SUPPLEMENTARY INFORMATION

Supplementary Methods

Identifying genes induced by YF-17D in most vaccinees.

The raw Affymetrix microarray probe data was assembled into probe sets representing individual genes based on the updated UniGene Build 199, January 16, 2007 to yield a list of 20,078 genes based on a previously published method instead of using Affymetrix predefined probe sets. *R* (http://www.r-project.org/) was used to assemble the probe sets in combination with RMA pre-processing, which includes global background adjustment and quantile normalization. Values below a minimum threshold of normalized fold change in expression of 0.01 for microarrays were reset to that threshold. Gene expression at time points post-vaccination were converted to fold changes by subtracting the pre-vaccination day 0 expression value. Genes with fold change in expression patterns that were similar among most subjects within a trial over time, were detected by identifying genes with normalized Log_2 transformed fold change gene expression values >0.5 or <0.5 in > 60% of the subjects, at days 3 or 7 and then tested for statistical significance by ANOVA adjusted with the Benjamini and Hochberg False Discovery Rate method with a cutoff of 0.05 in Genespring (Agilent Technologies).

One-way ANOVA was used to test for differences on the expression levels of each gene among days 0, 3 and 7. This test does not depend on the number of genes since it is run independently for each gene. Benjamini and Hochberg False Discovery Rate method depends on the number of tests performed and a pre-selection filter may affect the multiple testing corrections. However, we believe that the pre-filtering cut-off used was very low (only a Log_2 transformed fold change gene expression values of 0.5 or 41% increase or decrease on the gene expression levels in at least 3/5 of subjects) and necessary only to remove genes that did not fluctuate with time, which are often unexpressed/low expressed genes. Therefore, we believe that our pre-selection filter did not compromise our findings. Nevertheless, we explored testing the whole dataset by ANOVA without any pre-selection filter. This resulted in a list of 22 genes (**Supplementary Table 1**). This low number of genes is absolutely expected. A gene list with 20,000 genes will require for the gene with lowest *P*-value given by ANOVA an adjusted *P* -value lower than 0.0000025 (0.05/20,000) and for the gene with the second

Nature Immunology: doi:10.1038/ni.1688

lowest *P* -value, an adjusted *P*-value lower than 0.000005 (0.05/(20,000/2)). However we were curious as to whether this selection criterion were too stringent since it was not detecting the increased transcription of CD38 and IP-10, which we knew to be increased by flow cytometry and ELISA, data respectively (**Fig. 2** & **Supplementary Fig. 1**). Pre-filtering allowed the detection of these genes that had already been "verified" at the protein level and identified an additional subset a genes with close biological interactions with the 22 genes already selected (**Supplementary Table 1**). Furthermore, this expanded list of genes suggested a role for complement, which was verified by ELISA (**Supplementary Fig. 4**), and also had many more genes that could be verified by RT-PCR (**Supplementary Table 1**). Therefore while it is likely that omitting the pre-filtering step may result in a more rigorous statistical analysis, we feel that it may be too stringent and exclude potentially biologically relevant genes.

Identifying genes that correlate with magnitudes of immune responses.

Genes, whose expression correlated with the magnitude of the T-cell responses were identified by comparing the % of $CD38^+$ HLA-DR^+ (activated) $CD8^+$ T-cells to the normalized Log_2 transformed gene expression values. Genes with > 25% of the subjects having >0.5 or <0.5 change were analyzed by the Linear Model (lm) function in *R* to identify genes with a slope *P*-value <0.05. A predictive model of T cell responses was generated using ClaNC run within *R*. Principle Component Analysis (PCA) to visually reduce and summarize gene expression variance among the subjects was conducted in Genespring. The student t-test was performed in Prism to test whether genes displayed a significant difference between subjects when they were grouped by T cell responses. Gene networks and functional relationships were analyzed with Ingenuity Pathways Analysis (Ingenuity Systems) and the DAVID Bioinformatics Database (http://david.abcc.ncifcrf.gov/home.jsp). Transcription factor binding sites of gene lists were analyzed in TOUCAN v3.0.2 (http://homes.esat.kuleuven.be/~saerts/software/toucan.php) using the TRANSFAC v7.0 database of eukaryotic transcription factors. Binding site motifs were scanned for in the DNA sequence 2000 bases upstream through 200 bases downstream flanking the first exon of each gene with a double prior of 0.1 and the genomic background noise model based on the third order Markov Model

Nature Immunology: doi:10.1038/ni.1688

from the Human Eukaryotic Promoter Databse (Human EPD 3). RT-PCR genes from the Applied Biosystems Custom TaqMan Gene Expression plate, was normalized to the average C_t value of the housekeeping genes 18S rRNA (Hs99999901_s1), *ACTB* (Hs99999903_m1), and *B2M* (Hs99999907_m1) and then the difference in normalized C_t value between day 3 and 7 versus day 0 was calculated. Correlation between the fold changes in microarray and RT-PCR data were calculated using Prism. Genes believed to change with time post vaccination were tested for statistical significance by ANOVA adjusted False Detection Rate method with a cutoff of 0.05 in Genespring, as with the microarray data. For the CD8 predictive model, correlation between microarray and RT-PCR data for each individual gene was analyzed using Prism. For analysis of data from the experiment of stimulating PBMCs with YF-17D, we selected for genes that were up or down regulated by a factor of 0.5 fold in the Log_2 scale, after either 3 or 12 hours of stimulation with YF-17D, compared to cells cultured in media alone. We performed the student *t*-test for comparing YF-17D to media alone at 3 and 12 hours. The genes commonly modulated in both independent trials are analyzed for statistically overrepresented transcription factor binding sites in TOUCAN v3.0.2 using the TRANSFAC v7.0 public database of eukaryotic transcription factors.

Discriminant Analysis via Mixed Integer Programming.

There are five fundamental steps in discriminant analysis: (i) determine the data for input and the predictive output classes; (ii) gather a training set of data (including output class) from human experts or from laboratory experiments. Each element in the training set is an entity with corresponding known output class; (iii) determine the input attributes to represent each entity; (iv) identify discriminatory attributes and develop the predictive rules; (v) validate the performance of the predictive rules.

Utilizing the technology of large-scale discrete optimization and support-vector machines, we have developed novel predictive models[16,17] DAMIP, that simultaneously include the following features: the ability to classify any number of distinct groups; the ability to incorporate heterogeneous types of attributes as input; a high-dimensional data transformation that eliminates noise and errors in biological data; constraints to limit the rate of misclassification, and a reserved-judgment region that provides a safeguard

Nature Immunology: doi:10.1038/ni.1688

against over-training (which tends to lead to high misclassification rates from the resulting predictive rule); and successive multi-stage classification capability to handle data points placed in the reserved judgment region.

In our analysis, each Trial forms a data set. The entity in the dataset is an individual vaccinee, and the measurable attributes for each entity consists of the time measurement of gene array data described in the data collection part. For the T cell analysis, the group in each Trial is determined by the magnitude of the $CD8^+$ T cell response. There are about 800 total measurable gene attributes (of mixed time points) for each entity. We have 2 groups of vaccinees in the T cell analysis ("high" group and "low" group). Each experiment consists of the following two parts: a) Develop a classification rule using a training dataset (Trial 1), b) Use the rule developed from the training set to predict the group status of independent unknown entities (from Trial 2). The experiment is then repeated using Trial 2 as the training set and Trial 1 for blind prediction. For the B cell analysis, the group in each Trial is determined by the magnitude of the neutralizing antibody titers. Again there are the "high" and "low" groups. There are about 1600 total measurable gene attributes (of mixed time points) for each entity.

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 entities 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 patients, the blind test provides a further measurement of its 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, ten-fold cross validation is performed. In the ten-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 nine of the ten 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

Nature Immunology: doi:10.1038/ni.1688

measure of correct classification of the ten 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. This rule is then applied to each entity in the blind data to predict its group status. The percent of correct prediction of the blind data entities 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 ten 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 patients in each clinical Trial is rather small, we set each signature set to contain at most 5 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.

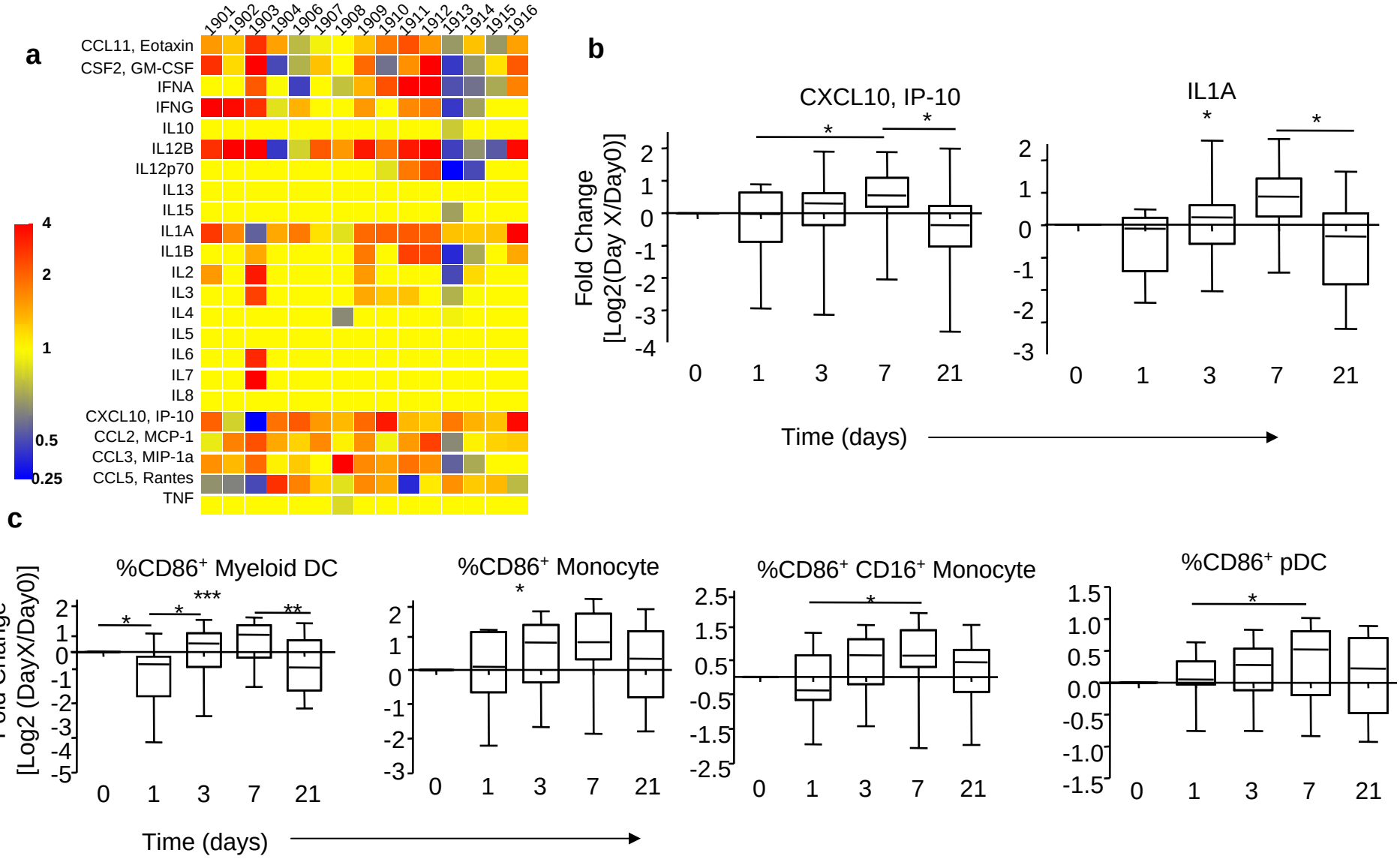
Nature Immunology: doi:10.1038/ni.1688

To perform a blind test, simply input each entity 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, 2008[17], 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.

With regards to why Trial 1 -> Trial 2 DAMIP predictions were significantly more successful for the antibody titer predictions that the Trial 2 -> Trial 1, since As Trial 2 is smaller than Trial 1, one possible explanation for the discrepancies in predictive power is that the ranges of individual variability in genetic responses is more completely captured in Trial 1 than Trial 2. In other words, out of the ranges of responses that humans can make, Trial 2 may contain a subset of those found in Trial 1. Therefore while Trial 1 based models only need to interpolate predictions for Trial 2, Trial 2 based models may need to extrapolate predictions for some of the Trial 1 subjects. “Interpolation” and “extrapolation” are traditionally thought of in terms of polynomial functions, in which case extrapolating data is associated with greater uncertainty and greater likelihood of inaccurate prediction.

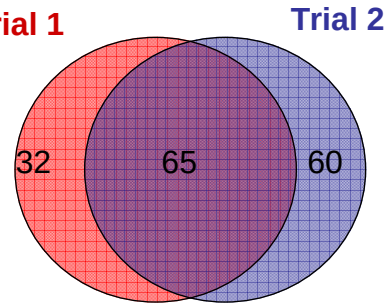
Nature Immunology: doi:10.1038/ni.1688



Supplementary Figure 1. Cytokine and dendritic cell responses to YF-17D. (a) The maximum fold change in cytokine expression out of days 3 or 7 is calculated and depicted as a heat map with GeneSpring software. (b) Out of the cytokines that are induced by vaccination, IP-10 and IL1A are significantly upregulated on day 7. Data were normalized using the pre-vaccination cytokine level [*i.e.* $Log_2(Cd)-Log_2(C0)$, where Cd is the cytokine concentration on day d]. (c) The percentage of $CD86^+$ myeloid dendritic cells, plasmacytoid dendritic cells, total monocytes, or inflammatory $CD16^+$ monocytes is first calculated for each day. The Log2 transformed values for the percentages of $CD86^+$ cells were normalized relative to baseline levels. The change in the percentage of $CD86^+$ positive cells is then calculated for each day relative to day 0 and tested for significance. The determination of significant changes was based on ANOVA followed by Tukey's multiple test comparison on the 15 subjects of Trial 1. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Nature Immunology: doi:10.1038/ni.1688

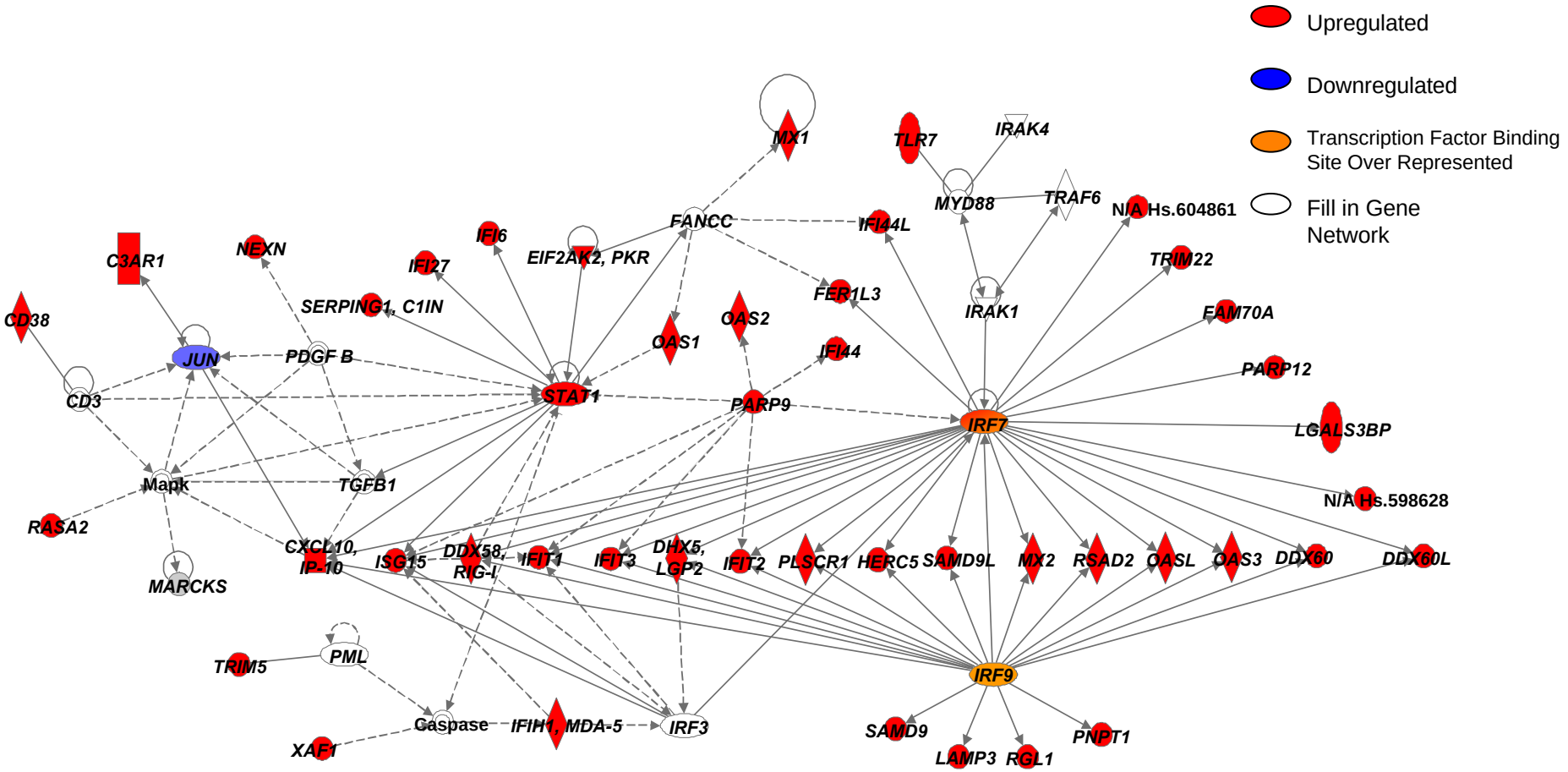
a



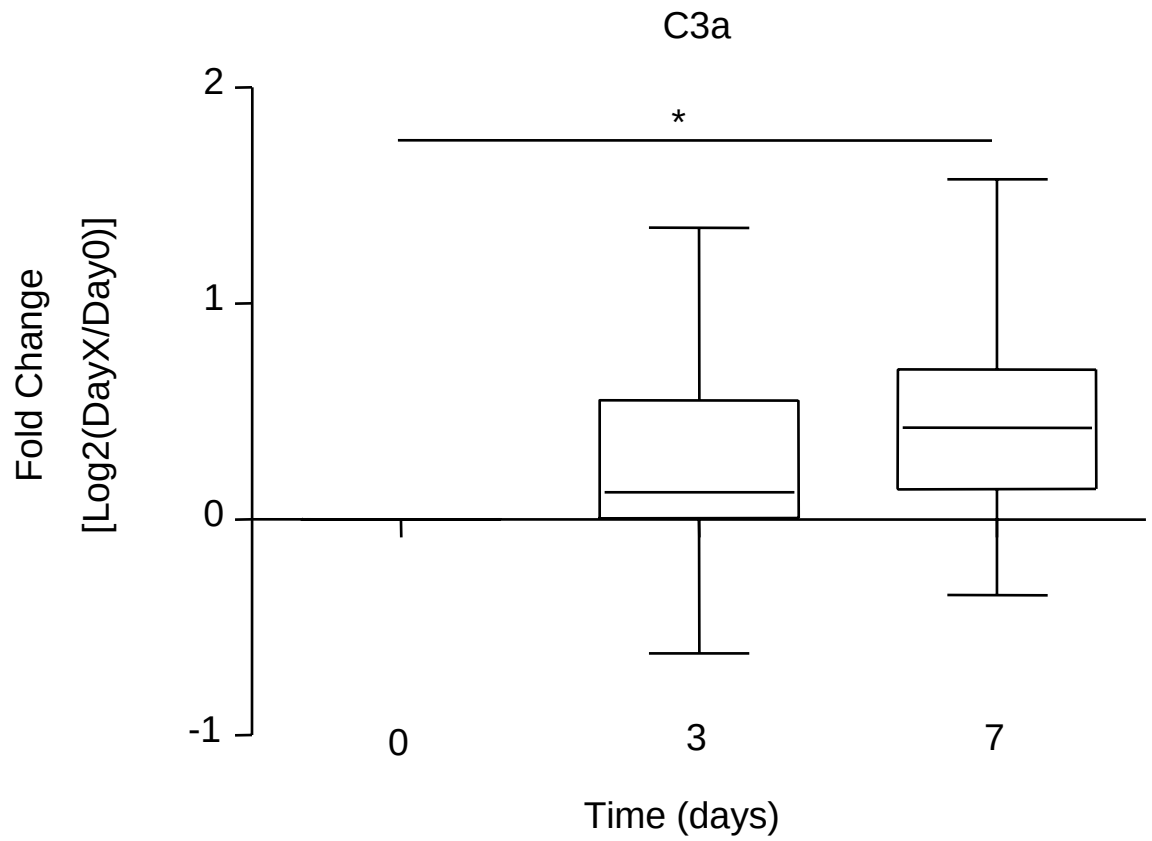
b

	Trial 1			Trial 2			Combined		
Gene Ontology Term	Count	%	*P*-value	Count	%	*P*-value	Count	%	*P*-value
Immune response	29	31.9	1.90E-13	39	32.8	2.70E-19	24	39.3	2.60E-14
Response to virus	11	12.1	6.20E-12	11	9.2	6.20E-11	11	18	4.60E-14
Response to pest, pathogen or parasite	21	23.1	2.10E-12	25	21	3.50E-14	18	29.5	3.90E-13
Inflammatory response	7	7.7	6.50E-04	10	8.4	6.70E-06	5	8.2	4.40E-03
Innate immune response	3	3.3	4.60E-02	4	3.4	8.90E-03	3	4.9	2.00E-02
Cell motility	NA	NA	NA	NA	NA	NA	4	6.6	3.90E-02
Biopolymer metabolism	NA	NA	NA	NA	NA	NA	15	24.6	5.50E-02
Programmed cell death	8	8.8	2.20E-02	9	7.6	2.40E-02	5	8.2	1.00E-01
Regulation of programmed cell death	6	6.6	3.30E-02	8	6.7	6.80E-03	NA	NA	NA
Cellular defense response	3	3.3	9.50E-02	NA	NA	NA	NA	NA	NA
Chemotaxis	NA	NA	NA	6	5	1.10E-03	NA	NA	NA
Taxis	NA	NA	NA	6	5	1.10E-03	NA	NA	NA
Humoral immune response	NA	NA	NA	5	4.2	1.90E-02	NA	NA	NA
Regulation of I-kappaB kinase/NF-kappaB cascade	NA	NA	NA	3	2.5	9.90E-02	NA	NA	NA

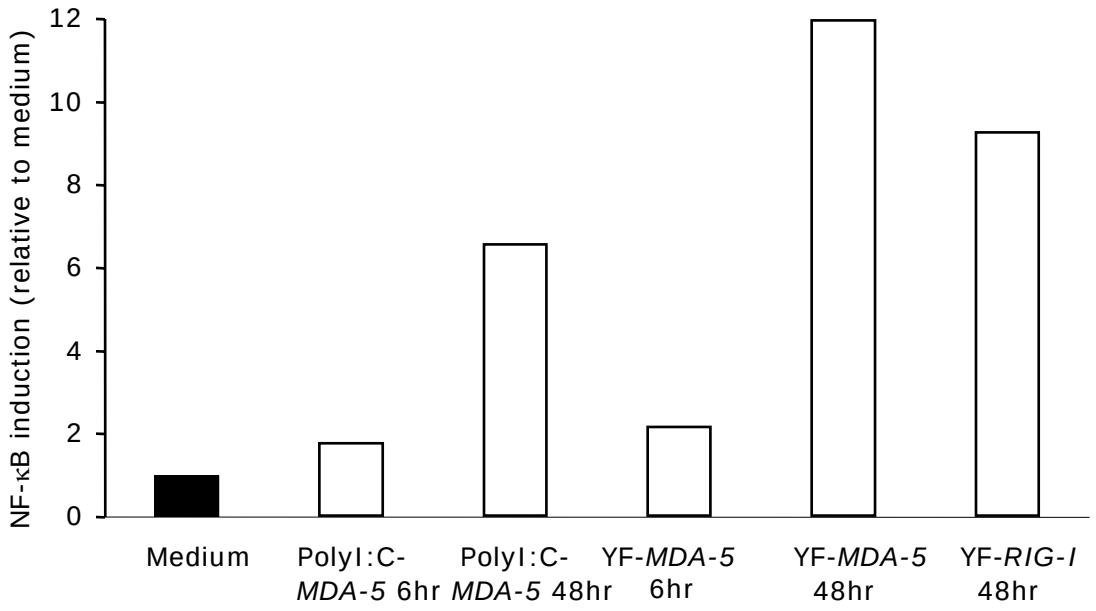
Supplementary Figure 2. Identification of commonly induced genes in two independent vaccine trials. (a) The fold change in expression is calculated for each gene on days 3 and 7 relative to day 0. Genes with Log_2 fold changes > 0.5 or < -0.5 in at least 60% of subjects are then selected. The linear expression values for these genes are then analyzed for significance in GeneSpring. (b) Genes with a Benjamini and Hochbery False Discovery Rate less than 0.05 for each trial are then compared. The genes identified as being significantly changed on days 3 or 7 are analyzed level 4 Gene Ontology terms using DAVID to identify associations among the genes. The results are based on 15 subjects in Trial 1 and 10 subjects in Trial 2.



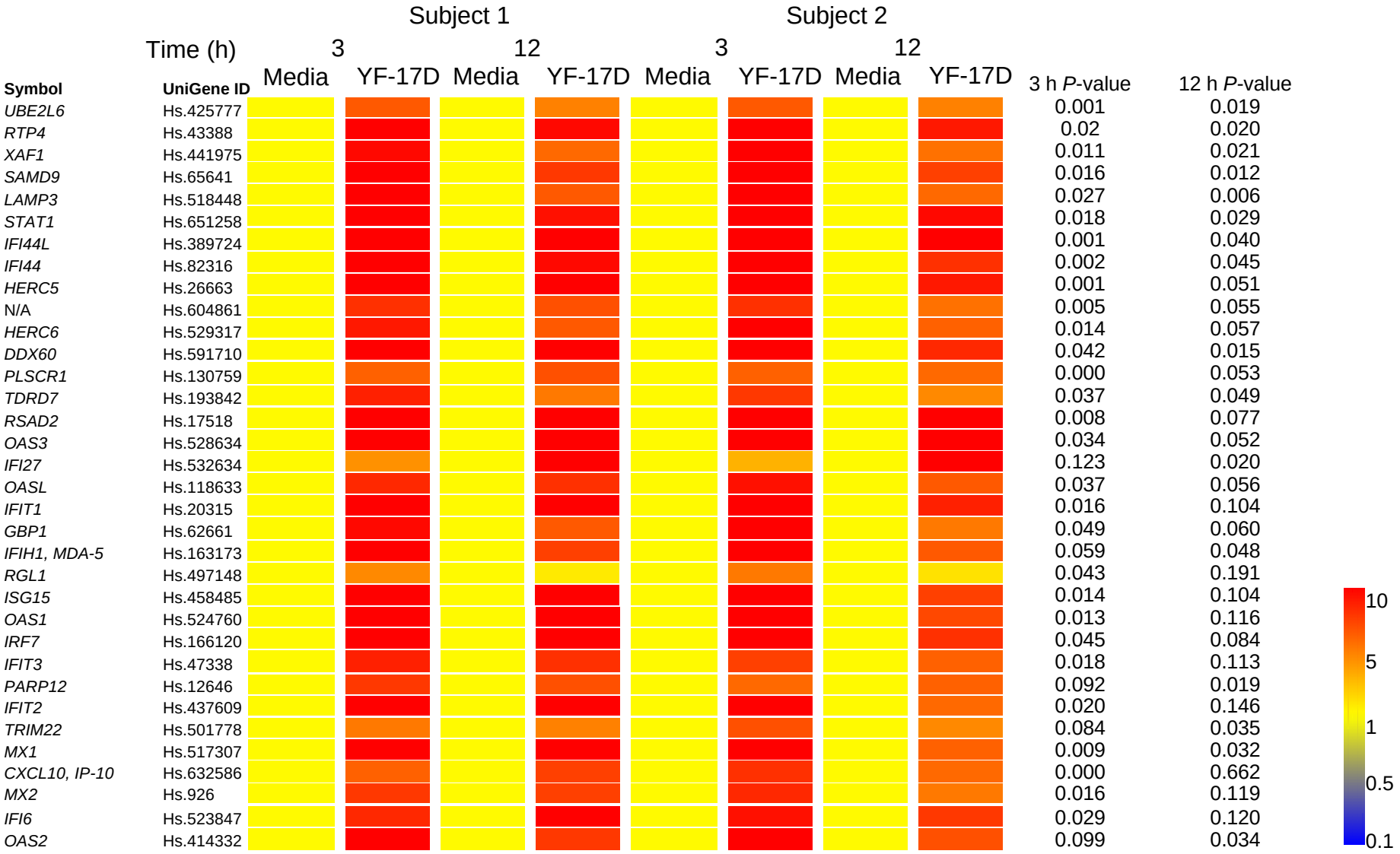
Supplementary Figure 3. Network of anti-viral genes in response to YF-17D. Ingenuity Pathways Analysis of genes identified in **Fig. 1b** as being regulated significantly in two independent trials and supplemented with transcription factor binding motif information from TOUCAN for *IRF7* and *IRF9* (**Supplementary Table 2** online).



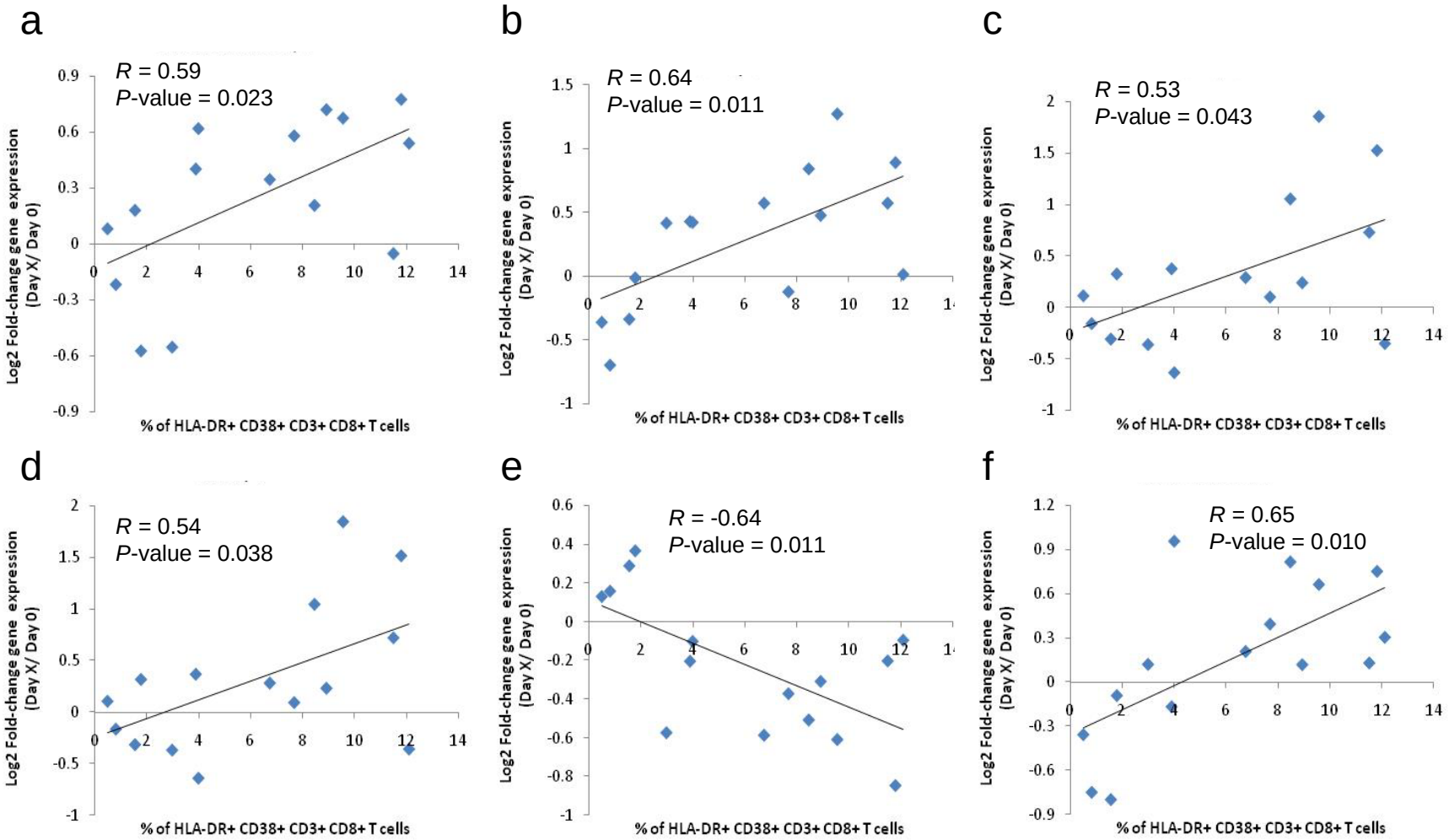
Supplementary Figure 4. Induction of complement C3a by YF-17D. Plasma concentrations of C3a is measured by ELISA to confirm activation of the complement pathways. The determination of significant changes was based on ANOVA followed by Tukey's multiple test comparison on the 10 subjects of Trial 2. * $P < 0.05$.



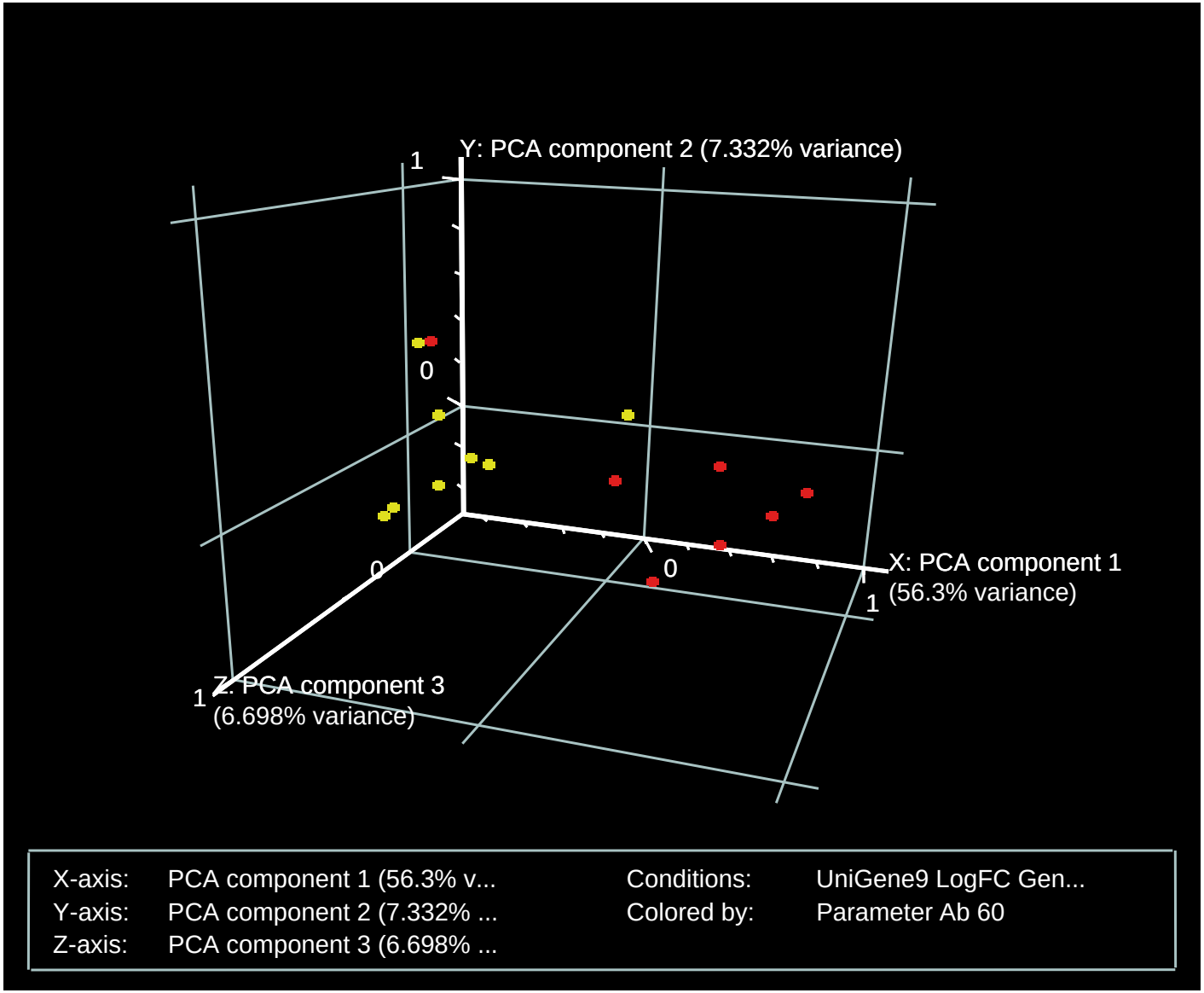
Supplementary Figure 5. YF-17D induces *NF-κB* activation via *RIG-I* and *MDA-5*. Human embryonic fibroblasts (HEK293 cell line) were co-transfected with plasmids encoding luciferase driven by an *NF-κB* promoter, plus a plasmid encoding either *MDA-5* or *RIG-I* for 24 hr. Then cells were stimulated with poly-IC or YF-17D for 6 hr or 48hr. *NF-κB* induction was detected by luciferase activity. Representative of 2 independent experiments.



Supplementary Figure 6. Induction of anti-viral genes in PBMCs stimulated in vitro with YF-17D. PBMCs from 2 healthy unvaccinated donors were isolated and plated at 1 x 10^6 cells per well in 48-well plates with 1 ml RPMI with 10% FBS and penicillin/streptomycin. The cells were cultured in the presence or absence of YF-17D at a MOI of 1. After 3 and 12 hours, RNA was isolated from the cells and processed for microarray analysis. For these experiments, the Affymetrix Human Genome 133A 2.0 Array was used. This microarray contains a subset of genes found on the Human 133 Plus 2.0 Array, which was used in the analysis of the vaccinees. We selected for genes that were up or down regulated by a factor of 0.5 fold in the Log_2 scale, after either 3 or 12 hours of stimulation with YF-17D, compared to cells cultured in media alone. The student t-test was used to compare YF-17D to media alone at 3 and 12 hours.



Supplementary Figure 7. Correlation coefficients and *P*-values of stress response genes that correlate with the magnitude of the CD8$^+$ T cell response. (a) Calreticulin at Day3. **(b)** Protein disulfide isomerase family A, member 5 at Day3. **(c)** Protein disulfide isomerase family A, member 4 at Day3. **(d)** Protein disulfide isomerase family A, member 4 at Day7. **(e)** Nuclear receptor subfamily 3, group C, member 1(glucocorticoid receptor) at Day3. (f) Eukaryotic translation initiation factor 2 alpha kinase 4 at Day7. The data are from the 15 subjects of Trial 1.



Supplementary Figure 8. Genomic signatures that correlate with the magnitude of the antibody response. Genes with a Log_2 fold change of > 0.5 or < -0.5 in greater than 25% of the subjects are first selected. Next the slope *P*-value of the day 60 antibody titers versus Log_2 fold change in gene expression was calculated for each remaining gene. Those genes with $P < 0.05$ are identified as having a significant relationship between early gene expression changes and later antibody responses. Unsupervised principle component analysis of the gene expression for each subject on both days 3 and 7 reveals that subjects could be segregated based on antibody titers above and below 170. Data are from the 15 subjects of Trial 1.

Supplementary Table 1

Symbol	Select Aliases	ANOVA Only	Pre-Filtered	RT-PCR	RT-PCR *P*-value	ANOVA Only vs RT-PCR	Pre-Filtered vs RT-PCR
OASL		1	1	1	0.00048	Confirmed	Confirmed
PARP12		0	1	1	0.00656	Not Detected but Confirmed	Confirmed
PLSCR1		0	1	1	0.00056	Not Detected but Confirmed	Confirmed
EIF2AK2	*PKR*	1	1	1	0.00377	Confirmed	Confirmed
KLHDC7B		1	1	0	NA	Not Tested	Not Tested
IFIH1	*MDA-5*	0	1	1	0.00218	Not Detected but Confirmed	Confirmed
IRF7		1	1	1	0.00010	Confirmed	Confirmed
RSAD2		1	1	1	0.00064	Confirmed	Confirmed
DDX58	*RIG-I*	0	1	1	0.01599	Not Detected but Confirmed	Confirmed
TDRD7		0	1	0	NA	0	Not Tested
IFIT1		0	1	0	NA	0	Not Tested
HERC5		1	1	1	0.00016	Confirmed	Confirmed
SIGLEC1		1	1	1	0.00005	Confirmed	Confirmed
MS4A4		0	1	0	NA	0	Not Tested
TRIM5		0	1	0	NA	0	Not Tested
SERPING1	*C1IN*	0	1	1	0.00838	Not Detected but Confirmed	Confirmed
PNPT1		0	1	0	NA	0	Not Tested
IFI44L		1	1	1	0.00000	Confirmed	Confirmed
OAS2		1	1	1	0.00575	Confirmed	Confirmed
UBE2L6		0	1	0	NA	0	Not Tested
RTP4		0	1	0	NA	0	Not Tested
FAM70A		1	1	1	0.41473	Not Confirmed	Not Confirmed
IFIT2		0	1	0	NA	0	Not Tested
XAF1		0	1	1	0.00005	Not Detected but Confirmed	Confirmed
TLR7		0	1	1	0.04090	Not Detected but Confirmed	Confirmed
ISG15		1	1	1	0.00005	Confirmed	Confirmed
FBXO6		0	1	0	NA	0	Not Tested
IFIT3		0	1	0	NA	0	Not Tested
CD38		0	1	1	0.03109	Not Detected but Confirmed	Confirmed
SAMD9L		0	1	0	NA	0	Not Tested
RNF36		0	1	0	NA	0	Not Tested
RGL1		1	1	0	NA	Not Tested	Not Tested
FER1L3		0	1	0	NA	0	Not Tested
TRIM22		0	1	0	NA	0	Not Tested
LGALS3BP		1	1	0	NA	Not Tested	Not Tested
MX1		1	1	1	0.00005	Confirmed	Confirmed
PARP9		0	1	1	0.00117	Not Detected but Confirmed	Confirmed
PARP14		0	1	1	0.09838	0	Not Confirmed
LAMP3		1	1	1	0.37645	Not Confirmed	Not Confirmed
MARCKS		0	1	0	NA	0	Not Tested
IFI6		1	1	0	NA	Not Tested	Not Tested
OAS1		0	1	1	0.00116	Not Detected but Confirmed	Confirmed
JUN		0	1	1	0.00007	Not Detected but Confirmed	Confirmed
OAS3		0	1	1	0.00188	Not Detected but Confirmed	Confirmed
HERC6		1	1	0	NA	Not Tested	Not Tested
IFI27		1	1	0	NA	Not Tested	Not Tested
DDX60L		0	1	0	NA	0	Not Tested

EPSTI1		1	1	1	0.00070	Confirmed	Confirmed
DHX58	*LGP2*	1	1	1	0.04016	Confirmed	Confirmed
C3AR1		0	1	1	0.07323	0	Not Confirmed
DDX60		0	1	0	NA	0	Not Tested
N/A		1	1	0	NA	Not Tested	Not Tested
CDKN1C		0	1	0	NA	0	Not Tested
N/A		0	1	0	NA	0	Not Tested
GBP1		0	1	1	0.07072	0	Not Confirmed
NEXN		0	1	0	NA	0	Not Tested
CXCL10	*IP10*	0	1	1	0.22878	0	Not Confirmed
N/A		0	1	0	NA	0	Not Tested
STAT1		0	1	0	NA	0	Not Tested
N/A		0	1	0	NA	0	Not Tested
N/A		0	1	1	0.05277	0	Not Confirmed
SAMD9		0	1	0	NA	0	Not Tested
CMPK2		0	1	0	NA	0	Not Tested
IFI44		1	1	1	0.00000	Confirmed	Confirmed
MX2		0	1	1	0.00601	Not Detected but Confirmed	Confirmed
		22	65	33			

Detected by Filter and p < 0.05 by RT-PCR	13	26
Detected by Filter and p > 0.05 by RT-PCR	2	7
Detected by Filter but Not Tested by RT-PCR	7	32
Not Detected by Filter but Confirmed by RT-PCR	13	0
Not Detected by Filter and Not Confirmed by RT-PCR	30	0

Supplementary Table 1: Strategies used to identify genes induced by YF-17D in the majority of vaccinees.

Footnote: This table compares the genes identified using 2 independent strategies: a 'pre-filtering strategy,' and a strategy of testing the entire database with ANOVA without pre-filtering. In the 'pre-filtering' strategy, genes with normalized Log2 transformed fold change gene expression values >0.5 or <0.5 in > 60% of the subjects at days 3 or 7 were identified, and then tested for statistical significance by ANOVA adjusted with the Benjamini and Hochberg False Discovery Rate method with a cutoff of 0.05 in Genespring. This analysis revealed a set of 65 genes that were commonly induced in both Trials 1 and 2 (indicated by a '1' in the column entitled 'Pre-Filtered'). Of these 65 genes, the ones we chose to validate by RT-PCR are indicated by a '1' in the column entitled 'RT-PCR' The RT-PCR *P*-values, and the results of this validation process, are indicated in the columns 'RT-PCR', and 'Pre-Filtered versus RT-PCR', respectively. In the second strategy not involving pre-filtering, we tested the entire dataset using ANOVA (column entitled 'ANOVA only'), and this yielded 22 genes which were a subset of the 65 genes identified via the pre-filtering method. The majority of these genes were confirmed by RT-PCR (column entitled 'ANOVA vs RT-PCR'). Importantly, many genes that were not included in this subset of 22 genes were also confirmed by RT-PCR. Furthermore, the genes encoding CD38 and IP-10, which we have demonstrated to be expressed at the protein level (**Fig. 2** and **Supplementary Fig. 1**), were not present amongst the 22 genes. Thus, while this second method of analysis omitting the pre-filtering step may result in a more rigorous statistical analysis, it may be too stringent and exclude potentially biologically relevant genes.

Supplementary Table 2: Two transcription factors induced by YF-17D in two independent trials.

Description	**Feature Name**	**Factor Name**	***N***	***P*-value**
Interferon-Stimulated Response Element	M00258-V$ISRE_01	IRF9	27	2.49E-06
Interferon Regulatory Factor 7	M00453-V$IRF7_01	IRF7	30	7.64E-04
Sterol Regulatory Element-Binding Protein 1	M00220-V$SREBP1_01	SREBF1	15	0.005390915

Footnote: The 65 genes which were found to be induced by YF-17D in **Fig. 1b** were imported into TOUCAN for transcription factor binding site (TFBS) analysis, and 44 out of the 65 genes were recognized by TOUCAN. The TRANSFAC v7.0 database of eukaryotic transcription factors was used as the reference for transcription factor binding site motifs. Binding site motifs were scanned in the DNA sequence 2000 bases upstream through 200 bases downstream flanking the first exon of each gene with a double prior of 0.1 and the genomic background noise model based on the third order Markov Model for the Human Eukaryotic Promoter Databse. The TFBSs found to have a statistically overrepresented frequencies including: interferon-stimulated response element (*ISRE*), interferon regulatory factor 7 (*IRF7*), and sterol regulatory element-binding protein 1 (*SREBF1*).

Supplementary Table 3: RT-PCR confirmation of the genes induced by YF-17D.

Symbol	Gene ID	TaqMan Assay	*P*-value
IFI44	Hs.82316	Hs00197427_m1	0.0000
IFI44L	Hs.389724	Hs00199115_m1	0.0000
EIF2AK2, PKR	Hs.131431	Hs00169345_m1	0.0038
MX1	Hs.517307	Hs00182073_m1	0.0001
SIGLEC1	Hs.31869	Hs00224991_m1	0.0004
ISG15	Hs.458485	Hs00192713_m1	0.0001
XAF1	Hs.441975	Hs00213882_m1	0.0001
HERC5	Hs.26663	Hs00180943_m1	0.0002
IRF7	Hs.166120	Hs00242190_g1	0.0001
RSAD2	Hs.17518	Hs00369813_m1	0.0006
SERPING1, C1IN	Hs.384598	Hs00163781_m1	0.0084
PARP9	Hs.518200	Hs00230231_m1	0.0012
OAS1	Hs.524760	Hs00242943_m1	0.0012
EPSTI1	Hs.546467	Hs00264424_m1	0.0007
JUN	Hs.525704	Hs99999141_s1	0.0001
OAS3	Hs.528634	Hs00196324_m1	0.0019
PARP12	Hs.12646	Hs00224241_m1	0.0066
PLSCR1	Hs.130759	Hs00275514_m1	0.0006
OASL	Hs.118633	Hs00388714_m1	0.0005
MX2	Hs.926	Hs00159418_m1	0.0060
OAS2	Hs.414332	Hs00159719_m1	0.0057
PARP14	Hs.518203	Hs00393814_m1	0.984
DHX58, LGP2	Hs.55918	Hs00225561_m1	0.0402
DDX58, RIG-I	Hs.190622	Hs00204833_m1	0.0160
TLR7	Hs.443036	Hs00152971_m1	0.0409
GBP1	Hs.62661	Hs00266717_m1	0.0707
CD38	Hs.479214	Hs00277045_m1	0.0311
IFIH1, MDA-5	Hs.163173	Hs00223420_m1	0.0022
STAT1	Hs.651258	Hs00234829_m1	0.0528
C3AR1	Hs.591148	Hs00269693_s1	0.0732
LAMP3	Hs.518448	Hs00180880_m1	0.3765
CXCL10, IP-10	Hs.632586	Hs00171042_m1	0.2288
FAM70A	Hs.437563	Hs00215705_m1	0.4147

Footnote: Of the 65 genes induced in most vaccinees (**Fig. 1**), we selected 33 genes for RT-PCR analysis. We assayed 10 day 3 versus day 0 and 15 day 7 versus day 0 time points from the 15 subjects in Trial 1. This revealed that 26 genes also have significant modulation as measured by RT-PCR. The *P*-values are testing the result of ANOVA on the RT-PCR data for the fold changes on days 0, 3, and 7.

Supplementary Table 4: The genes validated by ClaNC as being predictive of CD8+ T cell responses from **Fig. 4**.

Symbol	Gene Name	UniGene ID	GeneBank	Day
C1QB	Complement component 1, q subcomponent, B chain	Hs.8986	CA307782	3
EIF2AK4	Eukaryotic translation initiation factor 2 alpha kinase 4	Hs.412102	BM978043	7
MEF2A	MADS box transcription enhancer factor 2, polypeptide A	Hs.268675	Y16312	7
SLC2A6	Solute carrier family 2, member 6	Hs.244378	AJ011372	7
ALDH16A1	Aldehyde dehydrogenase 16 family, member A1	Hs.355398	BU741307	7
ALDH3B1	Aldehyde dehydrogenase 3 family, member B1	Hs.523841	BC014168	3
ASGR2	Asialoglycoprotein receptor 2	Hs.16247	CR594935	3
ASGR2	Asialoglycoprotein receptor 2	Hs.16247	CR594935	7
ATP6V1E1	ATPase, H+ transporting, lysosomal 31kDa, V1 subunit E1	Hs.517338	AW804839	3
BIRC3	Baculoviral IAP repeat-containing 3	Hs.127799	BQ004306	7
BNIP3L	BCL2/adenovirus E1B 19kDa interacting protein 3-like	Hs.131226	NM_004331	7
BCKDK	Branched chain ketoacid dehydrogenase kinase	Hs.513520	AF026548	3
CAMKK2	Calcium/calmodulin-dependent protein kinase kinase 2, beta	Hs.297343	NM_172226	3
CALR	Calreticulin	Hs.515162	BM806569	3
CRAT	Carnitine acetyltransferase	Hs.12068	AI809851	3
CTSB	Cathepsin B	Hs.520898	NM_001908	3
CD69	CD69 molecule	Hs.208854	AU309880	3
N/A	CDNA clone IMAGE:5271145	Hs.385760	BC038776	7
N/A	CDNA FLJ20387 fis, clone KAIA4452	Hs.636439	AK000394	7
N/A	CDNA: FLJ20905 fis, clone ADSE00244	Hs.612877	AK024558	3
CENPB	Centromere protein B, 80kDa	Hs.516855	BM703471	3
CXCR6	Chemokine (C-X-C motif) receptor 6	Hs.34526	CR624554	3
CXCR7	Chemokine (C-X-C motif) receptor 7	Hs.471751	BX111686	3
CXCR7	Chemokine (C-X-C motif) receptor 7	Hs.471751	BX111686	7
DEFA4	Defensin, alpha 4, corticostatin	Hs.591391	NM_001925	7
EMILIN2	Elastin microfibril interfacer 2	Hs.532815	AF270513	7
ETV3	Ets variant gene 3	Hs.352672	AF218540	3

Supplementary Table 4 Continued

Symbol	Gene Name	UniGene ID	GeneBank	Day
EIF4G3	Eukaryotic translation initiation factor 4 gamma, 3	Hs.467084	AF012072	7
FBXO15	F-box protein 15	Hs.465411	DB522515	7
GPR18	G protein-coupled receptor 18	Hs.631765	AW574811	7
GBGT1	Globoside alpha-1,3-N-acetylgalactosaminyltransferase 1	Hs.495419	CR622726	3
GAA	Glucosidase, alpha; acid	Hs.1437	AL043560	3
GAS2L1	Growth arrest-specific 2 like 1	Hs.322852	BC001782	3
HEATR3	HEAT repeat containing 3	Hs.647381	AW802598	7
HBA1	Hemoglobin, alpha 1	Hs.449630	AA331275	7
HBB	Hemoglobin, beta	Hs.523443	BP424559	3
HBB	Hemoglobin, beta	Hs.523443	BP424559	7
HBZ	Hemoglobin, mu	Hs.647389	CR597411	7
HTRA4	HtrA serine peptidase 4	Hs.322452	AL574735	3
FLJ10847	Hypothetical protein FLJ10847	Hs.232054	AI014423	7
SLC47A1	Hypothetical protein LOC731157	Hs.551062	AF150372	7
IMPDH1	IMP (inosine monophosphate) dehydrogenase 1	Hs.534808	BU687473	3
JUN	Jun oncogene	Hs.525704	NM_002228	3
C8orf82	Chromosome 8 open reading frame 82	Hs.105685	AA532638	3
C8orf82	Chromosome 8 open reading frame 82	Hs.105685	AA532638	7
MYL4	Myosin, light chain 4, alkali; atrial, embryonic	Hs.463300	AJ706934	7
NANS	N-acetylneuraminic acid synthase (sialic acid synthase)	Hs.522310	AA639295	7
NRGN	Neurogranin (protein kinase C substrate, RC3)	Hs.524116	NM_006176	3
NAPRT1	Nicotinate phosphoribosyltransferase domain containing 1	Hs.493164	BM674162	3
NAPRT1	Nicotinate phosphoribosyltransferase domain containing 1	Hs.493164	BM674162	7
NP	Nucleoside phosphorylase	Hs.75514	AW519082	3
NUDT14	Nudix (nucleoside diphosphate linked moiety X)-type motif 14	Hs.526432	CA775837	3
PNPLA6	Patatin-like phospholipase domain containing 6	Hs.631863	DN993154	3

Supplementary Table 4 Continued

Symbol	Gene Name	UniGene ID	GeneBank	Day
PRAM1	PML-RARA regulated adaptor molecule 1	Hs.465812	AW135236	3
PRAM1	PML-RARA regulated adaptor molecule 1	Hs.465812	AW135236	7
RAB8B	RAB8B, member RAS oncogene family	Hs.389733	NM_016530	3
RGS1	Regulator of G-protein signalling 1	Hs.75256	BU783195	3
NDRG2	Selenium binding protein 1	Hs.632460	CN262111	7
STK17A	Serine/threonine kinase 17a (apoptosis-inducing)	Hs.268887	NM_004760	3
SMARCD3	SMARC, subfamily d, member 3	Hs.647067	CA449683	3
SLC16A5	Solute carrier family 16, member 5	Hs.592095	AI953766	3
SLC2A6	Solute carrier family 2, member 6	Hs.244378	AJ011372	3
SLC25A13	Solute carrier family 25, member 13 (citrin)	Hs.489190	AJ496569	7
SLC39A11	Solute carrier family 39 (metal ion transporter), member 11	Hs.221127	BQ017291	7
SAT2	Spermidine/spermine N1-acetyltransferase 2	Hs.10846	CK821652	3
SAT2	Spermidine/spermine N1-acetyltransferase 2	Hs.10846	CK821652	7
SPON2	Spondin 2, extracellular matrix protein	Hs.302963	DB319294	7
N/A	Transcribed locus	Hs.642649	BE464165	3
TBC1D7	TBC1 domain family, member 7	Hs.484678	BF111612	3
TEP1	Telomerase-associated protein 1	Hs.508835	CD623678	3
THAP11	THAP domain containing 11	Hs.632200	BP395356	3
ADSSL1	Transcribed locus	Hs.375179	AA927922	7
ZEB1	Transcribed locus	Hs.593418	AI806174	7
ASGR2	Transcribed locus	Hs.595979	H47090	7
ASGR2	Transcribed locus	Hs.595979	H47090	3
CPEB3	Transcribed locus	Hs.603218	AI123721	7
N/A	Transcribed locus	Hs.604822	AI370631	3
N/A	Transcribed locus	Hs.607204	AI862844	3

Supplementary Table 4 Continued

Symbol	Gene Name	UniGene ID	GeneBank	Day
N/A	Transcribed locus	Hs.649837	AA528126	3
N/A	Transcribed locus	Hs.651406	AA600976	3
N/A	Transcribed locus	Hs.652017	CA844149	3
N/A	Transcribed locus	Hs.652017	CA844149	7
N/A	Transcribed locus	Hs.652922	CA313785	3
N/A	Transcribed locus	Hs.604290	AI281031	7
TCEAL4	Transcription elongation factor A (SII)-like 4	Hs.194329	BF718552	3
TMEM176A	Transmembrane protein 176A	Hs.647116	BM663079	3
TMOD1	Tropomodulin 1	Hs.494595	AK095748	7
TUFM	Tu translation elongation factor, mitochondrial	Hs.12084	AA983218	3
TNFSF14	Tumor necrosis factor (ligand) superfamily, member 14	Hs.129708	AY028261	3
FERMT3	UNC-112 related protein 2	Hs.180535	BF975449	3
ULK2	Unc-51-like kinase 2 (C. elegans)	Hs.168762	NM_014683	7
WDR40A	WD repeat domain 40A	Hs.651274	AA446117	7
ZFP82	Zinc finger protein 545	Hs.558734	BU618382	3
ZNF606	Zinc finger protein 606	Hs.652113	BM713422	3
ZSWIM5	Zinc finger, SWIM-type containing 5	Hs.135673	BQ448086	7
ZYX	Zyxin	Hs.490415	CB160586	3