Systems biology approach predicts immunogenicity of the yellow fever vaccine in humans

Troy D Querec[#1], **Rama S Akondy**[#1], **Eva K Lee**[2], **Weiping Cao**[1], **Helder I Nakaya**[1], **Dirk Teuwen**[3], **Ali Pirani**[4], **Kim Gernert**[4], **Jiusheng Deng**[1], **Bruz Marzolf**[5], **Kathleen Kennedy**[5], **Haiyan Wu**[5], **Soumaya Bennouna**[1], **Herold Oluoch**[1], **Joseph Miller**[1], **Ricardo Z Vencio**[5], **Mark Mulligan**[1,6], **Alan Aderem**[5], **Rafi Ahmed**[1], and **Bali Pulendran**[1,7]

[1]Emory Vaccine Center, Yerkes National Primate Research Center, 954 Gatewood Road, Atlanta, Georgia 30329, USA

[2]Center for Operations Research in Medicine & Healthcare, School of Industrial & Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, USA

[3]Sanofi Pasteur, 2 avenue Pont Pasteur, Lyon Cedex 07, France

[4]BimCore, Emory University School of Medicine, 1510 Clifton Road, Atlanta, Georgia 30322, USA

[5]Institute for Systems Biology, 1441 North 34th Street, Seattle, Washington 98103-8904, USA

[6]The Hope Clinic, 603 Church Street, Decatur, Georgia 30030, USA

[7]Department of Pathology & Laboratory Medicine, Emory University, 1364 Clifton Road, Atlanta, Georgia 30322, USA

[#] These authors contributed equally to this work.

Abstract

A major challenge in vaccinology is to prospectively determine vaccine efficacy. Here we have used a systems biology approach to identify early gene 'signatures' that predicted immune responses in humans vaccinated with yellow fever vaccine YF-17D. Vaccination induced genes that regulate virus innate sensing and type I interferon production. Computational analyses identified a gene signature, including complement protein C1qB and eukaryotic translation

initiation factor 2 alpha kinase 4—an orchestrator of the integrated stress response—that correlated with and predicted YF-17D CD8[+] T cell responses with up to 90% accuracy in an independent, blinded trial. A distinct signature, including B cell growth factor *TNFRS17*, predicted the neutralizing antibody response with up to 100% accuracy. These data highlight the utility of systems biology approaches in predicting vaccine efficacy.

The yellow fever vaccine (YF-17D) is one of the most effective vaccines ever made[1]; in the past 65 years, it has been administered to over 600 million people globally. YF-17D was developed empirically in the 1930s by Max Theiler, who attenuated the pathogenic Asibi strain of yellow fever virus[2]. A single injection of YF-17D induces a broad spectrum of immune responses, including cytotoxic T lymphocytes (CTLs), a mixed T helper type I (T_H1)-T_H2 profile, and neutralizing antibodies that can persist for up to 30 years[1]. The mechanism of protection is thought to be mediated by neutralizing antibodies, although cytotoxic T cells also likely to be important[3]. Yet, despite its success, little is known about the mechanisms by which YF-17D induces these effective immune responses.

Because of its longstanding use and efficacy, we proposed that using YF-17D as a model to understand the early immune mechanisms— frequently termed the 'innate response'— underlying this efficacy would be of value in designing new vaccines against other infections. Recent advances have demonstrated a fundamental role for the innate immune system, particularly Toll-like receptors (TLRs) and antigen-presenting cells such as dendritic cells (DCs), in controlling adaptive immune responses[4,5]. Consistent with this, it was recently shown that YF-17D infects DCs[6] and signals through multiple TLRs on distinct subsets of these DCs[7]. Such immunological 'deconstruction' of the mechanisms responsible for the efficacy of an established model vaccine such as YF-17D should provide insights into the design of new vaccines against emerging infections and global pandemics.

The goal of the present study was to perform a multivariate analysis of the innate immune responses in humans after vaccination with YF-17D to identify innate immune signatures that are sufficient to predict the subsequent adaptive immune response. To do this, we used high-throughput technologies, such as gene expression profiling, multiplex analysis of cytokines and chemokines, and multiparameter flow cytometry, combined with computational modeling. Although such tools have changed prognosis and therapy response prediction in oncology[8-10] and are beginning to be applied to identifying signatures of infections[11], they have not yet, to our knowledge, been applied to vaccinology.

RESULTS

YF-17D vaccination induces a network of antiviral genes

We vaccinated 15 healthy humans who had not been previously vaccinated with YF-17D and acquired blood samples at various time points. First, we studied the protein cytokine response in the blood of vaccinees at days 0, 1, 3, 7 and 21 after vaccination, using a 24-plex Luminex assay. Only the chemokine IP-10 (CXCL10, A003787) and the cytokine interleukin 1α (IL-1α) were significantly induced at any given time point, relative to their expression on day 0 ($P < 0.05$; **Supplementary Fig. 1a,b** online). Next we evaluated the

frequency and activation status of antigen-presenting cells, including DCs and monocytes, in the blood at various times after vaccination. There were increases in the percentages of CD86$^+$ myeloid DCs, plasmacytoid DCs, monocytes and CD14$^+$CD16$^+$ inflammatory monocytes at day 7 after vaccination, compared to that on day 0 or 1 (**Supplementary Fig. 1c**).

To gain a global perspective of the innate response to YF-17D, we performed transcriptional profiling of total peripheral blood mono-nuclear cells (PBMCs) from the 15 subjects (trial 1). For this analysis, we used the Affymetrix Human Genome U133 Plus 2.0 Array. The baseline normalized log$_2$ gene expression values were first filtered on the basis of the criterion that >60% of the subjects either upregulated or downregulated those genes by at least a factor of ±0.5 on days 3 or 7. The differential expression of these genes over time was analyzed for statistical significance by one-way analysis of variance (ANOVA); *P*-values were calculated for each gene over the time course of days 0, 1, 3, 7 and 21 by combining the data for all the subjects. The calculations were performed on the log$_2$-fold change in gene expression for day *d* versus day 0. To limit the detection of false positives, the *P*-values were adjusted by the Benjamini and Hochberg false-discovery-rate method with a cutoff of 0.05. This resulted in a list of 97 genes modulated by YF-17D vaccination (**Supplementary Fig. 2a** online). To confirm these results, we performed a similar analysis in an independent second trial of ten subjects who were vaccinated 1 year later with YF-17D. From this second trial (trial 2), we identified a list of 125 YF-17D-modulated genes, of which 65 were also identified in the initial trial (**Supplementary Fig. 2a**). Analyzing the dataset by an independent method, we ran an ANOVA on the entire dataset without any prefiltering. We obtained 22 genes, which were a subset of the 65 genes identified using the first strategy (**Supplementary Table 1** and Methods online, which includes a detailed discussion of both methods). However, this second method excluded many genes that could be independently verified by RT-PCR or even at the protein level (**Supplementary Table 1**).

Using the DAVID Bioinformatics Database (http://david.abcc.ncifcrf.gov/) we analyzed the Gene Ontology terms associated with the doubly confirmed set of 65 genes, which revealed an enrichment of genes related to various immunological responses, cell motility and biopolymer metabolism (**Supplementary Fig. 2b**). Those genes were then imported into TOUCAN (http://homes.esat.kuleuven.be/~saerts/software/toucan.php) for transcription factor binding site (TFBS) analysis, and 44 out of the 65 genes were recognized. The TFBSs found to have statistically over-represented frequencies included the interferon-stimulated response element (ISRE), interferon regulatory factor 7 (*IRF7*) binding site and sterol regulatory element–binding protein 1 (*SREBF1*) binding site (**Supplementary Table 2** online). Visualization of gene networks with Ingenuity Pathways Analysis supplemented with the TOUCAN transcription factor motif information revealed a closely interacting network of 50 interferon and antiviral genes, including *IRF7, OAS1, OAS2, OAS3* and *OASL*; genes involved in viral recognition, including *TLR7* (ref. 12), *DDX58* (RIG-I), *IFIH1* (MDA-5), *DHX58* (LGP2)[13] and *EIF2AK2* (PKR); and genes mediating antiviral immunity, such as *CXCL10* (IP-10), *MX1*, and the complement genes *SERPING1* (C1IN) and *C3AR1* (**Fig. 1a** and **Supplementary Fig. 3** online). Consistent with this, C3a, a product of the

classical, alternative, and mannan-binding lectin complement enzymatic pathways and an anaphylatoxin with chemotactic properties, was increased at day 7 (**Supplementary Fig. 4** online). Furthermore, YF-17D was observed to signal through RIG-I and MDA-5 to induce NF-βB activation (**Supplementary Fig. 5** online).

To depict gene expression in an organized fashion, we first categorized those 65 genes into sub-lists based on gene comment and summary information available through DAVID. The kinetics of expression of these gene sub-lists are presented as heat maps of baseline normalized expression (**Fig. 1b**). There was good agreement between trial 1 and trial 2 on the relative change of expression of each gene. Some genes changed as early as days 1 and 3, but the peak change for most genes was reached on day 7. The largest category contained genes with a clear role in interferon and innate antiviral responses, such as *IRF7* and *STAT1*. Other notable categories included genes in the complement pathway and ubiquitination and/or ISGylation (modification of proteins by addition of interferon stimulatory gene (ISG) products). For an independent verification of these genes, we assayed 10 day 3/day 0 and 15 day 7/day 0 changes in trial 1 by RT-PCR. A significant correlation ($P < 0.0001$) existed between the microarray data and RT-PCR results (**Fig. 1c** and **Supplementary Table 3** online). To test whether the RT-PCR data would independently measure significant changes in gene expression after YF-17D vaccination, a subset of 33 genes of greatest interest from the original microarray data were tested for relative RT-PCR expression by one-way ANOVA over time. Of the 33 genes, 26 had a P-value less than 0.05, confirming the microarray data (**Fig. 1d**).

Induction of this gene signature in response to YF-17D could have resulted from recruitment of specific cell types containing abundant transcripts for these genes, rather than *de novo* induction of gene expression. To determine whether YF-17D induced *de novo* expression of genes in PBMCs, we stimulated PBMCs *in vitro* with YF-17D for 3 or 12 h and then evaluated gene expression. Of the 65 genes induced *in vivo*, 34 were reproducibly and significantly induced ($P < 0.05$; **Supplementary Fig. 6** online). This result demonstrated that YF-17D was able to modulate the expression of these genes in a fixed population of cells. Taken together, this analysis revealed that the innate immune response to YF-17D vaccine was characterized by induction of IP-10 and IL1A (IL-1α) (**Supplementary Fig. 1a,b**), upregulation of CD86 on DCs and monocytes (**Supplementary Fig. 1c**), induction of a 'network' of genes mediating interferon-related antiviral responses (**Fig. 1a** and **Supplementary Fig. 3**), and complement activation (**Supplementary Fig. 4**).

Variable CD8+ T cell and antibody responses

We then evaluated the antigen-specific CD8+ T cell response and neutralizing antibody titers induced by vaccination. During the response to vaccination with YF-17D, activated CD8+ T cells transiently upregulate HLA-DR, CD38 and Ki-67 (a protein expressed during the cell cycle) and downregulate the antiapoptotic protein Bcl-2, and that the peak of expansion occurs at 2 weeks[14]. During this study, we also mapped a newly identified HLA-A0201-specific epitope in YF-17D; tracking CD8+ T cells by *in vitro* flow cytometry using tetramers made with this epitope revealed that antigen-specific CD8+ T cells appeared at the same time as the HLA-DR+CD38+ population (data not shown), and they constituted a

subset of HLA-DR$^+$CD38$^+$ cells at 2 weeks after vaccination (**Fig. 2a**). Also, the magnitude of the epitope-specific CD8$^+$ T cell responses in HLA-A2$^+$ vaccinees was directly proportional (r^2 = 0.724, *P* < 0.0001) to the size of their HLA-DR$^+$CD38$^+$ population (**Fig. 2b**). Together these data support the use of HLA-DR and CD38 to measure the magnitude of the YF-17D-specific CD8$^+$ T cell response.

In addition, these CD8$^+$ T cells expressed markers of T cell activation and function typical of effector T cells, including granzyme B, CD27, CD28 and CCR5 (**Fig. 2c**) and low abundances of CD45RA, CCR7 and CD127, when compared to naive CD8 T cells (**Fig. 2c**). Analysis of CD8$^+$ T cell activation by percentage of CD38$^+$ HLA-DR$^+$ cells at day 15 after vaccination showed, unexpectedly, that even with this highly effective vaccine, immune responses varied among individuals by more than tenfold (**Fig. 2d**). Notably, the magnitude of the CD8$^+$ T cell response at day 15 had a strong correlation with the magnitude of the response at later time points, such as day 30 (Pearson *r* = 0.9135; *P* < 0.0001 (two-tailed)). Similarly, the neutralizing antibody titers also varied considerably among the 15 individuals (**Fig. 2e**).

Signatures that predict antigen-specific CD8$^+$ T cell responses

We asked whether early signatures of innate immune activation could predict the subsequent T cell response. Notably, neither the induction of IP-10 or IL1A (IL-1α) nor the upregulation of CD86 on antigen-presenting cells (**Supplementary Fig. 1**) correlated with the magnitude of the CD8$^+$ T cell response. Furthermore, there was no correlation between the expression of the genes identified in the gene expression analysis described above (**Fig. 1a**) and the magnitude of the CD8$^+$ T cell response (data not shown). Therefore, we sought to identify an early gene signature that correlated with the magnitude of the CD8$^+$ T cell response in the 15 individuals in the first trial. We identified 839 genes that correlated with the magnitude of the CD8$^+$ T cell response (Methods and **Fig. 3**). As indicated by analysis in DAVID, these genes were largely associated with metabolism and immunological responses (**Table 1**). To visualize how well the genes identified by the relative expression and *P*-value cutoffs sorted the subjects in terms of CD8$^+$ T cell responses, we performed unsupervised principal component analysis. The genes segregated the subjects into two subgroups, with an activated CD8$^+$ T cell cutoff of 3% CD38$^+$HLA-DR$^+$ (**Fig. 3a**). GeneSpring's (http://www.chem.agilent.com/en-us/products/software/lifesciencesinformatics/genespringgx/) standard correlation with average linkage hierarchical clustering analysis confirmed that the subjects segregated into two groups on the basis of gene expression and the cutoff point was approximately 3% CD8$^+$ T cell activation (**Fig. 3b**).

However, the real test of such a signature is the extent to which it can truly predict the immune response in an independent trial. To this end, we determined whether the gene signature identified in trial 1 could predict the magnitude of the CD8$^+$ T cell response in trial 2 (and vice versa). To do this, we used two independent classification methods, called classification to nearest centroid (ClaNC)[15] and discriminant analysis via mixed integer programming (DAMIP)[16,17]. ClaNC has been previously shown to successfully develop predictive transcriptional cancer models[15]. Using the ClaNC model, we first determined the minimum number of genes in our signature of 839 genes (**Fig. 4**) required to correctly

classify vaccinees in trial 1 into the high (>3%) and low (o3%) CD8$^+$ T cell responders (**Fig. 4a,b**). This unsupervised model was first developed by plotting the error rates in this classification versus the number of genes (**Fig. 4a**). Zero errors in cross-validations were obtained with 10 to 45 genes per CD8$^+$ T cell category (**Fig. 4a**). Next, we used the signature identified in trial 1 to classify the vaccinees in trial 2 into high (>3%) versus low (o3%), CD8$^+$ T cell responders (**Fig. 4b**). Using less than 20 genes yielded error rates oscillating around 50%, which is no better than would be produced by chance; increasing the number of genes in the models stabilized the overall error rates at 20% (**Fig. 4b**). A minimum subset of 48 genes was needed to reach the minimum error rate (**Fig. 4b** and **Supplementary Table 4** online); the requirement for as many as 48 genes to accurately classify 15 subjects suggested over-training, however.

Therefore, we used as a second approach the DAMIP classification model, a general-purpose optimization-based predictive modeling framework and computational engine, which is a very powerful supervised-learning classification approach in predicting various bio-medical and biobehavioral phenomena[16], owing to the universal consistency of the resulting classification rules and their ability to classify with high prediction accuracy even among small training sets[17]. Furthermore, DAMIP is a discrete support vector machine coupled with a powerful feature selection module, and it has been proven in earlier studies to produce superior classification accuracy when compared to traditional quadratic or linear discriminant analysis[18]. We first trained the DAMIP model using trial 1 to obtain an unbiased estimate of correct classification. This was then followed by a blind test to predict the response of the subjects in trial 2. Specifically, trial 1 consisted of ten subjects in the high group and five in the low group, and trial 2 consisted of five subjects in the high group and five in the low group (**Fig. 3a,b**).

DAMIP allows the user to input the desired misclassification rate, and the classification system will then return predictive rules (each with the associated set of discriminatory patterns) that satisfy the input misclassification rate. In our analysis, setting the training error rate to be 20%, eight independent signature (discriminatory) sets, each associated with a predictive rule, were generated (**Table 2**). Each predictive rule was generated by a signature set with only two or three discriminatory genes, and each produced an unbiased estimate of 93% correct classification in tenfold cross-validation (**Table 2**). Using these predictive rules generated from trial 1, we performed blind tests on trial 2. To evaluate the consistency of the classification rules, in addition to singlefold blind test we also carried out tenfold blind tests. In the singlefold prediction, the prediction accuracy of trial 2 status was at least 80% among all rules produced by these eight independent signature sets, with some signatures reaching blind prediction rates of 90% (**Table 2**). The ten-fold blind prediction showed a similar trend, with prediction accuracies ranged from 80-88%. Examination of each singlefold and tenfold pair revealed that the prediction rates between them were within 5%, thus validating that each classification rule obtained from trial 1 was highly consistent and stable in the trial 2 blind-prediction process. Several genes, including *EIF2AK4* (A000827) and *SLC2A6*, were present in several signature sets of the DAMIP model and were also present in the ClaNC model (**Supplementary Table 4**). Notably, training on trial 2 and testing on trial 1 yielded several predictive signatures, which also contained *EIF2AK4* and *SLC2A6* (**Table 2**).

Many of the genes contained in the DAMIP and ClaNC signatures were verifiable using RT-PCR (**Table 3**). Although gene expression data across various time points were all input into the predictive model, most of the discriminatory signature sets consisted of only day 7 expression relative to day 0. Specifically, among the 22 rules (**Table 2**), only 6 rules involved signature sets that include different time measurements (day 3). Notably, we identified signature sets that provided at least 87% of prediction accuracy (**Table 2**). Although it may be convenient to select the best rules on the basis of the best prediction accuracy for future biological investigation, we caution against premature elimination of those results that offer 70% prediction rate, as some of the most commonly used diagnostic tests, such as the Pap smear, produce similar prediction rates.

Finally, the repeated representation of *EIF2AK* on multiple DAMIP model signatures and in the ClaNC model raised the possibility that this gene could have a key function in mediating CD8[+] T cell responses to YF-17D. *EIF2AK4* (also called GCN2 (mammalian general control nonderepressible 2)) serves a function in the so-called 'integrated stress response' by regulating translation in response to various stress signals from the environment. It does so by phosphorylating the α-subunit of translation initiation factor 2 (eIF2α)[19], which results in the shutdown of translation of most proteins in the cell. In contrast, the expression of proteins responsible for damage repair is increased by a process that involves redirection of these mRNAs from polysomes to discrete cytoplasmic foci known as 'stress granules' for transient storage[20]. Consistent with that, YF-17D induced phosphorylation of eIF2α (**Fig. 5a**) and the formation of stress granules (**Fig. 5b**). Moreover, several other genes encoding molecules involved in the stress-response pathway, including calreticulin, protein disulfide isomerase, the glucocorticoid receptor and c-Jun[20-22], were upregulated in response to YF-17D, and this correlated with the CD8[+] T cell response (**Supplementary Figure 7** online).

Signatures that predict antibody responses

To further strengthen the DAMIP results, we carried out predictions on the B cell antibody responses (**Table 4**). For the B cell analysis, we sought to identify an early gene signature that correlated with the magnitude of the neutralizing antibody response in the 15 individuals in the first trial. Here, trial 1 consisted of six subjects in the high group and nine in the low group, and trial 2 consisted of four subjects in the high group and six in the low group (**Supplementary Fig. 8** online). Genes that correlated with the magnitude of the neutralizing antibody response at day 60 were identified as was done for CD8[+] T cells (described above and in Methods). To visualize how well the genes identified by the relative expression and *P*-value cutoffs sorted the subjects in terms of the antibody responses, unsupervised principal component analysis was performed. The genes segregated the subjects into two subgroups with a neutralizing antibody titer cutoff of 170 (**Supplementary Fig. 8**). We then applied the DAMIP model to determine gene signatures that could predict the antibody response in trial 2. In trial 2, because antibody titers at day 60 were not available, we used the titers at day 90. As before (**Table 2**), we summarized those results with tenfold cross-validation scores of at least 80%. Here, whereas the classification rules from trial 1 uniformly predicted all the trial 2 cases correctly (resulting in singlefold blind prediction of 100%), the rules developed using trial 2 resulted in at most 80% singlefold

blind prediction accuracy (**Table 4**). We note that *TNFRSF17*, a receptor for the B cell growth factor BLyS-BAFF (ref. 23; A000383), was present in all the predictive signature sets of the DAMIP model, and several genes, including *KBTBD7* and *BEND4*, appeared in multiple signature sets (**Table 4**). Notably, many of these genes could be verified using RT-PCR (**Table 5**). These two independent analyses of T cells and B cell responses confirmed that the DAMIP method is suitable for identifying predictive signature sets. For both T cell and B cell analysis, we note that the classification rules generated from trial 1 provided higher blind prediction accuracy for trial 2 data than did the reverse analysis. This may be partly because trial 1 consisted of a slightly larger sample size.

DISCUSSION

Here, we have adopted an interdisciplinary approach using multiplex cytokine analysis, flow cytometry and microarray transcriptional profiling to characterize signatures of YF-17D vaccine responses. Because the high numbers of genes in microarray analysis increase the likelihood of false positives, we verified the observed transcriptional profiles with a second independent study using different subjects vaccinated a year later with a new vaccine lot. Our results indicated that several innate immune mechanisms are induced by YF-17D and that some signatures can be used to predict the strength of the adaptive immune response.

Of the 24 cytokines assayed, IP-10 and IL-1α were significantly induced after vaccination. This is consistent with similar results obtained during other flavivirus infections, such as dengue, West Nile virus and tick-borne encephalitits[24-26]. Thus, IP-10 and IL1A (IL-1α) are reliable markers of YF-17D vaccination, and they may play an integral role in responses to other flaviviruses. We performed a comprehensive microarray analysis to identify genomic signatures that correlated with the immune response. This analysis revealed molecular events observed in innate immune control of viruses. In particular, molecules involved in innate sensing of viruses, such as TLR7 (refs. 4,12), cytoplasmic receptors of 2,5′-OAS family members 1, 2, 3 and L, RIG-I, and MDA-5, as well as transcription factors that regulate type I interferons (*IRF7, STAT1*), were induced; consistent with this, YF-17D was also shown to signal through RIG-I and MDA-5. In addition we also detected the upregulation of *ISG15* and of *HERC5* and *UBE2L6*, which participate in ISGylation[27-29]. The four upregulated genes that are involved in ubiquitination may also be recruited into the ISGylation pathway, or they may remain as part of the ubiquitin pathway, where they form part of a negative feedback loop to downregulate the abundance of specific proteins[29]. Furthermore, there was also upregulation of LGP2, which negatively regulates the response mediated by RIG-I and MDA-5 (ref. 13). Thus, YF-17D vaccination induced a gene signature characteristic of viral infections; however there was no correlation between the induction of such genes and the magnitude of the CD8 T[+] cell response (data not shown).

A different signature was successful in predicting the CD8[+] T cell response. *C1QB* was a key positive predictor of T cells in the ClaNC model; this is consistent with the upregulation of *C3AR1* and *C1IN* and increased plasma C3a concentrations. Consistent with this, deficiencies in C1q, C3, C4, factor B, factor D, CR1 and CR2 each individually increase mortality, and diminish T cell and antibody responses, against the closely related flavivirus West Nile in mice[30]. In addition, two factors, *SLC2A6* (GLUT1) and *EIF2AK4*, were present

in the predictive signatures identified using two independent classification models. *SLC2A6* belongs to a family of membrane proteins that regulate glucose transport and glycolysis in mammalian cells[31]. Notably, in the signature derived in the ClaNC model, several other family members, *SLC16A5, SLC25A13, SLC39A11*, were also represented, suggesting a possible role for glucose metabolism in regulating the CD8[+] T cell response. Although the putative roles of such proteins in regulating immunity is not yet known, recent work suggests that, in T cells, CD28 signaling regulates glucose metabolism through expression of GLUT1 (ref. 32). *EIF2AK4* (also known as mammalian general control non-derepressible-2 (GCN2)) regulates protein synthesis in response to environmental stresses by phosphorylating the α-subunit of initiation factor 2 (eIF2α)[19]. In this stress response, the expression of proteins responsible for damage repair is increased, whereas translation of constitutively expressed proteins is aborted by redirection of these mRNAs from polysomes to discrete cytoplasmic foci known as stress granules for transient storage[20]. Consistent with this, YF-17D induced the phosphorylation of eIF2α and formation of stress granules. Moreover, several other genes involved in the stress response pathway, including calreticulin, protein disulfide isomerase and the glucocorticoid receptor JUN[19-22], were modulated in response to YF-17D and correlated with the CD8[+] T cell response. Recent work has shown an antiviral effect of *EIF2AK4* against RNA viruses[33], but the consequence of this for adaptive immunity is not known. It is thus tempting to speculate that the induction of the integrated stress response in the innate immune system might regulate the adaptive immune response to YF-17D, and perhaps other vaccines or microbial stimuli. Finally, in the case of antibody responses, the gene for *TNFRSF17*, a receptor for the B cell growth factor BLyS-BAFF[23,34], was key in the predictive signatures of the DAMIP model. Notably, BLyS-BAFF is thought to optimize B cell responses to B cell receptor- and TLR-dependent signaling[34].

We stress that the aforementioned signatures do not predict the efficacy of the YF-17D vaccine but rather its immunogenicity. YF-17D is highly efficacious, since epidemiological studies indicate that this vaccine confers protection in 80–90% of vaccinees[3]; the mechanism of protection is believed to be neutralizing antibodies, although cytotoxic T cells are also believed to play a role. To our knowledge, there is no epidemiological data on the magnitude of the antigen-specific CD8[+] T cell responses or neutralizing antibody titers necessary for protection against infection. Therefore, the relevance of the 'high' versus 'low' T cell and antibody responses for protection against infection with yellow fever is at present unclear. However, the goal of this study was to use YF-17D simply as a model to provide methodological evidence that critical parameters of protective immunity (that is, CD8[+] T cell and antibody responses) can indeed be predicted early after vaccination. The identification of gene signatures that correlate with, and are capable of predicting, the magnitudes of the antigen-specific CD8[+] T cell and neutralizing antibody responses provides the first methodological evidence that vaccine-induced immune responses can indeed be predicted. This in turn suggests that such approaches could be used predict the immunogenicity and/or protective efficacy of emerging vaccines. Whereas these findings may be applicable to other live attenuated vaccines, whether the same signatures identified in this study would be effective in predicting the immunogenicity of other vaccines, such as subunit vaccines or conjugate vaccines, remains to be determined. However, we propose that

it should be the goal of vaccine manufacturers to develop subunit and killed vaccines that do have signatures closer to those of YF-17D. This may be achieved by targeting several innate immunity signaling pathways, as YF-17D does.

In summary, we have demonstrated that systems biology approaches not only permit the observation of a global picture of vaccine-induced innate immune responses but can also be used to predict the magnitude of the subsequent adaptive immune response and uncover new correlates of vaccine efficacy. Using two independent trials, we found the DAMIP method useful in determining these correlates. This argument is further strengthened by examining independently both T cell and B cell responses using the DAMIP method. Further application of such approaches may be of interest to vaccine development in several ways. For example, different comparisons, such as vaccine responders versus vaccine nonresponders or good versus poor vaccines, may help to identify possible innate correlates of protection, previously unrecognized mechanisms of vaccine action, and early screening strategies of multiple vaccine candidates, hence facilitating research and development efforts. The recent setback with the Merck HIV vaccine[35] underscores the imperative for such approaches in predicting the immunogenicity and protective capacity of vaccines.

METHODS

Clinical study organization

The research was approved by the Emory University Institutional Review Board. Enrolled volunteers were healthy, aged 18 to 45, and signed a written informed consent form. Potential volunteers were excluded from participating in the study if they were pregnant or if they had been vaccinated previously with YF-17D. Blood samples for multiplex analysis of cytokines, innate immune cell and microarray analysis were collected in citrate-buffered cell preparation tubes (CPTs; Vacutainer; BD) at days 0, 1, 3, 7 and 21 after vaccination. PBMCs were frozen in DMSO with 10% FBS and stored at −80 °C. For T cell and antibody assays, blood was collected in citrate-buffered CPTs on days 0, 15 and 60. The tubes of blood were processed according to the manufacturer's protocol.

Multiplex analysis

Plasma samples from CPTs were stored at −80 °C before cytokine analysis. Assays were performed with the Beadlyte Human 22-Plex Multi-Cytokine Detection System with the addition of interferon-α2 and IL-1 receptor-α Beadmates to make a 24-plex assay (Upstate). Samples were run in duplicate following the manufacturer's protocol on a Bio-Plex Luminex-100 station (Bio-Rad). Data were normalized using the prevaccination cytokine level (that is, $\log_2 C_d - \log_2 C_0$, where C_d is the cytokine concentration on day d). The data were tested for significance in Prism by one-way ANOVA followed by the Tukey post hoc test.

Flow cytometric analysis. PBMCs from all time points for an individual were thawed, stained and acquired in parallel. Monocytes were gated as HLA-DR$^+$CD14$^+$ with the addition of CD16 to delineate the subpopulation of inflammatory monocytes. Myeloid DCs were gated as lineage cocktail HLA-DR$^+$CD11c$^+$, and plasmacytoid DCs were gated as

lineage cocktail HLA-DR$^+$CD123$^+$. CD86 expression was used to indicate the percentage of activated antigen-presenting cells within each population. The log$_2$-transformed values for the percentages of CD86$^+$ cells were normalized relative to baseline values. For T cell activation, after gating on the CD8$^+$CD3$^+$ T cells, we calculated the percentage of CD38$^+$HLA-DR$^+$ cells. Antibodies were obtained from BD Biosciences (HLA-DR, 340690; lineage cocktail, 340546; CD11c, 559877; CD14, 555399; CD123, 340545; CD86, 555658). The data were tested for significance in Prism by one-way ANOVA followed by the Tukey post hoc test.

Assay for yellow fever virus (YFV) neutralizing antibodies

Serum or plasma samples were heated to 56 °C for 30 min to inactivate complement. YFV neutralizing antibodies were measured by cytopathic effect (CPE) (trial 1) or by plaque reduction neutralization test (PRNT) (trial 2). In brief, for neutralizing antibodies by CPE, plasma dilutions in triplicate were incubated with 1,000 plaque-forming units of YFV at 37 °C for 1 h in 96-well flat-bottomed plates. Five thousand Vero cells were added to each well and the plates stained with crystal violet after 4 d. The last dilution that showed an intact monolayer of Vero cells with no CPE was used as the antibody titer. For the PRNT, various dilutions of the sera were incubated overnight at 4 °C with 200 plaque-forming units of YFV. Vero cell monolayers in drained six-well plates were incubated with this virus-serum mixture for 1 h at 37 °C. The wells were overlaid with a mix of agarose and 2XM199 medium and plaques counted 3–4 d later using neutral red. Because the CPE and PRNT assays have different scales of neutralizing antibody titers, the results between the two trials were normalized by their medians; that is, normalized subject X value in trial 2 = (trial 1 median/trial 2 median) × subject X value in trial 2.

RNA isolation and microarray and RT-PCR data generation

After PMBC isolation from CPTs, 2×10^6 cells were lysed in 1 ml of TRIzol (Invitrogen) and stored at −80 °C. After all time points were collected for a subject, the samples were thawed, and the RNA isolation proceeded according to the manufacturer's protocol. Total RNA sample quality was evaluated by spectrophotometer to determine quantity, protein contamination and organic solvent contamination, and an Agilent 2100 Bioanalyzer was used to check for RNA degradation. Two-round *in vitro* transcription amplification and labeling was performed starting with 50 ng intact, uncontaminated total RNA per sample, following the Affymetrix protocol. After hybridization on Human U133 Plus 2.0 Arrays for 16 h at 45 °C and 60 r.p.m. in a Hybridization Oven 640 (Affymetrix), slides were washed and stained with a Fluidics Station 450 (Affymetrix). Scanning was performed on a seventh-generation GeneChip Scanner 3000 (Affymetrix), and Affymetrix GCOS software was used to perform image analysis and generate raw intensity data. Initial data quality was assessed by background level, 3′ labeling bias, and pairwise correlation among samples. For this analysis, we used Affymetrix Human Genome U133 Plus 2.0 Array, but instead of using Affymetrix's sequence clusters to define genes, which is based on the UniGene database build 133, 20 April 2001, gene sequence clusters were based on the updated UniGene build 199, 16 January 2007, to yield a list of 20,078 genes. For RT-PCR analysis, Applied Biosystems constructed a custom TaqMan Gene Expression Plate Assay for 48 genes in their database. Two-step RT-PCR was performed. Values obtained by RT-PCR of genes

from the custom TaqMan Gene Expression plate (Applied Biosystems) were normalized to the average cycling threshold value of the 'housekeeping' genes encoding 18S rRNA (Hs99999901_s1), β-actin (Hs99999903_m1) and β2-microglobulin (Hs99999907_m1), and then the difference in normalized cycling threshold values between days 3 and 7 versus day 0 was calculated. Significance was determined by one-way analysis of variance over days 0, 3, and 7.

In vitro stimulation of human PBMCs with YF-17D

PBMCs from two healthy, unvaccinated donors were isolated and plated at 1×10^6 cells per well in 48-well plates with 1 ml RPMI with 10% FBS and penicillin plus streptomycin. The cells were cultured in the presence or absence of YF-17D at a multiplicity of infection of 1. After 3 and 12 h, RNA was isolated from the cells and processed for microarray analysis. For these experiments, the Affymetrix Human Genome 133A 2.0 Array was used. This microarray contains a subset of genes found on the Human 133 Plus 2.0 Array, which was used in the analysis of the vaccinees. The analysis was performed as described in the Supplementary Methods.

Data analysis

Full details are in Supplementary Methods. Immunofluorescence, immunoblot analysis and ELISA. BHK cells were cultured on cover slips in 24-well plate and stimulated with YF-17D. Cells were fixed with 3.7% formaldehyde and permeabilized with 0.5% saponin (Sigma). Cells were then incubated with anti-TIAR (C-18) (Santa Cruz 1749, 1:50) for 2 h at room temperature. After washing, cells were incubated with donkey anti-goat secondary antibody coupled to fluorescein isothiocyanate (Santa Cruz 2024, 1:100). F-actin structure was visualized using BODIPY 558/568 phalloidin (Invitrogen) and coverslips were mounted using ProLong Gold antifade reagent with 4,6-diamidino-2-phenylindole (DAPI; Invitrogen). Immunofluorescence signal was detected using a LSM510 confocal microscope (Zeiss), and images were captured and analyzed using the Zeiss LSM Image Browser. For immunoblot analysis, human total PBMC or BHK cells were lysed with 100 μl of M-PER mammalian protein extraction reagent (Pierce) containing Halt pro-tease inhibitor, EDTA and phosphatase inhibitor (Pierce). Equal amounts of protein were subjected to SDS-PAGE and transferred onto PVDF membranes. The blot was detected with anti-eIF2α and anti-phospho-eIF2α (Cell Signaling 9722) and developed with horseradish peroxidase-conjugated secondary antibody (Cell Signaling, 3597). Signals were visualized using SuperSignal West Pico chemiluminescent substrate (Pierce). C3a in plasma was measured by ELISA (Quidel A015).

Accession codes

UCSD-Nature Signaling Gateway (http://www.signaling-gateway.org): A003787 and A000827. Gene Expression Omnibus: microarray data, GSE13486.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Monath TP. Yellow fever vaccine. Expert Rev. Vaccines. 2005; 4:553–574. [PubMed: 16117712]

2. Theiler M, Smith HH. The use of yellow fever virus modified by in vitro cultivation for human immunization. J. Exp. Med. 1937; 65:787–800. Rev. Med. Virol. 10, 6-16; discussion 13-15 (2000). [PubMed: 19870634]

3. Monath, TP.; Cetron, M.; Teuwen, DE. Vaccines. 5th ed.. Saunders Elsevier; Philadelphia: 2008. Yellow fever vaccine.

4. Takeuchi O, Akira S. Recognition of viruses by innate immunity. Immunol. Rev. 2007; 220:214–224. [PubMed: 17979849]

5. Steinman RM, Bancherau J. Taking dendritic cells into medicine. Nature. 2007; 449:419–426. [PubMed: 17898760]

6. Barba-Spaeth G, Longman RS, Albert ML, Rice CM. Live attenuated yellow fever 17D infects human DCs and allows for presentation of endogenous and recombinant T cell epitopes. J. Exp. Med. 2005; 202:1179–1184. [PubMed: 16260489]

7. Querec T, et al. Yellow fever vaccine YF-17D activates multiple dendritic cell subsets via TLR2, 7, 8, and 9 to stimulate polyvalent immunity. J. Exp. Med. 2006; 203:413–424. [PubMed: 16461338]

8. Alizadeh AA, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature. 2000; 403:503–511. [PubMed: 10676951]

9. Potti A, et al. Genomic signatures to guide the use of chemotherapeutics. Nat. Med. 2006; 12:1294–1300. [PubMed: 17057710]

10. Sorlie T, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc. Natl. Acad. Sci. USA. 2001; 98:10869–10874. [PubMed: 11553815]

11. Ramilo O, et al. Gene expression patterns in blood leukocytes discriminate patients with acute infections. Blood. 2007; 109:2066–2077. [PubMed: 17105821]

12. Diebold SS, Kaisho T, Hemmi H, Akira S, Reis e Sousa C. Innate antiviral responses by means of TLR7-mediated recognition of single-stranded RNA. Science. 2004; 303:1529–1531. [PubMed: 14976261]

13. Rothenfusser S, et al. The RNA helicase Lgp2 inhibits TLR-independent sensing of viral replication by retinoic acid-inducible gene-I. J. Immunol. 2005; 175:5260–5268. [PubMed: 16210631]

14. Miller JD, et al. Human effector and memory CD8+ T cell responses to smallpox and yellow fever vaccines. Immunity. 2008; 28:710–722. [PubMed: 18468462]

15. Dabney AR. Classification of microarrays to nearest centroids. Bioinformatics. 2005; 21:4148–4154. [PubMed: 16174683]

16. Lee EK. Large-scale optimization-based classification models in medicine and biology. Ann. Biomed. Eng. 2007; 35:1095–1109. [PubMed: 17503186]

17. Brooks JP, Lee EK. Analysis of the consistency of a mixed integer programming-based multi-category constrained discriminant model. Ann. Oper. Res. 2008; 164:1–20.

18. Gallagher RJ, Lee EK, Patterson DA. Constrained discriminant analysis via 0/1 mixed integer programming. Ann. Oper. Res. 1997; 74:65–88.

19. Richter JD, Sonenberg N. Regulation of cap-dependent translation by eIF4E inhibitory proteins. Nature. 2005; 433:477–480. [PubMed: 15690031]

20. Kedersha N, Anderson P. Mammalian stress granules and processing bodies. Methods Enzymol. 2007; 431:61–81. [PubMed: 17923231]

21. Ron D, Walter P. Signal integration in the endoplasmic reticulum unfolded protein response. Nat. Rev. Mol. Cell Biol. 2007; 8:519–529. [PubMed: 17565364]

22. Kaufman RJ. Stress signaling from the lumen of the endoplasmic reticulum: coordination of gene transcriptional and translational controls. Genes Dev. 1999; 13:1211–1233. [PubMed: 10346810]

23. Cancro MP. The BLyS/BAFF family of ligands and receptors: key targets in the therapy and understanding of autoimmunity. Ann. Rheum. Dis. 2006; 65(suppl. 3):iii34–iii36. [PubMed: 17038469]

24. Chen JP, et al. Dengue virus induces expression of CXC chemokine ligand 10/IFN-gamma-inducible protein 10, which competitively inhibits viral binding to cell surface heparan sulfate. J. Immunol. 2006; 177:3185–3192. [PubMed: 16920957]

25. Shirato K, Kimura T, Mizutani T, Kariwa H, Takashima I. Different chemokine expression in lethal and non-lethal murine West Nile virus infection. J. Med. Virol. 2004; 74:507–513. [PubMed: 15368509]

26. Atrasheuskaya AV, Fredeking TM, Ignatyev GM. Changes in immune parameters and their correction in human cases of tick-borne encephalitis. Clin. Exp. Immunol. 2003; 131:148–154. [PubMed: 12519399]

27. Wong JJ, Pung YF, Sze NS, Chin KC. HERC5 is an IFN-induced HECT-type E3 protein ligase that mediates type I IFN-induced ISGylation of protein targets. Proc. Natl. Acad. Sci. USA. 2006; 103:10735–10740. [PubMed: 16815975]

28. Arimoto K, Konishi H, Shimotohno K. UbcH8 regulates ubiquitin and ISG15 conjugation to RIG-I. Mol. Immunol. 2008; 45:1078–1084. [PubMed: 17719635]

29. Staub O. Ubiquitylation and isgylation: overlapping enzymatic cascades do the job. Sci. STKE. 2004; 2004:pe43. [PubMed: 15304665]

30. Mehlhop E, Diamond MS. Protective immune responses against West Nile virus are primed by distinct complement activation pathways. J. Exp. Med. 2006; 203:1371–1381. [PubMed: 16651386]

31. Zhao FQ, Keating AF. Functional properties and genomics of glucose transporters. Curr. Genomics. 2007; 8:113–128. [PubMed: 18660845]

32. Frauwirth KA, et al. The CD28 signaling pathway regulates glucose metabolism. Immunity. 2002; 16:769–777. [PubMed: 12121659]

33. Berlanga JJ, et al. Antiviral effect of the mammalian translation initiation factor 2alpha kinase GCN2 against RNA viruses. EMBO J. 2006; 25:1730–1740. [PubMed: 16601681]

34. Woodland RT, Schmidt MR, Thompson CB. BLyS and B cell homeostasis. Semin. Immunol. 2006; 18:318–326. [PubMed: 16931037]

35. Pantaleo G. HIV-1 T-cell vaccines: evaluating the next step. Lancet Infect. Dis. 2008; 8:82–83. [PubMed: 18060842]
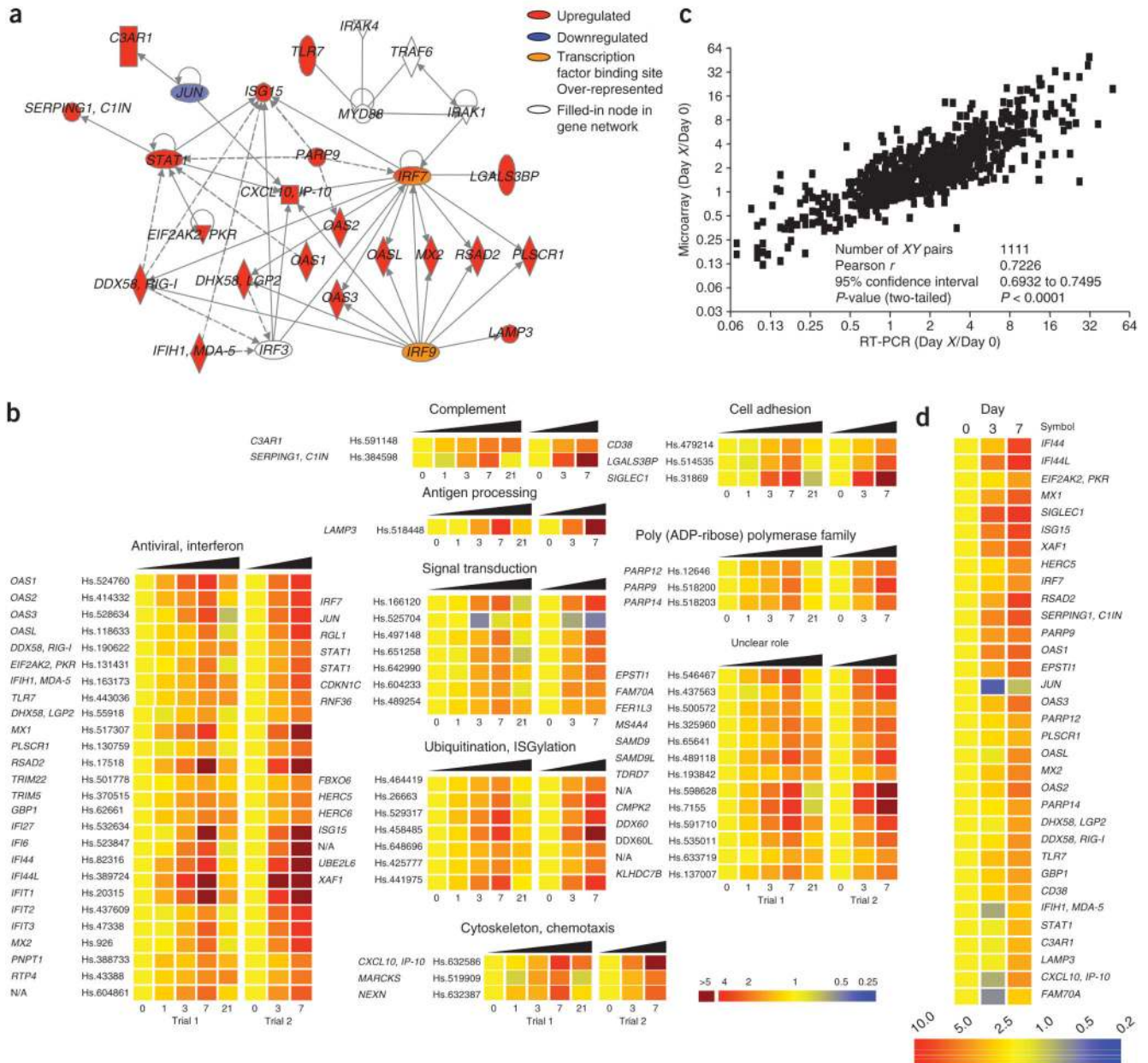
Figure 1.
Genomic signatures of innate immune responses to YF-17D. **(a)** Ingenuity Pathways Analysis of a subset of genes identified as being regulated significantly (Benjamini and Hochberg false-discovery rate, o0.05) in two independent trials and supplemented with transcription factor binding motif information from TOUCAN for *IRF7* and *IRF9* (complete network, **Supplementary Fig. 3**). **(b)** Heat map showing kinetics of changes in expression of common genes identified in two independent trials sorted into categories based on DAVID Bioinformatics Database gene descriptions. The heat map colors represent the average expression among the subjects for each time point (given in days at the bottom of each column). **(c)** Changes in relative gene expression have significant correlations between microarray and RT-PCR analysis. Each point represents a single gene at a given time point.

(d) Analysis of 33 genes identified as being significantly modulated by microarray analysis reveals that 26 genes also have significant modulation as measured by RT-PCR ($P < 0.05$). The heat map represents the gene expression by RT-PCR on days 3 and 7 as a multiple of that on day 0. All genes and time points were first normalized to the average cycling threshold value of expression of the housekeeping genes for 18S ribosomal *RNA, ACTB* (β-actin) and *B2M* (β2-microglobulin). The gene expression on days 3 and 7 as a multiple of that on day 0 was then calculated and imported into GeneSpring for heat map production. Data from **a,b** are derived from trials 1 and 2, with 15 and 10 subjects, respectively. Data from *c,d* are from trial 1, with 15 subjects.
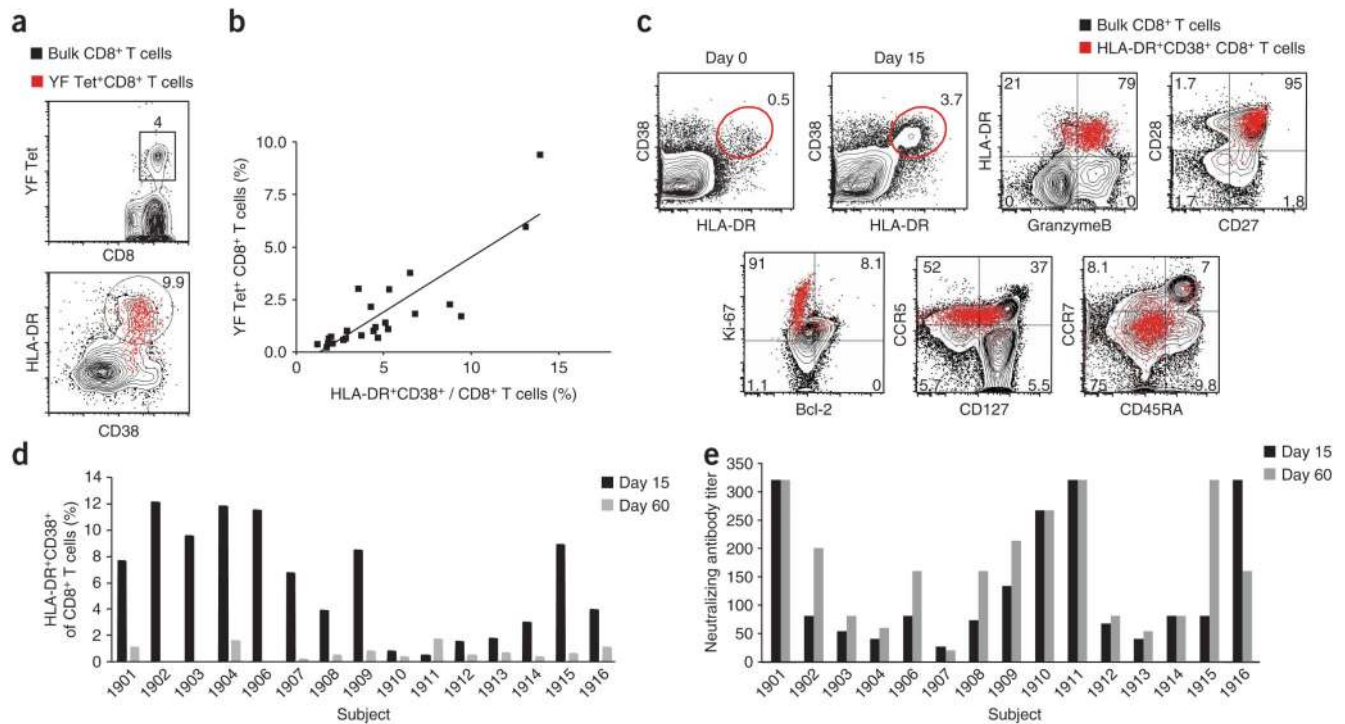
Figure 2.
Variations in the magnitudes of the antigen-specific CD8$^+$ T cell and neutralizing antibody responses to YF-17D. *(a)* Flow cytometry for expression of HLA-DR with CD38, on gated CD3$^+$CD8$^+$ T cells isolated from blood of YF-17D vaccinees. The red dots and numbers indicate the yellow-fever specific CD8$^+$ T cells that stained with the HLA-A2-restricted tetramer (YF-Tet$^+$). **(b)** Correlation between YF-Tet$^+$ T cells and HLA-DR$^+$CD38$^+$CD3$^+$CD8$^+$ T cells. **(c)** Flow cytometry analysis of granzyme B, CD27, CD28, Bcl-2, Ki67, CD127, CCR5, CD45RA and CCR7 in the blood of YF-17D subjects from trial 1. HLA-DR$^+$CD38$^+$CD8$^+$ T cells (in regions outlined for plots of days 0 and 15) have effector phenotype (red dots) on day 15. **(d,e)** Graph of flow cytometry data comparing day 15 and day 60 CD8$^+$ T cell activation and neutralizing antibody titers from 15 subjects in trial 1.
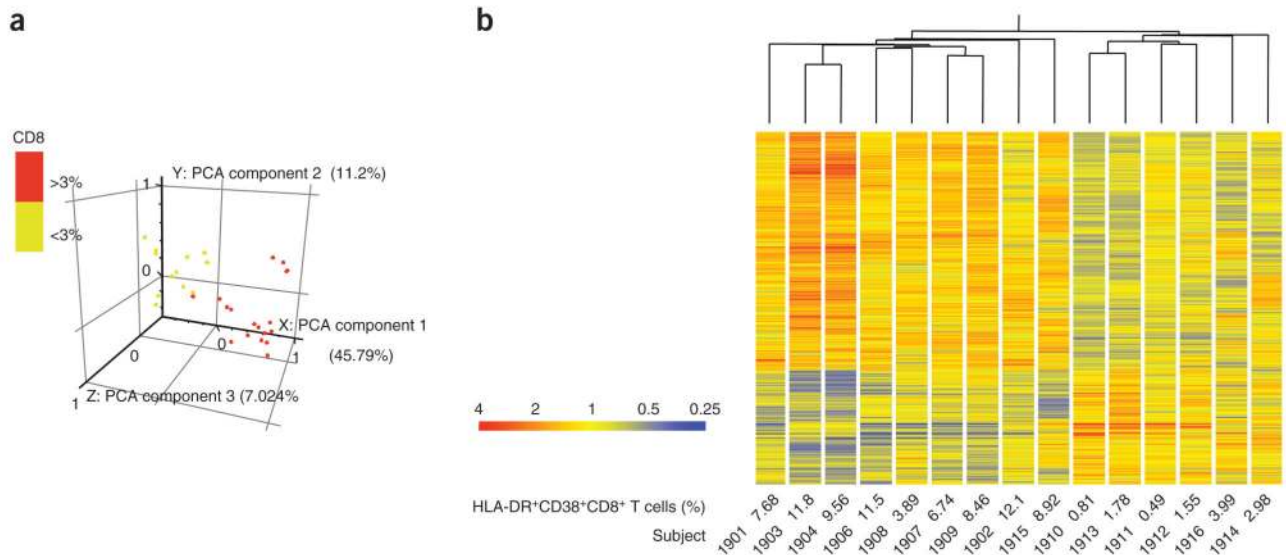
Figure 3.
Genomic signatures that correlate with the magnitude of the CD8[+] T cell response. Genes with a $\log_2$-fold change of >0.5 or −0.5 in more than 25% of the 15 subjects of trial 1 were first selected, for day 3 versus day 0 and separately for day 7 versus day 0. Next, the slope of the *P*-value of the percentage of activated CD8[+] T cells versus the $\log_2$-fold change in gene expression was calculated for each remaining gene. Those genes with *P* < 0.05 were identified as having a significant relationship between early gene expression changes and later CD8[+] T cell responses. **(a)** Unsupervised principal component analysis of the gene expression for each subject on both days 3 and 7 revealed that subjects could be segregated on the basis of CD8[+] T cell responses above and below 3%. **(b)** A standard correlation cluster analysis in GeneSpring confirmed the segregation of T cell responses into two groups with an approximate cutoff of 3–4% activation.
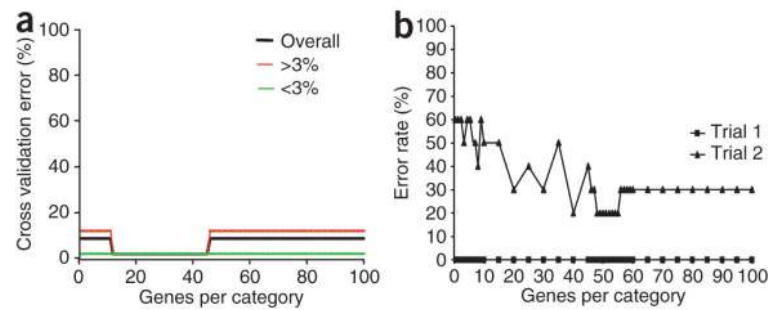
Figure 4.
Genomic signatures that predict the magnitude of the CD8$^+$ T cell responses, using the ClaNC model. The genes identified as having a relationship to the subsequent T cell responses, as described in **Figure 3**, were analyzed by ClaNC to develop a predictive model of CD8$^+$ T cell responses based on a subset of genes. **(a)** A process of leave-one-out cross-validation testing the predictive strengths of subsets of genes for ClaNC gene models. **(b)** The ClaNC gene models developed through cross validation on the first trial of 15 subjects was tested on both trials of 15 and 10 subjects to determine the error rates.
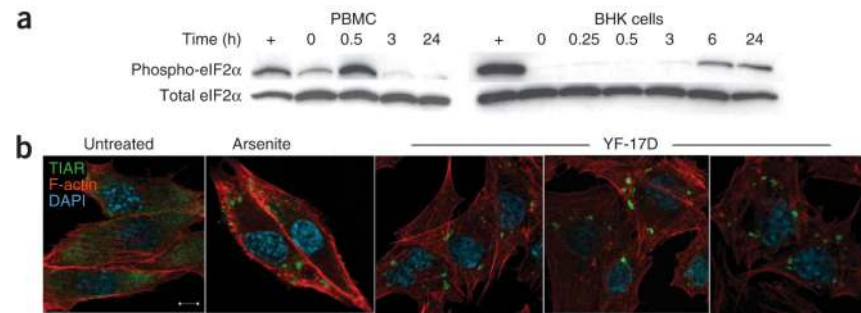
Figure 5.
YF-17D induces eIF2α phosphorylation and stress granule formation. **(a)** Immunoblot on lysates from human total PBMC or baby hamster kidney cells were treated with 0.5 mM arsenite for 30 min or YF-17D for the indicated lengths of time. Cell extracts were prepared and probed for eIF2α phosphorylation (top) as well as for total eIF2α abundance (bottom). **(b)** Fluorescence microscopy of baby hamster kidney cells treated with 0.5 mM arsenite for 30 min or YF-17D (multiplicity of infection 2) overnight before fixing and staining for cytotoxic granule-associated RNA-binding protein-like 1 (TIAR; green). Cells were counterstained with BODIPY 558/568 phalloidin for F-actin (red) and DAPI for nuclei (blue). Scale bars, 5 μm. Results are representative of two independent experiments.

Table 1

Genomic signatures that correlate with the magnitude of the CD8$^+$ T cell response

Gene ontology term	Count	Percentage	*P*-value
Cellular metabolism	292	42.9	1.40×10^{-4}
Primary metabolism	281	41.3	3.60×10^{-4}
Macromolecule metabolism	183	26.9	1.10×10^{-3}
Protein localization	34	5	1.50×10^{-3}
Response to pest, pathogen or parasite	30	4.4	8.10×10^{-3}
Response to other organism	31	4.6	1.00×10^{-2}
Establishment of localization	118	17.4	1.30×10^{-2}
Viral genome replication	4	0.6	2.70×10^{-2}
Regulation of cellular physiological process	125	18.4	3.50×10^{-2}
Cell organization and biogenesis	62	9.1	3.60×10^{-2}
Transport	106	15.6	3.80×10^{-2}
Regulation of metabolism	95	14	6.80×10^{-2}
Nitrogen compound metabolism	19	2.8	8.20×10^{-2}
Negative regulation of physiological process	29	4.3	8.40×10^{-2}
Response to wounding	19	2.8	8.50×10^{-2}

Genes identified in **Figure 3** were analyzed by DAVID for associations with particular Gene Ontology terms. The *P*-values refer to how significant an association a particular gene ontology term has with the gene list.

Table 2

Genomic signatures that predict the magnitude of the CD8+ T cell responses using the DAMIP model

Gene name	Gene symbol	Gene ID	DAMIP model predictive signatures																						
			Train on trial 1, test on trial 2								Train on trial 2, test on trial 1			Train on trial 1													
			1	2	3	4	5	6	7	8	1	2	3	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Solute carrier family 2 (facilitated glucose transporter), member 6	SLC2A6	Hs.244378 Day 7	×	×	×	×	×	×	×	×	×	×	×										×	×	×	×	
Eukaryotic translation initiation factor 2 alpha kinase 4	EIF2AK4	Hs.412102 Day 7	×	×	×	×	×		×	×													×				×
Integrin, alpha L (antigen CD11A)	ITGAL/LFA-1	Hs.174103 Day 7			×												×		×	×	×	×					
C-terminal binding protein 1	CTBP1	Hs.208597 Day 7				×													×								
Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein	YWHAE	Hs.513851 Day 3				×	×															×			×		
Transcribed locus		Hs.619443 Day 7						×			×							×		×	×	×	×	×	×	×	×	×
Protein phosphatase 1, regulatory (inhibitor) PPP1R14A subunit 14A		Hs.631569 Day 3							×																		×
Family with sequence similarity 62 member B	FAM62B	Hs.649908 Day 7		×						×							×									×	
Transcribed locus		Hs.42650 Day 7									×	×												×			
Accuracy of 10-fold cross-validation (%)			93	93	93	93	93	93	93	93	93	90	100	90	90	100	100	100	100	90	90	90	90	90	90	100	100
Accuracy of 1-fold blind prediction (%)			80	80	80	80	80	90	90	90	87	87	80	87	87	80	73	73	73	73	73	73	73	87	73	80	73
Accuracy of 10-fold blind prediction (%)			81	80	81	80	81	85	85	88	84	84	76	84	84	76	72	75	71	71	75	71	75	84	73	76	70

This table summarizes the classification rules that have tenfold cross-validation prediction of at least 80%. Tenfold cross validation on trial 1 resulted in eight different DAMIP predictive signatures, each of which had a tenfold unbiased estimate of 93% prediction rate in trial 1. *SLC2A6* and *EIF2AK4* are represented in several signatures. Blind prediction of rules developed from trial 1 on trial 2 data produced prediction accuracies in the 80–90% range. Tenfold blind predictions were also carried out to evaluate the consistency of the classification rules obtained by subsets of training data only. Here, for trial 1, rule 1 in the tenfold blind test, nine of the ten resulting rules resulted in 80% correct prediction on the blind data from trial 2, and one rule resulted in 90% correct prediction. Thus, the average unbiased prediction rate on the blind data was 81%. However, when we generated the classification rules using the entire training set, they predicted the blind data with an accuracy of 80% singlefold blind test). Conversely, 14 different discriminatory predictive signatures were obtained when trial 2 was used as the training set, with unbiased classification rates in the range 90–100%. *EIF2AK4* and *SLC2A6* were

also represented in these models. Blind prediction on independent trial 1 yielded 73–87% prediction accuracy. Although gene expression data across various time points were all input into the predictive model, most of the discriminatory signature sets (16 out of 22) consisted of only the day 7 expression relative to the day 0.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 3

RT-PCR confirmation of 15 genes used in CD8[+] Tcell activation prediction models

Symbol	UniGene	TaqMan assay	Day	Model	Pearson r	P-value
RGS1	Hs.75256	Hs0017526_m1	3	ClaNC	0.8924	0.0005
CD69	Hs.208854	Hs0015399_m1	3	ClaNC	0.8837	0.0007
ALDH3B1	Hs.523841	Hs00997594_m1	3	ClaNC	0.8117	0.0002
CXCR7	Hs.471751	Hs00664172_s1	3	ClaNC	0.788	0.0068
C1QB	Hs.8986	Hs00608019_m1	3	ClaNC	0.7803	0.0077
ASGR2	Hs.16247	Hs00154160_m1	7	ClaNC	0.7202	0.0025
JUN	Hs.525704	Hs99999141_s1	3	ClaNC	0.7184	0.0193
CXCR7	Hs.471751	Hs00664172_s1	7	ClaNC	0.7078	0.0032
ATP6V1E1	Hs.517338	Hs00762211 S1	3	ClaNC	0.6841	0.0049
ASGR2	Hs.16247	Hs00154160 m1	3	ClaNC	0.6056	0.0167
SLC2A6	Hs.244378	Hs00214042_m1	7	ClaNC, DAMIP	0.5494	0.0339
MEF2A	Hs.268675	Hs00271535_m1	7	ClaNC	0.5423	0.0368
CTBP1	Hs.208597	Hs00179922_m1	7	DAMIP	0.4634	0.0819
ITGAL	Hs.174103	Hs00158238_m1	7	DAMIP	0.4517	0.091
EIF2AK4	Hs.412102	Hs00383836_m1	7	ClaNC, DAMIP	0.4124	0.1266

The Pearson r is calculated for the log$_2$-fold change microarray data versus the relative RT-PCR measurements on either Day 3/Day 0 or Day 7/Day 0 with data points from each of the subjects samples assayed.

Table 4

Genomic signatures that predict the magnitude of the neutralizing antibody responses using the DAMIP model

Gene name	Gene symbol	Gene ID	DAMIP model predictive signatures														
			Train on trial 1, test on trial 2										Train on trial 2, test on trial 1				
			1	2	3	4	5	6	7	8	9	10	1	2	3	4	5
BEN domain-containing 4	BEND4	Hs.120591	×			×	×	×	×	×		×	×			×	
Transcribed locus		Hs.139006						×			×	×					
6-Phosphofructo-2-kinase/fructose-2,6-biphosphatase 3	PFKFB3	Hs.195471		×													
Tumor necrosis factor receptor superfamily, member 17	TNFRSF17	Hs.2556	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
Tumor protein D52	TPD52	Hs.368433				×			×		×			×		×	
Transcribed locus		Hs.481166					×	×		×		×					
Kelch repeat and BTB (POZ) domain containing 7	KBTBD7	Hs.63841	×	×	×	×	×	×	×	×	×	×			×		×
Transcribed locus		Hs.649726							×	×	×						×
Nucleosome assembly protein 1-like 2	NAP1L2	Hs.66180			×							×					
Accuracy of 10-fold cross-validation (%)			80	80	80	87	87	80	80	80	80	80	89	89	89	89	89
Accuracy of 1-fold blind prediction (%)			100	100	100	100	100	100	100	100	100	100	73	73	73	73	80
Accuracy of 10-fold blind prediction (%)			97	99	94	92	96	98	92	93	93	94	72	71	75	70	79

Analysis of signatures that predict the neutralizing antibody responses. Here all the discriminatory predictive signature sets turned out to consist of day 7 gene expression only. Further, training on trial 1 produces a high blind prediction accuracy on trial 2. *TNFRSF17* was present in all the predictive signature sets of the DAMIP model, and several genes, including *KBTBD7* and *BEND4* appeared in several signature sets.

Table 5

RT-PCR validation of genes in the DAMIP models for signatures that predict neutralizing antibody titers

Symbol	UniGene	Day	Pearson r	P-value
BEND4	Hs.120591	7	0.764	0.00002
KBTBD7	Hs.63841	7	0.543	0.02510
TNFRSF17	Hs.2556	7	0.784	0.000001
TPD52	Hs.368433	7	0.530	0.00667