

Global Analyses of Human Immune Variation Reveal Baseline Predictors of Postvaccination Responses

John S. Tsang,^{1,2,*} Pamela L. Schwartzberg,^{1,3,*} Yuri Kotliarov,¹ Angeliqe Biancotto,¹ Zhi Xie,^{1,9} Ronald N. Germain,^{1,4} Ena Wang,^{1,5,10} Matthew J. Olnes,^{1,6} Manikandan Narayanan,² Hana Golding,⁷ Susan Moir,⁸ Howard B. Dickler,¹ Shira Perl,¹ Foo Cheung,¹ The Baylor HIPC Center, and The CHI Consortium

¹Trans-NIH Center for Human Immunology, Autoimmunity and Inflammation, National Institutes of Health, Bethesda, MD 20892, USA

²Systems Genomics and Bioinformatics Unit, Laboratory of Systems Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA

³Genetic Disease Research Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA

⁴Lymphocyte Biology Section, Laboratory of Systems Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA

⁵Department of Transfusion Medicine, Clinical Center, National Institutes of Health, Bethesda, MD 20892, USA

⁶Hematology Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA

⁷Laboratory of Retrovirus Research, US Food and Drug Administration, Bethesda, MD 20892, USA

⁸Laboratory of Immunoregulation, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA

⁹Present address: Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou 510275, People's Republic of China

¹⁰Present address: Research Branch, Sidra Medical and Research Centre, Qatar

*Correspondence: john.tsang@nih.gov (J.S.T.), pams@nhgri.nih.gov (P.L.S.)
<http://dx.doi.org/10.1016/j.cell.2014.03.031>

SUMMARY

A major goal of systems biology is the development of models that accurately predict responses to perturbation. Constructing such models requires the collection of dense measurements of system states, yet transformation of data into predictive constructs remains a challenge. To begin to model human immunity, we analyzed immune parameters in depth both at baseline and in response to influenza vaccination. Peripheral blood mononuclear cell transcriptomes, serum titers, cell subpopulation frequencies, and B cell responses were assessed in 63 individuals before and after vaccination and were used to develop a systematic framework to dissect inter- and intra-individual variation and build predictive models of postvaccination antibody responses. Strikingly, independent of age and pre-existing antibody titers, accurate models could be constructed using pre-perturbation cell populations alone, which were validated using independent baseline time points. Most of the parameters contributing to prediction delineated temporally stable baseline differences across individuals, raising the prospect of immune monitoring before intervention.

INTRODUCTION

The development of accurate models that predict biological responses is one of the major goals of systems biology. Such models have the potential to increase our understanding of pathophysiology and contribute to the development of improved therapeutics (Kitano, 2002; Schadt, 2009). The human immune system provides an excellent context for developing such approaches: many immune cells and molecular components are readily accessible from blood, permitting collection of samples from individuals across multiple time points, followed by in-depth data generation and analyses (Davis, 2008; Pulendran et al., 2010). Furthermore, it is increasingly clear that the immune system and inflammation contribute not only to the pathogenesis of autoimmune and infectious disease, but also to cancer, cardiac disease, diabetes, obesity, neurodegeneration, and other chronic illnesses (Germain and Schwartzberg, 2011). Thus, a more comprehensive and quantitative understanding of how immune responses are orchestrated, together with identification of predictive parameters of effective versus damaging responses, could have implications for the prevention and treatment of diverse diseases.

Building quantitative models often involves the application of perturbations to the system and comprehensive measurements of the initial and resulting states (Chuang et al., 2010). Although advances in high-throughput technologies have made such measurements more routine, utilization of appropriate and ethical perturbations in humans is often a challenge. Here, the

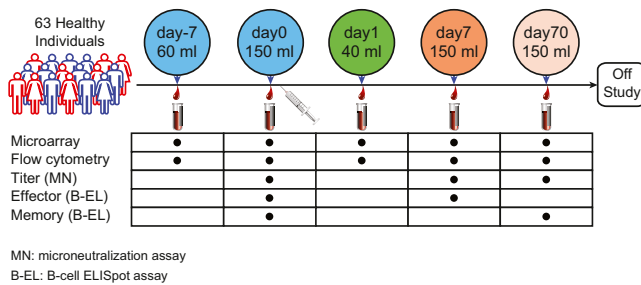


Figure 1. Study Design

Study design indicating blood collections and assays performed. Each subject was vaccinated with the seasonal and pandemic H1N1 influenza vaccines right after the day 0 blood draw.

immune system again offers an advantage, as it is amenable to experimental manipulation. The inactivated influenza vaccine, in particular, is used routinely in healthy and ill populations (Fiore et al., 2009) and provides an attractive perturbation for global data collection and systematic modeling. Upon vaccination, the immune system responds with coordinated changes that reflect the activation and interaction of distinct cell populations and pathways, culminating in the generation of short-lived plasma cells and the formation of germinal centers, from which high-affinity long-lived antibody-producing plasma and memory B cells derive (Pulendran and Ahmed, 2011). By one week post-immunization, a strong but transient plasmablast response can be detected in the blood (Cox et al., 1994; Pulendran et al., 2010), accompanied by increased antibodies in the serum (de Jong et al., 2003). Accordingly, transcriptional profiling of peripheral blood mononuclear cells (PBMCs) revealed substantial changes on days 1, 3, and 7 postvaccination, reflecting both early innate immune activation and day 7 plasmablast responses (Bucasas et al., 2011; Nakaya et al., 2011; Obermoser et al., 2013). Thus, influenza vaccination provides an excellent model of coordinated immune activity involving innate and adaptive responses.

While perturbation analysis is a cornerstone of systems biology, another critical factor for building models in humans is natural population variation. Differences in genetics and environment result in substantial diversity in molecular and cellular states among individuals before and after perturbation. Through correlation analysis, heterogeneity among individuals provides raw ingredients to infer functional relationships among system components—links that cannot be drawn if the parameters analyzed have insufficient diversity in a population. For example, intersubject variation in PBMC gene expression after vaccination has helped to identify postvaccination transcript correlates for antibody responses to yellow fever or influenza vaccination (Gaucher et al., 2008; Nakaya et al., 2011; Querec et al., 2009). However, with the exception of age, how intersubject differences at baseline contribute to outcome has not been well examined. A better characterization of immune variation in healthy individuals is critical not only for the identification of correlates and model building, but also for biomarker development, the definition and characterization of pathological states, and eventually, personalized medicine.

Here, we present a computational framework that utilizes vaccination and multiplexed measurements (gene expression, high density analyses of cell populations, and cellular and serological responses) to quantify baseline and response heterogeneity in a cohort of individuals and systematically identify correlates, build predictive models of vaccination response quality, and infer functional connectivities in the immune system. Using antibody responses as an exemplar endpoint, our analyses confirmed previously reported post-vaccination transcriptome correlates (Gaucher et al., 2008; Nakaya et al., 2011; Querec et al., 2009). Importantly, after accounting for the influence of pre-existing serology, age, ancestry, and gender, we have successfully constructed predictive models and have identified correlates of antibody responses based on prevaccination parameters alone. The robustness and translational potential of these findings is emphasized by our demonstration that the parameters playing essential roles in accurate prediction were cell subsets with temporally stable baseline values within individuals, raising the prospect of predicting the quality of immune responses in the clinic. The data and analytic framework presented provide a potential resource for studying human immunity in health and disease.

RESULTS

In-Depth Analysis of Human Immune Status before and after Vaccination

As a first step toward developing a systems-level understanding of human immunity, we generated a database of immunological measurements using samples drawn from healthy volunteers before and after administration of the 2009 seasonal and pandemic H1N1 (pH1N1) vaccines (Fiore et al., 2009; Table S1 available online and Figure 1). To facilitate the evaluation of intra-individual variation, two baseline blood samples were obtained: one a week before and the other immediately prior to vaccination. Responses were evaluated on days 1, 7, and 70 postvaccination to examine innate, adaptive, and long-term responses, respectively (Figure 1). Purified PBMCs and sera were frozen to allow subsequent assessments to be performed at the same time for a given individual, thereby minimizing batch effects. For the present study, assays included antibody-forming cell responses, influenza-specific serum neutralization titers, multiple 15 color flow cytometric analyses (examining 126 cell subpopulations), and transcriptome analyses using microarrays (Figures 1 and S1A and S1B). Analyses were performed on the 63 subjects from whom we were able to obtain samples at all scheduled time points.

Substantial Immune Baseline Variations in Healthy Subjects

We first examined baseline variation in immune parameters. As expected for subjects from the general population, we found a wide range of baseline titers to the seasonal vaccine, with corresponding large variations in vaccine-specific memory B cell numbers (Figure S2A) (Sasaki et al., 2008). In contrast, most of the cohort was naive for the pH1N1 virus based on titers, despite some potential cross-reactive memory B cell responses,

consistent with previous observations (Hancock et al., 2009; Li et al., 2012).

Examination of 126 PBMC subsets (Biancotto et al., 2011) (Figure S1B and Table S2) revealed a wide range of frequencies for many cell populations on day 0 (Figure S2B). Cell subsets with the highest variability among individuals were typically characterized by expression of activation markers or cytokines, likely reflecting activation of cells within a broadly defined parent population. To exclude experimental variability as the primary source of variation, we analyzed multiple aliquots of a control frozen PBMC sample along with an independent group of samples over a 2-month period. We found that technical variability was minimal (average control sample replication $R^2 = 0.99$; Figure S1B).

Aside from experimental noise, variation can be attributed to the combined effects of baseline differences across individuals (intersubject baseline variation) and temporal changes around the baseline within subjects (intrasubject variation). We took advantage of our multiple baseline measurements (days -7 , 0 , and 70) to assess the relative contribution of these two types of variation for each cell subset. Day 70 was used as an additional baseline because we did not observe substantial postvaccination changes except for influenza-specific antibody titers and B cell memory responses (Figure 2A; Experimental Procedures). Among populations with high overall variability, many B cell subsets showed low within-subject variability (and therefore high temporal stability) over a period of more than 2 months (indicated in dark gray in Figure 2A).

Transcript abundance for day 0 PBMCs also showed substantial variation (Figure 2B), likely reflecting, in part, the different proportions of cell subsets among subjects. To better quantify this variation in a functional context, we developed an approach to summarize “pathway activity” for each subject based on whether a substantial number of genes in a given annotated pathway or function were expressed at different levels in the PBMC of the individual relative to the cohort average (see Experimental Procedures). In addition to innate and adaptive immune pathways (Figures S2C and 2C), some metabolic pathways were also highly variable (Figure S2C), perhaps reflecting changes in metabolic profiles that occur in immune cells upon activation and differentiation (Gerriets and Rathmell, 2012). Assessment of inter- versus intrasubject variation demonstrated diverse patterns, with subject-to-subject baseline differences (blue bars) contributing substantially to the observed variability of many genes and pathways (Figures 2B and 2C).

Coherent Changes in Immune Parameters Postvaccination

We next evaluated whether we could detect coherent changes (i.e., consistent across subjects) following vaccination. As expected, we observed strong but highly variable increases in influenza vaccine-specific IgG+ antibody-secreting cells on day 7 (Figure 3A). As previously reported (Bucasas et al., 2011; Nakaya et al., 2011; Obermoser et al., 2013), we also found coherent changes in gene expression, with largely nonoverlapping signatures on days 1 and 7 (Figures 3B and 3C). Overall, day 1 responses included pathways reflecting innate immune-cell activation, including interferon-related genes, as observed by

others (Bucasas et al., 2011; Nakaya et al., 2011; Obermoser et al., 2013) (Figure 3C and Table S3). By day 7, adaptive pathways were more predominant, including those strongly associated with plasmablasts, such as “endoplasmic reticulum (ER) stress” and “N-glycan biosynthesis” (see below) (Gass et al., 2004; Iwakoshi et al., 2003). Nonetheless, transcript changes associated with adaptive responses were also observed on day 1, including “B cell receptor signaling.”

We were particularly interested in evaluating changes in the frequency of cell subsets, the majority of which have not been examined in the context of vaccination. As with the transcriptome, changes in subset frequencies on days 1 and 7 were largely nonoverlapping (Figure 3D). Again, day 1 changes reflected activation of innate immune cells, including CD40+ and CD86+ monocytes (ID65 and ID67) and IFN α + plasmacytoid DCs (ID78), populations that have been linked to vaccination efficacy and responses to influenza (Fonteneau et al., 2003; van Duin et al., 2007). However, we also observed changes in adaptive cell populations, particularly those reflecting activation of CD8+ and CD4+ T cells, as well as both naive and memory B cells (Kaminski et al., 2012). Day 7 responses primarily reflected changes in adaptive cell populations, including plasmablasts (ID87), as expected, but also in a related CD20+ B subset that expressed CD27 and CD38 (ID96; Figure S3), as well as activated T cell populations. Thus, despite substantial baseline differences among subjects, there were clear coherent changes in cell populations and gene expression postvaccination, with days 1 and 7 involving distinct cell types and genes.

Antibody Titer and Cellular Responses Depend on Prevaccination Serological State

A major question in understanding human immunity is the extent to which responses to a perturbation are affected by pre-existing immune status. To address this question, we first assessed the effects of day 0 serology and memory B cell status on outcomes. To evaluate titer responses, we adopted the “fold-change from baseline” metric used by the FDA and World Health Organization for the evaluation of vaccine efficacy (Figures 4A and S4A). Because the day 70 over day 0 fold change in titer correlated with both the day 7 fold change in titer and effector B cell responses and thus reflected both short- and long-term responses (Figures S4B and S4C), we used day 70 titer responses as the endpoint for subsequent analyses. Interestingly, subjects with higher initial titers tended to have lower fold changes. Though not intuitively obvious, inverse correlations have been observed in other influenza vaccination studies (Bucasas et al., 2011; Furman et al., 2013; Sasaki et al., 2008).

Next, we clustered subjects using a combination of six baseline parameters: the four normalized baseline titers to the viruses targeted by the vaccines and the fractions of memory IgG+ B cells specific for the seasonal and pH1N1 vaccines (Experimental Procedures). Unsupervised clustering followed by robustness analysis (Experimental Procedures) revealed two separable groups containing 33 and 26 subjects. Group 1 individuals were generally low for all six parameters, whereas group 2 had mostly higher values though less uniformly so for pH1N1 titers (Figure 4B). Analyses of postvaccination transcript changes revealed no significant differences between the two

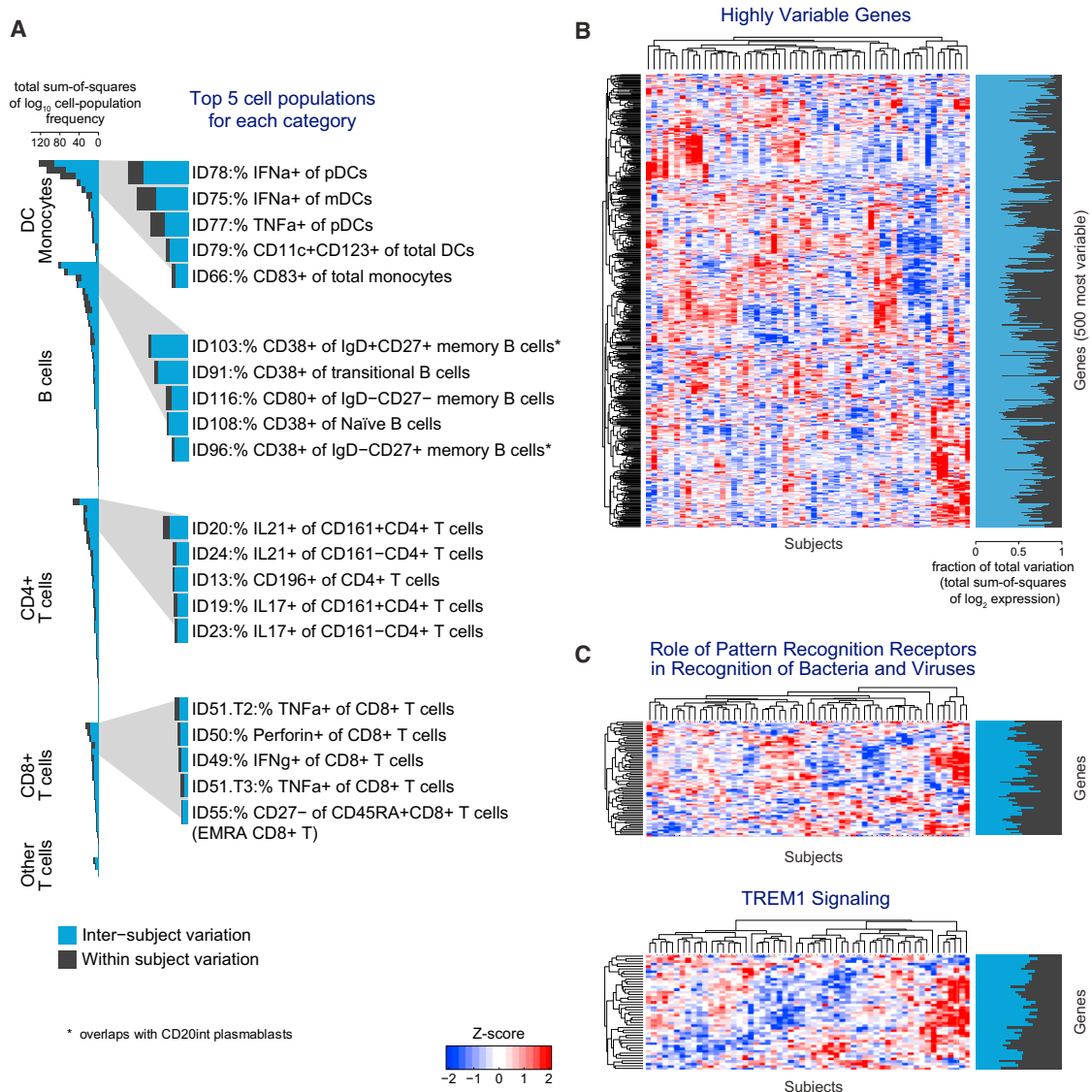


Figure 2. Inter- and Intrasubject Variation at Baseline

(A) Variation in cell population frequencies at baseline. The length of the bar denotes observed sum of squares (R^2). Blue and gray designate R^2 attributed to inter- and intrasubject variation, respectively.

(B) Variation in gene expression at baseline. Hierarchical-clustered heatmap of the 500 most variable genes and the relative proportion of inter- and intrasubject variation (Experimental Procedures).

(C) Clustered heatmaps of genes from two variable immune relevant pathways showing distinct patterns of expression heterogeneity across subjects (see Figure S2C and Experimental Procedures).

groups. However, the more naive group exhibited significantly greater increases in plasmablasts (ID87) and seasonal vaccine-specific ELISpot responses on day 7 (Figure 4C). Thus, both the plasmablast response on day 7 and the mean fold change in titers on day 70 were inversely correlated with initial titers.

Systematic Identification of Pre- and Postvaccination Predictors and Correlates

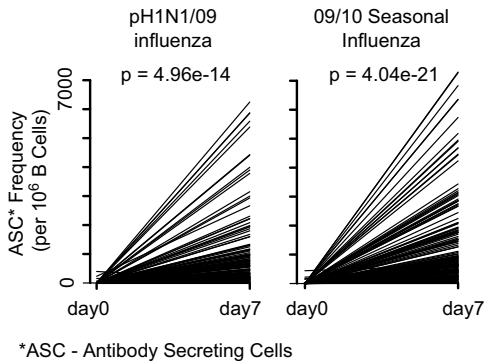
We next developed a systematic framework to take advantage of intersubject baseline and response variations to build predictive models of antibody responses. Our approach involved delineation

of the impact of intrinsic (e.g., age and gender), baseline, and response variables on outcome (Figure 5A). Given that most studies have focused primarily on linking postvaccination parameters to titers, we were especially interested in whether baseline parameters correlated with and could potentially predict outcome independent of age, gender, and pre-existing antibody titers.

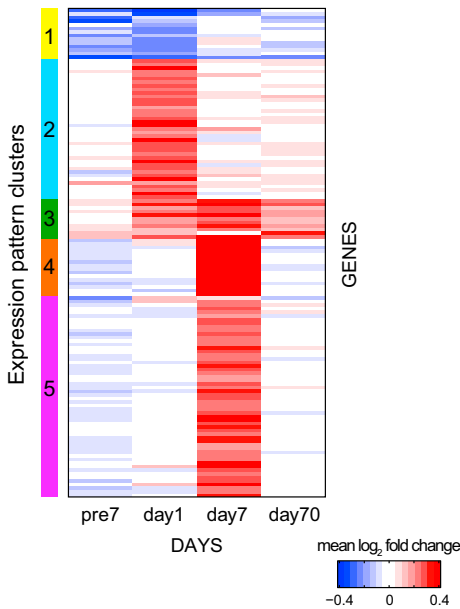
Defining Titer Response Endpoints

To account for the strong influence of baseline titer on day 70 responses, we devised two endpoint metrics. The first, maximum

A



B



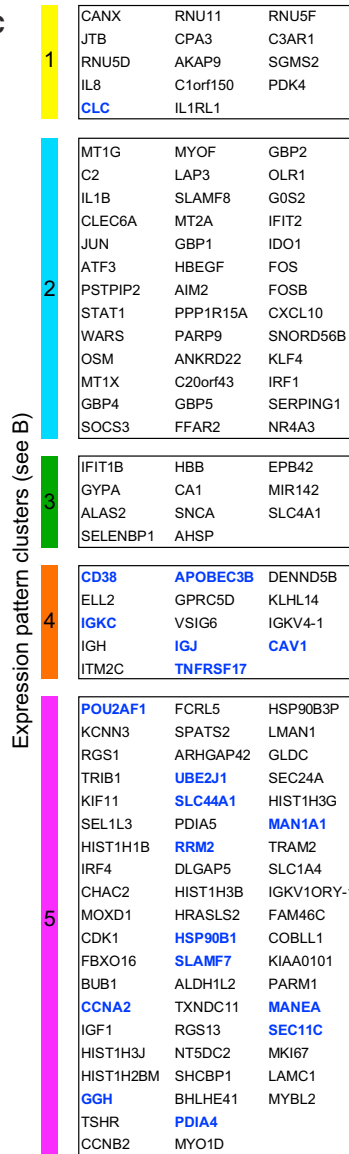
D

day1 vs day0

Cell population	P-value	
% CD40+ of total monocytes (ID65)	5.80E-09 ↑	■
% CD38+ of CD8+ T cells (ID41)	4.26E-05 ↑	■
% CD86+ of total monocytes (ID67)	1.39E-04 ↑	■
% CD86+ of Naïve B cells (ID110)	6.76E-04 ↑	■
% CD69+ of CD8+ T cells (ID39)	7.06E-04 ↑	■
% CD25+ of CD8+ T cells (ID40)	8.58E-04 ↑	■
% CD38+ of IgD+CD27+ memory B cells* (ID103)	1.00E-03 ↑	■
% IL2+ of CD4+ T cells (ID26)	1.15E-03 ↑	■
% HLA-DR+ of CD4+ T cells (ID10)	1.46E-03 ↑	■
% CD103+ of CD4+ T cells (ID31)	1.73E-03 ↑	■
% CD14+ of viable CD45+ cells (Total Monocytes) (ID64)	1.77E-03 ↑	■
% CD80+ of IgD-CD27+ memory B cells (ID97)	1.93E-03 ↑	■
% CD103+ of CD8+ T cells (ID44)	5.28E-03 ↑	■
% IFNα+ of pDCs (ID78)	6.11E-03 ↑	■
% HLA-DR+ of CD8+ T cells (ID42)	6.29E-03 ↑	■
% IL4+ of CD4+ T cells (ID28)	6.41E-03 ↑	■

* overlaps with CD20int plasmablasts

C



day1 vs day0

- * Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses
- * B Cell Receptor Signaling
- * Natural Killer Cell Signaling
- * Fcy Receptor-mediated Phagocytosis in Macrophages and Monocytes
- * Role of PKR in Interferon Induction and Antiviral Response
- * Activation of IRF by Cytosolic Pattern Recognition Receptors
- * Interferon Signaling

day7 vs day0

- * N-Glycan Biosynthesis
- * B Cell Receptor Signaling
- * T Cell Receptor Signaling
- * TGF-β Signaling
- * Endoplasmic Reticulum Stress Pathway

day7 vs day0

Cell population	P-value	
% CD21+ of plasmablasts (ID89)	6.70E-08 ↓	■
% CD27hi CD38hi of CD20- B cells (Plasmablasts) (ID87)	3.47E-07 ↑	■
% CD38+ of IgD-CD27+ memory B cells* (ID96)	1.39E-06 ↑	■
% CD86+ of CD20+ B cells (CD86+ activated mature B) (ID82)	7.59E-04 ↑	■
% CD38+ of CD4+ T cells (ID9)	1.35E-03 ↑	■
% CD86+ of Naïve B cells (ID110)	1.59E-03 ↑	■
% IgA+ of IgD-CD27+ memory B cells (ID99)	1.77E-03 ↓	■
% IL4+ of CD4+ T cells (ID28)	2.48E-03 ↑	■
% CD39+ of CD8+ T cells (ID43)	2.54E-03 ↑	■
% CD86+ of IgD-CD27- memory B cells (ID117)	3.74E-03 ↑	■

■ Innate ■ Adaptive

(legend on next page)

fold change (MFC), was defined as the maximum of the day 70 over day 0 fold change across all viral titers, based on normalized values (Experimental Procedures). The second, adjusted MFC (adjMFC), removed the nonlinear correlation between MFC and day 0 titers by first dividing the cohort into groups of subjects with similar day 0 titers and then adjusting the response within each group so that they were normalized and thus comparable (see Experimental Procedures and related approaches [Bucasas et al., 2011; Nauta, 2011]). Finally, for each metric, we defined “high” and “low” responders as subjects ranked above or below the top or bottom 20th percentile mark of the metric, respectively. We modeled the maximum response (Nakaya et al., 2011) because none of the cell population and gene expression parameters that we assessed targeted specific antigens. Thus, our goal was to determine whether models could be built to predict robust responses to at least one of the vaccine antigens.

Our analyses suggested that baseline titer alone should be a predictor of antibody responses postvaccination. However, responses were still highly variable within groups of individuals with similar baseline titers (Figures 4A and S4A). Thus, this variation (captured by adjMFC) could be utilized to determine baseline titer-independent correlates and predictive models. Though we were mainly interested in adjMFC, we also analyzed MFC to assess the effects of day 0 titers on prediction and associated predictive parameters.

Contributions of Intrinsic Factors

We first assessed the effects of intrinsic factors, including age, gender, and self-reported ancestry (Figure 5A), by using ANOVA analysis on all variables simultaneously. Consistent with previous data (Furman et al., 2013; Goodwin et al., 2006; Gross et al., 1995), age was the only correlate for both metrics (Figure 5B). As expected, initial titers were significantly correlated with MFC (Figure 5B), but not with adjMFC, for which more than 80% of the variance remained unexplained.

Method Overview for the Identification of Predictors and Correlates

We next conducted two types of analyses—predictive modeling and robust correlate identification—to assess whether any baseline or postvaccination parameters were associated with responses. Predictive modeling analyses pooled the high and low responders in our cohort and randomly divided them into nonoverlapping training and testing sets containing 75% and 25% of the subjects, respectively; the former was used for building models whose predictive performance was then evaluated in the *unseen* testing set (Experimental Procedures). Analyses were repeated 5,000 times using different randomly generated sets to

evaluate the robustness of the model construction method and to identify parameters that consistently contributed to prediction. When prediction was *not* possible, we also assessed whether we could identify parameters that consistently correlated with the response endpoints when different random subsets of subjects were used in the analysis (Experimental Procedures)—while an endpoint may have such robust correlates, these may be insufficient for reliable, quantitative prediction.

The first step for training predictive models involved selecting the top k parameters based on the parameters’ strength of correlation with the endpoint in the training set; we tested multiple values of k (see below). Our criterion for assessing robust prediction was to determine whether the area under the receiver operator characteristic curve (AUC) was above 0.5 in at least 75% of the 5,000 iterations (0.5 is the AUC level expected by chance). To reveal which parameters contributed to prediction, we identified those that were: (1) consistently ranked in the top k and therefore selected for model construction in a large number of iterations (% selected) and (2) weighted highly by the model in making its predictions (median weight) (Experimental Procedures).

Identification of Postvaccination Predictors and Robust Correlates

Given previous findings of postvaccination transcript correlates with antibody responses (Nakaya et al., 2011), we first examined data from days 1 and 7. To ensure that correlates were independent of intrinsic factors, we removed age, gender, and race effects from both cell population and gene expression data sets via linear regression (Experimental Procedures). Though models built using either gene expression or cell populations from day 1 were not predictive, interferon signaling correlated robustly with MFC (Figure 5C). Coherent changes in interferon-associated transcripts have been detected on days 1 and 7 in our and other related studies (Bucasas et al., 2011; Gaucher et al., 2008; Nakaya et al., 2011; Obermoser et al., 2013; Querec et al., 2009). In contrast, highly predictive models could be built using day 7 cell population or gene expression data for both adjMFC and MFC (Figures 5D and 5F). Consistent with previous findings (Nakaya et al., 2011), MFC could be predicted using as few as two or five genes, though more were required for adjMFC (Figure 5F). Notably, robustly correlated (Figure 5C) and predictive genes (Table S4) were enriched for ER stress, N-glycan biosynthesis, and cell-cycle pathways, whose gene members correlated strongly with plasmablast frequencies (see below). Accordingly, plasmablasts (ID87) and the overlapping CD20+ subpopulation (ID96) were the dominant cellular contributors (Figure 5E) (Withers et al., 2007).

Figure 3. Postvaccination Changes in Serologic, Cellular, and Transcriptomic Parameters

- (A) Increased antigen-specific antibody secreting cells on day 7 following vaccination measured by ELISpot.
 (B) Heatmap of genes from days –7, 1, 7, and 70 that changed significantly compared to day 0 (FDR < 0.05 and absolute log-fold change > 0.2; fold change from day 0 is indicated by color). Genes with similar patterns of change were grouped by clustering analysis (indicated by the color bars on the left).
 (C) Coherently changed genes in the clusters from (B). (Right) Representative immunological pathways enriched in coherently changing genes on days 1 and 7 (Experimental Procedures). Genes found to be predictive of or correlated with response titers in a previous influenza vaccination study (Nakaya et al., 2011) are marked in blue.
 (D) Coherently changed cell populations on days 1 and 7 postvaccination (FDR < 0.05 and > 10% change from day 0). Blue and light-brown boxes denote populations considered “innate” and “adaptive,” respectively.

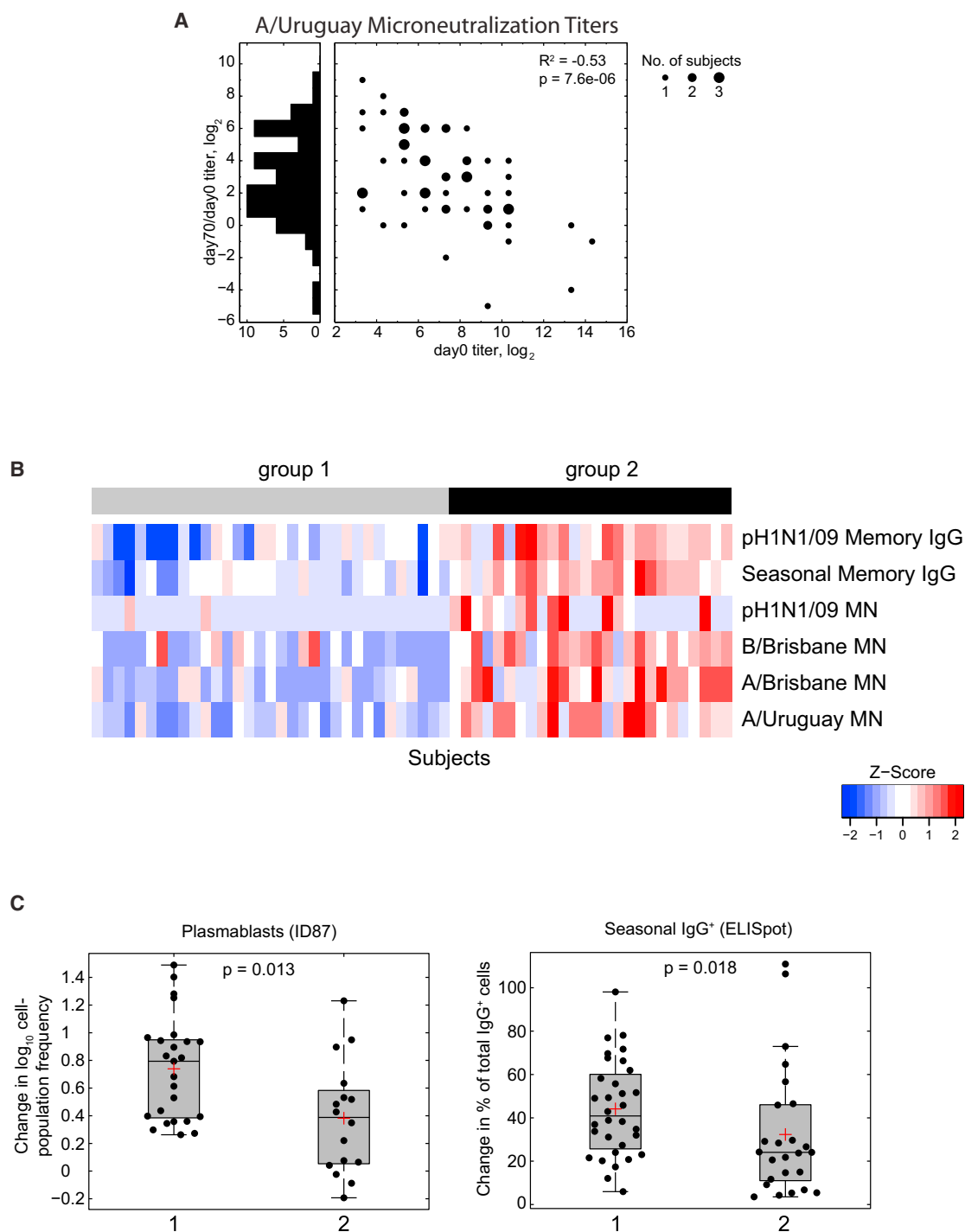
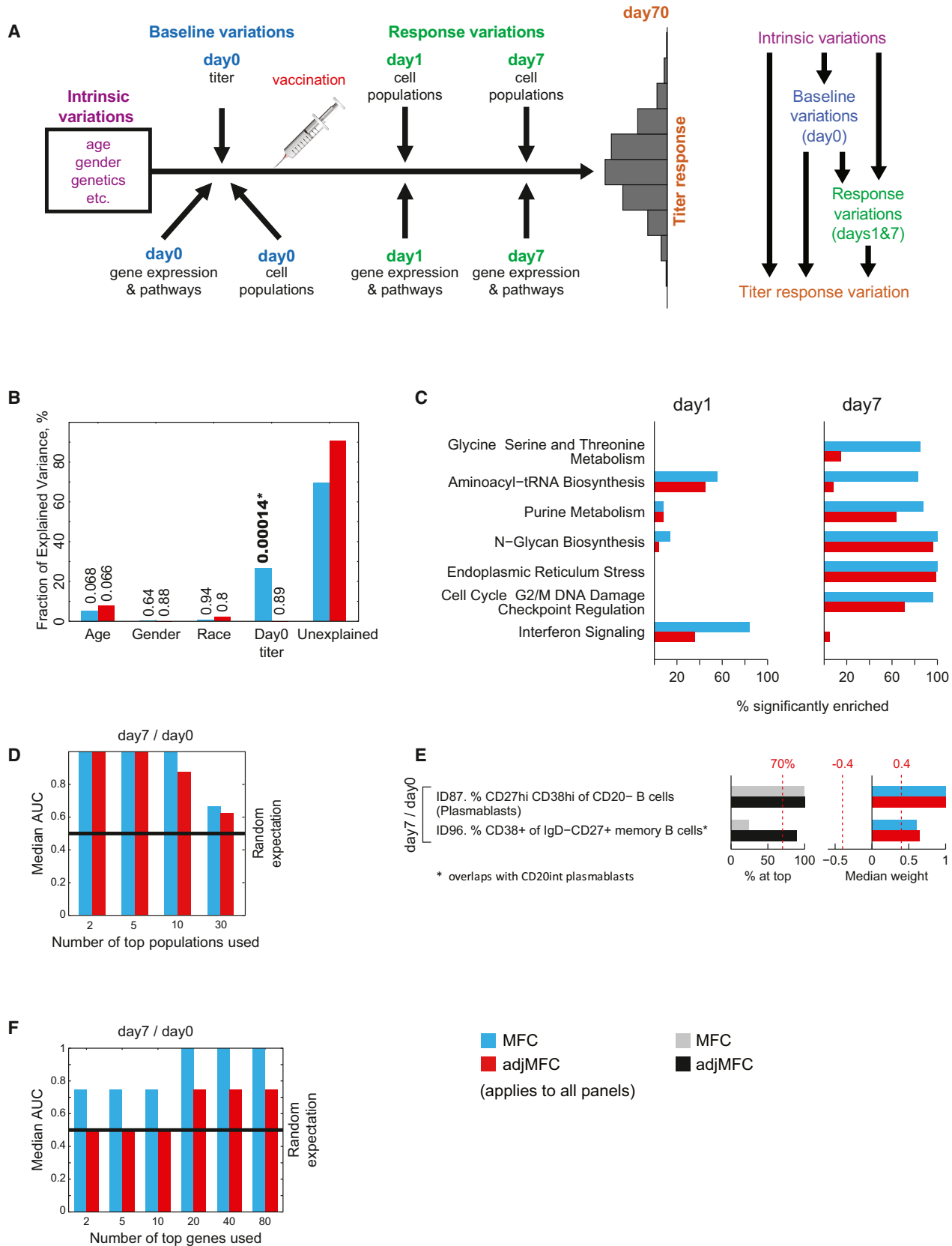


Figure 4. Postvaccination Responses Depend on Baseline Serology and Memory B Cell Status

(A) Relationship between initial and fold change (day 70/day 0) in titer for A/Uruguay after vaccination (see Figure S4 for other viruses). Histogram at left shows distribution of the titer responses.

(B) Heatmap showing two distinct groups of subjects identified based on robust clustering analysis of day 0 serologic and memory B cell variables (see Experimental Procedures)

(C) Postvaccination changes in plasmablast frequencies (left) and frequencies of IgG⁺ seasonal influenza vaccine-specific antibody secreting cells (right) on day 7 for the low (1) and high (2) initial titer groups from (B).



(legend on next page)

Identification of Prevaccination Predictors and Correlates

We next asked whether predictive models could be built using baseline (day 0) parameters alone. Again, we removed age, gender, and ancestry from cell population and gene expression data to ensure that models and correlates that we identified were independent of these factors (Experimental Procedures). Neither day 0 gene expression nor pathway activity (defined in Figure S2C) alone was predictive of either endpoint, though a number of robustly correlated genes and pathways could be identified (Figure 6A). Notably, genes associated with pattern recognition and interferon signaling robustly correlated to MFC but less so to adjMFC.

We then built and tested models for MFC and adjMFC using: (1) the frequency of 126 cell populations or (2) both the cell population frequencies and the activity score of 183 annotated pathways (Figure S2C). Cell frequency and transcript data were integrated using the pathway activity score because the numbers of pathways and cell populations were comparable, whereas the large number of genes could overwhelm signal from cell populations. Of note, predictive models could be built by using cell populations alone (Figure 6B); integrating pathway/transcript with cell population data performed substantially worse (best median AUC was 0.6 versus 0.78 using cell populations alone). Thus, day 0 cell frequencies alone provided the essential information for prediction.

We next identified cell populations that contributed most to prediction. Twelve cell populations were identified for adjMFC, including memory, naïve, and transitional B cells; CD4 effector memory T cells; IFN α + myeloid dendritic cells (mDC); and several activated T cell populations (Figure 6C and Experimental Procedures). Several of the B cell populations were CD38+, including CD20+IgD-CD27+CD38+ B cells (ID96), as well as transitional (ID91) and naïve (ID108) populations and an IgD+CD27+ memory population (ID103) that has been linked to early production of IgM (Shi et al., 2003). Only some of these populations—primarily the T cell subsets—contributed to the prediction of MFC (Figure 6C). Thus, by accounting for baseline serology using adjMFC, we uncovered additional baseline predictors.

Temporal Stability of Predictive Cell Populations

To evaluate the temporal stability of the day 0 predictive cell populations, we examined their intra- versus intersubject variation over time (Figure 2A). Except for ID105 (CD86+IgD+ memory B

cells), all of the B cell and some of the CD4+ T cell (including effector memory) predictive populations were remarkably stable within individuals over a period of more than 2 months (Figure 6D). Among these were B cell subsets that also exhibited some of the highest intersubject variation among all of the cell populations that we measured (e.g., ID103 and 91; Figure 2A), suggesting that these may be robust baseline markers for delineating intersubject differences. These stable B cell populations (ID91, 96, 103, 108) were all CD38+ and contributed to the prediction of adjMFC, but not MFC. Examining the predictive cell populations' temporal profile further confirmed their stability and ability to separate the high and low responders (Figures 6E, 6F, and S5). This was true even for ID96, which increased coherently across subjects at day 7 postvaccination and returned to lower levels by day 70. Together, these results suggest that many of the predictive markers that we uncovered reflected stable individual differences in baseline immune states that correlated with the capacity for robust responses to influenza vaccination.

Because some of the predictive populations were less temporally stable and thus may indicate transient events (e.g., an infection), we next assessed whether predictive models could be built using only the 56 most stable populations (those above the 80% stability mark; Figure 6D) as inputs to our 5,000 iteration cross-validation assessment. Even using a smaller k of 10 (versus 30 previously), adjMFC could be robustly predicted (median AUC = 0.78; Figure 6G). As expected, the populations contributing most to adjMFC prediction were the same stable ones identified above. Thus, the less stable populations were not essential for adjMFC prediction. In contrast, use of the stable cell populations decreased the predictive performance for MFC, which remained flat for all values of k . However, if the stable populations above the 80% mark in Figure 6D were omitted instead, performance remained high for MFC (median AUC = 0.83 for $k = 10$) but dropped substantially for adjMFC (median AUC = 0.56 for $k = 10$). Thus, the more transient signals provided the bulk of the predictive power for MFC, whereas that for adjMFC came from temporally stable parameters that may have captured the unique immune status of individuals.

We further evaluated the robustness of our results by building models using the entire cohort's day 0 cell population data as the training set ($k = 2, 5, 10, \text{ or } 30$) and then assessing whether these models could predict adjMFC using independent measurements from other baseline time points (day -7 or the more temporally removed day 70). Highly accurate prediction could still be

Figure 5. Predictive Modeling of Antibody Response and Identification of Correlates

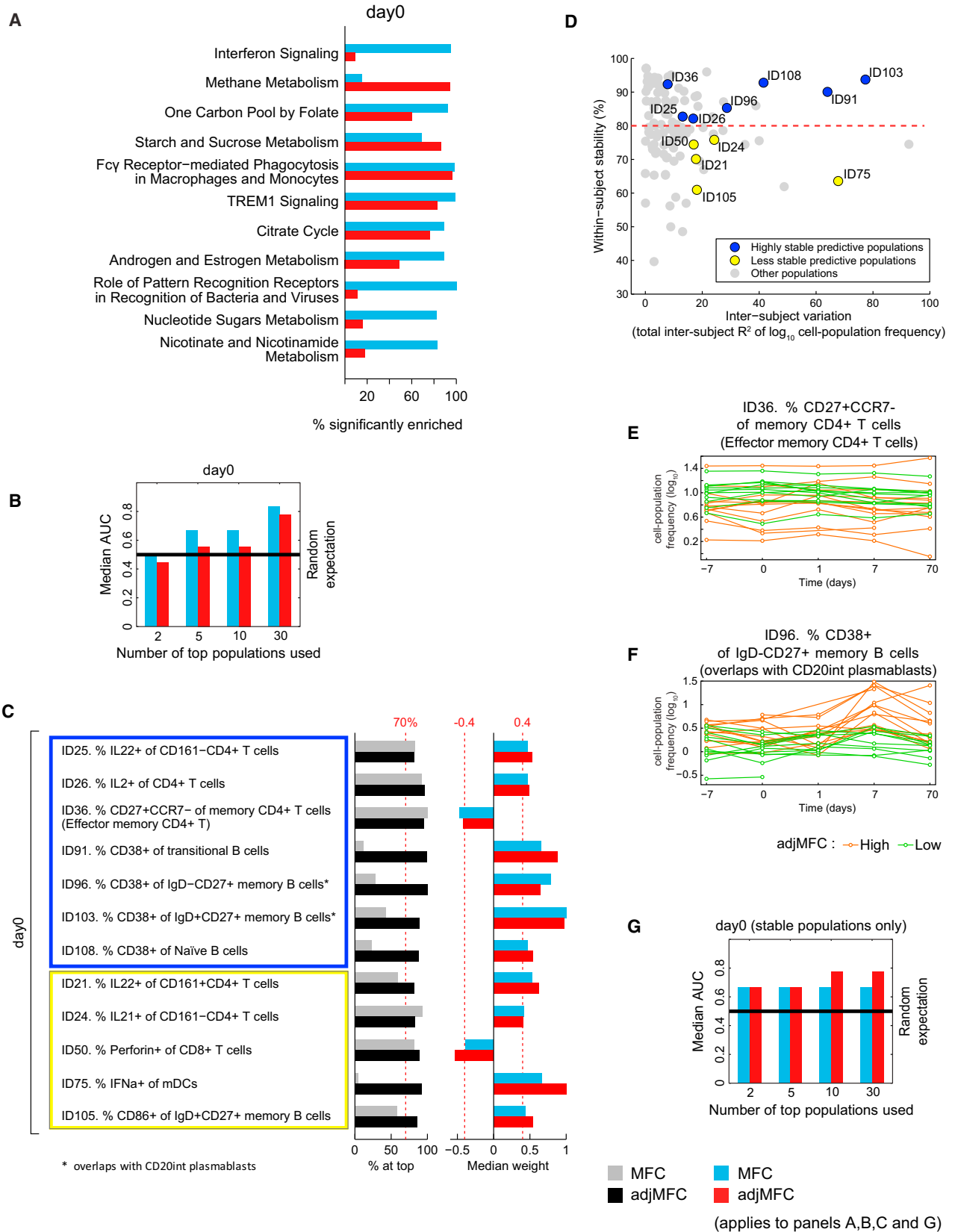
(A) Conceptual framework for analyzing contributors to titer response variation following vaccination. Variation in intrinsic factors (purple) could contribute to variation in baseline immune statuses (day 0, marked in blue) and postvaccination responses (days 1 and 7, marked in green and orange). Baseline variation could, in turn, contribute to response variation, and all sources of variation could together contribute to variation in titer responses. Contributions to titer variation were analyzed following the "flow" of time. (Right) The contributions of intrinsic variables were first analyzed, and their effects were removed from baseline (day 0) and response (days 1 and 7) parameters, which were then analyzed in a similar step-wise manner (Experimental Procedures).

(B) ANOVA modeling of day 70 titer response using age, gender, ethnicity, and baseline titers. Fraction of variance explained by each variable and the associated p values are shown for MFC and adjMFC.

(C) Robust transcriptomic correlates from days 1 and 7 postvaccination (shown as pathway enrichments) (Experimental Procedures). The percentage of times (in 100 random trials) that a pathway was enriched ($p < 0.05$) is shown. Pathways that were significant in more than 80% of the trials in at least one time point are shown.

(D) Prediction performance for MFC and adjMFC is shown for models built using different number of top cell populations (D) and genes (F) from day 7 (Experimental Procedures).

(E) Day 7 predictive cell populations (see Results and Experimental Procedures). The percentage of times (in 5,000 random trials) that a cell population was included in the model is shown. Median normalized weights of each population are shown on the right. The sign of the weight indicates the direction of correlation between the cell population and the titer endpoint.



(legend on next page)

made for all k 's, e.g., AUC = 0.91 and AUC = 0.86 for days –7 and 70 using $k = 10$, respectively. Thus, the predictive signals were remarkably stable over time.

Transcriptomic Signature of the Predictive Cell Populations

To obtain functional insight into the predictors, we took advantage of our simultaneous assessments of cell population frequencies and the PBMC transcriptome to identify transcripts and associated pathways whose abundance correlated with the frequency of predictive cell subsets (Experimental Procedures). As expected, the two predictive and coherently changing cell populations from day 7 (ID87 and ID96) were linked to known antigen-secreting cell/B cell/plasmablast-specific genes, including Ig chains, XBP1 and IRF4, as well as plasmablast-associated pathways such as ER stress (Figures 7B and 7D). These populations were also associated with N-glycan biosynthesis, consistent with the highly glycosylated state of immunoglobulins and the requirement for proper glycosylation for plasma cell development (Tulp et al., 1986). Many plasmablast-correlated genes were also ones that changed on day 7 (Figures 3B, 3C, and 7B [marked in purple and orange]) and included genes found to be predictive of or correlated with response titers in previous studies (e.g., TNFRSF17) (Gaucher et al., 2008; Nakaya et al., 2011; Querec et al., 2009). Thus, our analyses suggest that the day 7 coherent and predictive signatures in our as well as previous studies were likely drawn from a tight coupling between titer response and the degree of plasmablast expansion.

Our analyses also revealed that day 0 predictive populations were associated with distinct transcriptomic signatures (Figures 7A and 7C). Several mitochondria-related pathways were positively associated with ID108 (CD38+ naive B cells) and ID91 (CD38hi transitional B cells), perhaps reflecting changes in bioenergetic states associated with B cell development and activation. ID36 (effector memory CD4+ T cells) and ID50 (perforin+ CD8+ cells) also shared signatures, suggestive of common features of activation. Of note, several innate pathways (PRR signaling, TREM1 signaling, and interferon-related genes) were positively linked to IL22+ and IL2+ CD4+ T cells (ID25 and ID26) and were negatively linked to ID36 and ID50. The association of innate pathways with these adaptive subsets before vaccination suggests that activation of these pathways in these or other interacting cells may play important roles in determining response quality. Interestingly, ID96, a day 0 predictive B cell population that also coherently increased on day 7 (Figures 3D and 6F), showed largely distinct signatures on days 0 and 7 (Figures 7A–7D), suggesting differences in the cell population defined by those markers or distinct functional interactions with other cell populations post-vaccination. Thus, by linking intersubject variation in gene expres-

sion to cell subset frequencies, our approach reveals potential connections among components of the immune system.

DISCUSSION

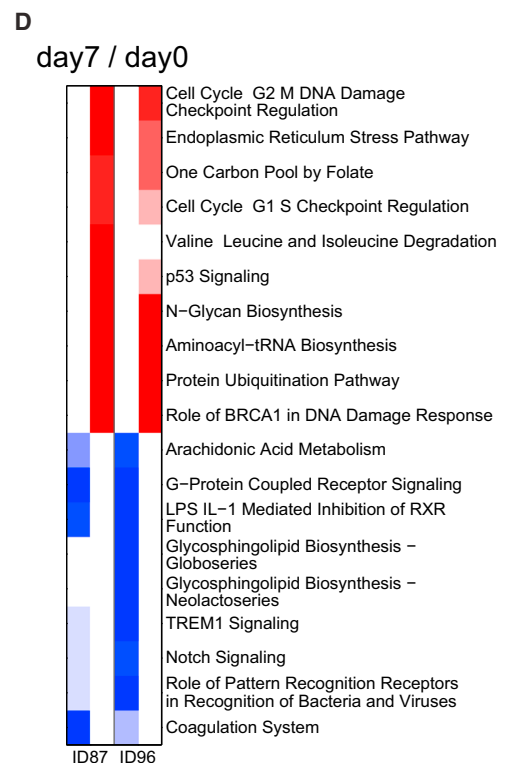
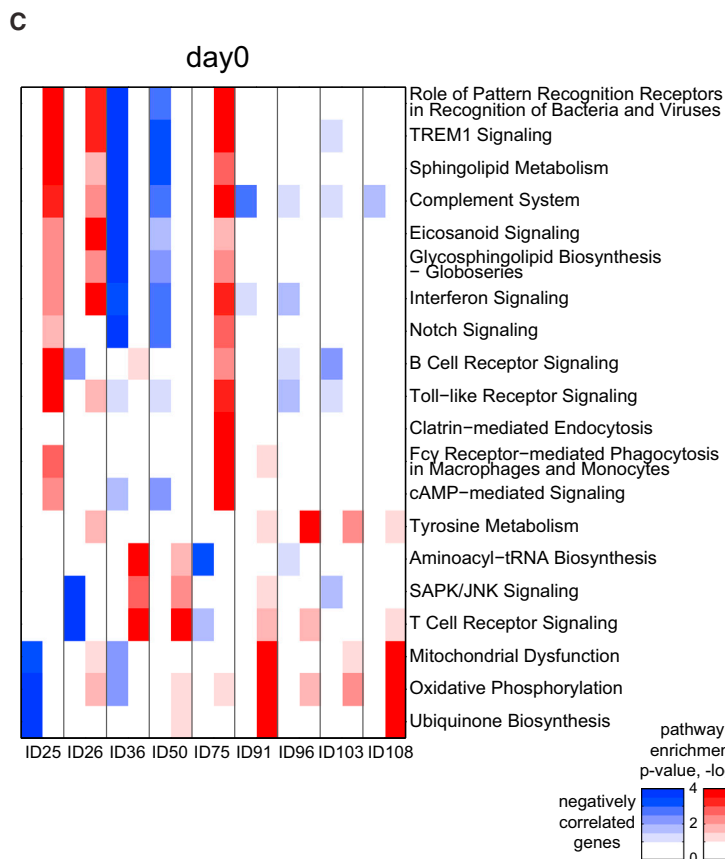
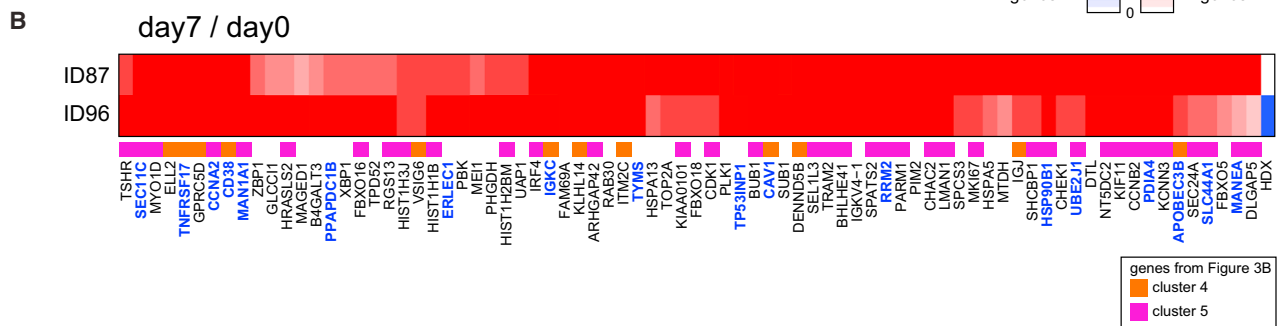
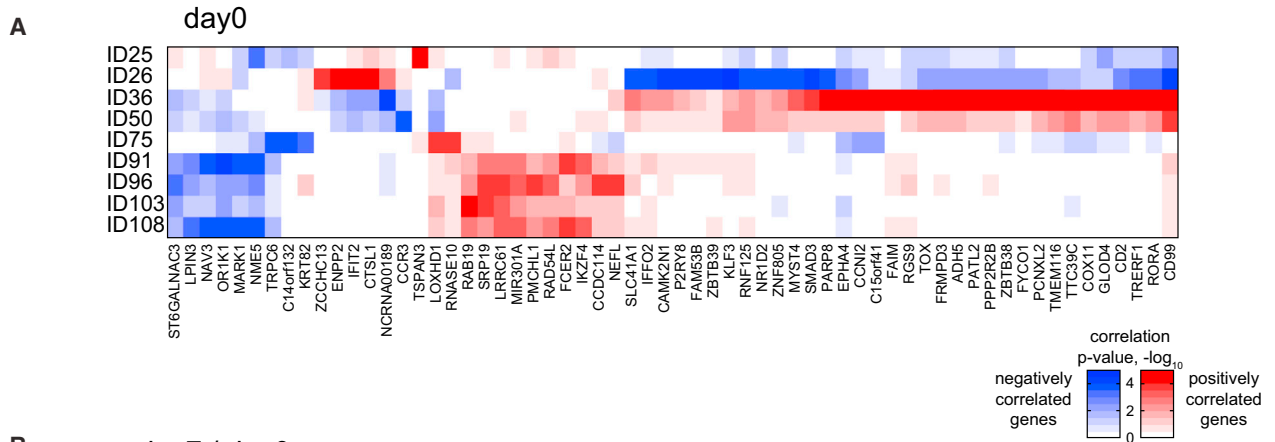
Using influenza vaccination as a model perturbation, we have gone beyond post hoc prediction and developed a framework that transforms multimodal data sets and natural population variation into predictive models that, unexpectedly, predicted serologic responses to influenza vaccination in our cohort using *baseline* PBMC subpopulation frequencies alone, independent of age, gender, initial serology, and the specificity of the cell populations for vaccine antigens. Notably, the essential cell subsets contributing to prediction were among the most stable within subjects over a time span of more than 70 days. Some of these also had the highest intersubject variation and thus were likely predictive because they delineated distinct baseline immune states among individuals. Our findings raise the possibility of using such measures as biomarkers of immune response potential in the clinic.

Despite substantial intersubject differences at baseline, we detected coherent changes in hundreds of transcripts postvaccination, as seen in recent vaccine studies (Bucasas et al., 2011; Gaucher et al., 2008; Nakaya et al., 2011; Querec et al., 2009). However, we also detected changes in the frequencies of cell populations, many of which have not been examined in the context of vaccination. Some of these changes reflect expected biology, with innate immune activation early and adaptive changes predominating by 1 week postvaccination. Nonetheless, the magnitude of responses was intimately linked to the prevaccination status of the individual, with lower initial titers being associated with larger fold increases in serum titers and plasmablast frequencies postvaccination. Although seemingly counterintuitive, such inverse correlations may reflect the current use of fold change to measure vaccine responses, differences in response timing (Henn et al., 2013), or inhibitory responses in pre-immune subjects (He et al., 2008). Thus, initial serology is an important yet potentially confounding variable that should be considered when exploring correlates and building predictive models.

A key question is whether some of our predictors captured the memory or activation status of the system due to prior exposure or whether they are indicative of the immune system's general propensity to respond. For example, the overall immune "threshold" for mounting a response may be determined by genetics and environmental factors such as the microbiota, contemporaneous stimulations by prevalent viruses such as CMV and EBV, or heterologous immunity due to degeneracy in T cell recognition (Münz et al., 2009; Welsh and Selin, 2002). Although it is possible that our observations were specific to the particular season or cohort, several lines of evidence suggest

Figure 6. Day 0 Predictors and Correlates

- (A) Robust correlates between day 0 gene expression and response titers (see Figure 5C).
 (B) Prediction performance using day 0 populations (as in Figure 5D).
 (C) Day 0 predictive cell populations analyzed as in Figure 5E.
 (D) Scatter plot of intrasubject stability (y) versus inter-subject variance (x) (see Methods). Predictive populations from (C) are highlighted in blue (more stable populations) and yellow (less stable).
 (E and F) Temporal profiles of two stable predictive cell populations for high (orange) and low (green) responders (based on adjMFC). Missing data points correspond to samples with low viability.
 (G) Predictive performance for day 0 using only the temporally stable populations (those above the 80% mark in Figure 6D).



(legend on next page)

that they have broader implications. First, the overall postvaccination responses of our cohort were consistent with those observed in other influenza vaccination cohorts from different seasons (Bucasas et al., 2011; Furman et al., 2013; Nakaya et al., 2011; Obermoser et al., 2013). Second, validation using independent measurements from other baseline time points (days –7 and 70) suggests that the essential predictive cell populations reflect *stable* baseline immune states as opposed to acute events. Indeed, initial analyses of four of our predictive B cell populations in healthy individuals from independent vaccine cohorts across multiple seasons (Obermoser et al., 2013) provides further support for their temporal stability (data not shown). Moreover, though the less stable predictive subsets (e.g., IFN α +mDC [ID75] and CD86+IgD+ memory B cells [ID105]) may reflect temporary environmental effects, removing them from our data only affected prediction of the initial titer-dependent endpoint (MFC), but not the baseline serology-independent adjMFC. Supporting the idea that B cell subsets can be useful response predictors, a correlation between prevaccination levels of class-switched memory B cells and serological responses to vaccination has been recently shown (Frasca et al., 2012). Another recent study has reported genetic determinants of immune cell population frequencies in humans (Orrù et al., 2013). Although B cell subset frequencies were not examined, it will be of interest to assess whether genetics regulates the frequencies of our temporally stable, predictive populations.

By associating changes in gene expression with changes in cell population frequencies, we have started to explore the nature of transcriptomic changes following vaccination. Changes in transcript abundance in PBMCs can originate from a combination of alterations in gene expression within individual cell populations and changes in cell population frequencies. Others have started to address these issues by sorting cell subsets followed by expression profiling, by correlating expression changes to known patterns of expression in major immune cell populations, or by computational deconvolution to infer cell-population-specific changes (Brandes et al., 2013; Nakaya et al., 2011; Shen-Orr et al., 2010). Here, we also found that changes in some cells were accompanied by transcript changes in the related cell populations. Tight correlations between plasmablast frequencies on day 7 and plasma cell-associated genes, such as immunoglobulins, ER stress, cell cycle, and N-glycan biosynthesis pathway genes, were involved. These plasmablast-correlated genes also included many that were previously associated with postvaccination response and predictive signatures (Nakaya et al., 2011; Querec et al., 2009). Thus, our results suggest that many of these transcriptomic changes can be accounted for by changes in plasmablast frequencies on day 7. Our observation that increases in plasmablasts were more prominent in subjects with low baseline titers, along with our integration of cell population, gene expression, and baseline immune status data, reveal that a sizable

component of the day 7 predictive transcriptome signature for MFC is likely derived from intersubject differences in prevaccination titers.

Unlike day 7 models, our baseline predictive models could only be built using cell population frequencies, suggesting that transcriptomic measurements of PBMCs were insufficiently sensitive to capture intersubject differences derived from the day 0 predictive populations, in contrast to the highly transcriptionally active plasmablasts. Whether expression data obtained from a simpler transcript pool, such as sorted B cell populations, provides better predictive power remains an important question. Correlating frequencies of cell populations with gene expression can also potentially reveal the internal state of that cell subset as well as that of other cell populations. Though most of our baseline predictive populations are adaptive, they are strongly correlated to genes and pathways with innate functions, some of which may reflect the (activation) states of functionally interacting cell populations. It will therefore be of interest to evaluate immune-activation differences (e.g., responses of PBMCs to TLR ligands) in high versus low responders to vaccination. Even on day 7, when the overlapping populations ID87 and ID96 were significantly correlated to largely the same genes and pathways, ID96 exhibited a stronger inverse association with several innate signaling pathways. These connections suggest distinct features for the CD20+ and CD20– populations of IgD-CD27+CD38+ B cells (Clutterbuck et al., 2006). Our results highlight the importance of evaluating and integrating multiple data sets for understanding human immune function.

Our data and analyses also serve as a potential resource for characterizing normal human immune variations and their implications for health and disease. One important goal of this study was to begin the building and analysis of the “normal human immunome” that characterizes the status of the apparently healthy immune system more broadly and fully than traditional assessments for future mining and reference. Such information would include clinical and family history, detailed cellular and molecular measurements of immune status, and microbiome status, as well as genome-wide genotypic information. Through the quantification of inter- as well as intraindividual variations, such a resource can provide essential baseline and comparative data for analyses of immune system behavior after perturbation and, through computational analyses, for de novo discovery of functional connections among molecular, cellular, and physiologic parameters of the human immune system in health or disease.

Our study highlights the importance of untangling correlation structures to identify the potential “root” correlates of responses as opposed to “passengers.” We have specifically shown that the strong effects of initial serology on titer outcome need to be properly modeled in order to identify whether and which baseline and response parameters contribute most to outcome. This may be particularly important in the context of influenza

Figure 7. Gene Expression and Pathway Enrichment Signatures of Predictive Cell Populations

Genes significantly correlated with predictive populations from (A) day 0 and (B) day 7. In (B), coherently changing genes from Figure 3 are highlighted by color, denoting the clusters from Figure 3B. Genes found to be predictive of or correlated with response titers in a previous influenza vaccination study (Nakaya et al., 2011) are marked in blue. Analyses were performed separately for positively and negatively correlated genes (in red and blue scales, respectively). Only genes significantly correlated with at least one population are shown (FDR < 0.01). Pathway enrichment analyses of gene correlates of (C) day 0 and (D) day 7 predictive cell populations (from Figure 6C).

vaccination, in which partial immunity is highly prevalent. Our adjMFC metric helped to reveal a number of temporally stable correlates not associated with MFC and thus may be useful in similar vaccine studies. One major constraint limiting predictive model construction and the discovery of novel functional relationships among parameters in a study like ours is statistical power, partially because the number of subjects, particularly the extreme responders, is far less than the thousands of parameters measured and the need to correct for multiple hypothesis testing. Meta-analyses across multiple studies will start to mitigate this issue but require the coordinated development and sharing of experimental and computational approaches in the research community to address issues ranging from controlling batch effects and standardizing quality control metrics to constructing predictive models. Achieving such integrated approaches across research centers will provide a major step forward in using computational approaches to understand the workings of the human immune system in health and disease.

EXPERIMENTAL PROCEDURES

For details and associated references, see [Extended Experimental Procedures](#).

Fasting blood samples were drawn between 8 and 11 AM from healthy volunteers on IRB-approved protocol (09-H-0239) and were processed within 30 min. PBMCs were cryopreserved or lysed and stored at -80°C for subsequent RNA extraction using miRNeasy (QIAGEN) and hybridization to HG 1.0 ST arrays (Affymetrix).

Frozen PBMCs from all time points for a subject were stained for five panels (Figure S1A) (Biancotto et al., 2011) using a common antibody mix on the same day. Cell frequencies were expressed as percent of parent population.

Virus-neutralizing titers (A/California/07/2009 [H1N1pdm09], H1N1 A/Brisbane/59/07, H3N2 A/Uruguay/716/07, and B/Brisbane/60/2001) were determined according to Hancock et al. (2009). Reported titers were the highest dilution that completely suppressed virus replication. Total and influenza-specific IgG/A frequencies of antibody-secreting cells were measured by ELISpot assays modified for frozen samples (Ho et al., 2011).

Microarray data were processed using APT (Affymetrix). Batch effects were removed using linear models in both raw and baseline-subtracted data. Frequencies of cell populations were transformed by \log_{10} (with zero values first set to 0.01). Samples and populations with low viability or cell numbers, respectively, were excluded.

Coherent changes from day 0 were computed using paired test in Limma (Smyth, 2004). Surrogate variable analysis (Leek and Storey, 2007) was used to remove hidden batch effects. The Wilcoxon test was used to compute coherent changes in cell population frequencies. False discovery rates were computed as described (Benjamini and Hochberg, 1995). Pathway enrichment analyses were performed using the geneSetTest function (Limma) and pathways annotated by Ingenuity (<http://www.ingenuity.com>).

To assess inter- and intrasubject variation, we used data from days -7 , 0, and 70 for each parameter and fitted an ANOVA model to obtain: (1) total R^2 , (2) the amount of variation explained by subject, and (3) the residual as an estimate of intrasubject variation.

AdjMFC and MFC are defined in the text. To take the maximum of individual viruses, we performed standardization by subtracting the median followed by dividing the maximum absolute deviation (MAD) within each virus. To compute adjMFC, we binned subjects based on their maximum baseline titers (across viruses) and then subtracted the median and divided by the MAD within each bin.

Contributions to titer response from age, gender, and ancestry were assessed using ANOVA. Effects were then removed from the cell population frequency and microarray data by retaining the residual after fitting a linear model for each probe and subset frequency. Cross-validation-based predictive modeling was done as described in the text and [Extended Experimental Procedures](#). We used diagonal linear discriminant analysis for cell frequency data and

when cell frequency and pathway status were combined. For transcript data alone, we used partial least square (for data dimension reduction due to the large number of genes) followed by linear discriminant analysis (PLS-LDA).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, five figures, four tables, and a complete author list and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.03.031>.

CONSORTIA

The members of the Baylor HIPC Center are Gerlinde Obermoser, Damien Chaussabel, and Karolina Palucka. The members of the CHI Consortium are Jinguo Chen, J. Christopher Fuchs, Jason Ho, Surender Khurana, Lisa R. King, Marc Langweiler, Hui Liu, Jody Manischewitz, Zoltan Pos, Jacqueline G. Posada, Paula Schum, Rongye Shi, Janet Valdez, Wei Wang, Huizhi Zhou, Daniel L Kastner, Francesco M. Marincola, J. Philip McCoy, Giorgio Trinchieri, and Neal S. Young.

AUTHOR CONTRIBUTIONS

Y.K., A.B., Z.X., R.N.G., E.W., and M.J.O. contributed equally to this work. M.N., H.G., S.M., H.B.D., S.P., and F.C. contributed equally to this work. J.S.T. directed, designed, and performed data analyses, interpreted results, and wrote the manuscript. P.L.S. designed the study, helped design experiments and analyses, interpreted results, and wrote the manuscript. Y.K., Z.X., M.N., and F.C. designed and performed computational analyses and helped prepare the manuscript. A.B. and E.W. designed and performed experiments, analyzed data, and helped prepare the manuscript. R.N.G. designed the study, helped design experiments and analyses, and helped prepare the manuscript. M.J.O. helped design, organize, and execute the clinical study. H.G. and S.M. designed experiments and analyzed data. H.B.D. helped design and organize the clinical study. S.P. helped organize and execute the clinical study.

ACKNOWLEDGMENTS

We thank M. Connors, K. Subbarao, R. Seder, M. Roederer, D. Douek, J. Benjink, I. Fraser, J. Mandl, members of the Systems Genomics and Bioinformatics Unit (NIAID), and M. Eichelberger (FDA) for helpful advice and feedback; the staff of the NIH Clinical Center for outstanding clinical support; and NIAID's high-performance computing team for outstanding support. This work is funded by the Intramural Programs of NIAID, NCI, NHLBI, NIAMS, NICHD, NIDDK, NINDS, NIEHS, NEI, and NHGRI.

Received: May 24, 2013

Revised: December 6, 2013

Accepted: March 24, 2014

Published: April 10, 2014

REFERENCES

- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J.R. Statist. Soc. B* 57, 289–300.
- Biancotto, A., Fuchs, J.C., Williams, A., Dagur, P.K., and McCoy, J.P., Jr. (2011). High dimensional flow cytometry for comprehensive leukocyte immunophenotyping (CLIP) in translational research. *J. Immunol. Methods* 363, 245–261.
- Brandes, M., Klauschen, F., Kuchen, S., and Germain, R.N. (2013). A top-down systems analysis identifies an innate feed-forward inflammatory circuit leading to lethal influenza infection. *Cell* 154, 197–212.
- Bucacas, K.L., Franco, L.M., Shaw, C.A., Bray, M.S., Wells, J.M., Niño, D., Arden, N., Quarles, J.M., Couch, R.B., and Belmont, J.W. (2011). Early patterns of gene expression correlate with the humoral immune response to influenza vaccination in humans. *J. Infect. Dis.* 203, 921–929.

- Chuang, H.Y., Hofree, M., and Ideker, T. (2010). A decade of systems biology. *Annu. Rev. Cell Dev. Biol.* 26, 721–744.
- Clutterbuck, E.A., Salt, P., Oh, S., Marchant, A., Beverley, P., and Pollard, A.J. (2006). The kinetics and phenotype of the human B-cell response following immunization with a heptavalent pneumococcal-CRM conjugate vaccine. *Immunology* 119, 328–337.
- Cox, R.J., Brokstad, K.A., Zuckerman, M.A., Wood, J.M., Haaheim, L.R., and Oxford, J.S. (1994). An early humoral immune response in peripheral blood following parenteral inactivated influenza vaccination. *Vaccine* 12, 993–999.
- Davis, M.M. (2008). A prescription for human immunology. *Immunity* 29, 835–838.
- de Jong, J.C., Palache, A.M., Beyer, W.E., Rimmelzwaan, G.F., Boon, A.C., and Osterhaus, A.D. (2003). Haemagglutination-inhibiting antibody to influenza virus. *Dev. Biol. (Basel)* 115, 63–73.
- Fiore, A.E., Shay, D.K., Broder, K., Iskander, J.K., Uyeki, T.M., Mootrey, G., Bresee, J.S., and Cox, N.J.; Centers for Disease Control and Prevention (2009). Prevention and control of seasonal influenza with vaccines: recommendations of the Advisory Committee on Immunization Practices (ACIP), 2009. *MMWR Recomm. Rep.* 58, 1–52.
- Fonteneau, J.F., Gilliet, M., Larsson, M., Dasilva, I., Münz, C., Liu, Y.J., and Bhardwaj, N. (2003). Activation of influenza virus-specific CD4+ and CD8+ T cells: a new role for plasmacytoid dendritic cells in adaptive immunity. *Blood* 101, 3520–3526.
- Frasca, D., Diaz, A., Romero, M., Phillips, M., Mendez, N.V., Landin, A.M., and Blomberg, B.B. (2012). Unique biomarkers for B-cell function predict the serum response to pandemic H1N1 influenza vaccine. *Int. Immunol.* 24, 175–182.
- Furman, D., Jovic, V., Kidd, B., Shen-Orr, S., Price, J., Jarrell, J., Tse, T., Huang, H., Lund, P., Maecker, H.T., et al. (2013). Apoptosis and other immune biomarkers predict influenza vaccine responsiveness. *Mol. Syst. Biol.* 9, 659.
- Gass, J.N., Gunn, K.E., Sriburi, R., and Brewer, J.W. (2004). Stressed-out B cells? Plasma-cell differentiation and the unfolded protein response. *Trends Immunol.* 25, 17–24.
- Gaucher, D., Therrien, R., Kettaf, N., Angermann, B.R., Boucher, G., Filali-Mouhim, A., Moser, J.M., Mehta, R.S., Drake, D.R., 3rd, Castro, E., et al. (2008). Yellow fever vaccine induces integrated multilineage and polyfunctional immune responses. *J. Exp. Med.* 205, 3119–3131.
- Germain, R.N., and Schwartzberg, P.L. (2011). The human condition: an immunological perspective. *Nat. Immunol.* 12, 369–372.
- Gerriets, V.A., and Rathmell, J.C. (2012). Metabolic pathways in T cell fate and function. *Trends Immunol.* 33, 168–173.
- Goodwin, K., Viboud, C., and Simonsen, L. (2006). Antibody response to influenza vaccination in the elderly: a quantitative review. *Vaccine* 24, 1159–1169.
- Gross, P.A., Hermogenes, A.W., Sacks, H.S., Lau, J., and Levandowski, R.A. (1995). The efficacy of influenza vaccine in elderly persons. A meta-analysis and review of the literature. *Ann. Intern. Med.* 123, 518–527.
- Hancock, K., Veguilla, V., Lu, X., Zhong, W., Butler, E.N., Sun, H., Liu, F., Dong, L., DeVos, J.R., Gargiullo, P.M., et al. (2009). Cross-reactive antibody responses to the 2009 pandemic H1N1 influenza virus. *N. Engl. J. Med.* 361, 1945–1952.
- He, X.S., Holmes, T.H., Sasaki, S., Jaimes, M.C., Kemble, G.W., Dekker, C.L., Arvin, A.M., and Greenberg, H.B. (2008). Baseline levels of influenza-specific CD4 memory T-cells affect T-cell responses to influenza vaccines. *PLoS ONE* 3, e2574.
- Henn, A.D., Wu, S., Qiu, X., Ruda, M., Stover, M., Yang, H., Liu, Z., Welle, S.L., Holden-Wiltse, J., Wu, H., et al. (2013). High-resolution temporal response patterns to influenza vaccine reveal a distinct human plasma cell gene signature. *Sci. Rep.* 3, 2327.
- Ho, J., Moir, S., Wang, W., Posada, J.G., Gu, W., Rehman, M.T., Dewar, R., Kovacs, C., Sneller, M.C., Chun, T.W., et al. (2011). Enhancing effects of adjuvanted 2009 pandemic H1N1 influenza A vaccine on memory B-cell responses in HIV-infected individuals. *AIDS* 25, 295–302.
- Iwakoshi, N.N., Lee, A.H., and Glimcher, L.H. (2003). The X-box binding protein-1 transcription factor is required for plasma cell differentiation and the unfolded protein response. *Immunol. Rev.* 194, 29–38.
- Kaminski, D.A., Wei, C., Qian, Y., Rosenberg, A.F., and Sanz, I. (2012). Advances in human B cell phenotypic profiling. *Front. Immunol.* 3, 302.
- Kitano, H. (2002). Computational systems biology. *Nature* 420, 206–210.
- Leek, J.T., and Storey, J.D. (2007). Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* 3, 1724–1735.
- Li, G.M., Chiu, C., Wrarmert, J., McCausland, M., Andrews, S.F., Zheng, N.Y., Lee, J.H., Huang, M., Qu, X., Edupuganti, S., et al. (2012). Pandemic H1N1 influenza vaccine induces a recall response in humans that favors broadly cross-reactive memory B cells. *Proc. Natl. Acad. Sci. USA* 109, 9047–9052.
- Münz, C., Lünemann, J.D., Getts, M.T., and Miller, S.D. (2009). Antiviral immune responses: triggers of or triggered by autoimmunity? *Nat. Rev. Immunol.* 9, 246–258.
- Nakaya, H.I., Wrarmert, J., Lee, E.K., Racioppi, L., Marie-Kunze, S., Haining, W.N., Means, A.R., Kasturi, S.P., Khan, N., Li, G.M., et al. (2011). Systems biology of vaccination for seasonal influenza in humans. *Nat. Immunol.* 12, 786–795.
- Nauta, J. (2011). *Statistics in clinical vaccine trials* (Heidelberg, Germany: Springer).
- Obermoser, G., Presnell, S., Domico, K., Xu, H., Wang, Y., Anguiano, E., Thompson-Snipes, L., Ranganathan, R., Zeitner, B., Bjork, A., et al. (2013). Systems scale interactive exploration reveals quantitative and qualitative differences in response to influenza and pneumococcal vaccines. *Immunity* 38, 831–844.
- Orrù, V., Steri, M., Sole, G., Sidore, C., Virdis, F., Dei, M., Lai, S., Zoledziewska, M., Busonero, F., Mulas, A., et al. (2013). Genetic variants regulating immune cell levels in health and disease. *Cell* 155, 242–256.
- Pulendran, B., and Ahmed, R. (2011). Immunological mechanisms of vaccination. *Nat. Immunol.* 12, 509–517.
- Pulendran, B., Li, S., and Nakaya, H.I. (2010). Systems vaccinology. *Immunity* 33, 516–529.
- Querec, T.D., Akondy, R.S., Lee, E.K., Cao, W., Nakaya, H.I., Teuwen, D., Pirani, A., Gernert, K., Deng, J., Marzolf, B., et al. (2009). Systems biology approach predicts immunogenicity of the yellow fever vaccine in humans. *Nat. Immunol.* 10, 116–125.
- Sasaki, S., He, X.S., Holmes, T.H., Dekker, C.L., Kemble, G.W., Arvin, A.M., and Greenberg, H.B. (2008). Influence of prior influenza vaccination on antibody and B-cell responses. *PLoS ONE* 3, e2975.
- Schadt, E.E. (2009). Molecular networks as sensors and drivers of common human diseases. *Nature* 461, 218–223.
- Shen-Orr, S.S., Tibshirani, R., Khatri, P., Bodian, D.L., Staedtler, F., Perry, N.M., Hastie, T., Sarwal, M.M., Davis, M.M., and Butte, A.J. (2010). Cell type-specific gene expression differences in complex tissues. *Nat. Methods* 7, 287–289.
- Shi, Y., Agematsu, K., Ochs, H.D., and Sugane, K. (2003). Functional analysis of human memory B-cell subpopulations: IgD+CD27+ B cells are crucial in secondary immune response by producing high affinity IgM. *Clin. Immunol.* 108, 128–137.
- Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* 3, Article3.
- Tulp, A., Barnhoorn, M., Bause, E., and Ploegh, H. (1986). Inhibition of N-linked oligosaccharide trimming mannosidases blocks human B cell development. *EMBO J.* 5, 1783–1790.
- van Duin, D., Allore, H.G., Mohanty, S., Ginter, S., Newman, F.K., Belshe, R.B., Medzhitov, R., and Shaw, A.C. (2007). Prevacine determination of the expression of costimulatory B7 molecules in activated monocytes predicts influenza vaccine responses in young and older adults. *J. Infect. Dis.* 195, 1590–1597.
- Welsh, R.M., and Selin, L.K. (2002). No one is naive: the significance of heterologous T-cell immunity. *Nat. Rev. Immunol.* 2, 417–426.
- Withers, D.R., Fiorini, C., Fischer, R.T., Ettinger, R., Lipsky, P.E., and Grammer, A.C. (2007). T cell-dependent survival of CD20+ and CD20- plasma cells in human secondary lymphoid tissue. *Blood* 109, 4856–4864.