Supplementary Information

Materials and Methods

Human subjects and influenza vaccines

This study is comprised of subjects vaccinated with TIV during the 3 consecutive influenza seasons in, 2008-09 (Trial 1, 28 subjects), 2007-08 (Trial 2, 9 subjects) and 2009-10 (Trail 3, 30 subjects). One trial comprised of subjects immunized with LAIV during 2008-09 influenza annual season was also included in this study. Young healthy adults, 18 to 50 years old, were vaccinated with one dose of either TIV (Fluarix®, GlaxoSmithKline Biologicals, in 2007-08 and 2008-09 seasons and Fluvirin®, Novartis Vaccines and Diagnostics Limited, in 2009-10 season) or LAIV (FluMist, MedImmune), following current guidelines for influenza vaccination. **Supplementary Table 10** shows the influenza virus strains contained in the vaccines. Written informed consent was obtained from each subject with institutional review and approval from the Emory University Institutional Review Board.

Cell and plasma isolation

Four blood samples were obtained from influenza vaccinees: at baseline, and at days 3, 7 and 30 (range, day 28 to 32) post-vaccination. Peripheral blood mononuclear cells (PBMC) and plasma were isolated from fresh blood (CPTs; Vacutainer® with Sodium Citrate; BD), following the manufacturer's protocol. PBMCs were frozen in DMSO with 10% FBS and stored at –80 ^{o}C and then transferred on the next day to liquid nitrogen freezers (–210 ^{o}C). Plasma samples from CPTs were stored at –80 ^{o}C. Trizol (Invitrogen) was used to lyse fresh PBMCs (1 ml of Trizol to ~1.5 x 10^{6} cells) and to protect RNA from degradation. Trizol samples were stored at –80 ^{o}C.

Hemagglutination inhibition assays

Hemagglutination inhibition (HAI) titers were determined based on the standard WHO protocol, as previously described[1]. Briefly, plasma samples were treated with receptor destroying enzyme (RDE; Denka Seiken Co.) by adding of 1 part plasma to 3 parts RDE and incubating at 37 ^{o}C

Nature Immunology doi:10.1038/ni.2067

overnight. The following morning, the RDE was inactivated by incubating the samples at 56 ^{o}C for one hour. The samples were then serially diluted with PBS in 96 well v-bottom plates (Nunc, Rochester, NY) and 4 HA units of either the H1N1, H3N2, or influenza B virus was added to each well. After 30 minutes at room temperature, 50 ul of 0.5% turkey RBCs (Rockland Immunochemicals) suspended in PBS with 0.5% BSA was added to each well and the plates were shaken manually. After an additional 30 minutes at room temperature, the plasma titers were read as the reciprocal of the final dilution for which a button was observed. Negative and positive control plasma samples for each virus were used for reference (data not shown).

B cell ELISPOT experiments

Direct ELISPOT to enumerate both total and influenza specific IgG, IgM and IgA plasmablasts present in fresh PBMC samples were essentially done as previously described[2,3]. Briefly, 96-well ELISPOT filter plates (Millipore, MAHA N4510) were coated overnight with either the influenza vaccine (same vaccine as used for donor vaccination) at a dilution of 1/20 in PBS or with goat anti-human Ig (Caltag). Plates were washed and blocked by incubation with complete RPMI containing 10% FCS at 37°C for 2 hrs. Purified and extensively washed PBMCs were added to the plates in dilution series and incubated for 6 hrs or overnight. Plates were washed with PBS followed by PBS containing 0.05% Tween and then incubated with a biotinylated anti-huIgG, anti-huIgM or anti-huIgA antibody (Caltag) and incubated for 1.5 hrs at room temperature. After washing, the plates were incubated with an avidin-D-HRP conjugate (Vector Laboratories) and finally developed using AEC substrate (3 amino-9 thylcarbozole, Sigma). Developed plates were scanned and analyzed using an automated ELISPOT counter (Cellular Technologies Ltd.).

Flow cytometry analysis of B cells

Flow cytometry analysis was performed on whole blood, as previously described[2]. Briefly, 300-400 µl blood was incubated with the appropriate antibodies for 30 minutes at room temperature. Red blood cells were then lysed by incubation with FACS lysing Solution (Beckton Dickinson)

for twice for 4 minutes at RT prior to analysis. Antibodies against CD3, CD20, CD38 and CD19 were purchased from Pharmingen, while anti-CD27 that was purchased from eBiosciences. ASCs were gated and isolated as $CD19^{+}CD3^{-}CD20^{lo/-}CD27^{high}CD38^{high}$ cells. Flow cytometry data was analyzed using FlowJo software.

Multiplex analysis

Plasma samples from CPTs were stored at -80°C prior to cytokine analysis. Luminex assays were performed with the Beadlyte Human 10-Plex Multi-Cytokine Detection System (Upstate). Plasma samples were run in duplicate following the manufacturer's protocol on a Bio-Plex Luminex-100 station (Bio-Rad). Data were normalized using the cytokine level at baseline (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 by t-test ($p < 0.05$).

Microarray experiments

Total RNA from fresh PBMCs (~1.5 x 10^6 cells) was purified using Trizol® (Invitrogen, Life Technologies Corporation) according to the manufacturer's instructions. All RNA samples were checked for purity using a ND-1000 spectrophotometer (NanoDrop Technologies) and for integrity by electrophoresis on a 2100 BioAnalyzer (Agilent Technologies). Two-round *in vitro* transcription amplification and labeling was performed starting with 50 ng intact, total RNA per sample, following the Affymetrix protocol. After hybridization on Human U133 Plus 2.0 Arrays (using GeneTitan platform, Affymetrix, or individual cartridges) for 16 h at 45 ^{o}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. CEL files from outlier samples were excluded and the remaining CEL files of all the samples belonging to the same influenza season trial were

grouped, and the microarray intensity data of probe sets (hereafter referred to "genes") were normalized by RMA, which includes global background adjustment and quantile normalization.

Identification of differentially expressed genes induced by vaccination

Gene expression fold-change of each subject was obtained by subtracting the day 3 or 7 $\log_2$ expression values by its corresponding baseline value. Genes were filtered out if no increase/decrease higher than 25% (1.25-fold) was observed on at least 20% of the vaccinees on day 3 or day 7 compared to baseline. Three independent statistical tests were applied to the remaining genes: One-way analysis of variance (ANOVA), Signal-to-noise (S2N) statistic and significance analysis of microarrays (SAM). Only genes identified by all 3 analyses were considered differentially expressed. Pathway and network analyses were performed using the Ingenuity Pathways Analysis (Ingenuity Systems) software.

One-way ANOVA was used to test for differences on the expression levels of each gene among days 0, 3 and 7, as previously described[4]. *P*-values were calculated for each gene over the time course 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.

S2N statistic[5] was independently used to identify genes differentially expressed between day 3 and day 7 post-vaccination compared to baseline. Signal-to-noise ratio was calculated for each gene (g) using the formula $S2N = [\mu_d(g) - \mu_0(g)] / [\sigma_d(g) + \sigma_0(g)]$, where μ_d and σ_d, are, respectively the mean and standard deviation of $\log_2$ expression values of all vaccinees on day *d*, and μ_0 and σ_0 are, respectively the mean and standard deviation of $\log_2$ expression values of all vaccinees at baseline. Measured S2N ratios were compared to calculated S2N values obtained in 5,000 data sets where a given class label was randomly permuted. For these comparisons, *p*-values represent the frequency at which the S2N statistic from randomly permuted data exceeded the measured S2N statistic[6]. A gene was considered differentially expressed if the *p*-value were lower than 0.05.

Nature Immunology doi:10.1038/ni.2067

Finally, SAM[7] method was applied to identify genes differentially expressed between day 3 and day 7 post-vaccination compared to baseline. The calculations were performed on the $\log_2$-fold change in gene expression for days 3 and 7 post-vaccination versus day 0, separately (SAM one class, 5,000 permutations). A FDR cutoff of 0.05 and local FDR of 0.10 were used to identify the differentially expressed genes.

Meta-analysis of immune cells

A meta-analysis was performed using publicly available microarray data of major types of PBMCs, i.e. T cells, natural killer cells, B cells and monocytes. We added to this, the microarray data of plasmacytoid dendritic cells (pDC) and myeloid dendritic cells (mDC) obtained from PBMCs of vaccinees (described in "Microarray of FACS-sorted immune cell subsets" section). Raw microarray CEL files were downloaded from NCBI GEO website (http://www.ncbi.nlm.nih.gov/sites/geo) or ArrayExpress website (http://www.ebi.ac.uk/arrayexpress/), processed by MAS5 normalization, and log2 transformed. The list of all GEO or ArrayExpress studies and all samples used in this meta-analysis can be found in **Supplementary Table 3**. Purification strategy, RNA processing method, and hybridization strategy can be found in the original publications. All samples were hybridized to Affymetrix Human Genome U133 Plus 2.0 Arrays or Affymetrix Human Genome U133A Arrays. Only probe sets presented in both platforms (i.e. belonging to Affymetrix Human Genome U133A Arrays) were used in the analyses.

Annotation of each sample was obtained from SOFT files and manually curated into defined cell subsets. Samples were removed from the analyses based on the severity of the disease or treatment and/or on the method of cell purification. Student's t-test ($p < 0.05$ and mean fold-change > 2) was used to identify differentially expressed genes between any given 2 cell subsets. In the PBMC meta-analysis, genes were classified as "highly expressed in subset X" if they were up-regulated in this subset X compared to at least 4 other subsets (out of 5 possible comparisons). Using this approach, we classified 9,397 genes as highly expressed in a specific PBMC subset (**Supplementary Table 4**). In the B cell subset meta-analysis, genes were classified as "highly expressed in B cell subset X" if they were: (1) "highly expressed in B cell

subset" in the PBMC analysis and (2) up-regulated in this B cell subset X compared to at least 2 other B cell subsets (out of 3 possible comparisons). Using this approach, we classified 509 genes as highly expressed in a specific B cell subset (**Supplementary Table 4**).

Next, we compared the set of highly expressed genes on each cell subset to the genes differentially expressed after TIV or LAIV vaccination and to the genes whose expression correlates with B cell responses. Fold enrichment was calculated using the formula ('common ZX'/'deg Z')/('subset X'/'total'), where 'common ZX' = number of genes differentially expressed or correlated to condition Z (TIV or LAIV vaccination or B cell responses) and also highly expressed in cell subset X; 'deg Z' = total number of genes differentially expressed or correlated to condition Z; 'subset X' = total number of genes highly expressed in cell subset X and 'total' = total number of genes in the chip common to both platforms (i.e. belonging to Affymetrix Human Genome U133A Arrays).

Identification of gene signatures that correlate to B cell responses

Genes, whose expression correlated with the magnitude of the B cell responses were identified using 2 criteria: Pearson correlation *P*-value < 0.05 and log2 fold-change between day X (3 or 7) and day 0 > 0.322 (25% increase) or < -0.322 (25% decrease) in at least 8 vaccinees (>25% of the subjects). B cells responses were defined as the log10 number of IgG-secreting cells at day 7 (ELISPOT) and the log2 fold-change of antibody titers at day 28 compared to day 0 (HAI assays).

DAMIP method

Discriminant Analysis via Mixed Integer Programming (DAMIP)[8] was used to identify the signatures that predict the antibody response to influenza vaccination. In our previous work[4], we described in detail how DAMIP can be used to identify signatures that predict the immunogenicity of the YF-17D vaccine. Three different analyses were performed by DAMIP. First, we ran DAMIP using two microarray data sets, comprised of subjects vaccinated with TIV during the 2008-09 influenza season (2008 trial) and during the 2007-08 influenza season (2007

Nature Immunology doi:10.1038/ni.2067

trial). On each data set, vaccinees were classified into two groups: "very high HAI responders" (HAI response 8-fold or higher) and "very low HAI responders" (HAI response 2-fold or lower). About 3,000+ total measurable gene attributes of mixed time stamps (i.e., genes in 2008 and 2007 trials whose expression at d3/d0 or d7/d0 correlated with HAI response) were used as input to the program. DAMIP then performed experiments consisting of the following two parts: a) Perform feature selection and establish classification rules (called here as "DAMIP gene signatures") using a training dataset (2008 Trial), b) Use the rule developed from the training set to blind predict the group status of independent unknown vaccinees (from 2007 TIV Trial). To ensure global solution, a combinatorial feature selection algorithm was developed. Using this approach, DAMIP model identified 271 sets of genes, each containing 2 to 4 genes, that result in > 80% 10-fold cross-validation accuracy, as well as > 80% blind prediction accuracy of the vaccinees, as low or high antibody responders on both trials (**Supplementary Table 6**). 12 of these sets of gene signatures offer up to 90% accuracy.

A second analysis was next performed by adding a third influenza trial, comprised of subjects vaccinated with TIV during 2009-10 influenza season (2009 trial). However, on this trial we used only 44 genes selected from the DAMIP gene signatures generated in the first analysis. Also, instead of microarrays, expression level of these 44 genes was determined by real-time RT-PCR. To avoid the identification of over-trained rules, DAMIP program was re-run using the 2008 trial as training set and the 2007 and 2009 trials as blind predictive sets. This approach identified 47 sets of genes that can correctly classify > 80% of the vaccinees, as very low or very high antibody responders in any of the 3 trials (**Supplementary Table 8**).

In the third analysis, we changed the cutoff used to define high and low responders. A vaccinee with 4-fold or higher increase in the HAI titers post-vaccination were classified as "high" responders and those with 2-fold or lower increase were classified as "low" responders. Using 2008 trial as a training set and 2007 and 2009 trials as blind predictive sets, DAMIP model was able to generate 42 gene signatures comprised of 3 to 4 discriminatory genes, with an unbiased estimate of correct classification higher than 80%, as determined by 10-fold cross-validation and 10-fold blind prediction (**Supplementary Table 9**).

Performance and validation of the rules is reported in: a) 10-fold cross validation which reports the unbiased estimate of classification correctness in the training stage, and b) in 1-fold

and 10-fold blind prediction of the independent Trial vaccinees which report the prediction accuracy of new and unknown data. While 10-fold cross validation offers the confidence interval and reliability of the rules generated and tested within the same Trial of subjects, the blind prediction offers a further measurement of its accuracy in predictive power, and practical usage across different independent Trials.

10-fold cross-validation: To obtain an unbiased estimate of the reliability and quality of the derived classification rules, 10-fold cross validation is performed. In the 10-fold cross validation procedure, the training set is randomly partitioned into ten subsets of roughly equal size. Ten computational experiments are then run, each of which involves a distinct training set made up of 9 of the 10 subsets and a test set made up of the remaining subset. The classification rule obtained via a given training set is applied to each point in the associated test set to determine to which group the rule allocates it. The process is repeated until each subset has been used once for testing. The cumulative measure of correct classification of the 10 experiments provides the unbiased estimation of correct classification.

1-fold blind predictions: In 1-fold blind prediction, a classification rule is first developed using all the training data using the selected set of discriminatory gene signatures. This rule is then applied to each vaccinee in the blind data to predict its group status. The percent of correct prediction of the blind data vaccinees is recorded, providing a measure of overall prediction accuracy.

10-fold blind predictions: In 10-fold blind predictions, we generate 10 classification rules as in 10-fold cross-validation – each rule is generated using nine of the ten subsets of the training sets. Then, the blind data (all of them) are tested on this rule. This process is repeated ten times, and the average cumulative prediction forms the unbiased prediction correctness of the blind data.

To develop the classification rule from a given training set, the training data is fed into the DAMIP model. The feature selection algorithm inside the model will then determine, out of the large set of gene measurements, a subset of genes – a discriminatory signature -- that may help to classify entities in the training set into the two groups. The classification rate associated with the signature set (obtained by performing 10-fold cross-validation using the selected

signature features) is then recorded. This "learning" process is repeated, each time an updated discriminatory signature set and associated classification rate are obtained and recoded. Users can pre-set the number of discriminatory gene measurements that are desired in each signature set. Since the number of vaccinees in each clinical Trial is rather small, we set each signature set to contain at most 4 gene attributes. Users can also pre-set an appropriate target value for the classification rate. Thus the machine will continue to learn (generate signatures sets and associated classification rates) and terminate when the target classification rate is achieved, or when it reaches a level of correct classification and cannot improve any further (in this case, it may not have achieved the pre-set target rate). In our study, the learning process was terminated when the resulting classification rate reached 80%. Developing a classification rule is computationally expensive due to the combinatorial nature of the feature selection process. However, once a rule has been obtained, it is easy and inexpensive to apply it to new unknown entities to predict group membership.

To perform a blind test, we simply input each vaccinee from an independent Trial and process it through the classification rule (obtained from a training set). This takes less than a second of CPU time.

In Brooks and Lee[9], it was proved the classification rule resulted from DAMIP is strongly universally consistent. It consistently results in low inter-group misclassification rates; it is insensitive to the specification of prior probabilities, yet capable of reducing misclassification rates when the number of training observations from each group is different. Further, the DAMIP rule is proved to be stable regardless of the proportion of training observations from each group.

Analysis of immune responses in *CamkIV*-/- mice

8-12 week old C57BL6 mice were purchased from Jackson Labs (Maine, VM). *CamkIV* -/- mice were generated as described[10] and backcrossed with C57BL6/J mice for more than 15 generations, in the animal facility at Duke University. Mice were immunized in the right and the left hamstring muscles with a 1:5 diluted human Fluvirin® (Novartis Vaccines and Diagnostics, Cambridge, MA) vaccine dose (3ug of HA of each strain included in the vaccine). Mice were bled at time points indicated for analysis of influenza-specific antibody responses. All animal

procedures were performed in accordance with guidelines established by IACUC at Duke University.

Antibody ELISA assays

96 well Nunc maxisorp plates (Nunc, Denmark) were coated with 100µl of 2µg/ml of Fluvirin 2009 or 2010 (Novartis Vaccines and Diagnostics, Cambridge, MA) overnight at 4^0C. Plates were washed 3 times with PBS/0.5% Tween 20 using a Biotek auto plate washer and blocked with 200µl of 4% non fat dry milk (Biorad, Hercules, CA) for 2 hours at room temperature. Serum samples from immunized mice at the indicated time points were diluted in 0.1% non-fat dry milk in PBS/0.5% Tween 20 (1:100) and incubated on blocked plates for 2 hours at room temperature. Detection antibodies were purchased from Southern Biotech (Birmingham, AL). The plates were washed 5 times post sample sera incubation, and anti-mouse IgG2c-horse radish peroxidase (HRP) conjugate (1:2000) and anti-mouse IgG1-HRP conjugate at (1:5000) in PBS/0.5% Tween 20, were added and plates were incubated for 2 hours at room temperature. Plates were washed 7 times with PBS/0.5% Tween 20, developed using 100µl per well of tetramethylbenzidine (TMB) substrate (BD Biosciences, San Diego, CA) and stopped using 2N H_2SO_4. Plates were analyzed using a BioRad plate reading spectrophotometer at 450nm with correction at 595nm. Results are represented as O.D values at 450nm at the indicated time points.

Microarray of FACS-sorted immune cell subsets

Frozen PBMCs from 6 TIV and 6 LAIV vaccinees at day 0 and day 7 post-immunization (24 samples) were thaw, washed with FACS buffer (PBS with 5% FBS) and counted. First, the following antibodies were used to separate ~2 x 10^6 PBMCs into a "DC" fraction (cell enriched by negative selection) and a "B cell and Monocyte" fraction (cell enriched by positive selection), accordingly to manufacturer's instructions (Miltenyi Biotech, Auburn, CA): biotin anti-CD56 (clone;B159, BD Biosciences), biotin anti-CD3 (clone;UCHT1, BD Biosciences), biotin anti-CD16 (clone;3G8, BD Biosciences), anti-CD19 MicroBeads (Miltenyi Biotech, Auburn, CA),

anti-CD14 MicroBeads (Miltenyi Biotech, Auburn, CA), anti-biotin MicroBeads (Miltenyi Biotech, Auburn, CA). Cells from "DC" fraction were washed 2x with FACS buffer (PBS with 5% FBS) and stained with FITC labeled lineage cocktail 1 (clones;SK7, 3G8, SJ25C1, L27, MφP9, NCAM16.2, BD Biosciences), PE labeled anti-CD123 (clone;9F5, BD Biosciences), PerCP labeled anti-HLA-DR (clone;L243(G46-6), BD Biosciences) and APC labeled anti-CD11c (clone;S-HCL-3, BD Biosciences). Cells from "B cell and Monocyte" fraction were washed 2x with FACS buffer (PBS with 5% FBS) and stained with PerCP labeled anti-HLA-DR (clone;L243(G46-6), BD Biosciences), FITC labeled anti-CD14 (clone; MφP9, BD Biosciences) and APC labeled anti-CD19 (clone;SJ25C1, BD Biosciences). Next, myeloid dendritic cells (lin-1$^-$ HLA-DR$^+$ CD11C^{high} CD123^{low}) and plasmacytoid dendritic cells (lin-1$^-$ HLA-DR$^+$ CD123^{high} CD11C^{low}) were sorted from "DC" fraction, and B cells (HLA-DR$^+$ CD19$^+$) and monocytes (HLA-DR$^+$ CD14$^+$) sorted from "B cell and Monocyte" fraction using FACS Aria cell sorter (BD Biosciences). Following sorting, cells (~500 to ~50,000) were pelleted and resuspended in Trizol® reagent (Invitrogen, Carlsbad, CA).

RNA extraction was performed using the Trizol® (Invitrogen, Life Technologies Corporation) according to the manufacturer's instructions. Concentrations of total RNA were determined using a Nanodrop spectrophotometer or via the Ribogreen RNA quantitation kit (Molecular Probes/Invitrogen). RNA purity was determined by Bioanalyzer 2100 traces (Agilent Technologies). Total RNA was amplified using the WT-Ovation Pico RNA Amplification system (NuGEN) according to the manufacturer's instructions. Next, cDNA was labeled using the Nugen FL-Ovation cDNA Biotin Module V2 kit, following the manufacturer's protocol. Hybridization to Affymetrix HT Human Genome U133A Array (Affymetrix, Inc) and washing was performed on the Affymetrix GeneChip Array Station (GCAS) automation platform. The arrays were scanned on the GeneChip HT Array Plate Scanner (Affymetrix 00-0332) and the data were processed using RMA normalization. Quality control tests were performed using Expression Console software (Affymetrix, Inc) and outlier samples excluded from subsequent analyses. Gene expression fold-change of each subject was obtained by subtracting the day 7 $\log_2$ expression values by its corresponding baseline value. For each cell subset and vaccine type, non-informative genes were filtered out if no increase/decrease higher than 25% (1.25-fold) was

Nature Immunology doi:10.1038/ni.2067

observed on at least 4 of the vaccinees on day 7 compared to baseline. SAM[7] method was applied to identify genes differentially expressed between day 7 post-vaccination compared to baseline. The calculations were performed on the $\log_2$-fold change in gene expression for day 7 post-vaccination versus day 0, separately (SAM one class, maximum permutations allowed). A FDR cutoff of 0.05 was used to identify the differentially expressed genes (**Supplementary Table 2**).

RT-PCR

RT-PCR experiments were performed to: (1) validate the differences detected by microarray analyses in 2007/08 and 2008/09 influenza trials, and (2) evaluate the efficacy of candidate predictors of immunogenicity in a third and independent influenza vaccine trial (2009/10 influenza trial). For these experiments, total RNA was reverse-transcribed using a High-Capacity cDNA Archive Kit Protocol (Applied Biosystems Inc.). Custom designed Low Density Array and Quantitative Real-Time PCR Analysis Low Density Arrays for 48 genes were custom designed by Applied Biosystems (**Supplementary Table 7**). All sequences were amplified using the Applied Biosystems 7900HT Sequence Detection System. Total 25 LDAs runs were performed for 194 human samples. Raw data were obtained using the SDS2.3 software (Applied Biosystems). Real-Time StatMinerTM software (Integromics Inc) was used to perform a quality control test for all runs. Gene expression data were normalized to the average Ct value of the housekeeping genes 18S rRNA, *ACTB*, *PPIA* and *B2M* and then the difference in normalized Ct value between day 3 and 7 versus day 0 was calculated. Correlation between the log2 fold-changes in microarray and RT-PCR data of 2007/08 and 2008/09 influenza trials were calculated by Pearson correlation. Log2 fold-change values between day 3 and 7 versus day 0 of 2009/10 influenza trial were used by DAMIP program. RT-PCR experiments were supported in part by the Virology & Drug Discovery Core of the Emory Center for AIDS Research (P30 AI050409).

***In vitro* stimulation of PBMCs**

Blood samples were collected from 2 healthy subjects and peripheral blood mononuclear cells (PBMCs) were isolated following the manufacturer's protocol (CPTs; Vacutainer® with Sodium Citrate; BD). PBMCs were plated (1 x 10^6 cells) in RPMI incomplete media and stimulated with FluMist® LAIV vaccine (10^4 fluorescent focus units/ ml), Fluarix® TIV vaccine (10 mcg/ ml), Yellow Fever YF17D vaccine (1 MOI) or media (negative control) at 37^{o}C. After 1h of incubation, FBS 10% and antibiotics (Sigma-Aldrich, A5955) were added to each well. PBMCs were then incubated at 37^{o}C for 24h and the RNA extracted using Trizol® (Invitrogen, Life Technologies Corporation), following the manufacturer's instructions. RNA was reverse-transcribed using SuperScript® III RT-PCR kit (Invitrogen). For each gene and biological replicate, real-time PCR reactions (BIO-RAD, iQ SYBR Green Supermix) were performed in triplicate in a MyiQ Single Color Real-Time PCR Detection System machine (BIO-RAD). RNA expression level of each gene was calculated by the formula $2^{-\Delta Ct}$, where ΔCt = mean Ct of the gene – mean Ct of the housekeeping gene *GAPDH*.

Western blot experiments

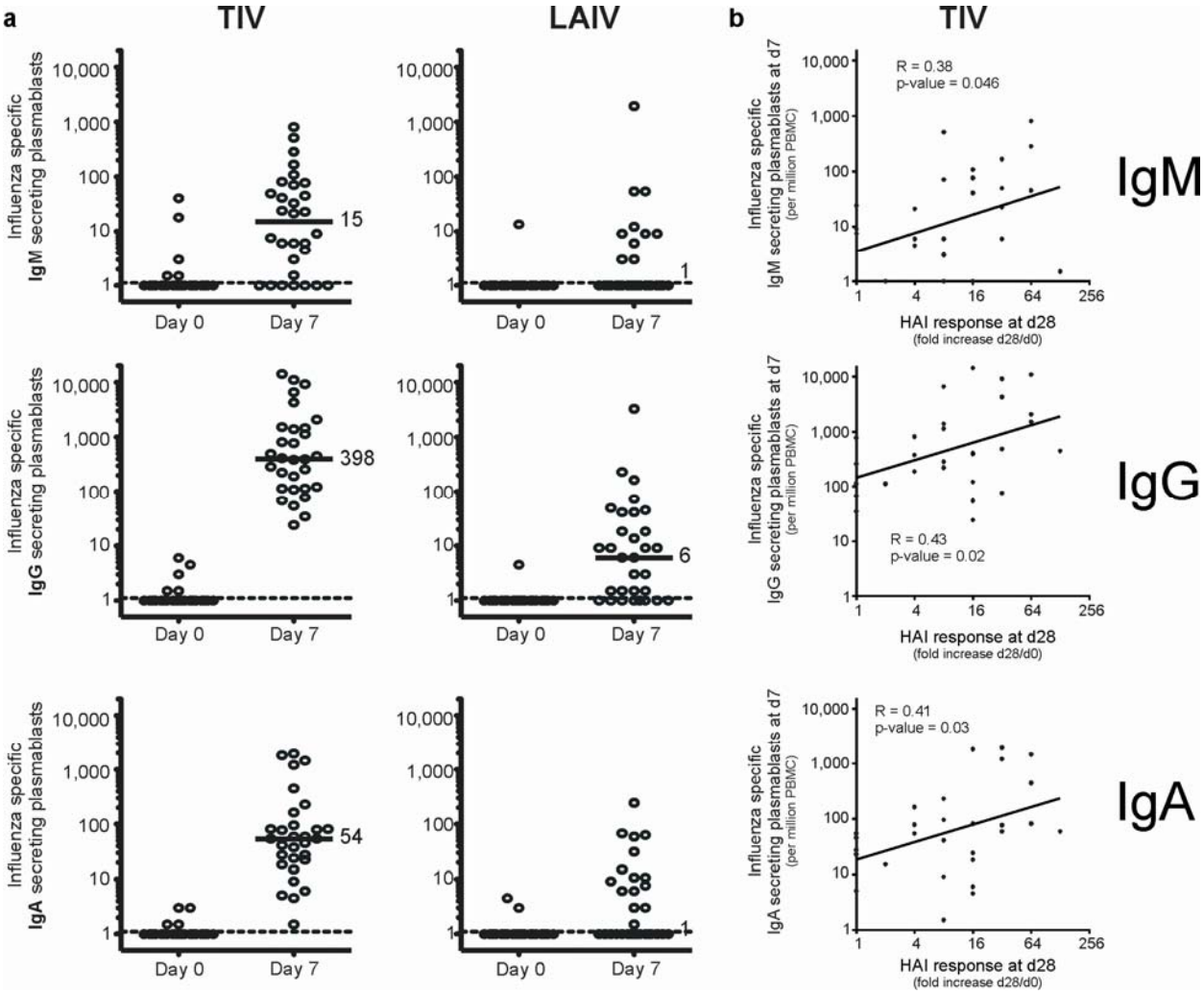
Splenocytes from C57BL/6 mice were stimulated *in vitro* with the indicated concentrations of TIV for 1 or 2 hours. Cells were harvested and lysed using M-PER mammalian protein extraction reagent (Pierce) containing Halt protease inhibitor, EDTA and phosphatase inhibitor (Pierce). Equal amounts of protein were subjected to SDS-PAGE and transferred onto Nitrocellulose membranes. Mouse phosphorylated and non-phosphorylated forms of CaMKIV were detected with anti-phospho-CaMKIV (p-CamKIV (Thr 196), rabbit poyclonal IgG, Cat# SC-28443-R, Lot# E1010, Santa Cruz Biotechnology) and anti-CaMKIV (Camk IV (H-5) mouse monoclonal IgG1, Cat# SC-55501, Lot#L3109, Santa Cruz Biotechnology) antibodies and the blot developed with horseradish peroxidase–conjugated secondary antibody (Anti-mouse IgG-peroxidase antibody, Cat# A9044; Lot# 127k4847, Sigma and Anti-rabbit IgG-peroxidase antibody, Cat# A0545; Lot# 038k4753, Sigma). Signals were visualized using SuperSignal West Pico chemiluminescent substrate (Pierce).

Human PBMCs were stimulated for the indicated time points with TIV (1 ug/ml). Stimulation with LPS (10 ng/mL for 30 minutes) was used as positive control[11]. Cells were harvested and

lysed using M-PER mammalian protein extraction reagent (Pierce) containing Halt protease inhibitor, EDTA and phosphatase inhibitor (Pierce). Equal amounts of protein were subjected to SDS-PAGE and transferred onto Nitrocellulose membranes. Human phosphorylated and non-phosphorylated forms of CaMKIV were detected with anti-phospho-CaMKIV (p-CamKIV (Thr 196), rabbit poyclonal IgG, Cat# SC-28443-R, Lot# E1010, Santa Cruz Biotechnology) and anti-CaMKIV (Camk IV (H-5) mouse monoclonal IgG1, Cat# SC-55501, Lot#L3109, Santa Cruz Biotechnology) antibodies and the blot developed with horseradish peroxidase–conjugated secondary antibody (Anti-mouse IgG-peroxidase antibody, Cat# A9044; Lot# 127k4847, Sigma and Anti-rabbit IgG-peroxidase antibody, Cat# A0545; Lot# 038k4753, Sigma). Signals were visualized using SuperSignal West Pico chemiluminescent substrate (Pierce).

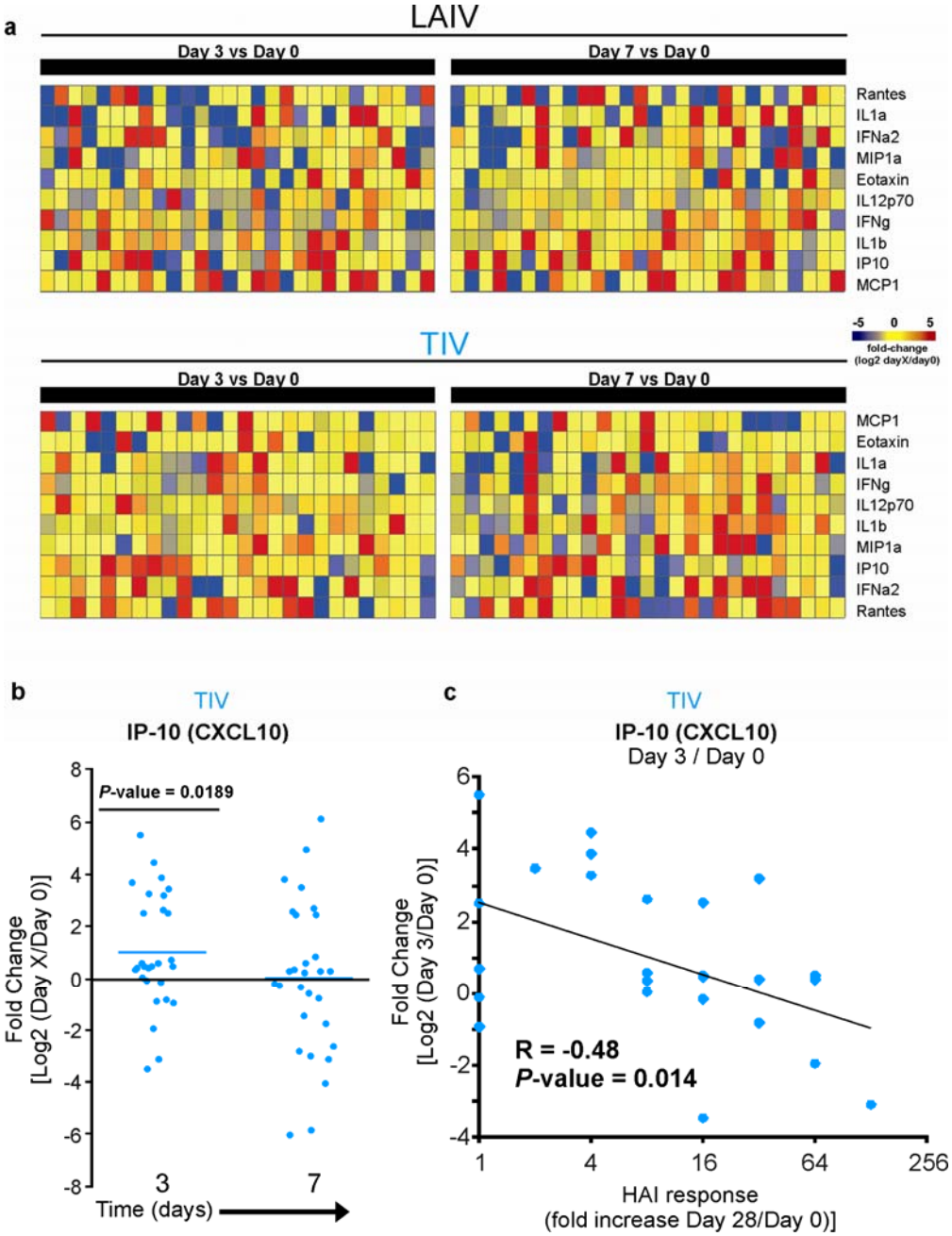
Nature Immunology doi:10.1038/ni.2067

Supplementary Figure 1. Influenza-specific plasmablasts induced by LAIV and TIV vaccination. (a) Influenza-specific IgM, IgG and IgA secreting plasmablasts were measured by ELISPOT in PBMCs of vaccinees at 0 (baseline) and 7 days after vaccination. Each sample was measured in duplicate, averaged and plotted as plasmablasts per million PBMCs. Median values are shown. (b) Pearson correlation between influenza-specific IgM, IgG and IgA secreting plasmablasts at day 7 and the plasma HAI antibody response at day 28 on TIV vaccinees.



Nature Immunology doi:10.1038/ni.2067

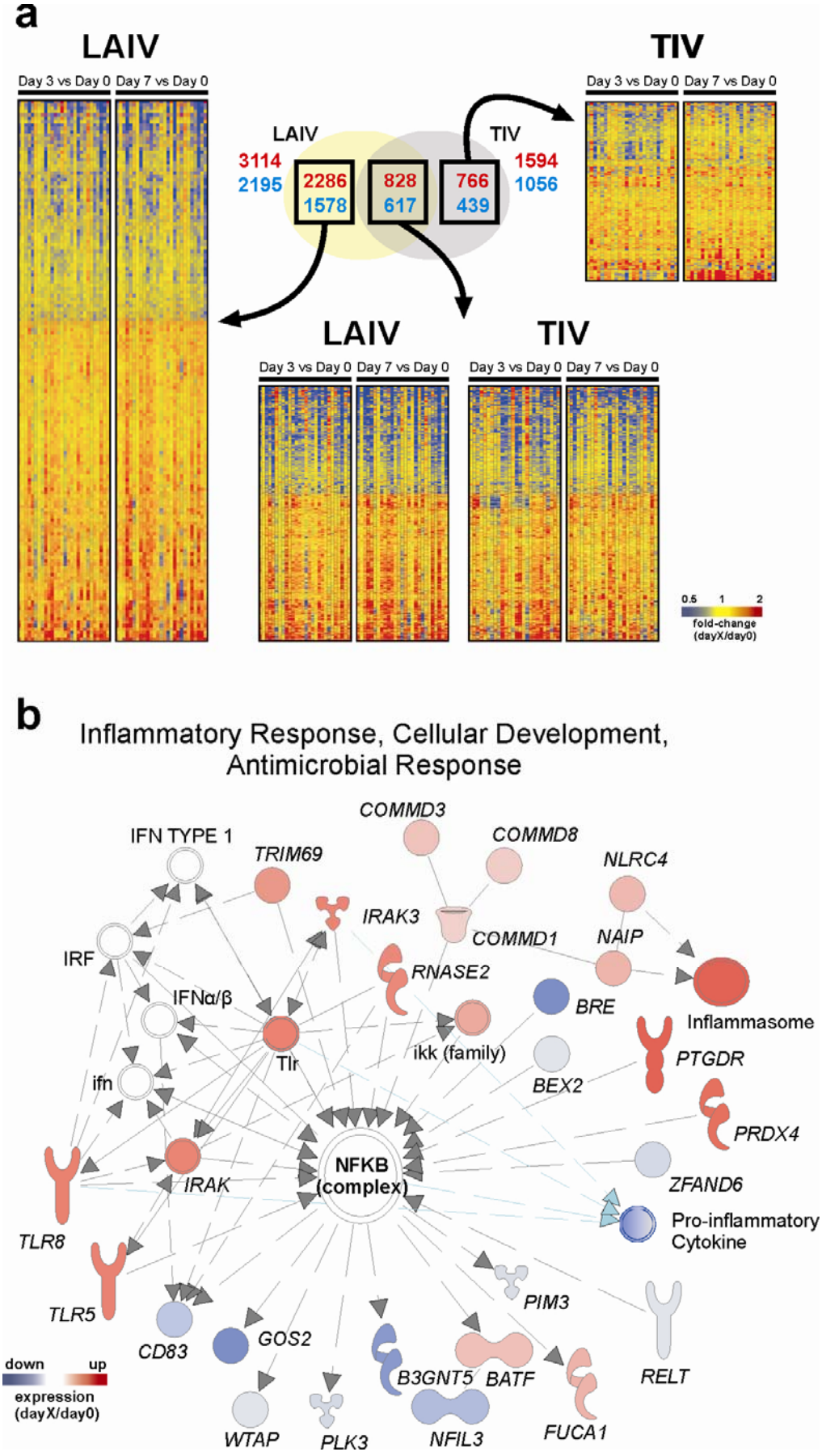
Supplementary Figure 2. Cytokine response in plasma to influenza vaccination. (a) The fold change in cytokine concentrations at days 3 or 7 post-vaccination compared to day 0 (baseline) is depicted as a heat map with Spotfire software. Each column represents one vaccinee. The cytokine names are shown to the right of the heat maps. (b) IP-10 is significantly induced by TIV vaccine on day 3. Data were normalized using the pre-vaccination cytokine level [i.e. Log2(Cd)-Log2(C0), where Cd is the cytokine concentration on day d]. (c) Statistically significant negative correlation between the fold change in the concentration of IP-10 (at day 3 relative to day 0) and the plasma antibody response at day 28 in TIV vaccinees.



Nature Immunology doi:10.1038/ni.2067

Supplementary Figure 3. Molecular signatures of the influenza vaccines. (a) Genes differentially expressed in the PBMCs of LAIV and TIV vaccinees. The numbers of up (red) and down (blue) regulated genes identified in TIV and LAIV vaccinees are shown in the Venn diagram. The heat maps show the fold-change expression of the differentially expressed genes (in rows) identified in both vaccines and those only identified in one vaccine. Fold-change expression was calculated for each vaccinee (in columns) by subtracting the log2 expression values of day 3 and day 7 to the corresponding baseline. (b) An "Inflammatory Response" network was identified by Ingenuity pathway analysis as one of the top 10 networks enriched in genes differentially expressed in both TIV and LAIV vaccinees. Solid and dashed lines respectively represent, direct and indirect interactions reported for the genes. The colors represent the "extreme" mean fold-change in gene expression on days 3 or 7 compared to day 0 in either TIV or LAIV vaccinees. "Extreme" mean fold-change is calculated by taking the minimum mean fold-change out of two or more values if the highest difference compared to baseline is negative, or taking the maximum mean fold-change if the highest difference is positive.

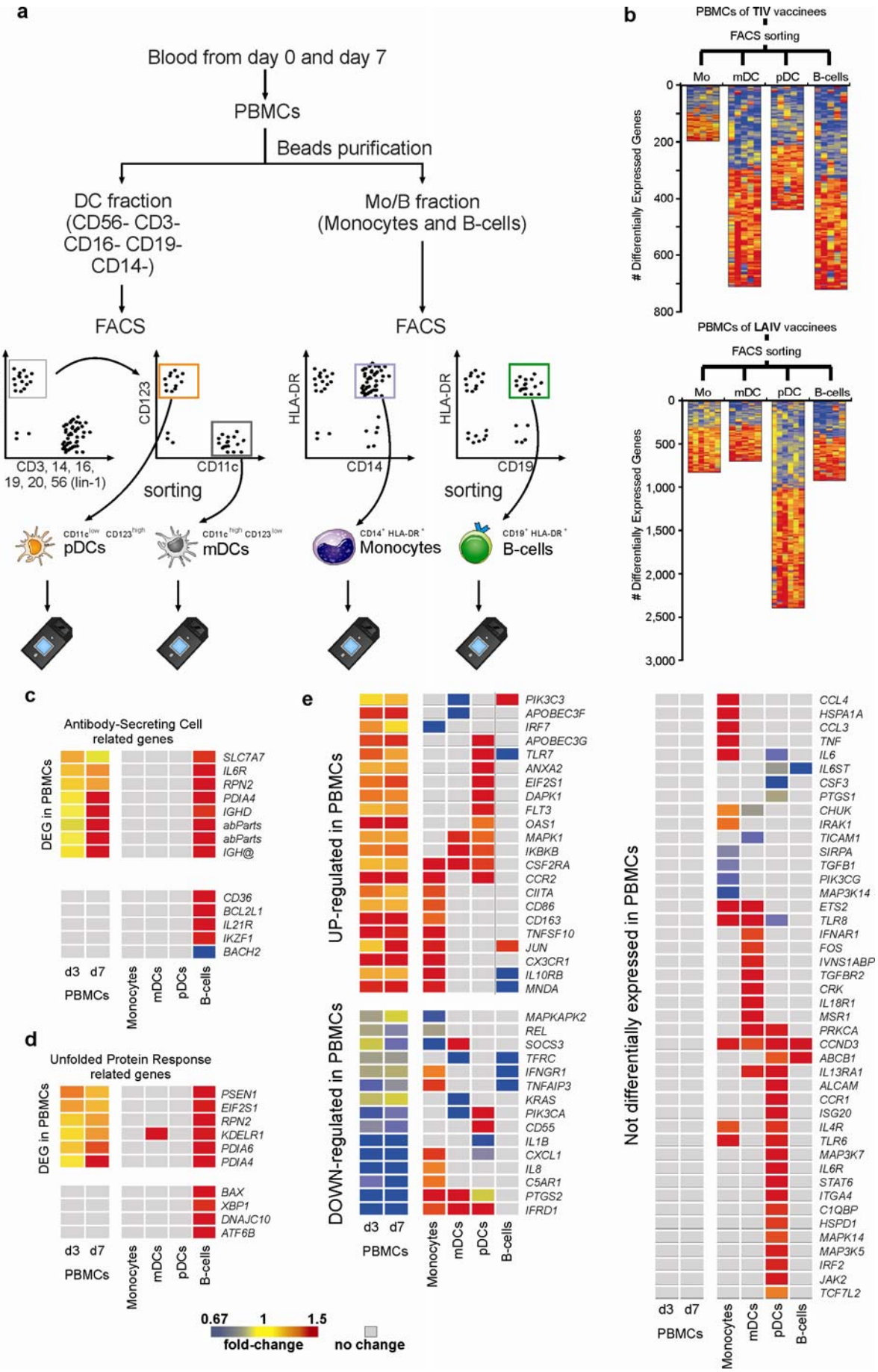
Nature Immunology doi:10.1038/ni.2067



Nature Immunology doi:10.1038/ni.2067

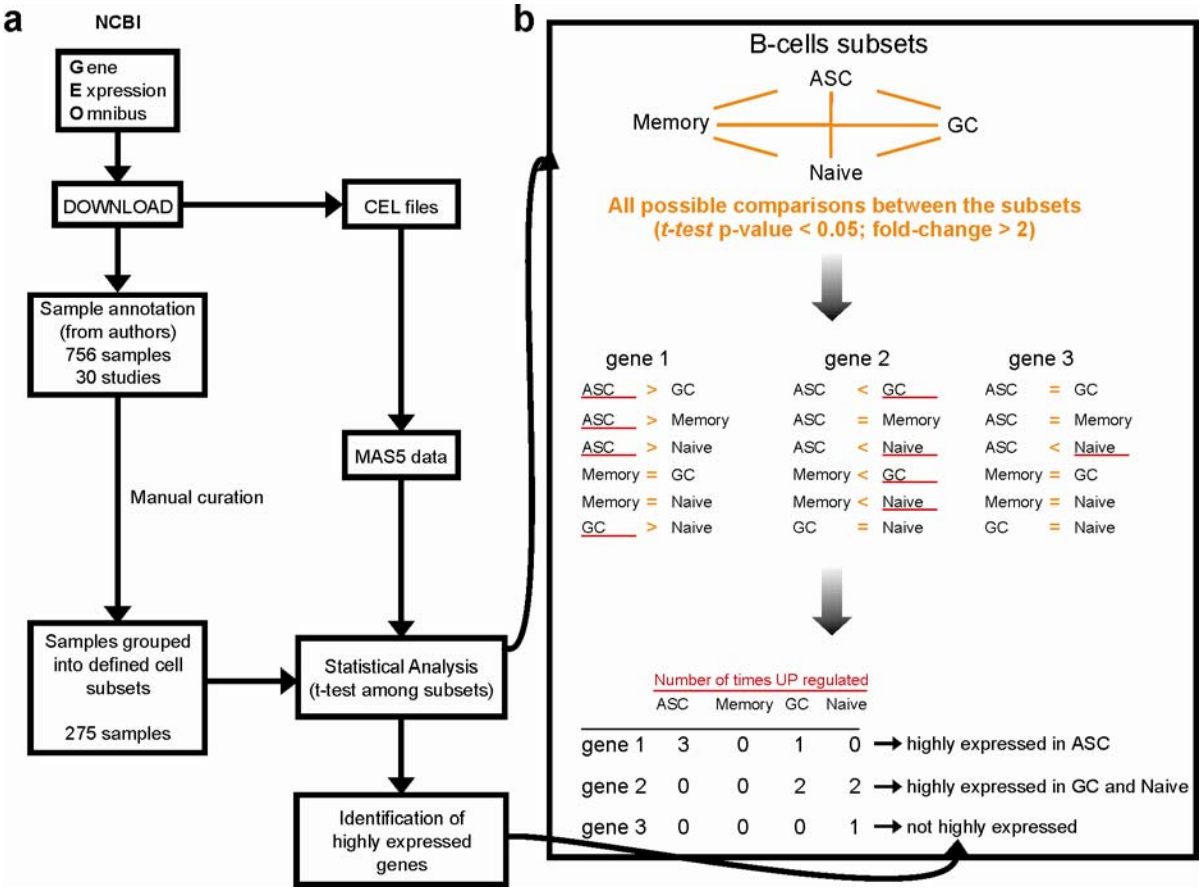
Supplementary Figure 4. Transcriptomic changes in specific cell subsets isolated from PBMCs of individuals vaccinated with LAIV or TIV. (a) Schema describing the microarray analysis of myeloid dendritic cells (mDCs), plasmacytoid dendritic cells (pDCs), Monocytes and B cells sorted from PBMCs of TIV and LAIV vaccinees at day 0 (baseline) and day 7 post-vaccination. (b) Number of differentially expressed genes of each cell subset after TIV (top) or LAIV (bottom) vaccination. (c) Heat map of genes associated with "antibody-secreting cells". The top heat map shows the genes identified as differentially expressed in the PBMC analysis and also in the sorting cell microarray analyses. The bottom heat map shows the differentially expressed genes only identified in sorting cell microarray analyses. The gray color represents genes with no statistically significant expression change on a given cell subset or on the PBMCs. (d) Heat map of genes associated to "Unfolded Protein Response". The top heat map shows the genes identified as differentially expressed in the PBMC analysis and also in the sorting cell microarray analyses. The bottom heat map shows the differentially expressed genes only identified in sorting cell microarray analyses. (e) Heat map of interferon-related genes. The heat map on the left shows the genes identified as differentially expressed in the PBMC analysis and also in the sorting cell microarray analyses. The heat map on the right shows the differentially expressed genes only identified in sorting cell microarray analyses. The gray color represents genes with no statistically significant expression change on a given cell subset.

Nature Immunology doi:10.1038/ni.2067



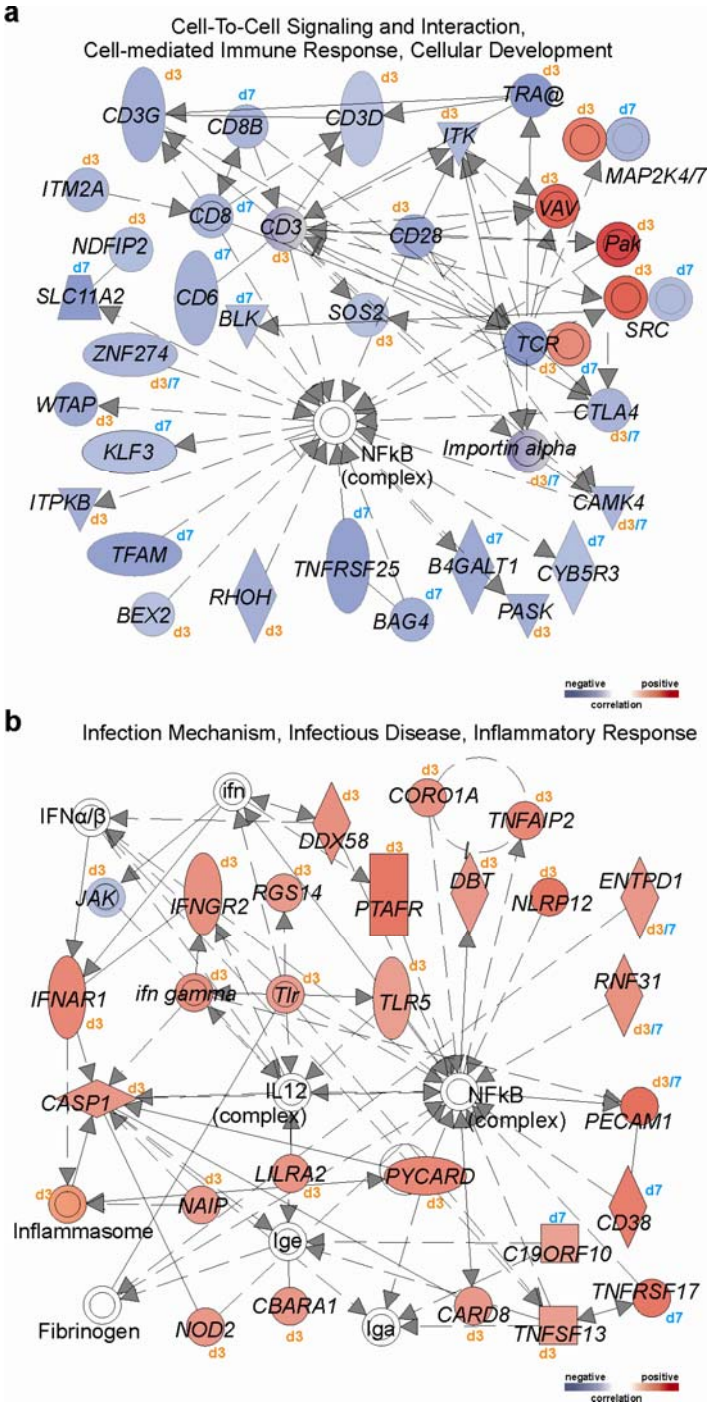
Nature Immunology doi:10.1038/ni.2067

Supplementary Figure 5. Strategy for identifying cell type specific gene signatures using meta-analysis. (a) Schema describing the meta-analysis used to identify the genomic signatures of subsets of immune cells. For each selected study, Affymetrix CEL files were downloaded from GEO database (http://www.ncbi.nlm.nih.gov/geo/), processed and analyzed as shown in (b). Annotation of each sample was obtained from SOFT files and manually curated into defined cell subsets. (b) Cell subsets were compared to each other by *t*-test (*P*-value < 0.05 and fold-change > 2) in order to identify genes highly expressed on each cell type.



Nature Immunology doi:10.1038/ni.2067

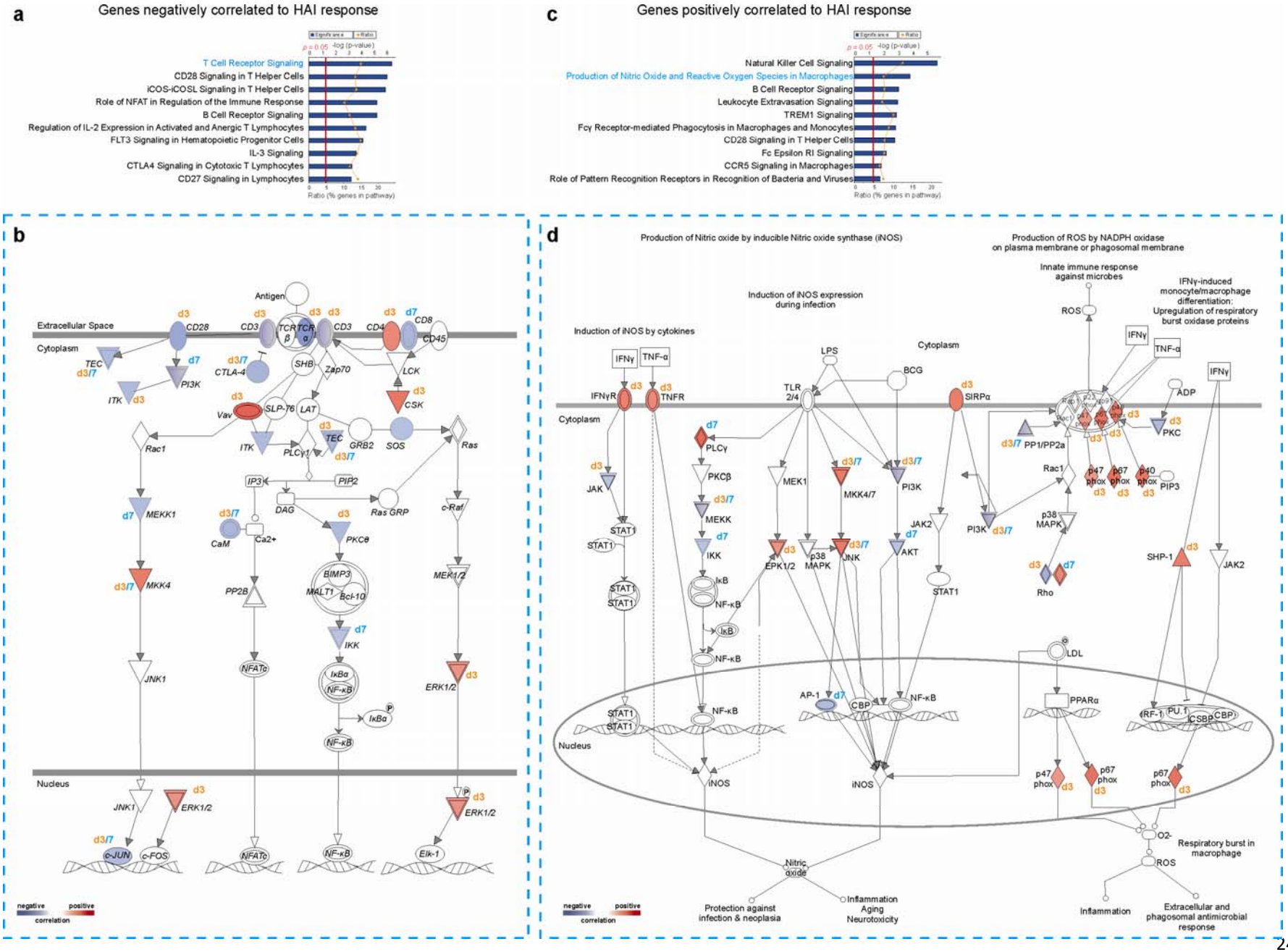
Supplementary Figure 6. Gene networks correlated to antibody response post-TIV vaccination. (a) Gene network identified by Ingenuity Pathway analysis enriched in genes negatively correlated to HAI antibody response. Solid and dashed lines represent, respectively direct and indirect interactions reported for the genes. The colors represent the correlation between day 3/ day 0 (light brown label) or day 7/ day 0 (blue label) fold-change expression and the day 28/ day 0 HAI antibody response. (b) Gene network identified by Ingenuity Pathway analysis enriched in genes positively correlated to HAI antibody response.



Nature Immunology doi:10.1038/ni.2067

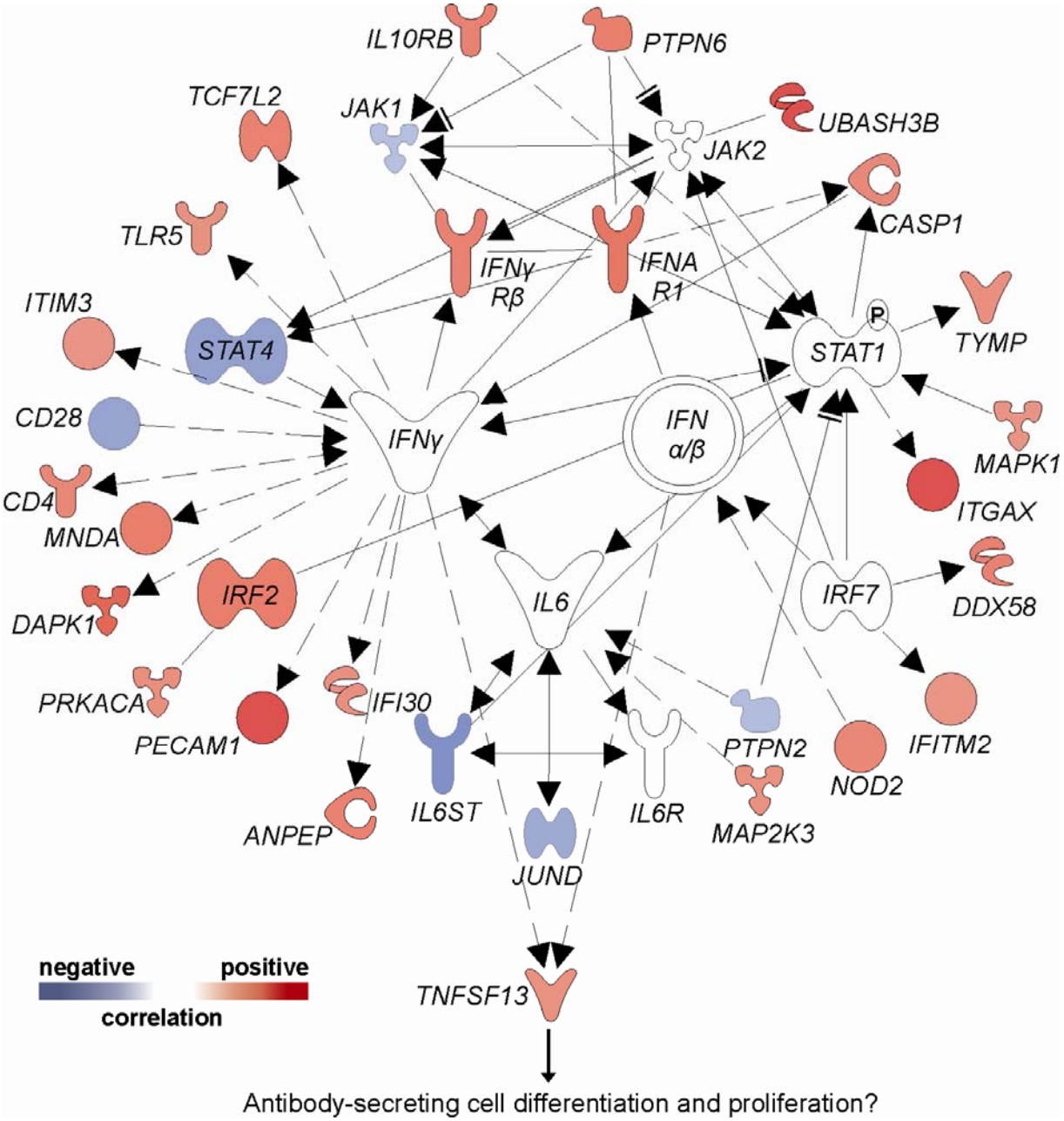
Supplementary Figure 7. Canonical pathways correlated to antibody response post-TIV vaccination. (a) Functional classification performed using Ingenuity Pathway Analysis of genes whose baseline normalized expression at either day 3 or day 7 is negatively correlated (Pearson, *P*-value < 0.05) to baseline-normalized antibody response at day 28 post-TIV vaccination. The horizontal axis on top and the blue bars represent the significance threshold of each canonical pathway. The red line indicates the threshold *P*-value equivalent to 0.05. The brown line and the horizontal axis on the bottom represent the percent of correlated genes in each pathway. Pathway written in blue letters is shown in panel S9b. (b) "T Cell Receptor Signaling" network (adapted from Ingenuity Pathway original network) is significantly enriched in genes whose expression is negatively correlated to antibody response (see legend of Fig. 2a). Genes with expression positively correlated to antibody response were also represented in this network. (c) Functional classification performed using Ingenuity Pathway Analysis of genes whose baseline normalized expression at either day 3 or day 7 is positively correlated (Pearson, *P*-value < 0.05) to baseline-normalized antibody response at day 28 post-TIV vaccination (see legend of Figure S9a). Pathway written in blue letters is shown in panel S9d. (d) "Production of Nitric Oxide and Reactive Oxygen Species in Macrophages" network (adapted from Ingenuity Pathway original network) is significantly enriched in genes whose expression is positively correlated to antibody response (see legend of Fig. 2a). Genes with expression negatively correlated to antibody response were also represented in this network.

Nature Immunology doi:10.1038/ni.2067



Nature Immunology doi:10.1038/ni.2067

Supplementary Figure 8. Interferon related genes that correlate with the antibody response to TIV. Genes related to the Interferon response whose expression on day 3 post-vaccination is correlated to antibody HAI response on day 28 post-vaccination. Solid and dashed lines represent, respectively direct and indirect interactions reported for the genes.



Nature Immunology doi:10.1038/ni.2067

References

1. Chen, G.L., Lamirande, E.W., Jin, H., Kemble, G. & Subbarao, K. Safety, immunogencity, and efficacy of a cold-adapted A/Ann Arbor/6/60 (H2N2) vaccine in mice and ferrets. *Virology* **398**, 109-114 (2010).
2. Wrammert, J.*, et al.* Rapid cloning of high-affinity human monoclonal antibodies against influenza virus. *Nature* **453**, 667-671 (2008).
3. Crotty, S.*, et al.* Cutting edge: long-term B cell memory in humans after smallpox vaccination. *J Immunol* **171**, 4969-4973 (2003).
4. Querec, T.D.*, et al.* Systems biology approach predicts immunogenicity of the yellow fever vaccine in humans. *Nat Immunol* **10**, 116-125 (2009).
5. Golub, T.R.*, et al.* Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**, 531-537 (1999).
6. Singh, D.*, et al.* Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell* **1**, 203-209 (2002).
7. Tusher, V.G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* **98**, 5116-5121 (2001).
8. Lee, E.K. Large-scale optimization-based classification models in medicine and biology. *Ann Biomed Eng* **35**, 1095-1109 (2007).
9. Brooks, J.P. & Lee, E.K. Analysis of the consistency of a mixed integer programming-based multi-category constrained discriminant model. *Annals of Operations Research* **174**, 147-168 (2010).
10. Wu, J.Y.*, et al.* Spermiogenesis and exchange of basic nuclear proteins are impaired in male germ cells lacking Camk4. *Nat Genet* **25**, 448-452 (2000).
11. Zhang, X.*, et al.* Calcium/calmodulin-dependent protein kinase (CaMK) IV mediates nucleocytoplasmic shuttling and release of HMGB1 during lipopolysaccharide stimulation of macrophages. *J Immunol* **181**, 5015-5023 (2008).