Published in final edited form as:

Sci Transl Med. 2019 August 28; 11(507): . doi:10.1126/scitranslmed.aaw4236.

A vaccine-induced gene expression signature correlates with protection against SIV and HIV in multiple trials

Philip K. Ehrenberg¹, Shida Shangguan^{1,2}, Biju Issac^{1,2}, Galit Alter³, Aviva Geretz^{1,2}, Taisuke Izumi^{1,2}, Christopher Bryant⁴, Michael A. Eller^{1,2}, Frank Wegmann⁵, Richard Apps⁶, Matthew Creegan^{1,2}, Diane L. Bolton^{1,2}, Rafick P. Sekaly⁷, Merlin L. Robb^{1,2}, Robert A. Gramzinski¹, Maria G. Pau⁵, Hanneke Schuitemaker⁵, Dan H. Barouch^{3,8}, Nelson L. Michael^{1,2}, Rasmi Thomas^{9,2,†}

¹U.S. Military HIV Research Program (MHRP), Walter Reed Army Institute of Research, Silver Spring, MD 20910, USA. ²Henry M. Jackson Foundation for the Advancement of Military Medicine, Bethesda, MD 20817, USA. ³Ragon Institute of MGH, MIT, and Harvard, Cambridge, MA 02139, USA. ⁴EMMES Corporation, Rockville, MD 20850, USA. ⁵Janssen Vaccines and Prevention B.V., 2329 Leiden, Netherlands. ⁶Center for Human Immunology, National Institutes of Health, Bethesda, MD 20814, USA. ⁷Case Western Reserve University, Cleveland, OH 44106, USA. ⁸Center for Virology and Vaccine Research, BIDMC, Harvard Medical School, Boston, MA 02115, USA. ⁹U.S. Military HIV Research Program (MHRP), Walter Reed Army Institute of Research, Silver Spring, MD 20910, USA.

Abstract

Current HIV vaccines are only partially efficacious, but this presents an opportunity to identify correlates of protection, providing new insight into mechanisms able to prevent acquisition. Two independent preclinical challenge studies in non-human primates (NHPs) previously showed partial efficacy of a mosaic adenovirus 26 (Ad26)-based HIV-1 vaccine candidate. To investigate the basis of this protection, we performed whole transcriptomics profiling by RNA sequencing

[†]Corresponding author. rthomas@hivresearch.org.

[#]These authors contributed equally to this work.

Author contributions: Author contributions are listed in alphabetical order. R.T. conceived and supervised the project. D.H.B., M.L.R., N.L.M., R.A.G. and R.T. contributed to the study rationale. A.G., B.I., C.B., P.K.E., R.A., R.T. and S.S. contributed to the data analyses and interpretation and manuscript preparation. D.L.B., G.A., M.A.E., M.C., P.K.E., R.P.S. and T.I. provided reagents and performed laboratory experiments/analyses. D.H.B., F.W., H.S. and M.G.P. provided samples and immune response data from the preclinical trials. All authors edited the manuscript.

Competing interests: M.G.P., H.S. and F.W. are employees of Janssen, pharmaceutical companies of Johnson & Johnson. D.H.B. received consulting payments from Merck, IGM Biosciences, and SQZ Biotech. D.H.B. is a co-inventor of HIV vaccine patents PCT/US2009/064999 "Antiviral vaccines with improved cellular immunogenicity" and PCT/US2009/060494 "Biochemically stabilized HIV-1 env trimer vaccine." M.G.P. is a coinventor on patents PCT/US2017/049817 and PCT/US2018/051274 "Methods for inducing an immune response against human immunodeficiency virus infection in subjects undergoing antiretroviral treatment." H.S. is a coinventor on PCT/US2017/049817. M.G.P. and H.S. are coauthors on patents PCT/US2018/043016 "Methods for safe induction of cross-clade immunity against human immunodeficiency virus infection in human" and PCT/US2015/051891 "Methods and compositions for inducing protective immunity against human immunodeficiency virus infection."

Data and materials availability: All data associated with this study are present in the paper or the Supplementary Materials. NGS sequencing data have been deposited in the European Nucleotide Archive repository (www.ebi.ac.uk/) under study accession number PRJEB31297.

Supplementary Materials:

stm.sciencemag.org/cgi/content/full/11/507/eaaw4236/DC1

(RNA-seq) in sorted lymphocytes from peripheral blood samples at different time points after vaccination but before challenge. We observed a transcriptional signature in B cells that associated with protection from acquisition of simian immunodeficiency virus (SIV) or the simian-human immunodeficiency virus (SHIV) in both studies. Strong antibody responses were elicited, and genes from the signature for which expression was enriched specifically associated with higher magnitude of functional antibody responses. The same gene expression signature also associated with protection in RV144, in the only human HIV vaccine trial to date that has shown efficacy, and in two additional NHP studies that evaluated similar canarypox-based vaccine regimens. A composite gene expression score derived from the gene signature was one of the top-ranked correlates of protection in the NHP vaccine studies. This study aims to bridge preclinical and clinical data with the identification of a gene signature in B cells that is associated with protection from SIV and HIV infection by providing a new approach for evaluating future vaccine candidates.

One Sentence Summary:

A gene signature in B cells associates with protective efficacy of different vaccine regimens against SIV and HIV infection.

INTRODUCTION

A highly efficacious preventive HIV vaccine in humans has been elusive to date. The RV144 HIV-1 vaccine showed partial efficacy in people from Thailand, but was insufficient for licensure (1). Mathematical modeling has demonstrated that with effective pre-exposure prophylaxis the ability to reduce HIV risk by even 50% could be an effective public health approach to the HIV pandemic (2). Some vaccine regimens have shown 50% or more efficacy in the nonhuman primate (NHP) simian immunodeficiency virus (SIV) or the chimeric simian HIV (SHIV) challenge models, and correlates of infection risk that are linked to vaccine-evoked immune responses have been identified (3-5). With human clinical trials being time consuming and expensive, preclinical trials in NHPs are invaluable. There are clear differences in virus exposure and transmission between humans and NHPs, but being able to test a product rapidly from bench to bedside makes the NHP model an attractive alternative to other animal models. Steps are being taken to understand events leading from human exposure to infectious virus to productive infection in order to design a NHP model that most accurately imitates natural HIV infection (6). Although the protection observed in NHPs can be used as evidence that a vaccine regimen may be effective and deserves further development, strong correlates that predict immunogenicity and protection in clinical trials are also needed. Recent advances in next-generation sequencing (NGS) technologies promote a more prominent role for gene expression analysis both in the optimization of vaccination regimens and in gaining understanding of the immunological basis for protection (7-9).

Protective efficacy of 50% was previously identified in an SIV challenge study with an adenovirus 26 (Ad26) prime followed by a purified envelope glycoprotein (gp140) boost vaccine regimen (3). These findings were replicated in the SHIV challenge model in an independent NHP study (4). We proposed that the partial efficacy observed could be due to

variation in host gene expression elicited by vaccination. Here, we used whole transcriptional profiling in different lymphocyte subsets to find a gene signature that correlated with protection in multiple studies.

RESULTS

A B cell signature correlates with protection in multiple trials

We analyzed peripheral blood mononuclear cells (PBMCs) that were available from rhesus monkeys (09-11 study) immunized with the Ad26/gp140 vaccine regimen, where 50% of the animals remained uninfected after six intrarectal SIVmac251 challenges (3). We also included samples that were available from a second NHP preclinical trial (13-19 study) where rhesus monkeys were immunized with the same vaccine that showed 33% efficacy after six intrarectal SHIV-SF162P3 challenges (4). We extracted total RNA at different time points after vaccination but prior to challenge from four sorted lymphocyte populations of CD4⁺ T cells, CD8⁺ T cells, natural killer (NK) cells and B cells (fig. S1). We then generated whole transcriptome RNA sequencing (RNA-seq) data from sorted populations in 21 animals from the Ad26/gp140 arms of the SIV and SHIV preclinical challenge trials (10-12). Differential expression analysis found no overlapping genes that were significant (absolute fold change > 1.5, $P < 0.05$, $q < 0.10$) when comparing samples from animals that were protected from infection or succumbed to infection at the end of six challenges. We performed gene set enrichment analysis (GSEA) in both studies to identify whether previously defined sets of genes (gene signatures) were associated with infection status at the end of six challenges (13-15). We screened signatures from the Molecular Signatures Database (MSigDB) previously observed to be associated with biological phenotypes for enrichment in monkeys that remained uninfected after the challenge [normalized enrichment score (NES) 1.4, $P < 0.001$] (13). We identified one gene set that was significantly enriched in B cells from uninfected animals in both the 09-11 and 13-19 studies (NES 1.4, $P < 0.001$) (Fig. 1A and data file S1). This gene set consisting of 200 genes was previously generated from gene expression datasets from human immunogenicity trials of influenza vaccine, when comparing B cells to monocytes at day 7 after vaccination (16). Of the 200 genes enriched in humans, only 140 genes were detected in the RNA-seq data generated in the NHP studies. Enrichment was greatest in the 13-19 study sampled at 2 weeks after vaccination (NES=2.13, $P < 0.001$), but was still present in the 09-11 study analyzed 20 weeks after the last vaccine dose (NES=1.48, $P < 0.001$). Other than the B cell subset, there were no enriched gene sets that were common between the two studies in the other lymphocyte populations (NES 1.4, $P < 0.001$) (data file S1).

Given that we observed the same signal in two independent NHP studies, we asked whether this signature was enriched in the human RV144 HIV-1 vaccine efficacy trial (1). Although the vaccine regimen in the RV144 study was different from those of the NHP preclinical trials and included a canarypox virus ALVAC prime with a gp120 protein boost, we observed the same B cell gene signature enrichment in the uninfected compared with the infected RV144 vaccinees in total PBMCs from 172 donors 2 weeks after vaccination (NES=2.04, $P < 0.001$) (Fig. 1B) (1). We also observed enrichment of this signature in two independent SIV challenge studies that showed some vaccine protection and used pox-

protein regimens similar to RV144 (NES=2.66, $P<0.001$), although significance was slightly decreased for the ALVAC-SIV/gp120 study, where gene expression was measured before the last vaccination (NES=1.59, $P<0.002$) (Fig. 1C and D) (17, 18). The total number of enriched genes associating with protection from infection in the B cell signature varied between the different studies (table S1).

A composite gene expression score is a predictor of SIV/SHIV acquisition in NHPs

Because this B cell signature associated with protection in multiple studies, we investigated whether the enriched genes in the signature from the Ad26/gp140 arm of the 09-11 study (3) could predict acquisition in the Ad26/gp140 arm of the 13-19 study (4). There were 58 enriched genes in the B cell signature in the 09-11 study, of which only 53 passed the gene expression filtering step in the 13-19 study. We used a composite gene expression score (GES) consisting of an average of standardized expression of 53 normalized genes from the 09-11 training dataset to build our prediction model. Internal validation using optimism corrected bootstrapping determined an area under the curve (AUC) of 0.84, suggesting that this classifier was able to discriminate infected from uninfected animals in the 09-11 study. This model, consisting of 53 expressed genes, was able to predict protection in the 13-19 test dataset (AUC=0.86) (Fig. 2A). We then tested this prediction model in an independent arm of the 13-19 study that had received an additional Ad26 boost and was previously shown to be the most protective, with 67% of the animals remaining uninfected after the last challenge (4). We performed RNA-seq on sorted B cells from animals of this arm of the 13-19 study and computed the composite GES for the 53 genes that were enriched in the 09-11 study. The classifier from the 09-11 study was able to predict acquisition in the independent Ad26/Ad26+gp140 validation dataset with an AUC=0.75 (Fig. 2B). The GES was not able to predict protection in the Ad26/MVA+gp140 arm of the 13-19 study (AUC=0.44), which had previously demonstrated 42% efficacy, lower than the Ad26/Ad26+gp140 arm (fig. S2) (4). We validated the association of higher GES with reduced acquisition by quantitative polymerase chain reaction (qPCR) in the 09-11 cohort (fig. S3). In addition to testing infection as a dichotomous outcome, a time-to-event analysis of the number of SHIV exposures before infection was performed combining animals from the two arms of the 13-19 study. This survival analysis showed that monkeys with a higher GES significantly associated with longer time to infection [hazard ratio (HR)=0.01, $P=0.005$]. This association however did not reach significance in the 09-11 study (HR=0.04, $P=0.06$) (Fig. 2C).

GES is an independent correlate of protection in NHPs

Both studies previously identified functional antibody-based correlates of protection (3, 4). We tested whether the B cell gene signature was enriched with increased magnitude of cellular and antibody immune responses in both studies (table S2). The immune correlates were measured at specific time points and analyzed as categorical variables. When independently testing each individual immune response for enrichment of the B cell gene signature we observed that it associated with higher antibody-dependent cellular phagocytosis (ADCP), percent tier 2 neutralization, and cellular responses to Pol as measured by enzyme-linked immunosorbent spot (ELISPOT) in the 09-11 study (NES 1.4, $P<0.001$) (table S3). Higher magnitudes of ELISPOT to gp140, ADCP, Fc antibody polyfunctionality, interferon- γ , immunoglobulin G (IgG) breadth, and antibody-dependent

neutrophil phagocytosis (ADNP) also showed significant enrichment of the B cell gene signature in the combined arms of the 13-19 study (NES 1.4, $P<0.001$) (table S3). ADCP was the only immune response that associated with the B cell gene signature in both NHP studies (NES=1.99-2.21, $P<0.001$) (Fig. 3, A and B). Given that the GES was a good predictor of SIV or SHIV acquisition, we tested its relative ranking by Cox regression with respect to other immune measures in the two NHP studies. We used an elastic net method to generate a variable importance in projection (VIP) plot for the immune responses along with the GES and ranked them within each study. In both studies the GES was present as one of the top-ranked correlates of protection even after accounting for other potential correlates (Fig. 3, C and D). These findings were further supported by a univariate analysis of the GES on time to infection in the 13-19 study, but did not reach statistical significance in the 09-11 study (13-19: HR=0.29, $P=0.005$; 09-11: HR=0.19, $P=0.06$) (tables S4 and S5). As this gene signature was seen in B cells, we asked whether the GES correlated with frequency of B cells as a fraction of peripheral blood lymphocytes in the animals within either of the studies, but observed no significant correlation.

Specific genes in the B cell signature are associated with protection against SIV/SHIV

We performed a network analysis to identify relationships between the 53 enriched genes in the B cell signature and their known role and function. We analyzed these relationships based on co-expression and genetic interactions annotated in curated network databases (19). The genes in the B cell signature either directly or indirectly interact genetically and affect functions such as metabolism, cell signaling, cell differentiation, migration, proliferation, inflammation, immune response, apoptosis and phagocytosis (Fig. 4A and data file S2). Univariate analyses of the association of the 53 genes with time to infection identified the most protective genes with HR 0.5 that associated with protection in both studies (Fig. 4B). Genes associating with protection in both NHP studies were *TNFSF13*, *BHLHE40*, *SDCBP*, and *OGFRL1*. To evaluate the effect of these four host genes on SIV/SHIV acquisition in a multivariate analysis, we used logistic regression with a least absolute shrinkage and selection operator (Lasso) penalty. The only genes that remained in the model and associated with reduced risk of infection in both the 09-11 and 13-19 studies were *OGFRL1* and *TNFSF13*. Expression of *TNFSF13* also correlated with the greatest number of immune responses that were specific to each study, including ADCP and Gag-specific ELISPOT in the 09-11 study and antibody-dependent complement deposition (ADCD), ADNP, IgG breadth and gp140-specific ELISPOT in the 13-19 study (fig. S4).

DISCUSSION

Several NHP challenge studies using different vaccine strategies have shown partial protection against SIV or SHIV acquisition (3, 4, 17, 18). To date there is only one human study that showed partial protective efficacy against HIV (1). Correlates of protection in the human RV144 study (1) included Env-specific binding antibodies, viral genetic signatures, and specific host alleles (20-23). Given the partial protection observed in NHP and humans, we asked whether there were patterns of vaccine-induced host gene expression that associated with protection across vaccine regimens and platforms. Although the regimens and time points after vaccination varied between the preclinical and clinical trials, we

hypothesized that the protection observed in multiple studies was due to a common gene signature and could be identified by unbiased whole-transcriptome-based analyses. We used an NGS-based transcriptome approach and identified enrichment of a gene set (16) that correlated with protection from infection in the two studies and other NHP studies. This signature was also protective in the human RV144 trial. That this gene set was first identified in the context of influenza vaccine immunogenicity (16), and has now been shown to associate with protective efficacy of HIV vaccination, suggests these genes may be broadly involved in responses to vaccination, although the particular changes that are protective likely differ between viruses. We hypothesize that the gene signature is vaccine-induced, as it was not observed in the placebo participants in the RV144 study (1). It will however be important to evaluate baseline expression of this gene signature before vaccination, as such variation has demonstrated detectable effects on influenza vaccine responses when using systems biology approaches (24).

As enriched genes in this B cell signature were predictive of protection against both SIV and SHIV infection, we defined a composite GES by averaging *Z*-scores of the expression of the 53 enriched genes for each animal in any given study. The GES was one of the top correlates of protection in the NHP studies and represents a new tool for assessing HIV vaccine candidates. This will allow testing the predictive capacity of this gene signature in other vaccine studies that are completed or ongoing (25) and may aid the empirical design of vaccines with increased efficacy. The signature we identified also furthers understanding of the immunological basis for protection conferred by the present vaccines studied. The signature correlated with functional Fc-mediated phagocytosis responses in NHP, supporting the importance of this mechanism which was also indicated in a recent study (26).

Our gene network analyses suggest that many of the enriched genes in the signature have direct or ancillary roles in both innate and adaptive immune responses and are functionally associated. Specific genes in the signature directly associating with decreased risk of acquisition in both studies were *TNFSF13*, *BHLHE40*, *SDCBP*, and *OGFRL1*. *TNFSF13* belongs to the tumor necrosis factor ligand superfamily, *BHLHE40* has been shown to suppress IL10, *SDCBP* encodes the syntenin protein whose overexpression inhibits HIV-1 production, and *OGFRL1* is a less-characterized paralog of the opioid growth factor (27-29). The only gene to have an independent protective effect and correlate with maximum number of immune responses in both the NHP studies was *TNFSF13*, which is also known as a proliferation-inducing ligand (*APRIL*). *TNFSF13* is a ligand for *TNFRSF17/BCMA* and *TNFSF13B/TACI*, both of which are involved in numerous B cell functions including proliferation, class switching, affinity maturation, and antibody production (30). *TNFRSF17* expression has previously been described to correlate with increased magnitude of vaccine-induced antibody titers against influenza and yellow fever (16, 31, 32). Together, these data suggest that *TNFSF13* signaling correlates with protective vaccine-induced responses against several viruses. *TNFSF13* is typically produced by innate immune cells, but can also be induced in activated B cells (28, 33, 34). It is not likely that *TNFSF13* alone is responsible for the protection observed with acquisition, as the association of the GES with decreased risk of SIV/SHIV acquisition was stronger than the single gene. Instead, the nature and function of this protein suggest that it may be a potential master regulator of the vaccine-induced protection, in concert with other genes enriched in these studies.

There are limitations to this study, including the small sample size of the 09-11 study, which we think contributed to some less significant associations than in the combined 13-19 study. Further, these analyses were limited to transcriptomics of peripheral blood cells. Other noteworthy limitations were the differences between the 09-11 and 13-19 studies including the antigen (SIV/HIV insert), challenge (SIVmac251/SHIV SF162), vaccine strategy (number/timing of immunization), adjuvant (AS01B/alum) and time points at which RNA-seq was performed (20 weeks/2 weeks after last vaccination). All these have the potential to be confounders in a cross-prediction analysis by decreasing the signal to background noise ratio. Despite these limitations, we were able to use the NHP RNA-seq data to identify an overlapping gene set in B cells that was associated with protection against SIV and SHIV challenge and that was then validated in a human trial of HIV-1 infection.

The B cell signature identified here also showed association with functional antibody-based responses that have previously been observed to be correlates of protection against SIV (3, 4). Prospective interrogation of this transcriptional signature would be of value in the two ongoing HIV vaccine efficacy studies, evaluating both ALVAC (HVTN 702) and Ad26-based (HVTN 705) HIV vaccine regimens. More broadly, our data demonstrate the potential to discover protective correlates using an approach that mines transcriptomic data in multiple preclinical and clinical trials. Single-gene testing methods might be too stringent for studies consisting of small sample sizes and hence using a gene set analysis might be more feasible for preclinical and clinical trials. In addition, screening gene sets previously observed to be associated with biological phenotypes for association with protection in our dataset avoids limitations of current mechanistic understanding and focuses on processes of known biological perturbations. The potential significance of this approach is shown by our identification of a gene signature in B cells that associates with protection against both SIV and HIV across different vaccine platforms.

MATERIAL AND METHODS

Study design:

For discovery analysis cryopreserved PBMCs from 21 rhesus monkeys were selected from the vaccine arms of two studies where partial efficacy was observed, namely, the Ad26/gp140 arm of the 09-11 SIV challenge study (n=10) (3) and the Ad26/gp140 arm of the 13-19 SHIV challenge study (n=11) (4). We measured whole transcriptome gene expression by RNA-seq at timepoints after vaccination but before challenge to identify genes or gene sets that associated with protection from infection in both arms. We performed validation studies using animals from the additional arms of the 13-19 study where efficacy was observed, including the Ad26/Ad26 + gp140 (n=12) and Ad26/MVA + gp140 (n=9) arms. Sample sizes were based on availability of PBMC samples.

Flow cytometry.

PBMCs were thawed and stained with a cocktail of antibodies including CD3 Alexa Fluor 700 (clone SP34-2, BD Pharmingen), CD4 Brilliant Violet 605 (clone OKT4, BioLegend), CD8 Brilliant Violet 785 (clone RPA-T8, BioLegend), CD14 Pacific Blue (clone M5E2, BD Pharmingen), CD16 phycoerythrin (PE)-Cy7 (clone 3G8, BD Pharmingen), CD20

fluorescein isothiocyanate (FITC) (clone 2H7, BD Pharmingen), NKG2A PE (clone Z199, Beckman Coulter), and LIVE/DEAD Aqua viability dye (Life Technologies) to differentiate lymphocyte populations. The stained cells were analyzed and sorted using a BD FACS Aria II, producing pure bulk populations of CD4+ T cells, CD8+ T cells, NK cells, and B cells. The gating strategy is shown in fig. S1.

RNA extraction and sequencing.

Total RNA was extracted from sorted lymphocytes using the Single Cell RNA Purification kit (Norgen Biotek Corp). NGS RNA-seq was performed using SMART-Seq technology as described previously (10-12). Briefly, 2.5 ng of sample RNA input was processed using the SMARTer Stranded Total RNA-Seq Pico Input Mammalian kit (Takara Bio Inc) per the manufacturer's instructions, with spike-ins of exogenous control RNA (ThermoFisher Scientific). External ERCC spike-in was added to each sample to a final dilution of 1:260,000. Final library PCR amplifications used 13 cycles. Quantity and quality of amplified cDNA libraries were determined using the Qubit double-stranded DNA HS kit (ThermoFisher Scientific) and a High Sensitivity DNA chip (Agilent). MiSeq quantitation was performed using a 50 cycle MiSeq Reagent Kit v2 and MiSeq instrument (Illumina). The final 2 nM sample dilutions were then adjusted for any deviations from expected read percentages. We performed NGS on pooled fractions with either the HiSeq 2500 (with TruSeq PE Cluster cBot v3 and 200-cycle TruSeq SBS v3 kits), or NextSeq 500 (with a 300-cycle NextSeq High Output v2 kit) instrument (all Illumina) per the manufacturer's instructions for paired-end output targeting 50 million reads per sample.

Sequence analysis.

Paired-end sequencing data was converted to fastq, and adapters were masked using *bcl2fastq* v2.17.1.14 software (Illumina). The raw fastq files were examined for quality using FastQC v0.11.5, and reads were trimmed using Trimmomatic v0.36 when the average quality score of a 15-base pair sliding window fell below Q30. Trimmed reads with minimum length of 50 base pairs were retained (35). A median of 20% raw reads was discarded when using these parameters. Paired-end reads passing filter were aligned to the NHP genome assembly (*Macaca mulatta*, NCBI version Mmul 8.0.1) using HiSat2 v2.1.0 (36). Gene expression quantification was performed using HTSeq-count v0.9.1 (37) and a trimmed mean of M values--based normalization of the read counts was carried out using limma and edgeR v3.18.1 packages (38, 39). Genes were filtered on the basis of expression within a cell subset if 50% of the samples within a comparison group had normalized read counts per million >1. On average, a total of 11146 genes were expressed per animal.

Microarray data.

Published normalized, log-transformed microarray data from two vaccine studies, DNA-SIV/ALVAC + gp120 (18) and ALVAC-SIV/gp120 (17), were downloaded from the Gene Expression Omnibus database (accession codes GSE108011 and GSE72624, respectively). For the DNA-SIV/ALVAC + gp120 study (n=12), we analyzed microarray data at 1 week after the second boost time point. For the ALVAC-SIV/gp120 study (n=27), we analyzed data at the time point 24 hours after the third immunization (before the last boost). Microarray data were also available for 172 vaccinated individuals from the sample set

described in the RV144 immune correlates study at time point 2 weeks after the last vaccination (20, 40). All microarray datasets used in this study were generated on the HumanHT-12 v4.0 expression BeadChip platform (Illumina). For the RV144 study, the raw expression data were normalized, log-transformed and background-subtracted as described (40). The samples in each study were grouped into outcomes after challenge or infection status after immunization (1, 17, 18).

Immune response data.

Functional immune response data for the 09-11 and 13-19 studies have been described previously (3, 4). Additional assays including IgG breadth and ADNP were analyzed. IgG binding activity was measured against 15 SIV antigens in the 09-11 study and 22 HIV antigens in the 13-19 study. IgG breadth was calculated by the sum of the measurements whose binding activity was above the median of the measurement across all animals in the trial. ADNP was assessed in the 13-19 study by the measurement of the uptake of antibody-opsonized, antigen-coated fluorescent beads by primary neutrophils. Biotinylated A244 gp120 was used to saturate the binding sites on 1- μ m fluorescent neutravidin beads (Invitrogen). Excess antigen was removed by washing the beads, which were then incubated with NHP antibody samples for 20 min at 37°C. Leukocytes were isolated from blood collected from HIV seronegative donors after ACK lysis of red blood cells. After opsonization, the freshly isolated leukocytes were added, and the cells were incubated for 1 hour at 37°C to allow phagocytosis. The cells were then stained for CD66b to identify neutrophils and then fixed, and the extent of phagocytosis was measured by flow cytometry on a Stratadigm S1000EXi flow cytometer. The data are reported as a phagocytic score as described previously (41).

Quantitative polymerase chain reaction.

Gene expression data were confirmed by qPCR. Because only small amounts of RNA were available, we used the Biomark platform (Fluidigm) for validation on a subset of genes based on assay availability and robustness. Briefly, cDNA was synthesized by reverse transcription (RT) from 2.5 ng of RNA extracted from sorted B cells using an RT preamplification reaction mix, composed of TaqMan Fast Virus 1-Step Master Mix (Applied Biosystems) and a pooled 96 target TaqMan assay mix (ThermoFisher Scientific). RT was performed in three steps: cDNA synthesis (50°C for 10 min), denaturation (95°C for 2 min), and preamplification (18 cycles of 95°C for 15 s and 60°C for 4 min). Duplicate technical replicates of cDNA samples were assayed as described previously (42), using the GE 96x96 standard v2 protocol. The cycle threshold (Ct) values were validated and exported using the Biomark gene expression real-time PCR analysis software (Fluidigm).

Statistical analyses.

Two-class differential gene expression was performed using the edgeR package (38), comparing uninfected to infected animals in the 09-11 and 13-19 discovery studies. Genes were considered significant based on threshold of absolute fold change >1.5, $P < 0.05$ and $q < 0.10$. To reduce the stringent limitations of single gene differential expression analyses due to multiple testing corrections, normalized RNA-seq transcription profiles from different cell subsets were analyzed using the GSEA method (13). Genes were analyzed against the

C2 canonical pathways, C7 immunologic signatures, and the blood transcription modules in the MSigDB (13, 15, 32). For each GSEA comparison we tested whether gene sets were significantly associated with protection against challenge. Signatures were considered significantly protective if genes were enriched in the uninfected versus infected animals using a threshold of NES 1.4 and $P < 0.001$. Signatures derived only from primates were further down-selected if they overlapped in the Ad26/gp140 arms of the 09-11 and 13-19 studies.

For further analysis, a total of 58 leading-edge genes (enriched genes) in the B cell signature from the 09-11 study were identified to associate with SIV acquisition by GSEA and were selected for downstream processing as described below. A composite GES of enriched genes in the 09-11 study was computed by averaging Z -scores of the normalized gene expression for each monkey. We generated prediction models for infection status using the composite GES from 53 of the 58 enriched genes from the 09-11 study. Five of the enriched genes (*ALDH3B1*, *CD302*, *RTN1*, *BST1* and *SLC27A3*) were not considered as they did not pass the edgeR filtering step in the 13-19 study. We also computed a GES for the same 53 genes in the different arms of the 13-19 study. Bootstrap optimism-corrected area under receiver operating characteristics (ROC) curves were used for internal validation of model discrimination. The performance of the classifier was assessed using AUC with 95% confidence interval (CI; an AUC of 0.5 indicates random discrimination and 1 indicates perfect discrimination capabilities). We then assessed the performance of this classifier on the Ad26/gp140 test and Ad26/Ad26 + gp140 validation datasets of the 13-19 study. We performed a similar prediction analysis using a GES of 51 enriched genes from the 09-11 study and tested in the Ad26/MVA + gp140 arm of the 13-19 study, as two of the enriched genes (*CREB5*, *RRAGD*) did not pass the filtering step. We analyzed protection, defined as number of challenges required for infection by Cox proportional hazard regression using the GES as a continuous variable for both the 09-11 and 13-19 studies. The GES from the two arms of the 13-19 study (Ad26/gp140 and Ad26/Ad26 + gp140) were combined for this analysis and all other analyses mentioned below. All uninfected animals were censored at the last challenge. The proportionality assumption for the Cox model was acceptable on the basis of the test of independence between the scaled Schoenfeld residuals and time.

We further evaluated the gene expression B cell signature and GES by comparing them to other immune responses generated from the 09-11 and 13-19 studies. Using GSEA, we tested the effect of the B cell gene signature on the magnitude of immune response as a categorical variable based on the median values. We also used a penalized multivariable model to assess what combination of GES and other immune response variables was jointly associated with infection risk. The embedded method with elastic net regularization, which combines both L1 and L2 penalties, was used to test the relative ranking of the GES with respect to all other 15 and 11 immune measures for the 09-11 and 13-19 studies, respectively, with time to infection as outcome (43). The optimal values of the elastic net penalty α and the tuning parameter λ were selected using five-fold cross validation. We ranked candidate features using the VIP score, which reflects the magnitude of the contribution of an individual feature to the generated model after accounting for other features. We generated VIP plots of all features of the 09-11 and 13-19 studies by averaging the variable importance score from 1000 repeated feature selection processes. Time to event

was analyzed by univariate Cox proportional hazard regression for all immune correlates along with the GES for both studies to test their independent effects on acquisition.

To query the 53 genes in the GES against published networks for possible interactions we used the GeneMANIA application (v3.4) on Cytoscape (v3.2.1) with default parameters (19, 44). Networks for these 53 queried genes were predicted on the basis of the most-weighted interaction scores using default categories of association data in GeneMANIA, namely correlations in expression and genetic interactions. The genes within these predicted networks were then manually separated into groups based on their function as ascertained from published literature. Univariate Cox proportional hazard regression was performed for all 53 enriched genes for both studies to test their effects on time to infection. Forest plots of HR with 95% CI were generated for all 53 genes from both studies to select the most significant variables in a univariate analysis. A multivariate analysis by the Lasso method was used to downselect the genes that were most associated with infection status (45). We selected the tuning parameter through a five-fold cross validation. The algorithm was run 1000 times, and genes that were repeatedly retained more than 80% of the time were selected. Correlations between immune responses and the 53 genes were investigated by Spearman correlation.

A two-sided P value of <0.05 was considered statistically significant for all statistical analyses described above. All descriptive and inferential statistical analyses were performed using R 3.4.1 GUI 1.70 build (7375) v3.0 and higher, and GraphPad Prism 7 statistical software packages (GraphPad Software).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

We thank K. May (MHRP) for performing RNA-seq for the NHP studies and N. Rahman (MHRP) for the ad hoc bioinformatics support. We acknowledge P. Dawson (EMMES Corporation) for support related to statistical analysis. The views expressed are those of the authors and should not be construed to represent the positions of the U.S. Army or the U.S. Department of Defense (DOD).

Funding: This work was supported by a cooperative agreement (W81XWH-07-2-0067) between the Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc. and the DOD. This research was funded in part by the U.S. National Institute of Allergy and Infectious Disease. D.H.B. received research funding from the NIH, DOD, Gates Foundation, amfAR, Gilead, Janssen, and Novavax.

REFERENCES AND NOTES

1. Rerks-Ngarm S, Pitisuttithum P, Nitayaphan S, Kaewkungwal J, Chiu J, Paris R, Prensri N, Namwat C, de Souza M, Adams E, Benenson M, Gurunathan S, Tartaglia J, McNeil JG, Francis DP, Stablein D, Birx DL, Chunsuttiwat S, Khamboonruang C, Thongcharoen P, Robb ML, Michael NL, Kunasol P, Kim JH, Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. The New England journal of medicine 361, 2209–2220 (2009). [PubMed: 19843557]
2. Medlock J, Pandey A, Parpia AS, Tang A, Skrip LA, Galvani AP, Effectiveness of UNAIDS targets and HIV vaccination across 127 countries. Proc Natl Acad Sci U S A 114, 4017–4022 (2017). [PubMed: 28320938]

3. Barouch DH, Alter G, Broge T, Linde C, Ackerman ME, Brown EP, Borducchi EN, Smith KM, Nkolola JP, Liu J, Shields J, Parenteau L, Whitney JB, Abbink P, Ng'ang'a DM, Seaman MS, Lavine CL, Perry JR, Li W, Colantonio AD, Lewis MG, Chen B, Wenschuh H, Reimer U, Piatak M, Lifson JD, Handley SA, Virgin HW, Koutsoukos M, Lorin C, Voss G, Weijtens M, Pau MG, Schuitemaker H, Protective efficacy of adenovirus/protein vaccines against SIV challenges in rhesus monkeys. *Science* 349, 320–324 (2015). [PubMed: 26138104]
4. Barouch DH, Tomaka FL, Wegmann F, Stieh DJ, Alter G, Robb ML, Michael NL, Peter L, Nkolola JP, Borducchi EN, Chandrashekar A, Jetton D, Stephenson KE, Li W, Korber B, Tomaras GD, Montefiori DC, Gray G, Frahm N, McElrath MJ, Baden L, Johnson J, Hutter J, Swann E, Karita E, Kibuuka H, Mpendo J, Garrett N, Mngadi K, Chinyenze K, Priddy F, Lazarus E, Laher F, Nitayapan S, Pitisuttithum P, Bart S, Campbell T, Feldman R, Lucksinger G, Borremans C, Callewaert K, Roten R, Sadoff J, Scheppler L, Weijtens M, Feddes-de Boer K, van Manen D, Vreugdenhil J, Zahn R, Lavreys L, Nijs S, Tolboom J, Hendriks J, Euler Z, Pau MG, Schuitemaker H, 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, 232–243 (2018). [PubMed: 30047376]
5. Barouch DH, Liu J, Li H, Maxfield LF, Abbink P, Lynch DM, Iampietro MJ, SanMiguel A, Seaman MS, Ferrari G, Forthal DN, Ourmanov I, Hirsch VM, Carville A, Mansfield KG, Stablein D, Pau MG, Schuitemaker H, Sadoff JC, Billings EA, Rao M, Robb ML, Kim JH, Marovich MA, Goudsmit J, Michael NL, Vaccine protection against acquisition of neutralization-resistant SIV challenges in rhesus monkeys. *Nature* 482, 89–93 (2012). [PubMed: 22217938]
6. Fennessey CM, Keele BF, Using nonhuman primates to model HIV transmission. *Curr Opin HIV AIDS* 8, 280–287 (2013). [PubMed: 23666391]
7. Goodwin S, McPherson JD, McCombie WR, Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* 17, 333–351 (2016). [PubMed: 27184599]
8. Byron SA, Van Keuren-Jensen KR, Engelthaler DM, Carpten JD, Craig DW, Translating RNA sequencing into clinical diagnostics: opportunities and challenges. *Nat Rev Genet* 17, 257–271 (2016). [PubMed: 26996076]
9. O'Connor D, Pollard AJ, Characterizing vaccine responses using host genomic and transcriptomic analysis. *Clin Infect Dis* 57, 860–869 (2013). [PubMed: 23728145]
10. Ramskold D, Luo S, Wang YC, Li R, Deng Q, Faridani OR, Daniels GA, Khrebtkova I, Loring JF, Laurent LC, Schroth GP, Sandberg R, Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat Biotechnol* 30, 777–782 (2012). [PubMed: 22820318]
11. Picelli S, Faridani OR, Bjorklund AK, Winberg G, Sagasser S, Sandberg R, Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* 9, 171–181 (2014). [PubMed: 24385147]
12. Yarzabek B, Zaitouna AJ, Olson E, Silva GN, Geng J, Geretz A, Thomas R, Krishnakumar S, Ramon DS, Raghavan M, Variations in HLA-B cell surface expression, half-life and extracellular antigen receptivity. *Elife* 7, (2018).
13. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102, 15545–15550 (2005). [PubMed: 16199517]
14. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstrale M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D, Groop LC, PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 34, 267–273 (2003). [PubMed: 12808457]
15. Godec J, Tan Y, Liberzon A, Tamayo P, Bhattacharya S, Butte AJ, Mesirov JP, Haining WN, Compendium of Immune Signatures Identifies Conserved and Species-Specific Biology in Response to Inflammation. *Immunity* 44, 194–206 (2016). [PubMed: 26795250]
16. Nakaya HI, Wrangmert J, Lee EK, Racioppi L, Marie-Kunze S, Haining WN, Means AR, Kasturi SP, Khan N, Li GM, McCausland M, Kanchan V, Kokko KE, Li S, Elbein R, Mehta AK, Aderem A, Subbarao K, Ahmed R, Pulendran B, Systems biology of vaccination for seasonal influenza in humans. *Nat Immunol* 12, 786–795 (2011). [PubMed: 21743478]

17. Vaccari M, Gordon SN, Fourati S, Schifanella L, Liyanage NP, Cameron M, Keele BF, Shen X, Tomaras GD, Billings E, Rao M, Chung AW, Dowell KG, Bailey-Kellogg C, Brown EP, Ackerman ME, Vargas-Inchaustegui DA, Whitney S, Doster MN, Binello N, Pegu P, Montefiori DC, Foulds K, Quinn DS, Donaldson M, Liang F, Lore K, Roederer M, Koup RA, McDermott A, Ma ZM, Miller CJ, Phan TB, Forthal DN, Blackburn M, Caccuri F, Bissa M, Ferrari G, Kalyanaraman V, Ferrari MG, Thompson D, Robert-Guroff M, Ratto-Kim S, Kim JH, Michael NL, Phogat S, Barnett SW, Tartaglia J, Venzon D, Stablein DM, Alter G, Sekaly RP, Franchini G, Adjuvant-dependent innate and adaptive immune signatures of risk of SIVmac251 acquisition. *Nat Med* 22, 762–770 (2016). [PubMed: 27239761]
18. Vaccari M, Fourati S, Gordon SN, Brown DR, Bissa M, Schifanella L, Silva de Castro I, Doster MN, Galli V, Omsland M, Fujikawa D, Gorini G, Liyanage NPM, Trinh HV, McKinnon KM, Foulds KE, Keele BF, Roederer M, Koup RA, Shen X, Tomaras GD, Wong MP, Munoz KJ, Gach JS, Forthal DN, Montefiori DC, Venzon DJ, Felber BK, Rosati M, Pavlakis GN, Rao M, Sekaly RP, Franchini G, HIV vaccine candidate activation of hypoxia and the inflammasome in CD14(+) monocytes is associated with a decreased risk of SIVmac251 acquisition. *Nat Med* 24, 847–856 (2018). [PubMed: 29785023]
19. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T, Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13, 2498–2504 (2003). [PubMed: 14597658]
20. Haynes BF, Gilbert PB, McElrath MJ, Zolla-Pazner S, Tomaras GD, Alam SM, Evans DT, Montefiori DC, Karnasuta C, Sutthent R, Liao HX, DeVico AL, Lewis GK, Williams C, Pinter A, Fong Y, Janes H, DeCamp A, Huang Y, Rao M, Billings E, Karasavvas N, Robb ML, Ngauy V, de Souza MS, Paris R, Ferrari G, Bailer RT, Soderberg KA, Andrews C, Berman PW, Frahm N, De Rosa SC, Alpert MD, Yates NL, Shen X, Koup RA, Pitisuttithum P, Kaewkungwal J, Nitayaphan S, Rerks-Ngarm S, Michael NL, Kim JH, Immune-correlates analysis of an HIV-1 vaccine efficacy trial. *The New England journal of medicine* 366, 1275–1286 (2012). [PubMed: 22475592]
21. Li SS, Gilbert PB, Tomaras GD, Kijak G, Ferrari G, Thomas R, Pyo CW, Zolla-Pazner S, Montefiori D, Liao HX, Nabel G, Pinter A, Evans DT, Gottardo R, Dai JY, Janes H, Morris D, Fong Y, Edlefsen PT, Li F, Frahm N, Alpert MD, Prentice H, Rerks-Ngarm S, Pitisuttithum P, Kaewkungwal J, Nitayaphan S, Robb ML, O'Connell RJ, Haynes BF, Michael NL, Kim JH, McElrath MJ, Geraghty DE, FCGR2C polymorphisms associate with HIV-1 vaccine protection in RV144 trial. *J Clin Invest* 124, 3879–3890 (2014). [PubMed: 25105367]
22. Rolland M, Edlefsen PT, Larsen BB, Tovanabutra S, Sanders-Buell E, Hertz T, deCamp AC, Carrico C, Menis S, Magaret CA, Ahmed H, Juraska M, Chen L, Konopa P, Nariya S, Stoddard JN, Wong K, Zhao H, Deng W, Maust BS, Bose M, Howell S, Bates A, Lazzaro M, O'Sullivan A, Lei E, Bradfield A, Ibitamuno G, Assawadarachai V, O'Connell RJ, deSouza MS, Nitayaphan S, Rerks-Ngarm S, Robb ML, McLellan JS, Georgiev I, Kwong PD, Carlson JM, Michael NL, Schief WR, Gilbert PB, Mullins JI, Kim JH, Increased HIV-1 vaccine efficacy against viruses with genetic signatures in Env V2. *Nature* 490, 417–420 (2012). [PubMed: 22960785]
23. Prentice HA, Tomaras GD, Geraghty DE, Apps R, Fong Y, Ehrenberg PK, Rolland M, Kijak GH, Krebs SJ, Nelson W, DeCamp A, Shen X, Yates NL, Zolla-Pazner S, Nitayaphan S, Rerks-Ngarm S, Kaewkungwal J, Pitisuttithum P, Ferrari G, McElrath MJ, Montefiori DC, Bailer RT, Koup RA, O'Connell RJ, Robb ML, Michael NL, Gilbert PB, Kim JH, Thomas R, HLA class II genes modulate vaccine-induced antibody responses to affect HIV-1 acquisition. *Science translational medicine* 7, 296ra112 (2015).
24. Tsang JS, Schwartzberg PL, Kotliarov Y, Biancotto A, Xie Z, Germain RN, Wang E, Olnes MJ, Narayanan M, Golding H, Moir S, Dickler HB, Perl S, Cheung F, Baylor HC, Consortium CHI, Global analyses of human immune variation reveal baseline predictors of postvaccination responses. *Cell* 157, 499–513 (2014). [PubMed: 24725414]
25. Hsu DC, O'Connell RJ, Progress in HIV vaccine development. *Hum Vaccin Immunother* 13, 1018–1030 (2017). [PubMed: 28281871]
26. Ackerman ME, Das J, Pittala S, Broge T, Linde C, Suscovich TJ, Brown EP, Bradley T, Natarajan H, Lin S, Sassic JK, O'Keefe S, Mehta N, Goodman D, Sips M, Weiner JA, Tomaras GD, Haynes BF, Lauffenburger DA, Bailey-Kellogg C, Roederer M, Alter G, Route of immunization defines multiple mechanisms of vaccine-mediated protection against SIV. *Nat Med* 24, 1590–1598 (2018). [PubMed: 30177821]

27. Yu F, Sharma S, Jankovic D, Gurram RK, Su P, Hu G, Li R, Rieder S, Zhao K, Sun B, Zhu J, The transcription factor Bhlhe40 is a switch of inflammatory versus antiinflammatory Th1 cell fate determination. *J Exp Med* 215, 1813–1821 (2018). [PubMed: 29773643]
28. Mackay F, Schneider P, Rennert P, Browning J, BAFF AND APRIL: a tutorial on B cell survival. *Annu Rev Immunol* 21, 231–264 (2003). [PubMed: 12427767]
29. Gordon-Alonso M, Rocha-Perugini V, Alvarez S, Moreno-Gonzalo O, Ursa A, Lopez-Martin S, Izquierdo-Useros N, Martinez-Picado J, Munoz-Fernandez MA, Yanez-Mo M, Sanchez-Madrid F, The PDZ-adaptor protein syntenin-1 regulates HIV-1 entry. *Mol Biol Cell* 23, 2253–2263 (2012). [PubMed: 22535526]
30. Mackay F, Schneider P, Cracking the BAFF code. *Nat Rev Immunol* 9, 491–502 (2009). [PubMed: 19521398]
31. Querec TD, Akondy RS, Lee EK, Cao W, Nakaya HI, Teuwen D, Pirani A, Gernert K, Deng J, Marzolf B, Kennedy K, Wu H, Bennouna S, Oluoch H, Miller J, Vencio RZ, Mulligan M, Aderem A, Ahmed R, Pulendran B, Systems biology approach predicts immunogenicity of the yellow fever vaccine in humans. *Nat Immunol* 10, 116–125 (2009). [PubMed: 19029902]
32. Li S, Roupheal N, Duraisingham S, Romero-Steiner S, Presnell S, Davis C, Schmidt DS, Johnson SE, Milton A, Rajam G, Kasturi S, Carlone GM, Quinn C, Chaussabel D, Palucka AK, Mulligan MJ, Ahmed R, Stephens DS, Nakaya HI, Pulendran B, Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. *Nat Immunol* 15, 195–204 (2014). [PubMed: 24336226]
33. Chu VT, Enghard P, Riemekasten G, Berek C, In vitro and in vivo activation induces BAFF and APRIL expression in B cells. *J Immunol* 179, 5947–5957 (2007). [PubMed: 17947668]
34. Cotugno N, De Armas L, Pallikkuth S, Rinaldi S, Issac B, Cagigi A, Rossi P, Palma P, Pahwa S, Perturbation of B Cell Gene Expression Persists in HIV-Infected Children Despite Effective Antiretroviral Therapy and Predicts H1N1 Response. *Front Immunol* 8, 1083 (2017). [PubMed: 28955330]
35. Bolger AM, Lohse M, Usadel B, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120 (2014). [PubMed: 24695404]
36. Kim D, Langmead B, Salzberg SL, HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 12, 357–360 (2015). [PubMed: 25751142]
37. Anders S, Pyl PT, Huber W, HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169 (2015). [PubMed: 25260700]
38. Robinson MD, McCarthy DJ, Smyth GK, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140 (2010). [PubMed: 19910308]
39. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43, e47 (2015). [PubMed: 25605792]
40. Fourati S, Ribeiro SP, Pereira Lopes F, Blasco Tavares, Talla A, Lefebvre F, Cameron M, Kaewkungwal J, Pitisuttithum P, Nitayaphan S, Rerks-Ngarm S, Kim JH, Thomas R, Gilbert PB, Tomaras GD, Koup RA, Michael NL, McElrath MJ, Gottardo R, Sekaly RP, Integrated systems approach defines the antiviral pathways conferring protection by the RV144 HIV vaccine. *Nat Commun* 10, 863 (2019). [PubMed: 30787294]
41. Darrah PA, Patel DT, De Luca PM, Lindsay RW, Davey DF, Flynn BJ, Hoff ST, Andersen P, Reed SG, Morris SL, Roederer M, Seder RA, Multifunctional TH1 cells define a correlate of vaccine-mediated protection against *Leishmania major*. *Nature medicine* 13, 843–850 (2007).
42. Johnson S, Eller M, Teigler JE, Malveste SM, Schultz BT, Soghoian DZ, Lu R, Oster AF, Chenine AL, Alter G, Dittmer U, Marovich M, Robb ML, Michael NL, Bolton D, Streeck H, Cooperativity of HIV-Specific Cytolytic CD4 T Cells and CD8 T Cells in Control of HIV Viremia. *J Virol* 89, 7494–7505 (2015). [PubMed: 25972560]
43. Zou H, Hastie T, Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67, 301–320 (2005).
44. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT, Maitland A, Mostafavi S, Montojo J, Shao Q, Wright G, Bader GD, Morris Q,

- The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res* 38, W214–220 (2010). [PubMed: 20576703]
45. Tibshirani R, Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 58, 267–288 (1996).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 1. A gene expression signature seen in B cells is associated with protection in multiple SIV/HIV vaccine studies.

(A) The same B cell-based gene signature was significantly enriched in the Ad26/gp140 arms of the 09-11 (n=10) and 13-19 (n=11) NHP studies in the uninfected compared with infected monkeys, (B) in the human RV144 vaccine study (n=172), (C and D) and in two independent RV144-like pox-protein vaccine regimens in NHP (n=39).

Fig. 2. A subset of genes from the B cell gene signature predicts protection and is associated with reduced SIV/SHIV infection.

(A) Fifty-three enriched genes from the 09-11 study B cell signature were used to generate a composite GES consisting of the average of standardized normalized gene expression. The receiver operator characteristic (ROC) curve illustrates the discriminating ability of the GES of enriched genes from the Ad26/gp140 arms of the 09-11 training dataset (AUC=0.84; 95% CI: 0.52 to 1) and the 13-19 test dataset (AUC=0.86; 95% CI: 0.56 to 1) to predict protection from acquisition of SIV (n=10) and SHIV (n=11), respectively. (B) The model from the 09-11 study was also able to predict protection from acquisition of SHIV in the Ad26/Ad26 + gp140 validation arm of the 13-19 study (n=12) by logistic regression (AUC=0.75; 95% CI: 0.41 to 1). (C) Effect of the GES as a continuous variable by Cox proportional hazards ratio (HR) is shown for both NHP studies.

Fig. 3. The B cell gene signature and GES associated with immune correlates of protection after vaccination.

(A and B) ADCP is the only functional immune response that associates with the protective enriched B cell gene signature in both the 09-11 (n=10) and 13-19 (n=23) studies. (C and D) Relative ranking of immune responses (blue) and GES (red) based on time to infection in the 09-11 and 13-19 studies were generated by averaging the VIP scores across 1000 repeated feature selection processes. Standard error bars of the average of the VIP scores are shown for each feature. *week 96; #week 64.

Fig. 4. Functional relationships of the 53 enriched genes from the B cell gene signature. (A) Shown is a gene interaction network derived from the protective B cell gene signature. The network edges represent co-expression (purple) and genetic interactions (green). Red nodes represent the 53 enriched genes, and cyan squares indicate interacting network genes. (B) Effect of expression of the 53 genes individually on time to infection as a continuous variable in the 09-11 (n=10) and 13-19 (n=23) studies. The most protective genes with HR < 0.5 are shown in red, and of these the overlapping genes in the 09-11 and 13-19 studies are boxed. *, P<0.05.