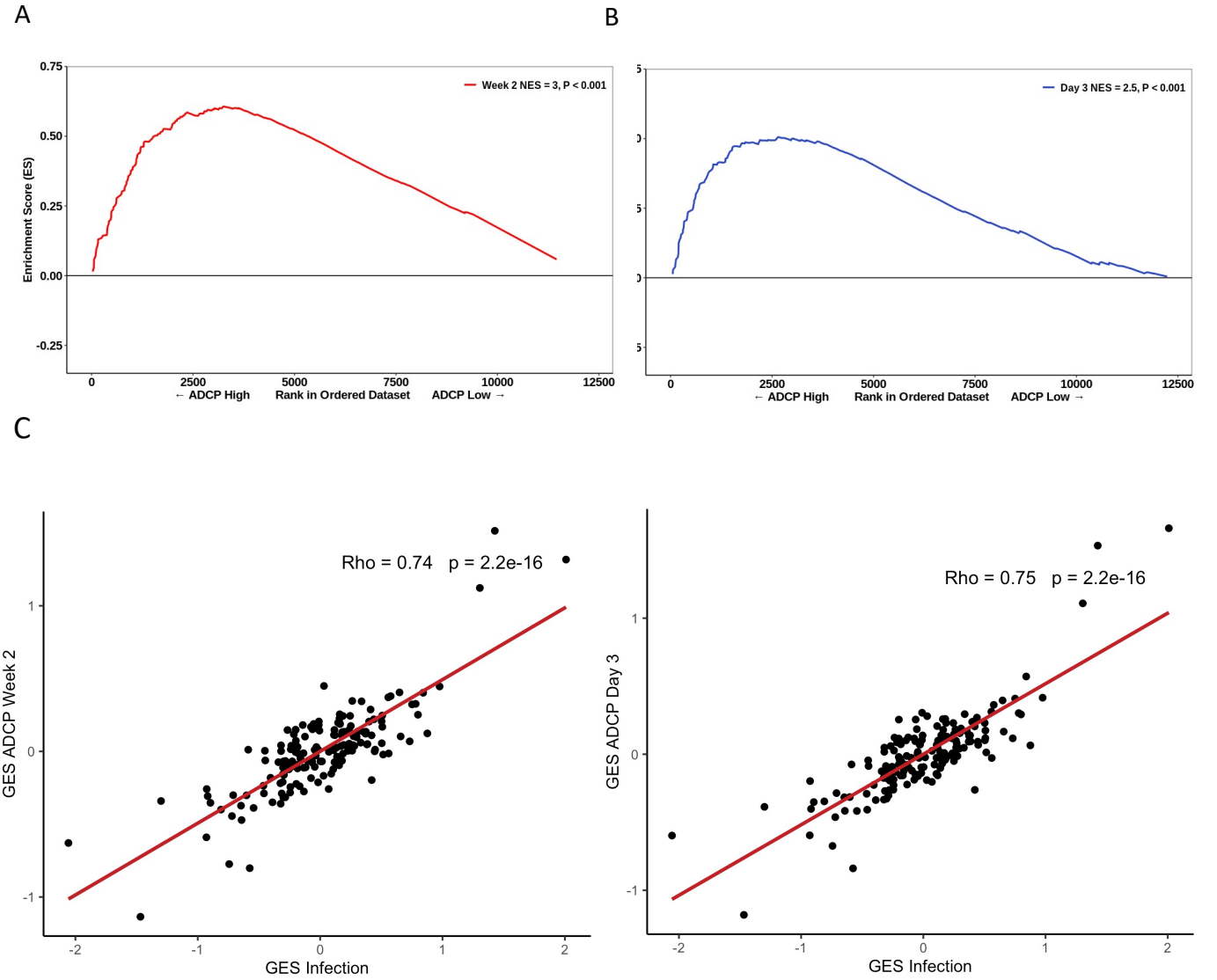


Fig. 3. Strong relationship between functional ADCP responses in a human vaccine trial and the protective RV144 signature. The geneset that associated with protection in an efficacy study was also enriched with higher magnitude of ADCP measured two weeks after vaccination in an immunogenicity trial that employed the RV144 vaccine regimen. NES from RNA-seq data at timepoints (A) 2 weeks (118 enriched genes) (N = 24) and (B) 3 days (93 enriched genes) (N = 21) post the RV144 vaccine regimen in the RV306 trial are indicated. (C) GES computed from the enriched genes associating with ADCP correlated strongly with the protective GES in the RV144 study (N =170).

was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 1
and is also made available for use under a CC0 license.

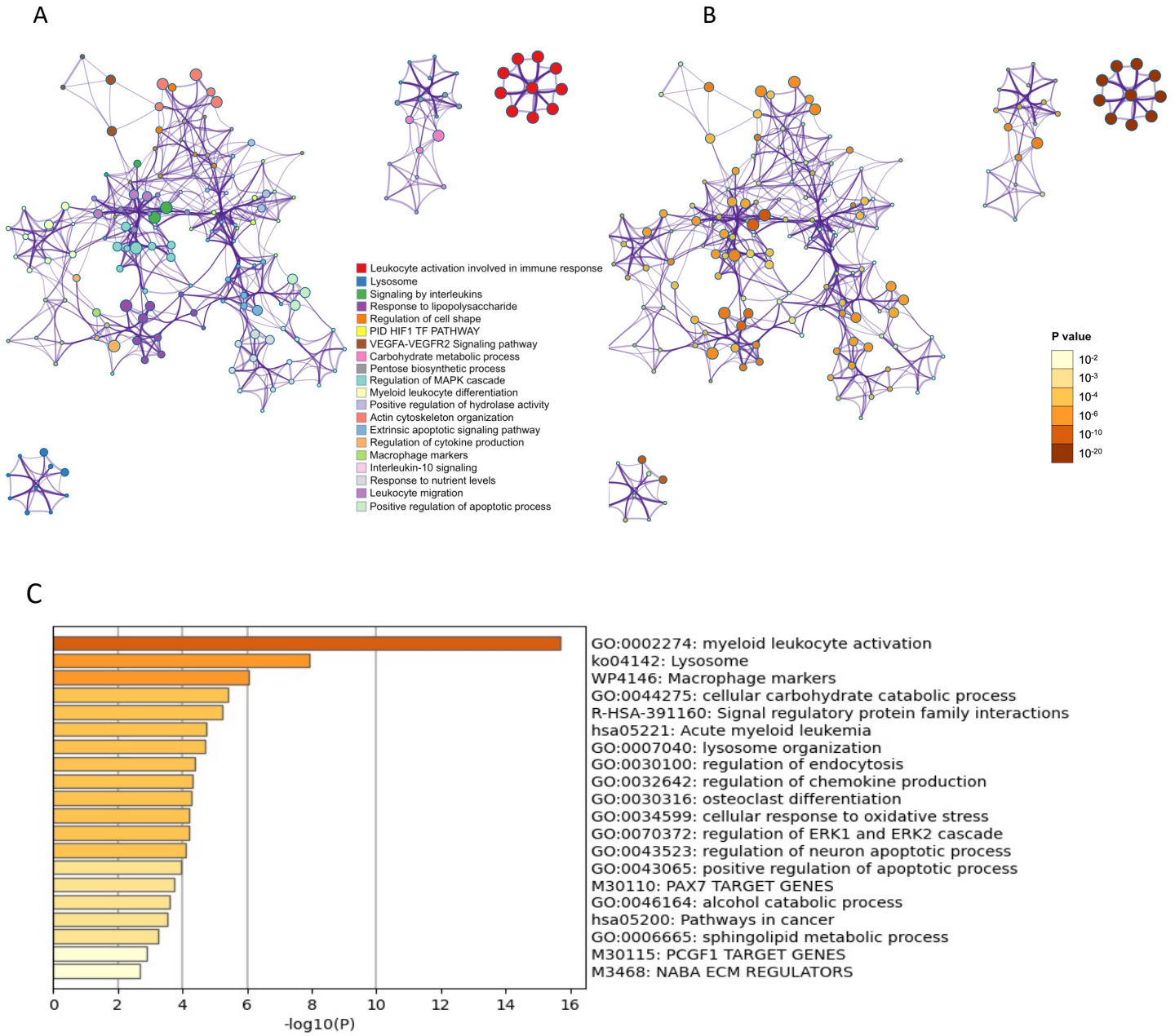


Fig. 4. Pathway analyses of the enriched genes in the different vaccine studies. A meta-analysis of pathways including enriched genes with reduced infection or higher ADCP was performed. (A) The top related pathways and their (B) corresponding significance are indicated for the RV144 infection (N = 172), 09-11 (infection & ADCP) (N = 10), two arms of 13-19 (infection & ADCP) (N = 23), and RV306 ADCP (2 timepoints) (N = 45) enriched genes. Each term is represented by a circle node, where its size is proportional to the number of input genes that fall into that term, and its color represents its cluster identity. Terms with a similarity score > 0.3 are linked by an edge (the thickness of the edge represents the similarity score). The network is visualized with Cytoscape (v3.1.2) with "force-directed" layout and with edge bundled for clarity. (C) Clustering of the 63 enriched genes that associated with reduced infection in the RV144 study.

was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 1
and is also made available for use under a CC0 license.

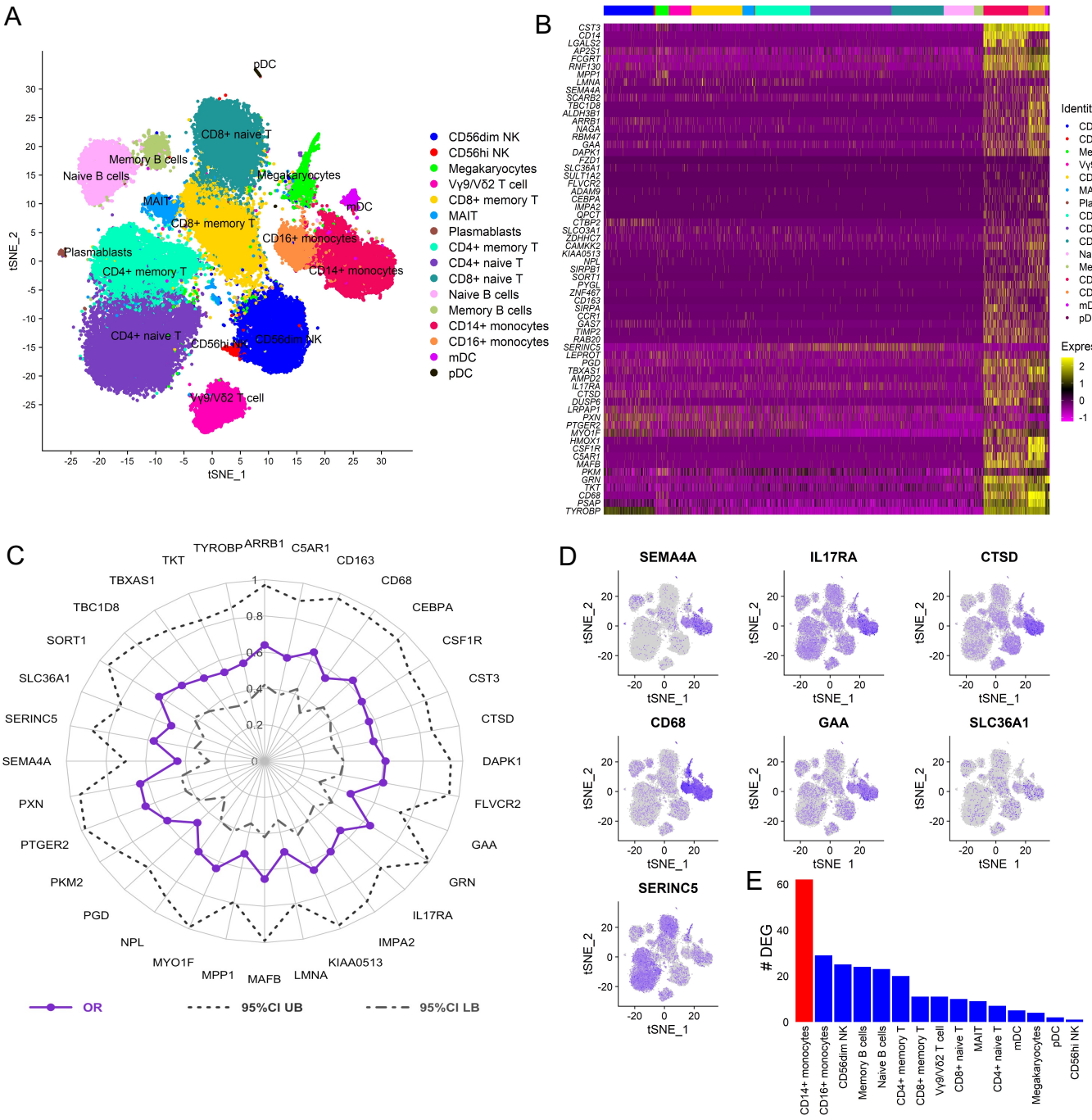


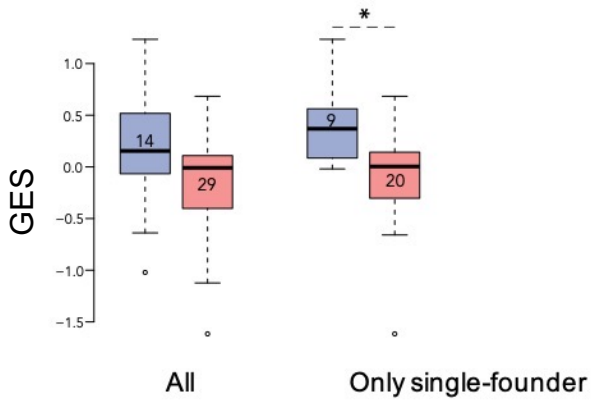
Fig. 5. Cellular origin of the RV144 signature. Single cell CITE-seq in vaccinated participants (N = 12) that employed the RV144 vaccine regimen (day 3 after last vaccination) identified expression of the genes in the signature in cells from the myeloid lineage. (A) Clustering based on cell surface expression of CITE-seq data (B) Heat map of the mRNA expression of the 63 genes from the RV144 signature from single cells. Columns represent single cells from different protein cell subsets and rows the mRNA gene expression (C) Radar plot showing significant genes in the signature that associated with decreased risk of infection in RV144 ($P<0.05$, $q<0.1$) (N =170). (D) Feature plots of the expression of the most protective genes show that SEMA4A, IL17RA, CTSD, CD68 and GAA were mainly expressed in monocytes. (E) CD14+ monocytes had the highest number of differentially expressed genes (DEG) when comparing high versus low ADCP (2 weeks after vaccination) from single cell CITE-seq vaccinated participants that employed the RV144 vaccine regimen (day 3 after last vaccination).

was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 1
and is also made available for use under a CC0 license.

Table 1. Gene signature associates with vaccine protection in multiple trials.

Study	Vaccine Regimen	Species	Partial protection	N	Method	Protective signature
09-11	Ad26/gp140	NHP	Y	10	RNA-seq	Y
13-19	Ad26/gp140	NHP	Y	11	RNA-seq	Y
13-19	A26/Ad26 +gp140	NHP	Y	12	RNA-seq	Y
13-19	Ad26/MVA+gp140	NHP	Y	9	RNA-seq	N
	ALVAC-SIV/gp120	NHP	Y	27	Microarray	Y
	DNA-SIV/ALVAC+gp120	NHP	Y	12	Microarray	Y
RV144	ALVAC/gp120	Human	Y	170	Microarray	Y
HVTN 505	DNA/rAd5	Human	N	42	RNA-seq	N

was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 1
and is also made available for use under a CC0 license.



Supplementary Fig. 1. Association of the GES with HIV-1 breakthrough infections in a human vaccine trial. GES was significantly higher in the placebo (blue) compared to vaccine recipients (red) among the RV144 participants with single-founder breakthrough infection ($P<0.05$). Numbers indicate the number of participants plotted. Asterisk indicate significant pairwise differences by Mann-Whitney test.