COMPASS identifies T-cell subsets correlated with clinical outcomes

Lin Lin¹, Greg Finak¹, Kevin Ushey¹, Chetan Seshadri², Thomas R Hawn², Nicole Frahm¹, Thomas J Scriba³, Hassan Mahomed³, Willem Hanekom³, Pierre-Alexandre Bart⁴, Giuseppe Pantaleo⁴, Georgia D Tomaras⁵, Supachai Rerks-Ngarm⁶, Jaranit Kaewkungwal⁷, Sorachai Nitayaphan⁸, Punnee Pitisuttithum⁹, Nelson L Michael¹⁰, Jerome H Kim¹⁰, Merlin L Robb¹¹, Robert J O'Connell¹², Nicos Karasavvas¹², Peter Gilbert¹, Stephen C De Rosa^{1,13}, M Juliana McElrath^{1,2,13} & Raphael Gottardo¹

Advances in flow cytometry and other single-cell technologies have enabled high-dimensional, high-throughput measurements of individual cells as well as the interrogation of cell population heterogeneity. However, in many instances, computational tools to analyze the wealth of data generated by these technologies are lacking. Here, we present a computational framework for unbiased combinatorial polyfunctionality analysis of antigen-specific T-cell subsets (COMPASS). COMPASS uses a Bayesian hierarchical framework to model all observed cell subsets and select those most likely to have antigen-specific responses. Cell-subset responses are quantified by posterior probabilities, and human subject-level responses are quantified by two summary statistics that describe the quality of an individual's polyfunctional response and can be correlated directly with clinical outcome. Using three clinical data sets of cytokine production, we demonstrate how COMPASS improves characterization of antigen-specific T cells and reveals cellular 'correlates of protection/immunity' in the RV144 HIV vaccine efficacy trial that are missed by other methods. COMPASS is available as open-source software.

Recent technological advances in flow and mass cytometry assays have transformed the field of immunology by enabling large numbers of parameters to be quantified at the single-cell level in a high-throughput fashion. Many studies and clinical trials now rely on these assays to provide multiparameter, single-cell measurements, which enable a more comprehensive understanding of cellular function than methods that measure one parameter at a time. In particular, single-cell analyses by intracellular cytokine staining (ICS)—a type of flow cytometry assay (Fig. 1)—have become important tools to characterize subsets of antigen-specific T cells capable of simultaneously producing multiple effector cytokines and other functional markers, known as polyfunctional T cells¹. Polyfunctional T cells have an important role in protective immunity and nonprogression of diseases, and their presence correlates with favorable clinical outcomes^{2–4}. Vaccination in humans can generate broad T-cell cytokine responses^{5,6}, making polyfunctional T-cell subsets attractive as potential biomarkers of protection against disease. However, statistical tools for analyzing the complexity of these immune responses are lacking.

Although many analytic tools exist for cytometry-based assays^{7–9}, very few tools have been developed specifically for high-dimensional ICS data analysis. Existing strategies are basic and low dimensional, ranging from *ad hoc* rules based on fold-changes¹⁰, Hotelling's T^2 statistics¹¹ and 2×2 contingency tables^{12,13} to simple graphical displays of summary statistics⁷. In most ICS assays, the frequencies (and thus cell counts) of antigen-specific subsets are very small (<0.1% of total T cells), making robust statistical analysis difficult. This difficulty increases as the cells are further partitioned based on co-expression of multiple cytokines. In addition, multiple comparisons across subjects (persons in the clinical study) and cell subsets must be adjusted for and can further reduce statistical power. Given the lack of analytical tools, studies have been limited to formal statistical comparisons of only a few functions (secreted cytokines), and most of the work on polyfunctionality has been phenomenological.

To address some of these limitations, Finak *et al.*¹⁴ proposed a framework called mixture models for single-cell assays (MIMOSA). MIMOSA is based on mixtures of beta-binomials and rigorously analyzes count data derived from ICS assays but was mainly developed for univariate analysis of cell subsets, such as cells expressing a single function or a specified combination of functions.

¹Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. ²Division of Allergy and Infectious Diseases, University of Washington, Seattle, Washington, USA. ³South African Tuberculosis Vaccine Initiative, Institute of Infectious Disease and Molecular Medicine and School of Child and Adolescent Health, University of Cape Town, Cape Town, South Africa. ⁴Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland. ⁵Duke Human Vaccine Institute, Duke University Medical Center, Durham, North Carolina, USA. ⁶Department of Disease Control, Ministry of Public Health, Nonthaburi, Thailand. ⁷Data Management Unit, Faculty of Tropical Medicine, Mahidol University, Ratchathewi, Bangkok, Thailand. ⁸Thai Component, Armed Forces Research Institute of Medical Sciences, Bangkok, Ratchathewi, Bangkok, Thailand. ⁹Vaccine Trials Center, Faculty of Tropical Medicine, Mahidol University, Ratchathewi, Bangkok, Thailand. ¹⁰US Military HIV Research Program, Walter Reed Army Institute of Research, Silver Spring, Maryland, USA. ¹¹US Army Military HIV Research Program, Walter Reed Army Institute of Research; Henry M. Jackson Foundation, Bethesda, Maryland, USA. ¹²US Army Medical Component, Armed Forces Research Institute of Medical Sciences, Ratchathewi, Bangkok, Thailand. ¹³Department of Laboratory Medicine, University of Washington, Seattle, Washington, USA. Correspondence should be addressed to R.G. (rgottard@fredhutch.org).

Received 17 April 2014; accepted 4 March 2015; published online 25 May 2015; doi:10.1038/nbt.3187

Although MIMOSA includes a multivariate version, Finak *et al.*¹⁴ were solely interested in identifying positive immune responses irrespective of the qualitative aspect of the response, and as a result, the output is still univariate (probability of response). Furthermore, to apply MIMOSA to multivariate data, the authors assumed that a measurable antigen-specific response exists across all functional subsets. However, this assumption is incorrect in practice. Different antigens usually induce very different functional profiles, and many of the possible functional cell subsets are not expected to be associated with antigen specificity. MIMOSA cannot jointly model all subsets to identify distinct antigen-specific responses, and this is particularly important to do as the number of definable subsets grows exponentially with the number of cytokines analyzed. As an example, an ICS experiment measuring seven functions can define 128 Boolean cell subsets, but only a fraction of those are expected to signify responsiveness to a specific antigen. A possible solution would be to model and test each subset separately, but this is computationally intensive, ignores the dependence between subsets and leads to multiple testing problems. In the interest of decreasing the number of variables while taking into account the degree of functionality, Larsen *et al.*¹⁵ introduced a polyfunctionality index (PI) that summarizes the polyfunctional profile into a single number. However, the PI uses empirical proportions, which are known to be extremely noisy when cell counts are small, and combines information from all cell subsets, including nonspecific ones, thus masking real signal. Indeed, many functional subsets that are not antigen-specific can have substantial background, that is, they can be present in substantial numbers in nonstimulated cells, and their inclusion makes the defined index more variable. In addition, by weighting the subsets by their magnitude, low-magnitude (but potentially important) subsets will be down-weighted significantly. Thus, the PI falls short of describing the true breadth of polyfunctionality, limiting its clinical utility. An ideal framework for the analysis of T cells should identify and quantify changes in possibly rare, antigen-specific cell subsets and permit the definition of different qualities of a polyfunctional response, such as through summary statistics that can be correlated with outcomes of interest.

To address these needs, we have developed COMPASS, which uses a Bayesian hierarchical mixture model to identify antigen-specific

changes across all observable T-cell subsets simultaneously, enabling the definition of subject- and cohort-level antigen-specific T-cell profiles that can be summarized and correlated with outcomes of interest. This method jointly models all cell subsets, regularizes small cell counts and allows information sharing across all subjects through the calculation of subset-specific posterior probabilities that can be used to automatically detect and quantify antigen-specific subsets. COMPASS uses a computationally efficient Markov Chain Monte Carlo algorithm to explore the space of all possible functional T-cell subsets and compute the subset-specific posterior probabilities. These probabilities naturally account for multiple testing¹⁶ and permit the derivation of a false-discovery-rate estimate through a direct posterior probability approach¹⁷. COMPASS provides two scores—functionality and polyfunctionality—that can be correlated with any clinical outcome of interest. These scores summarize, as a single number for each subject, the posterior probabilities of antigen-specific response across cell subsets. We apply this method to three ICS data sets—the RV144 HIV vaccine case-control study^{18,19}, a cross-sectional study of South African adolescents infected with *Mycobacterium tuberculosis* (MTB)²⁰ and a phase 1b trial from the HIV Vaccine Trials Network²¹—and demonstrate the ability of COMPASS to identify specific polyfunctional responses that are correlated with clinical outcomes of interest.

RESULTS

COMPASS enables unbiased quantification of all cell subsets

We reanalyzed the ICS data generated through the RV144 HIV vaccine case-control study with 92TH023-Env peptide pool *ex vivo* stimulation, as described in Online Methods and in a prior publication¹⁸. Expression of a set of six functions (interferon (IFN)- γ , TNF- α , interleukin (IL)-2, IL-4, IL-17 and CD40L) was measured in CD4⁺ T cells by ICS ($n = 226$ vaccine and $n = 36$ placebo recipients). We used COMPASS to define 64 functional cell subsets from the Boolean combinations of the individual cytokines expressed in each cell. Only 24 of these had non-negligible (over five cells in more than two subjects) cell counts and were considered for analysis. Arguably, if there are only two subjects with no more than five cells in any cell subset then it would be difficult to detect antigen-specific responses, and we feel

Figure 1 Overview of an ICS experiment. Blood samples are drawn from I subjects. A sample is split into aliquots that are subject to stimulation with antigen or are left unstimulated as negative controls. After stimulation, whole peripheral blood mononuclear cells are labeled with fluorophore-conjugated antibodies against phenotypic (e.g., CD4, CD3, CD8, AVID/Live/Dead) and functional (e.g., IFN- γ , IL-2, TNF- α , CD40L, IL-17, IL-4) markers (cells are permeabilized and labeled intracellularly for functional markers). The single-cell expression of each marker on each labeled cell is measured by flow cytometry. After acquisition, data are processed and distinct cell populations of interest identified by gating, which classifies each marker as either 'positive' (expressed) or 'negative' (not expressed). A COMPASS analysis assumes the 'gating' is given, and the COMPASS tool summarizes the number of cells expressing different combinatorial functional markers for specific phenotypic cell subsets (e.g., the number of CD4⁺ T cells expressing different combinations of cytokines). For M markers, this produces an I by 2^M matrix of counts. COMPASS simultaneously models the counts for the paired stimulated and unstimulated samples for each subject across all combinatorial functional cell subsets.

Figure 2 Polyfunctionality analysis in RV144.

(a) Heatmap of COMPASS posterior probabilities for the RV144 data set. Columns correspond to the different cell subsets modeled by COMPASS (shown are the 15 of 24 subsets with detectable antigen-specific responses that had over five cells in more than two subjects), color-coded by the cytokines they express (white = “off”, shaded = “on”, grouped by color = “degree of functionality”), and ordered by degree of functionality from one function on the left to five functions on the right (blue to pink).

Subsets with maximum posterior probabilities <0.005 were removed from the heatmap. Rows correspond to subjects (shown are only 226 vaccine recipients), which are ordered by their status: noninfected (top) and infected (bottom), and by FS within each group.

Each cell of the heatmap shows the probability that the corresponding cell-subset (column) exhibits an antigen-specific response in the corresponding subject (row), where the probability is color-coded from white (zero) to purple (one). (b) Box plots of functionality and PFSs stratified by HIV infection status in RV144 among 226 vaccine recipients. Noninfected individuals have higher scores than infected ones (Wilcoxon test $P = 0.03$ (FS), $P = 0.01$ (PFS)). Both scores are inversely correlated with infection (Table 1).

that these values should be reasonable for most ICS data sets. In addition, changing these cutoffs to be less stringent had no impact on the significance of the results or conclusions presented here and the values were selected purely for computational convenience to reduce the time it takes to fit a COMPASS model to cell subsets that contain no signal. A heatmap of posterior probabilities of antigen-specific responses from COMPASS for these T-cell subsets shows that, whereas some antigen-specific subsets were universally present in almost all individuals vaccinated (e.g., IL-2, CD40L), others exhibited heterogeneity across subjects (e.g., the four- and five-function subsets) (Fig. 2a). These findings support the need for unbiased polyfunctionality analyses that can provide insights on the possible association between polyfunctional antigen-specific T cells and vaccine efficacy.

ICS assays are often used to identify vaccine responders based on the magnitude of antigen-specific T-cell responses, although current approaches tend to be univariate. COMPASS can identify vaccine responders by computing response probabilities for each subject. We compared COMPASS with the two standard approaches for designating positivity—Fisher’s exact test¹² and MIMOSA¹⁴—using the primary endpoint in the original RV144 analysis¹⁸ as well as the multivariate MIMOSA and the PI. A receiver operating characteristic (ROC) analysis shows that COMPASS substantially increased sensitivity and specificity when discriminating between vaccine and placebo recipients compared to all other approaches (Supplementary Fig. 1).

In addition to subject-level response probabilities, COMPASS defines two scores that summarize a subject’s entire antigen-specific

polyfunctional profile into a single numerical value (Fig. 2b). The functionality score (FS) is defined as the proportion of antigen-specific subsets detected among all possible ones. The polyfunctionality score (PFS) is similar, but it weighs the different subsets by their degree of functionality, naturally favoring subsets with higher degrees of functionality, motivated by the observations that a higher-degree function has been correlated with good outcomes in certain vaccine studies^{5,6}. The FS was well correlated with the number of functions, when measured by multiplex bead array on the same set of samples from the RV144 case-control study (Fig. 3a; correlation $\rho = 0.69$, $P < 2.2 \times 10^{-16}$, 95% confidence interval (CI) = (0.62, 0.75)), suggesting that COMPASS detects true antigen-specific responses in T-cell subsets. Out of the 12 cytokines measured by a multiplex bead array, only four were also measured by ICS (IFN- γ , TNF- α , IL-2, IL-4). Subjects with a high FS also expressed many other cytokines not measured in the ICS assay. We also tried to restrict the analysis to the four most common cytokines, but it did not improve the correlation; in fact, the correlation decreased slightly to 0.65. The FS was also significantly correlated with HIV Env-specific antibody binding in plasma (Fig. 3b; correlation $\rho = 0.41$, $P = 9.87 \times 10^{-11}$, 95% CI = (0.30, 0.52)). The placebo recipients had a much smaller FS and PFS compared to those of vaccine recipients, and the PI was noisier than the FS and PFS (Supplementary Fig. 2). All together, these data indicate that our FS and PFS were good at capturing and summarizing a subject’s response to vaccination. These results were also confirmed with the RV144 pilot data, used to inform the selection of assays to be used for the case-control study, thus providing validation from an independent set of subjects (Supplementary Figs. 3–6). The FS and PFS for RV144, broken down by infection status (Fig. 2b), varied greatly by subject, and, on average, noninfected subjects had higher scores than infected individuals; these effects were stronger for PFS (Wilcoxon test $P = 0.03$ (FS), $P = 0.01$ (PFS)).

Figure 3 Functionality scores and multiplex bead array in RV144.

(a) Proportion of expressed cytokines as measured by multiplex bead array. (b) FS versus IgG antibody binding in RV144. A set of 12 cytokines was measured by multiplex bead array. The proportion of detectable secreted cytokines was calculated in each individual and compared to the FS. The FS is significantly correlated with overall cytokine production ($\rho = 0.68$, $P < 2.2 \times 10^{-16}$) and antibody binding ($\rho = 0.50$, $P = 3.02 \times 10^{-10}$). The fitted regression line from simple linear model is plotted in blue, and the 95% confidence interval is shown in gray.

Table 1 Estimated odds ratios for HIV-1 infection risk for the subset specific responses as determined by logistic regression models that adjust for baseline risk category and gender in the RV144 case-control study

Variable	Odds ratio (95% CI)	P-value	q-value
FS	0.62 (0.42, 0.91)	0.014	0.06
PFS	0.57 (0.38, 0.84)	0.005	0.05
IL-4 ⁺ IL-2 ⁺ CD40L ⁺	0.62 (0.43, 0.90)	0.013	0.06
TNF- α ⁺ IFN- γ ⁺ IL-4 ⁺ IL-2 ⁺ CD40L ⁺	0.58 (0.39, 0.86)	0.006	0.05

Only subsets with $P < 0.05$ are shown here. Odds ratio are per one s.d. for each variable and are adjusted for IgA level, gender and baseline behavioral risk score. Lower and upper limits of the 95% confidence intervals (CI) for the estimated ratios are also shown. Q-values are the FDR-adjusted P -values across all 17 considered variables.

COMPASS identifies correlates of infection risk in RV144

Among the 17 different types of immune assays and their 152 component variables used to assess correlates of infection risk in RV144, 6 assays (including ICS CD4⁺ T-cell responses) were chosen as primary variables for having optimal statistical power when adjusting for multiple comparisons¹⁸. The major findings in a multivariate model including all six primary variables were that gp70-V1V2-specific plasma IgG inversely correlated with infection rate (odds ratio 0.57, $P = 0.02$, $q = 0.08$), and Env-specific plasma IgA was a direct correlate of infection (odds ratio 1.54, $P = 0.03$, $q = 0.08$). No CD4⁺ T-cell cellular correlates were identified in the primary analysis in the original study¹⁸.

In contrast, using the same method as in the original correlate analysis¹⁸ but including the FS and PFS rather than the primary CD4⁺ T-cell endpoint variable, the PFS was inversely correlated with infection ($P = 0.005$, $q = 0.05$, **Table 1**), as was the FS but to a lesser extent ($P = 0.01$, $q = 0.06$, **Table 1**). This correlation was also supported by the fact that the number of functions (secreted cytokines) as measured by multiplex bead array was also inversely correlated with infection ($P = 0.013$, **Supplementary Table 1**). However, the multiplex bead array data do not permit polyfunctionality analysis as it is an assay performed on bulk cells, and as such, it cannot measure the co-expression of multiple cytokines at the single-cell level. Our PFS suggests that polyfunctional antigen-specific CD4⁺ T cells might have played a role in vaccine efficacy.

To identify the specific cell subsets that contributed most to this correlation, we evaluated each subset of interest as a correlate for

infection risk (**Fig. 2a**). Two antigen-specific T-cell subsets were significantly correlated with infection ($q \leq 0.1$, **Table 1**). It can be seen that the polyfunctional subset expressing CD40L, IL-2, IL-4, IFN- γ and TNF- α shows the most significance (odds ratio (OR) = 0.58, $P = 0.006$, $q = 0.05$) followed by a three-function subset (CD40L, IL-2 and IL-4; OR = 0.62, $P = 0.01$, $q = 0.06$). The five-function subset was as significant as the previously reported gp70-V1V2 correlate. Notably, these subsets were identified in an unbiased way rather than by limiting our analysis to very specific subsets based on expected biological relevance as was done in the original study¹⁸. Both subsets included IL-4 and CD40L, a cytokine and functional marker, respectively, both important for the CD4⁺ T cell–B cell interaction. Thus, these particular CD4⁺ T-cells subsets may contribute the T-cell help necessary for the antibody production detected as a correlate in the primary analysis.

All subsets, including nonsignificant ones, are presented in **Supplementary Table 2**. In addition, we have also assessed the magnitude of response for all functional subsets (**Supplementary Table 3**) as well as the PI and MIMOSA scores (**Supplementary Table 1**) as individual correlates of risk, and none of these variables were significant. The superiority of COMPASS was also supported by an ROC analysis, using the predicted infection probabilities from a logistic regression model (**Supplementary Fig. 7**).

MTB infection induces highly polyfunctional responses

We used COMPASS to investigate polyfunctionality in CD4⁺ T cells obtained from a South African study of 18 MTB-uninfected and 22 MTB-infected subjects. Subjects were classified as MTB-infected, as described in the Online Methods section and in the original study²⁰. T cells were stimulated with MTB-specific²² and MTB-nonspecific peptide pools, and a set of seven cytokines and functional markers was measured by ICS, with 22 Boolean combinations defined for COMPASS analysis.

As expected, CD4⁺ T-cell responses to MTB-specific proteins were largely absent from MTB-uninfected subjects (**Fig. 4**). *Ex vivo* stimulation with MTB-specific antigens induced strong antigen-specific responses spread across polyfunctional subsets of degree three and five in MTB-infected subjects. In contrast, MTB-uninfected subjects had weak responses in these subsets. The appearance of the five-function subset (IL-17 α , CD40L, IFN- γ , TNF- α , IL-2) in some MTB-uninfected

Figure 4 Polyfunctionality analysis of MTB response. (a) Heatmap of COMPASS posterior probabilities for the TB data set. Columns correspond to the different cell subsets modeled by COMPASS (shown are the six of 22 and 19 subsets with detectable antigen-specific response that had over five cells in more than two subjects), color-coded by the cytokines they express (white = “off”, shaded = “on”, grouped by color = “degree of functionality”), and ordered by degree of functionality from one function on the left to five functions on the right (blue to pink). Subsets with maximum posterior probabilities <0.1 were removed from the heatmap. Rows correspond to 40 subjects, which are ordered by level of IFN- γ release as measured by QFT. Subjects with positive QFT test results are labeled as TB-positive. Each cell shows the probability that the corresponding cell-subset (column) exhibits an antigen-specific response in the corresponding subject (row), where the probability is color-coded from white (zero) to blue (one) for MTB-specific stimulation and red (one) for MTB-nonspecific. Yellow indicates a response to both stimulations. (b) Box plots of FS and PFS stratified by TB positive and TB negative for both TB-specific and non-TB-specific stimulations.

Figure 5 Polyfunctionality analysis of HVTN078 vaccine response. **(a)** Heatmap of COMPASS posterior probabilities for the HVTN078 data. Columns correspond to the different cell subsets modeled by COMPASS (shown are the 22 of 26 subsets with detectable antigen-specific response and having over five cells in more than two subjects), color-coded by the cytokines they express (white = “off”, shaded = “on”, color = “degree of functionality”), and ordered by degree of functionality from one function on the left to five functions on the right. Rows correspond to 71 subjects ordered by treatment group: T1 (pale green): NYVAC+Ad5; T2–T4 (purple, aquamarine, rose), increasing doses of Ad5+NYVAC, and by FS within each group. Each cell of the heatmap shows the probability that a given cell-subset (column) has an antigen-specific response in the corresponding subject (column), where the probability is color-coded from white (zero) to purple (one). **(b)** Box plots of FS and PFS for CD4⁺ T cells stratified by treatment groups.

persons may represent a more sensitive marker of infection or may simply be a false-positive result. Further study will be required to distinguish among these possibilities. In contrast, the response to nonspecific mycobacterial peptides was concentrated over one subset expressing IFN- γ , IL-2, TNF- α and CD40L, and was independent of subject's MTB status. This response to nonspecific mycobacterial proteins was likely due to *Bacillus Calmette-Guérin* (BCG) vaccination or environmental exposure to nontuberculous mycobacteria.

COMPASS detected MTB-specific responses for some cell subsets that did not include IFN- γ , suggesting that such analysis may add value to standard Quantiferon test (QFT) analyses in identifying patients with MTB infection (Fig. 4a). Given this observation, we assessed whether COMPASS could be used to classify MTB-uninfected and MTB-infected subsets and how this would compare to a classification based on marginal IFN- γ response alone, essentially a surrogate for the QFT. Using COMPASS's polyfunctional profile to classify subjects as MTB-infected versus uninfected provided increased sensitivity and specificity compared to using IFN- γ alone or the PI, as shown by a ROC analysis comparing COMPASS to various summary statistics based on the IFN- γ response (Supplementary Fig. 8). The lack of total concordance between the QFT and ICS IFN- γ response may be explained by differences between fresh whole blood and frozen peripheral blood mononuclear cells (PBMCs) used for the two assays, the lack of TB7.7 in the ICS assay and measurement of IFN- γ expression (ICS) versus secretion (enzyme-linked immunosorbent assay, ELISA).

COMPASS identifies functional differences between vaccine regimens

As a third example, we applied COMPASS to ICS data from a clinical trial of HIV vaccine candidates (HVTN078)²³ to determine if a polyfunctionality analysis could unveil differences missed in the primary analysis. The combination of heterologous vectors has proven to be a good strategy to increase immune responses to vaccination²⁴, which is particularly relevant for HIV where there is not yet an effective, licensed vaccine. Bart *et al.*²¹ showed that priming with rAd5 followed by a boost with NYVAC resulted in an increased percentage of antigen-specific T cells producing IL-2 and/or IFN- γ compared to NYVAC followed by rAd5, but that higher doses of the rAd5 primer did not lead to any increased response. A heatmap of the posterior probabilities for all CD4⁺ T-cell Boolean subsets considered (Fig. 5a) supports the findings of Bart *et al.*²¹ and also shows that increasing dosage of rAd5 led to a decrease in overall response, which is

nicely summarized by the FS and PFS (Fig. 5b). Although two of the experimental groups T3 and T4, have comparable FS and PFS, it is notable that very few subjects in T4 produced antigen-specific T cells co-expressing IL-4, IFN- γ , TNF- α and CD40L compared to T3 (Fig. 5a). This was confirmed when the probability of response for this subset was plotted separately (Supplementary Fig. 9). This suggests that a difference in dosage led to substantial qualitative differences in immune responses after vaccination ($P = 0.004$ for Fisher's exact test on proportions of responders between T3 and T4).

DISCUSSION

These three examples highlight the ability of COMPASS to reveal differences in the quality of an immune response that were not evident using standard approaches. It remains uncertain as to whether specific cellular subsets of unique polyfunctional profiles have clinical significance, for example, in terms of mediating protection from disease. The RV144 data set provided an opportunity to investigate the association between subset detection and a clinical endpoint, HIV infection. Indeed, two vaccine-induced polyfunctional CD4⁺ T-cell subsets not identified in the prior analyses^{18,19} were shown to be associated with decreased risk of HIV infection. Although the original cellular endpoint in RV144 was not designed to detect polyfunctional responses, current tools used to assess polyfunctionality are inadequate and miss this association. In the MTB study, COMPASS was used to detect associations between cellular subsets and the presence or absence of MTB infection. In this case, the diagnostic methods for determining infection are not definitive and our approach has the potential for augmenting diagnostic methods. The third example, applied in the context of a phase I clinical trial, did not have a clinical endpoint, but the difference in one four-function subset between two identical vaccine regimens that differed only in dose raises questions about the functional significance of this subset, or at least, the potential of this subset as a future correlate of vaccine efficacy.

COMPASS analysis of the tuberculosis data set revealed two findings that would likely have been missed using standard methods of analyzing multiparameter FACS (fluorescence-activated cell sorting) data. First, we demonstrated the presence of a polyfunctional CD4⁺ T-cell subset recognizing MTB-specific proteins that did not express IFN- γ , yet was preferentially detected among MTB-infected subjects. IFN- γ release assays are the clinical standard for diagnosis and are still imperfect²⁵, so this observation may have implications for the design of improved diagnostics. Second, we demonstrated

the presence of a T-cell subset that simultaneously produced IL-17a, IFN- γ , CD40L, IL-2 and TNF- α in response to MTB-specific proteins among MTB-infected subjects. To our knowledge, this has not been previously reported, and it is not clear if such highly-polyfunctional T cells are important for protection from tuberculosis, even though the analysis of RV144 data presented here would suggest this is possible. The availability of COMPASS will make it easier to analyze phase 2 studies of TB vaccines currently underway for the association of these subsets with clinical outcomes. Finally, only a small proportion of MTB-infected individuals will eventually develop active tuberculosis, so the stratification of MTB-infected subjects into different T-cell response categories could be used to predict which individuals are at highest risk of progression to active TB disease.

The FS and PFS introduced in this paper provide unique T-cell response summaries that can be used to quantify a subject's immune response and be used as a biomarker to be correlated with a given outcome. Although the score definitions are related, and thus the scores are expected to be correlated, we believe that the two scores are complementary and provide more information than one score alone. For example, even though the correlation between the two scores in the RV144 case-control study is 0.95, we have shown that whereas the FS was better correlated with other cellular and antibody functions, as measured by multiplex bead array assay, the PFS provided a better correlate of protection. In addition, multivariate response profiles, based on the posterior probabilities returned by COMPASS, were helpful in singling out specific subsets that were associated with specific treatment groups or clinical variables. Other published studies have reported similar findings^{26–28}, reinforcing the need for unbiased, multivariate analyses. It should be noted, however, that as COMPASS analyzes and reports probabilities for all functional subsets, the results could be prioritized and/or summarized in reference to cytokine-based functionalities (e.g., Th1, Th2, Th17).

These findings reinforce the idea that the quality rather than the magnitude of T-cell responses is more important for determining the outcome of infection or response to vaccination²⁹. We think that for the purpose of summarizing the functionality or polyfunctionality over all subsets, it is preferable to define a variable that does not include the magnitude of the response; otherwise, the large-magnitude subsets would mask the low-magnitude ones. This is particularly important as the magnitude of high-degree (i.e., polyfunctional) subsets is typically smaller. This being said, our FS and PFS do take into account the magnitude of response up to a certain point, after which the subsets are treated equally (when the model judges that there is a significant increase upon stimulation and we are certain of the specificity of the response). Also, because our approach is based on a model, the relationship between the magnitude and the probability of antigen specificity takes into account the uncertainty in the observed cell counts (**Supplementary Fig. 10**), such that a cell subset with greater background will have lower response probabilities, even at comparable magnitude. Although we could use the actual proportion of antigen-specific cells in our calculation, we have found that it did not add any predictive value (**Supplementary Figs. 11 and 12 and Supplementary Table 4**). This could be explained by the fact that either the quality of the antigen-specific response has a more important effect on clinical outcome than the quantity of antigen-specific T cells, or that the estimates of proportions are too noisy to be meaningful when the overall proportion of antigen-specific cells is small, which is clearly the case for HIV and TB. However, if one wants to look at the magnitude, we think that it might be better quantified at the subset level, which can easily be done with COMPASS as it models all subsets jointly.

The development of statistically sound methods for the characterization of antigen-specific T cells from single-cell assays is becoming increasingly important, particularly as single-cell technologies improve and the number of functionally distinct subsets that can be defined from an experiment increases exponentially. As we have shown, standard approaches that ignore the multivariate nature of ICS data may miss rare but important signals. This problem will be exacerbated as new high-dimensional, single-cell technologies such as CyTOF, multiplexed-qPCR and RNA-seq become more widely used. Although our approach was developed primarily for flow cytometry-based ICS assays, it is directly applicable to CyTOF-based ICS experiments and should be generalizable to other assay technologies including multiplexed-qPCR and single-cell RNA-seq.

METHODS

Methods and any associated references are available in the [online version of the paper](#).

Accession codes. COMPASS is available as an R package (<http://github.com/RGLab/COMPASS>) and provides an interactive, web-based interface for visualizing all data and results. All results presented in this paper can be visualized through a web-tool available at: <http://rglab.github.io/COMPASS/>. All data are available here: <https://zenodo.org/record/17500#.VVEvB9NViko>.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

This work was funded by National Institute of Health grants (R01 EB008400), and grants (UM1 AI068635) and (UM1 AI068618) to the HIV Vaccine Trials Network (HVTN) and the Statistical Data Management Center (SDMC), the Human Immunology Phenotyping Consortium (U19 AI089986), and the Collaboration for AIDS Vaccine Discovery (OPP1032325). The work was also supported in part by an Interagency Agreement Y1-AI-2642-12 between the US Army Medical Research and Materiel Command and the National Institutes of Allergy and Infectious Diseases and by a cooperative agreement (W81XWH-07-2-0067) between the Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., and the US Department of Defense. We also acknowledge G. Peterson for technical assistance with all assays related to the TB data set.

AUTHOR CONTRIBUTIONS

R.G. directed the project. L.L., G.F. and R.G. conceived the research method. L.L. and R.G. developed the statistical model COMPASS. L.L., G.F. and K.U. developed the COMPASS R package, analyzed the data and contributed to interpretation. L.L., G.F. and R.G. wrote the manuscript. C.S. and S.C.D. generated the MTB dataset. C.S. also contributed to the interpretation of the MTB data. T.J.S., T.R.H., H.M. and W.H. contributed to the TB study design and provided overall manuscript feedback. N.F., S.C.D. and M.J.M. acquired and contributed all RV144 cellular data, and contributed to study design and data interpretation. G.D.T. acquired and contributed RV144 IgG data, contributed to study design and data interpretation. P.G., S.R.-N., J.K., S.N., P.P., N.L.M., J.H.K., M.L.R., R.J.O. and N.K. contributed to the RV144 study design and data interpretation. P.G. also contributed to the statistical analysis of the RV144 case-control data and data interpretation. N.F., S.C.D., P.G., P.-A.B., M.J.M. and G.P. contributed to the HVTN078 study design, helped acquire ICS results and contributed to data interpretation. All authors discussed the results and commented on the manuscript.

DISCLAIMER

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the Department of Defense.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

1. De Rosa, S.C. *et al.* Vaccination in humans generates broad T-cell cytokine responses. *J. Immunol.* **173**, 5372–5380 (2004).
2. Attig, S. *et al.* Simultaneous infiltration of polyfunctional effector and suppressor T cells into renal cell carcinomas. *Cancer Res.* **69**, 8412–8419 (2009).
3. Rodrigue-Gervais, I.G. *et al.* Dendritic cell inhibition is connected to exhaustion of CD8+ T cell polyfunctionality during chronic hepatitis C virus infection. *J. Immunol.* **184**, 3134–3144 (2010).
4. Ciuffreda, D. *et al.* Polyfunctional HCV-specific T-cell responses are associated with effective control of HCV replication. *Eur. J. Immunol.* **38**, 2665–2677 (2008).
5. Precopio, M.L. *et al.* Immunization with vaccinia virus induces polyfunctional and phenotypically distinctive CD8(+) T cell responses. *J. Exp. Med.* **204**, 1405–1416 (2007).
6. Darrah, P.A. *et al.* Multifunctional TH1 cells define a correlate of vaccine-mediated protection against *Leishmania major*. *Nat. Med.* **13**, 843–850 (2007).
7. Roederer, M., Nozzi, J.L. & Nason, M.X. SPICE: Exploration and analysis of post-cytometric complex multivariate datasets. *Cytometry A* **79**, 167–174 (2011).
8. Aghaeepour, N. *et al.* Critical assessment of automated flow cytometry data analysis techniques. *Nat. Methods* **10**, 228–238 (2013).
9. Lo, K., Brinkman, R.R. & Gottardo, R. Automated gating of flow cytometry data via robust model-based clustering. *Cytometry A* **73**, 321–332 (2008).
10. Trigona, W.L. *et al.* Intracellular staining for HIV-specific IFN- γ production: statistical analyses establish reproducibility and criteria for distinguishing positive responses. *J. Interferon Cytokine Res.* **23**, 369–377 (2003).
11. Nason, M. Patterns of immune response to a vaccine or virus as measured by intracellular cytokine staining in flow cytometry: hypothesis generation and comparison of groups. *J. Biopharm. Stat.* **16**, 483–498 (2006).
12. Horton, H. *et al.* Optimization and validation of an 8-color intracellular cytokine staining (ICS) assay to quantify antigen-specific T cells induced by vaccination. *J. Immunol. Methods* **323**, 39–54 (2007).
13. Proschan, M.A. & Nason, M. Conditioning in 2×2 tables. *Biometrics* **65**, 316–322 (2009).
14. Finak, G. *et al.* Mixture models for single-cell assays with applications to vaccine studies. *Biostatistics* **15**, 87–101 (2014).
15. Larsen, M. *et al.* Evaluating cellular polyfunctionality with a novel polyfunctionality index. *PLoS ONE* **7**, e42403 (2012).
16. Scott, J.G. & Berger, J.O. Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. *Ann. Stat.* **38**, 2587–2619 (2010).
17. Newton, M.A., Noueiry, A., Sarkar, D. & Ahlquist, P. Detecting differential gene expression with a semiparametric hierarchical mixture method. *Biostatistics* **5**, 155–176 (2004).
18. Haynes, B.F. *et al.* Immune-correlates analysis of an HIV-1 vaccine efficacy trial. *N. Engl. J. Med.* **366**, 1275–1286 (2012).
19. Rerks-Ngarm, S. *et al.* Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. *N. Engl. J. Med.* **361**, 2209–2220 (2009).
20. Mahomed, H. *et al.* Predictive factors for latent tuberculosis infection among adolescents in a high-burden area in South Africa. *Int. J. Tuberc. Lung Dis.* **15**, 331–336 (2011).
21. Bart, P. *et al.* rAd5 prime/NYVAC-B boost regimen is superior to NYVAC-B prime/rAd5 boost regimen for both response rates and magnitude of CD4 and CD8 T-cell responses. *Retrovirology* **9** (suppl. 2), 072 (2012).
22. Reddy, T.B.K. *et al.* TB database: an integrated platform for tuberculosis research. *Nucleic Acids Res.* **37**, D499–D508 (2009).
23. Bart, P.-A. *et al.* HIV-specific humoral responses benefit from stronger prime in phase Ib clinical trial. *J. Clin. Invest.* **124**, 4843–4856 (2014).
24. McElrath, M.J. & Haynes, B.F. Induction of immunity to human immunodeficiency virus type-1 by vaccination. *Immunity* **33**, 542–554 (2010).
25. Pai, M. *et al.* Gamma interferon release assays for detection of mycobacterium tuberculosis infection. *Clin. Microbiol. Rev.* **27**, 3–20 (2014).
26. Ewer, K.J. *et al.* Protective CD8+ T-cell immunity to human malaria induced by chimpanzee adenovirus-MVA immunisation. *Nat. Commun.* **4**, 2836 (2013).
27. Overstreet, M.G., Cockburn, I.A., Chen, Y.-C. & Zavala, F. Protective CD8 T cells against *Plasmodium* liver stages: immunobiology of an “unnatural” immune response. *Immunol. Rev.* **225**, 272–283 (2008).
28. Derrick, S.C., Yabe, I.M., Yang, A. & Morris, S.L. Vaccine-induced anti-tuberculosis protective immunity in mice correlates with the magnitude and quality of multifunctional CD4 T cells. *Vaccine* **29**, 2902–2909 (2011).
29. Hawkridge, T. *et al.* Safety and immunogenicity of a new tuberculosis vaccine, MVA85A, in healthy adults in South Africa. *J. Infect. Dis.* **198**, 544–552 (2008).

ONLINE METHODS

Data for these studies derive from three clinical protocols that were approved by the relevant institutional review boards. All study participants provided written informed consent for immune response exploratory analyses.

RV144 case-control data set. HIV-negative healthy volunteers enrolled in the RV144 trial (<https://www.clinicaltrials.gov/> registration number NCT00223080) and were vaccinated at weeks 0, 4, 12 and 24; their immune responses at week 26 were evaluated as immune correlates of infection risk through a case-control analysis^{18,19}. A total of 246 vaccinated subjects were used for this analysis: 41 subjects who acquired HIV-1 infection (cases) after week 26 and 205 frequency-matched subjects who did not become infected over the follow-up period (controls). A total of 17 types of immune assays were run on the case-control samples. In addition, these assays were also performed on random samples from 40 placebo recipients (20 cases and 20 controls). The analysis presented here focused on assessing polyfunctional HIV-1 envelope protein (Env)-specific T-cell response using ICS, one of the 17 available assays. As a means to validate our analysis and explore relationships between cellular and antibody functions, we also correlated our ICS responses with two other assays: a multiplex cytokine bead array measuring antigen-specific cytokine secretion in PBMCs, and a binding antibody multiplex assay measuring IgG binding to HIV-1 envelope proteins. After removing missing data due to assay failure, we ended up with 226 vaccinated subjects (38 infected, 188 noninfected) with complete data across the three assay types.

ICS. In this data set, a set of 6 functions—TNF- α , IFN- γ , IL-4, IL-2, CD40L and IL-17a—were measured at the single-cell level in CD4⁺ T cells in the presence and absence of stimulation with a peptide pool matching one of the HIV-1 envelope proteins contained in the vaccine (92TH023). Cell level data were extracted from raw data files using the analysis described in Haynes *et al.* (2012)¹⁸, leading to the definition of count data for the $2^6 = 64$ theoretical Boolean subsets for each subject and stimulation condition. However, the actual number of observed (i.e., nonempty) cell subsets is 59, and we further filtered out subsets that had fewer than six cells in fewer than three subjects, reducing that number to 24.

Multiplex cytokine bead array. A set of 12 cytokines (IFN- γ , IL-4, IL-2, IL-5, TNF- α , IL-10, TNF- β , IL-13, MIP1- β , GM-CSF, IL-3, IL-9) were measured using multiplex bead array technology. For each subject and cytokine, the response is defined as the difference in log concentration between the stimulated (92TH023-Env) and unstimulated samples. For each subject, individual cytokines were called positive/negative using the thresholds defined in the original study¹⁸.

IgG total binding. Binding IgG antibodies to the envelope (Env) protein of the HIV virus were measured using a binding antibody multiplex assay. Here, we used the IgG total binding response to Env defined as the mean binding IgG to multiple Env proteins¹⁸.

Pilot data. As part of the RV144 case-control study, the 17 assay types used in the case-control study were selected from 32 pilot assay types on the basis of reproducibility, ability to detect post-vaccine responses, and uniqueness of responses detected, from which six primary assay variables were selected for the correlate analysis. ICS and multiplex bead array were two of the assays types that were evaluated during the pilot phase. All samples used in the pilot phase were from noninfected subjects. For ICS, 36 placebo and 119 vaccine recipient (60 before and 59 after vaccination) were used. For multiplex bead array data, 30 placebo and 98 vaccine recipient (57 before and 41 after vaccination) samples were used.

TB data set. We obtained cryopreserved PBMC from an epidemiologic study of South African adolescents who were screened for the presence of latent tuberculosis infection (LTBI) using tuberculin skin testing and Quantiferon-TB GOLD in tube testing of whole blood²⁰. This data set includes 40 subjects, 22 MTB-infected and 18 MTB-uninfected. Subjects were classified as MTB-infected using both TB skin testing and the Quantiferon test in-tube gold (QTF-gold) that measures IL- γ release in whole-blood after stimulation with

ESAT-6, CFP-10 and TB7.7 peptides²⁵. PBMCs were plated at a density of 2×10^5 per well and stimulated for 6 h with either DMSO or peptide pools consisting of 15 mers overlapping by 12 peptides for the following mycobacterial proteins: ESAT-6, CFP-10, TB10.4, Ag85A and Ag85B at a final concentration of 1 μ g/ml. Cells were stained using a published panel in which we replaced MIP1- β and CD107a with IL-17a Alexa 700 and IL-22 PE Cy7 (refs. 12,30). Analysis of CD3⁺CD4⁺ events was performed in FlowJo (TreeStar Inc.) after first gating on single cell events, CD14⁻ events, live cells and lymphocytes. For the purposes of analysis, we pooled counts obtained after stimulation with ESAT-6 and CFP-10 and defined this as “MTB-specific” because these proteins are known to be absent in *Bacillus Calmette-Guérin* (BCG) and many environmental mycobacteria. By extension, counts obtained after stimulation with Ag85A, Ag85B and TB10.4 were pooled and defined as ‘MTB-nonspecific’ because these proteins are present in *M. tuberculosis* as well as BCG and many environmental mycobacteria. Seven functions were measured at the single-cell level in CD4⁺ T cells: TNF- α , IL- γ , IL-4, IL-2, CD40L, IL-17a and IL-22, leading to 128 theoretical subsets. However, the actual number of observed cell subsets was 79, and using the filtering criteria described above we reduced that number to 22 (MTB-specific) and 19 (MTB-nonspecific).

HVTN078. HVTN078 (refs. 21,23) is a randomized, double-blind phase 1b clinical trial (ClinicalTrials.gov registration number NCT00961883) to evaluate the safety and immunogenicity of heterologous prime/boost vaccine regimens (NYVAC-B/rAd5 versus rAd5/NYVAC-B) in healthy, HIV-1 uninfected, Ad5 seronegative adult participants. Eighty participants were enrolled into one of four groups receiving different combinations of NYVAC-B (New York Vaccinia (NYVAC) vector containing HIV-1BX08 gp120 and HIV-1IIIIB gag-pol-nef at a dose of 1×10^7 PFU (plaque-forming unit)) and rAd5 (HIV-1 recombinant adenoviral serotype 5 (rAd5) vector vaccine VRC-HIVADV014-00-VP ((HIV-1HXB2/NL4-3 Gag-Pol fusion; HIV-192RW020, HIV-1HXB2/Bal and HIV-197ZA012 Env), at three increasing doses (1×10^8 , 1×10^9 , 1×10^{10} PFU)). We refer to the four different groups as T1, T2, T3 and T4. In the T1 group, NYVAC was the prime with rAd5 as the boost, whereas subjects in T2-T4 received rAd5 as the prime and NYVAC as the boost with the three increasing doses of the prime. Here we used a subset of the ICS data generated through the trial measuring seven functions: IL- γ , IL-2, IL-4, TNF- α , MIP1- β , CD107a and CD40L in CD4⁺ T cells in the presence and absence of stimulation with HIV-1 peptides, leading to 128 theoretical subsets in 71 subjects (T1:29, T2:15, T3:13, T4:14). However, the actual number of observed cell subsets is 107, and using the filtering criteria described above we reduced that number to 26 subsets.

Statistical framework for modeling count data. Without loss of generality, we assume cell counts are obtained from I subjects under two conditions: (antigen)-stimulated and unstimulated as depicted in **Figure 1**. Let M denote the number of markers measured, then in theory, there are $K_M = 2^M$ possible Boolean combinations defining functional cell subsets, depending on whether the marker is expressed or not. We let K ($K \leq K_M$) denote the actual number of cell subsets considered for statistical analysis (i.e., after filtering empty and very sparse cell subsets). The observed counts for the K cell subsets in the stimulated and unstimulated samples are represented by n_{sik} and n_{uik} , $k = 1, \dots, K$, $i = 1, \dots, I$, respectively, or represented as vectors we have $n_{si} = (n_{s1i}, \dots, n_{sKi})'$ and $n_{ui} = (n_{u1i}, \dots, n_{uKi})'$. Then $N_{si} = \sum_k n_{sik}$ and $N_{ui} = \sum_k n_{uik}$ are the total number of cells for subject i in the stimulated and unstimulated samples, respectively (**Fig. 1**). Without loss of generality, we order the cell subsets such that the K^{th} category represents the subset where no cytokines are expressed, that is, the degree of functionality is zero.

For a given subject i , we jointly modeled the cell counts under the two conditions using multinomial distributions: $(n_{si} | p_{si}) \sim MN(N_{si}, p_{si})$ and $(n_{ui} | p_{ui}) \sim MN(N_{ui}, p_{ui})$, where p_{si} and p_{ui} are the unknown proportions for the stimulated and unstimulated paired samples, respectively. To detect a responding subject, as well as antigen-specific subsets within a subject, we considered two competing hypothesis: $H_0: p_{ui} = p_{si}$, and $H_a: \exists k \in \{1, \dots, K-1\}$ such that $p_{sik} > p_{uik}$. Under the null hypothesis H_0 , there is no difference in the proportion of cytokine producing cells between the stimulated and unstimulated samples, and thus the proportion vector parameter is shared across the two multinomial models. Under the alternative hypothesis H_a , some subsets show an increase in their proportions. We define the cell subsets that express at

least one function and are different under H_a as antigen-specific subsets as the change in proportion is induced by the antigen stimulation. The K^{th} null category is not considered here, as a change there would only reflect a change in some other category as the proportion vector has to sum to one. This framework allows each subject to be responding in none, some or all of the subsets, and jointly models all the subjects and subsets to allow information sharing to improve the power in detecting rare signals. The ultimate objective is to automatically identify antigen-specific cell subsets for each subject.

Statistical model for detecting antigen-specific polyfunctional T-cell subsets. To allow the automatic identification of antigen-specific cell subsets for each subject, we introduced a binary indicator, γ_{ik} associated with each subject i and each category k , such that if the category is antigen-specific, $\gamma_{ik} = 1$, otherwise $\gamma_{ik} = 0$. In other words, when $\gamma_{ik} = 1$, the (unknown) cell proportions (p_{sik}, p_{uik}) for the stimulated and unstimulated samples are different, otherwise they are equal. The distribution of cell proportions can easily be specified conditional on the latent indicators, which is shown in **Supplementary Methods**, “Priors.” Our implementation of the COMPASS model uses an optimized Markov chain Monte Carlo (MCMC) algorithm allowing full exploration of the joint posterior distribution (**Supplementary Methods**, “Posterior sampling”). Then statistical inference about responding subsets across subjects, subsets or polyfunctionality degree would be based on the posterior summaries of the latent indicators, such as using the posterior mean, and/or FDR-thresholded posterior probabilities. Our approach uses a Bayesian variable-selection prior, with a subset specific—but common across subjects—prior weight shared and estimated across all subjects. As a result, the inference automatically takes into consideration multiple comparisons¹⁶ across subjects.

Subject-level response probabilities. ICS assays are often used for determining vaccine responders based on the magnitude of antigen-specific T-cell responses. Current approaches for positivity tend to be univariate; a single cell-subset needs to be defined as an endpoint (e.g., cells expressing IL-2 and/or IL- γ). Because COMPASS uses a Bayesian approach to jointly model all cell subsets, any posterior summary of interest is readily available. COMPASS can classify vaccine responders or MTB-infected individuals by computing subject-level response probabilities, defined as the probability that at least two (disjoint) cell subsets exhibit an antigen-specific response in that subject. The rationale is that antigen stimulation in subjects with antigen-specific cells should induce changes in a variety of cell subsets, whereas nonspecific responses are expected to be sporadic and noisy. Note that even though our subject-level response probabilities are computed based on two or more responding cell subsets, we still define the alternative hypothesis H_a (described above) as at least one difference, as COMPASS can return posterior probabilities for each subset, that is, we really consider all possible alternative models.

Polyfunctionality and functionality score definition. We introduce the functionality score (FS) and polyfunctionality score (PFS) to summarize the response across the different cell subsets for each subject. The benefits of a single-number summary of subject-level response have been outlined elsewhere¹⁰. Primarily, it greatly facilitates statistical analysis, comparisons across treatment groups and correlation with outcome measures. We define the FS as the posterior mean of the average number of antigen-specific cell subsets among all measured subsets, irrespective of the degree of functionality:

$$FS_i = \sum_{k=1}^{K-1} \hat{\gamma}_{ik} / (K_M - 1)$$

where $\hat{\gamma}_{ik}$ is the posterior mean of γ_{ik} estimated using MCMC (**Supplementary Methods**, “Posterior sampling”). The FS ranges from zero to one and measures the proportion of distinct cell subsets that are expressed for a given subject among all possible subsets.

The PFS is the mean of the posterior probabilities, weighted by the degree of functionality of the corresponding subset, and normalized by the total number of possible cell subsets that could be observed, given the number of markers considered:

$$PFS_i = \left(\sum_{k=1}^{K-1} \hat{\gamma}_{ik} \times d(k) / \binom{M}{d(k)} \right) / (M \times (M+1)/2)$$

where $d(k)$ is the degree of functionality for cell subset k . The PFS ranges from zero to one. The normalization by the theoretical number of possible cell subsets facilitates the comparison of FS and PFS across different data sets, provided they utilize the same markers. It should be noted that the FS and PFS—based on the posterior probabilities—do not directly take into account the frequencies (i.e., magnitudes) of antigen-specific cells. The posterior probability summarizes all the evidence that a cell subset is antigen-specific by comparing the proportion of cytokine-positive cells in the stimulated sample to the corresponding proportion in the control sample. Once there is enough evidence that the subset is indeed antigen-specific (i.e., the probability is one), the actual proportion of antigen-specific cells, which can be estimated as the difference in proportion between the stimulated and unstimulated samples, is no longer relevant. In other words, two cell subsets with a probability of one will be treated equally in our approach, except for the degree of functionality that factors into the PFS even though they may have differing magnitude. Although, we could use the actual proportion of antigen-specific cells in our calculation (**Supplementary Methods**, “Posterior summary based on magnitude of response”), we have found that it did not improve our analysis, and that in fact it decreased some of the correlation reported here (**Supplementary Figs. 11 and 12 and Supplementary Table 4**).

The PFS will assign a higher score to subjects with antigen-specific cell subsets of higher degree, whereas the FS will assign a higher score to subjects that exhibit antigen specificity in more cell subsets irrespective of their degree of functionality. The two scores are complementary in the sense that the PFS emphasize the quality (i.e., the polyfunctionality) of antigen-specific cell subsets, whereas, the FS looks at the quantity of antigen-specific cell subsets. As an example, if we are considering six markers, two subjects may have the same PFS if one has a single, degree-six, antigen-specific cell subset (PFS = 0.286), whereas the other has antigen specificity in all subsets with degree < 4 (PFS = 0.286). The FS would distinguish between these, assigning a score of 0.016 to the first subject and a score of 0.651 to the latter.

Polyfunctionality index. The polyfunctionality index was calculated as described in Larsen *et al.*¹⁵ using both uncorrected and background-corrected cell frequencies with the tuning parameter q set to 1, which is the equivalent of the PFS score where all subsets are weighted by their degree of functionality. Basically, for the uncorrected PI score we simply used the stimulated frequencies, p_{sik} , whereas for the corrected PI score, we used the background-corrected frequencies defined as $\max(p_{sik} - p_{uik}, 0)$. The two scores are referred to as “PI” and “PI corrected” in **Supplementary Figures 1–12**. We have also tried other tuning parameter values including $q = 1.2$ and $q = 2$, and the results did not improve.

Correlate of risk analysis. All immune variables identified here were assessed as correlates of infection risk (CoR) by using the statistical methods specified in the original correlates study¹⁸. Briefly, for each immune biomarker, logistic regression accounting for the sampling design was used to estimate the odds ratio (OR) of infection, controlling for gender and baseline behavioral risk, and IgA levels.

RV144 primary analysis results. Six assays from the 17 different types of immune assays in the original RV144 study were selected as primary endpoints for optimal statistical power when adjusting for multiple comparisons. These selected primary variables included Env-specific plasma IgA, Env-specific plasma IgG binding avidity, gp70-V1V2-specific plasma IgG, neutralizing antibodies, ADCC (antibody dependent cell-mediated cytotoxicity) and Env-specific CD4⁺ T-cells. The major findings in a multivariate model including all six primary variables were that gp70-V1V2-specific plasma IgG inversely correlated with infection rate (odds ratio 0.57, $P = 0.03$, $q = 0.08$), and Env-specific plasma IgA was a direct correlate of infection (odds ratio 1.54, $P = 0.03$, $q = 0.08$).

30. De Rosa, S.C., Carter, D.K. & McElrath, M.J. OMIP-014: validated multifunctional characterization of antigen-specific human T cells by intracellular cytokine staining. *Cytometry A* **81**, 1019–1021 (2012).