

SCIENTIFIC REPORTS



OPEN

Transcriptomic signatures of cellular and humoral immune responses in older adults after seasonal influenza vaccination identified by data-driven clustering

Emily A. Voigt¹, Diane E. Grill², Michael T. Zimmermann², Whitney L. Simon¹, Inna G. Ovsyannikova¹, Richard B. Kennedy¹ & Gregory A. Poland¹

PBMC transcriptomes after influenza vaccination contain valuable information about factors affecting vaccine responses. However, distilling meaningful knowledge out of these complex datasets is often difficult and requires advanced data mining algorithms. We investigated the use of the data-driven Weighted Gene Correlation Network Analysis (WGCNA) gene clustering method to identify vaccine response-related genes in PBMC transcriptomic datasets collected from 138 healthy older adults (ages 50–74) before and after 2010–2011 seasonal trivalent influenza vaccination. WGCNA separated the 14,197 gene dataset into 15 gene clusters based on observed gene expression patterns across subjects. Eight clusters were strongly enriched for genes involved in specific immune cell types and processes, including B cells, T cells, monocytes, platelets, NK cells, cytotoxic T cells, and antiviral signaling. Examination of gene cluster membership identified signatures of cellular and humoral responses to seasonal influenza vaccination, as well as pre-existing cellular immunity. The results of this study illustrate the utility of this publically available analysis methodology and highlight genes previously associated with influenza vaccine responses (e.g., *CAMK4*, *CD19*), genes with functions not previously identified in vaccine responses (e.g., *SPON2*, *MATK*, *CST7*), and previously uncharacterized genes (e.g. *CORO1C*, *C8orf83*) likely related to influenza vaccine-induced immunity due to their expression patterns.

Worldwide, influenza affects 5–10% of adults annually, and results in an estimated 250,000 to 500,000 deaths¹. Influenza morbidity and influenza-associated deaths increase significantly with age^{2,3}, and more than 90% of influenza-associated deaths occur in individuals ≥ 65 years of age⁴. Although seasonal influenza vaccination offers protection against severe influenza disease, levels of protection vary between seasons, individuals, and age—tending to be lower in elderly populations^{5–12}. In fact, the effectiveness of seasonal trivalent inactivated influenza vaccination among community-dwelling older adults has been estimated to be only 30–40%^{6,12–14}. With the aging of populations in the U.S. and globally, it is imperative that influenza vaccine-induced immunity in older adults be better understood^{15–17}.

Systems vaccinology and vaccinomics, the application of systems biology to the study of vaccines, are a promising method to better understand human immune responses to vaccines from a holistic perspective^{18,19}. A seminal paper by Querec *et al.* in 2008 applied a systems biology approach to study yellow fever vaccine-induced immunity and identified novel genes involved in vaccine-induced antibody and CD8+ T cell responses whose expression levels could predict immunogenicity of the vaccine in subjects, setting a new standard for vaccine studies²⁰. Such systems biology approaches provide complementary insights to reductionist approaches by revealing novel interactions between immune system processes critical to developing immune responses to vaccines²¹. Systems biology approaches have been previously applied to the study of seasonal influenza vaccination

¹Mayo Clinic Vaccine Research Group, Mayo Clinic, Rochester, MN 55905, USA. ²Division of Biomedical Statistics and Informatics Mayo Clinic, Rochester, MN 55905, USA. Correspondence and requests for materials should be addressed to G.A.P. (email: poland.gregory@mayo.edu)

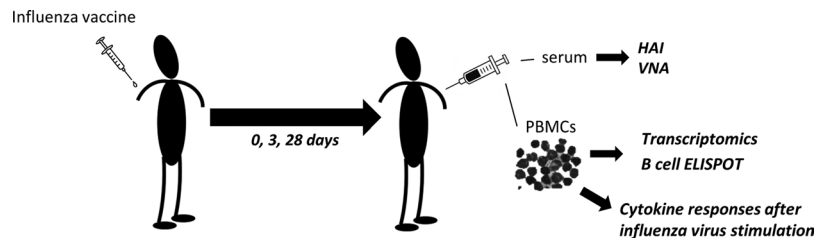


Figure 1. Study design. Adults were vaccinated with trivalent seasonal influenza vaccine. At Days 0, 3, and 28 post-vaccination, blood samples were drawn and serum antibody (HAI, VNA) titers measured. PBMCs were assayed for anti-influenza memory activity by influenza A/H1N1-specific B-cell, and mRNA transcriptomic profiles measured by NextGen sequencing. PBMCs were also stimulated *in vitro* by influenza A/H1N1 virus, and resulting cytokine secretion (IFN γ , IL1 β , IL-2, IL-4, IL-6, IL-8, IL-10, IL-12p70, IL-13, TNF α) was measured by ELISA.

in humans, resulting in the identification of important players in, and predictors of, immune responses. These include the frequencies of specific immune cell subsets, pre-existing immunity from previous influenza vaccination or infection, and gene expression signatures of immune response in *ex vivo* human immune cells^{9,22–28}.

Such systems studies of the human response to vaccination require complex analytical methods to mine important immune-related information out of large datasets. In particular, the in-depth study of transcriptional changes in peripheral blood mononuclear cells (PBMCs) post-vaccination may lead to a better understanding of the development of humoral and cellular immune responses after influenza vaccination; however, systems-level characterization of PBMC responses to vaccination requires analytical techniques to prune large transcriptomics datasets to the subset of biologically relevant genes. As transcriptomic datasets are large (thousands of genes at multiple time points), the identification of single genes as predictors of immune responses is challenging^{29,30}. This issue is exacerbated in biological situations where marginal associations are weak and/or noisy, leading to high rates of false-positive identifications. PBMCs also represent a complex mixture of immune cell types, each with its own changing pattern of gene expression, inherently providing additional complexity to transcriptomic datasets. As genes work within networks, not individually, effective analytical methods to identify important drivers of immunity that focus on groups of genes may better model the mechanisms of response³¹.

Weighted Gene Correlation Network Analysis (WGCNA) is a new data-driven clustering algorithm that can be used to identify clusters of genes that act similarly across individuals^{32,33}. This gene clustering method was developed in order to effectively study transcriptomic data from complex systems, such as human disease states and plant microbiome interactions^{32,34–36}. For biological scenarios where with low signal-to-noise ratios and weak marginal associations, WGCNA cluster analysis has been demonstrated to be more reproducible and less prone to finding false positives than marginal meta-analysis statistical techniques³⁷. WGCNA has been utilized to identify subsets of genes from transcriptomic datasets that are involved in the biological questions studied, while excluding genes that are likely unrelated^{38–40}. To date, WGCNA has been sparsely used in the study of human immunology, and thus further validation of this technique for such applications, and comparisons of results to those of previous systems studies of influenza vaccination in humans is essential. We tested the utility of WGCNA in analyzing transcriptomic profiles of PBMCs from older adults after seasonal influenza vaccination. The algorithm generated fifteen gene expression clusters, eight of which were highly enriched for immunity-related genes. These immune-relevant clusters had distinct and biologically interpretable functions, and cluster gene expression correlated with subject immune responses corresponding to those biological functions. These gene clusters allowed us to identify independent marker genes for the development of cellular (PBMC cytokine secretion) and humoral (serum antibody, B-cell ELISPOT) immunity. These results compared well with previous studies using alternative analysis approaches. Further study of these clusters identified likely involvement of specific immune cell subsets in the development of cellular, memory B-cell, and antibody immune responses.

Materials and Methods

The study population and laboratory methods described herein are similar or identical to those published in our previous studies^{41–46}.

Recruitment. A cohort of generally healthy recipients of the 2010/11 seasonal trivalent inactivated influenza vaccine (TIV; Fluarix; containing the A/California/7/2009 NYMC X-191 (H1N1), A/Victoria/210/2009 NYMC X-187 (H3N2; an A/Perth/16/2009-like virus), and B/Brisbane/60/2008 viral strains) was used for this study^{41,47}. Subjects were 50–74 years of age, and vaccination was administered by standard protocol into the deltoid muscle using a 1 inch needle. Vaccine lot #AFLUA524AA (GlaxoSmithKline) was used for all subjects. Blood samples (100 ml each) were collected before vaccination (Day 0) and at Days 3 and 28 post-vaccination (Fig. 1)⁴¹. Recruitment was performed at Mayo Clinic, Rochester, MN; a complete dataset was obtained from 138 individuals, and these datasets were used for further analysis.

Ethics Statement. Written informed consent was obtained from each participant, and the Mayo Clinic Institutional Review Board approved the study. Review Board approved the study. All methods were performed in accordance with the relevant guidelines and regulations.

The virus was propagated in nine-day-old embryonated chicken eggs obtained from Charles River Laboratories in Storrs, CT, USA. Mayo Clinic's Institutional Animal Care and Use Committee (IACUC) policy does not require review of research on unhatched embryonated eggs destroyed before hatching, in accordance with the Office of Laboratory Animal Welfare and the National Institutes of Health policy.

MDCK cells were obtained from the American Type Culture Collection.

Isolation of peripheral blood mononuclear cells. PBMCs were isolated from whole blood samples as described previously using BD Vacutainer® CPT™ Cell Preparation Tubes with sodium citrate⁴³. The cells were then resuspended in freezing medium, aliquoted, and cryopreserved for future use^{42–44}.

Growth of influenza virus. The influenza A/California/7/2009/H1N1-like virus strain used in this study was provided by the Centers for Disease Control and Prevention (Atlanta, GA). As previously described, the virus was propagated in embryonated chicken eggs and harvested from the allantoic fluid. 50% tissue culture infectious dose (TCID₅₀) measurement of virus stock titers was conducted by infection of MDCK cells with serial dilutions of the virus and addition of red blood cells (RBC) five days after infection. Hemagglutination (HA) was then determined in wells with successful virus replication^{42,44,48–50}.

Previous measurement of immune response to influenza vaccine. Hemagglutination-inhibition influenza A/H1N1 antibody (HAI) titers, virus-neutralizing influenza A/H1N1 antibody (VNA) titers, and influenza A/H1N1-specific B-cell ELISPOT counts for these subjects before and after vaccination were previously published by our group^{42,44–46}. Participants' influenza A/H1N1-specific HAI titers were measured from sera at each timepoint pre- and post-vaccination. Serum influenza A/H1N1-specific neutralizing antibody titers were measured by a cell-based microneutralization assay at each timepoint with influenza A/H1N1 virus stimulation (200 plaque-forming units per 5 µl), as previously described⁴⁶. Quantification of influenza A/H1N1-specific B cells (memory-like IgG B cells) was performed on subjects' PBMCs using the MabTech Human IgG ELISpot^{PLUS} Kit (Mabtech, Inc.; Cincinnati, OH), as described previously⁴⁴. Influenza A/H1N1 virus stock was diluted 1:50 (50,000 TCID₅₀/well) before coating the ELISPOT plates.

In vitro chemokine/cytokine assays. PBMCs were thawed, counted, and plated at 2×10^5 cells per well in 96 well plates. Six wells were plated per subject; three of these were stimulated with influenza A/California/7/2009/H1N1-like virus (MOI 0.5) and three left unstimulated. Two additional wells were stimulated with 5 µg PHA (positive control). Plates were incubated for 48 hours at 37 °C, after which supernatants were harvested, plated in a new 96-well plate, and frozen until use. The Meso Scale Diagnostics (MSD; Rockville, MD) V-PLEX Proinflammatory Panel 1 (human) Kit was used according to the manufacturer's protocol to measure 10 cytokines (IFN- γ , IL-10, IL-12p70, IL-13, IL-1 β , IL-2, IL-4, IL-6, IL-8, and TNF- α) in the aforementioned supernatants. Cytokines were categorized as T_H1 (IFN- γ , IL-2), T_H2 (IL-10, IL-4), and pro-inflammatory (TNF α , IL-6); the data for each cytokine were transformed to a uniform scale and then averaged within-category.

Statistical Methods

RNA-seq. The mRNA sequencing methods were identical to those published in previous transcriptomics studies⁵¹. Briefly, total RNA was extracted from each cryopreserved subject PBMC sample using RNeasy Plus mini kits and RNAProtect reagent (Qiagen; Valencia, CA). cDNA libraries were created in Mayo Clinic's Gene Sequencing Facility using the mRNA-Seq. 8 Sample Prep Kit (Illumina; San Diego, CA), and poly-A RNA was isolated using magnetic purification. Single-end read sequencing was performed using an Illumina HiSeq. 2000. Sequencing reads were aligned to the human genome build 37.1 using TopHat (1.3.3) and Bowtie (0.12.7). Quality control and normalization of the mRNA-sequencing gene counts data are as described by Ovsyannikova *et al.*⁵¹ Briefly, gene counts were normalized using Conditional Quantile Normalization⁵², and 14,197 genes with at least 32 counts at one of our three timepoints (Day 0, 3, or 28) were used in subsequent analyses.

WGCNA. Weighted Gene Coexpression Network Analysis (WGCNA) was used to create data-driven phenotype-agnostic gene clusters using the Day 28 normalized gene expression data^{33,53,54}. For all pairs of genes, a co-expression similarity matrix was calculated as $S_{ij} = |0.5 + 0.5 \cdot \text{cor}(x_i, x_j)|$. This signed similarity matrix was used to construct an adjacency matrix, calculated as $A_{ij} = S_{ij}^\beta$, with soft-thresholding to preserve the strength of the correlation between the genes. The parameter β was selected as the smallest value that achieves scale free topology ($\beta = 12$ for this analysis). The topological overlap dissimilarity was calculated from the adjacency matrix, and hierarchical clustering was used to define the gene clusters. Each gene cluster's eigengene is used as a summary of the cluster's gene expression activity, representing the first principal component of the gene expression levels within the cluster. Pearson's correlation was used to correlate subject immune phenotypes with each cluster's eigengene to identify which clusters were related to vaccination responses.

Gene enrichment analyses. We performed enrichment analysis using the RITAN⁵⁵ package (<https://github.com/MTZimmer/RITAN>) and leveraging previously published Blood Translation Modules (BTMs)⁵⁶.

Cluster-cluster interaction analysis. Databases of known regulatory interactions were examined using RITAN to further understand the biologic relationships between the individual genes found in the WGCNA gene clusters. Interactions from the Pathway Interaction Database⁵⁷, Transcription Factor Encyclopedia⁵⁸, a directed protein interaction network⁵⁹, HPRD⁶⁰, CCSB⁶¹, HumanNet⁶² (minimum score ≥ 0.0), and STRING⁶³ (minimum score ≥ 700) were accessed using RITAN. This integrated network resource provided a comprehensive summary of known interactions across the human proteome.

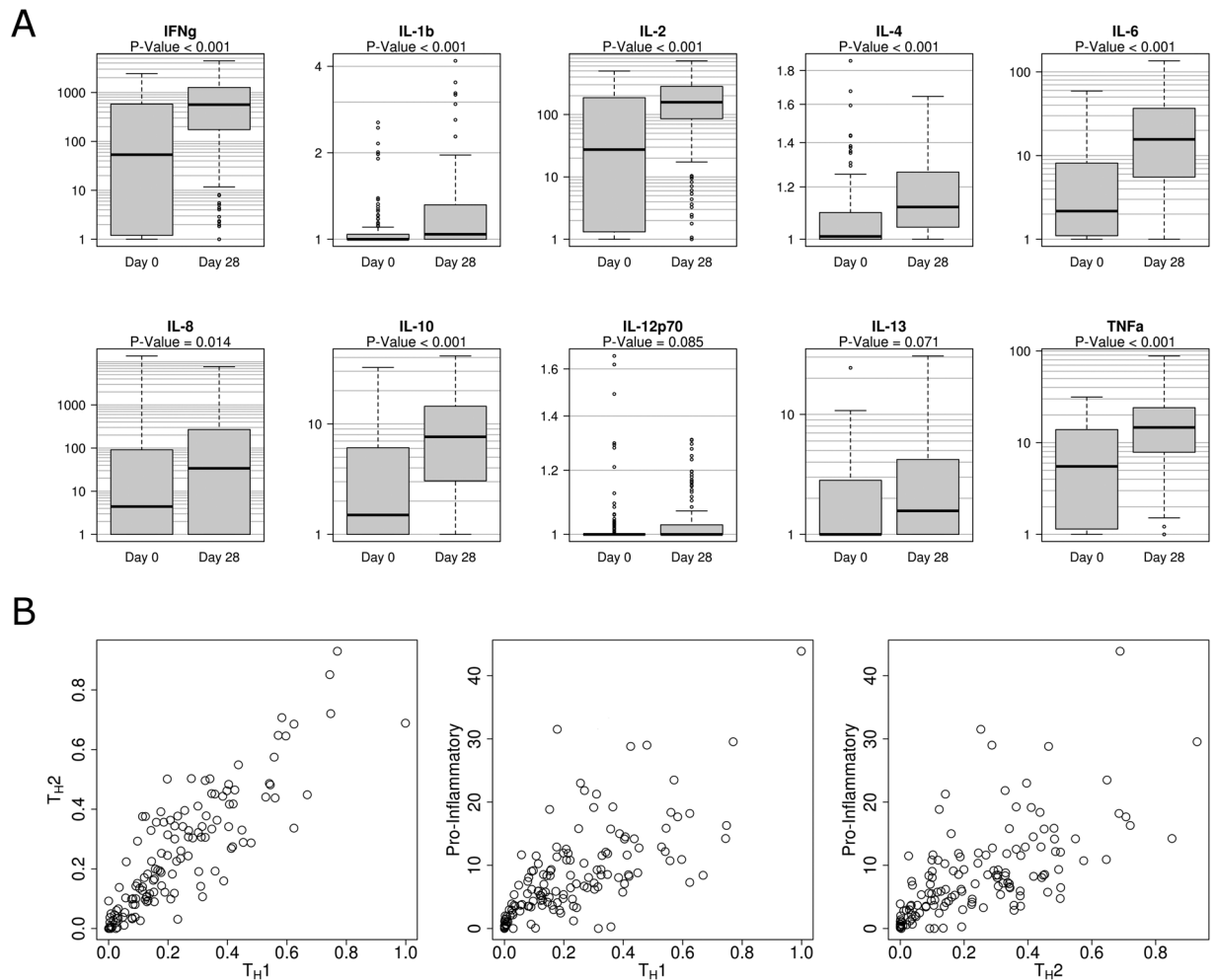


Figure 2. Influenza vaccination primes human PBMCs to secrete cytokines in response to influenza A/H1N1 virus stimulation. Panel A: Influenza A/H1N1-stimulated cytokine responses in PBMCs harvested prior to seasonal influenza vaccination were measurable, representing significant pre-existing immunity. However, cellular cytokine responses were significantly boosted above pre-vaccination levels for T_H1 (IFN- γ , IL-2), T_H2 (IL-4, IL-10), and proinflammatory (IL-6, TNF- α) cytokines. IL-1 β , IL-12p70, and IL-13 cytokines were secreted at low levels approaching the assay limit of detection. IL-8 showed high levels of assay variability. Panel B: T_H1, T_H2, and pro-inflammatory post-vaccination (Day 28) cytokine responses are highly correlated.

Using this interaction network, we computed a gene cluster interaction metric quantifying the extent to which the genes in two clusters are directly regulated by one another. The fraction of genes in one gene cluster that interacted with genes in the second gene cluster was calculated, then compared to the expected interaction of two randomly-generated gene clusters of the same size, as follows: for gene clusters $s^1 = \{g_1^1, g_2^1, \dots, g_n^1\}$ and $s^2 = \{g_1^2, g_2^2, \dots, g_m^2\}$ that contain $||s^1|| = n$ and $||s^2|| = m$ genes, let the neighbors of a gene be $A(g)$. Further, let the neighbors of genes in cluster s^1 that are also members of cluster s^2 be denoted $A(s^1)^2$ and U the cluster union. Finally, let the overlap between the two clusters be: $O(s^1, s^2) = U(A(s^1)^2, A(s^2)^1)$. The interactivity is formally defined as: $G = ||O(s^1, s^2)|| / ||U(s^1, s^2)||$. We used permutation tests to judge the significance of observed interactivity scores. To assess the significance of an individual cluster-cluster interactivity score, randomly selected groups of genes of the same respective sizes as the clusters were generated and interactivity was calculated. This was repeated 1,000 times, resulting in a randomly generated distribution of interactivity scores. We declared the original cluster-cluster interactivity value as significant when it was at least three median-based standard deviations away from the median of this empirical null distribution (p-value < 0.001).

Results

Vaccination boosts cytokine responses of human PBMCs. In order to assess PBMC cytokine recall responses in response to influenza vaccination, we sampled subject PBMCs immediately prior to vaccination and 28 days post-vaccination. Harvested PBMCs from 138 subjects were stimulated *in vitro* with A/H1N1 influenza virus, and supernatants were removed after 48 hours and assayed for secreted cytokines (Fig. 2).

PBMCs collected post-vaccination produced significantly more IFN- γ , IL-1 β , IL-2, IL-4, IL-6, IL-10, and TNF α in response to *in vitro* influenza A/H1N1 virus stimulation than PBMCs harvested prior to vaccination and identically stimulated (Fig. 2, panel A). Some individuals demonstrated high pre-vaccination PBMC cytokine

recall responses, which is evidence of pre-existing immunity to influenza A/H1N1. Levels of IL-1 β , IL-12p70, and IL-13 were deemed too low for meaningful analysis (i.e., post-stimulation samples below the assay limit of detection in 31%, 80%, and 20% of subjects, respectively). IL-8 responses were highly variable between replicates (standard deviation between replicates was more than triple that of other cytokines), and these data were accordingly removed from further analyses.

T_H1 cytokines (IFN- γ , IL-2) were more highly secreted than pro-inflammatory cytokines (IL-6, TNF α), which are in turn more highly secreted than T_H2 cytokines (IL-4, IL-10). Secreted cytokine levels in individuals are highly inter-correlated, with subjects typically showing consistently high (or low) secretion across T_H1, T_H2, and pro-inflammatory cytokines (Fig. 2, panel B). Secreted cytokines were not found to correlate significantly with VNA, HAI, or B-cell ELISPOT responses (data not shown).

WGCNA clusters transcriptomic data from human PBMCs after vaccination according to expression patterns across subject.

Gene expression (mRNA-Seq) data was previously collected from study subjects' PBMC samples collected before and after vaccination^{51,64}. This effort successfully measured expression levels of 14,197 genes across subjects before, three days after, and 28 days after vaccination. Full transcriptomic and immune datasets were finally available after strict quality control for 138 study subjects.

In order to identify genes that correlate highly with immune responses and are influential in vaccine-induced immunity, we clustered gene-expression data into groups of similarly behaving genes across our subjects using WGCNA³³. This data-driven clustering method used average-linkage hierarchical clustering to group genes according to their expression patterns across the subjects. WGCNA clustering of Day 28 gene-expression levels, based on the expression across the 138 subjects with full transcriptomic and immune datasets, resulted in the construction of 14 non-overlapping gene clusters (Supplementary Table S1). The size of these gene clusters ranged from 54 genes to 2,790 genes. For simplicity, we assigned a color to each gene cluster and refer to each specific cluster by color coding throughout this work. There were 1,112 genes that did not fit into other clusters and were grouped into a fifteenth "grey" cluster.

WGCNA gene clusters are highly enriched for specific immune cell types and/or immune-related functions.

To investigate the biologic relevance of the gene clusters created by WGCNA, gene enrichment analysis was conducted on the 100 genes in each cluster with the highest correlation with the cluster eigengene, using Blood Translation Modules (BTMs) as references. BTMs were developed from a systems-biology study of responses to five human vaccines and represent many of the known gene signatures of vaccine-related immunological processes and immune cell types⁵⁶. Results of the enrichment analysis are shown in Table 1. Similar enrichment results were obtained using the DAVID functional annotation and Gene Ontology analysis (data not shown)^{65–67}.

The WGCNA algorithm, acting solely on transcriptomic data, was able to separate clusters of genes with similar transcriptomic patterns across the 138 subjects (i.e., high responding vs. low responding subjects). These gene clusters were associated in turn with a variety of BTMs corresponding to PBMC functions (Table 1). The Tan cluster, for example, is highly enriched for genes involved in cell cycle and transcription, while the Greenyellow cluster is highly enriched for genes involved in B cell signatures and responses. The Yellow gene cluster contains a large number of genes related to monocytes, cell cycle and transcription, and inflammatory signaling (including innate antiviral responses). The Magenta cluster is highly enriched for genes involved in platelet activation, myeloid cells, and monocytes. The Salmon and Purple clusters are enriched for genes active in NK cells, and the Pink cluster appears to be involved in monocyte and immune activation. Finally, the Green and Black gene clusters are enriched for T cell differentiation and activation. The remaining clusters—Blue, Brown, Cyan, Turquoise, Red, and Grey—were not enriched for any BTM gene signatures with currently known functions.

Expression of genes in WGCNA-derived gene clusters correlates with different aspects of measured immune responses to vaccination.

To examine the possible function of genes within the WGCNA clusters, we examined the correlation between the expression of each gene and the subjects' measured immune outcomes: antibody responses (HAI/VNA)^{42,44,46}, memory B cell responses (B-cell ELISPOT)⁴⁴, and PBMC cytokine recall responses (secretion of IFN- γ , IL-2, IL-4, IL-6, IL-10, and TNF- α) (Supplementary Table S2). A heatmap of these correlations is shown in Fig. 3, left panel. The eigenvector of each cluster (first principal component of the cluster genes' transcriptional data, representing the overall expression behavior of genes in the cluster) was also tested for correlation with subjects' immune outcomes using Pearson's correlation test (Fig. 3, right panel).

The WGCNA clustering algorithm separated the 14,197-gene transcriptomic dataset into biologically meaningful clusters of similarly behaving genes. The WGCNA-derived gene clusters that were highly enriched for specific immune-related gene functions (Table 1) also exhibited significant correlations with relevant subject immune outcome measures. Gene clusters that were enriched for T cell activation, antiviral signaling, and monocyte activation proved to be correlated with PBMC cytokine recall responses (Red, Black, Green, Cyan, Pink, Yellow; 4,963 total genes). Gene clusters enriched for NK and T cell activity (Salmon and Purple; 321 genes total) were found to correlate with subject antibody responses (HAI and VNA titers).

The cluster most strongly associated with B-cell activity (Greenyellow; 135 genes) was appropriately found to strongly correlate with subject B-cell ELISPOT responses. Thus, through this clustering technique, the total transcriptomic dataset (14,197 genes) was narrowed to 135 similarly acting genes likely involved in the development and maintenance of influenza A/H1N1-specific B cells. Notably, these 135 genes are not significantly correlated with PBMC cytokine recall responses to influenza A stimulation. B cell ELISPOT activity is also more weakly correlated with another gene cluster enriched for T cell activity (Black cluster; 629 genes).

Cluster name	# genes	Enriched for: (Top 3 BTMs)	Strength of enrichment (p-value)	Summary class
Tan	115	mitotic cell cycle (M4.7)	$<10^{-25}$	Cell cycle
		cell cycle and transcription (M4.0)	$<10^{-25}$	
		cell cycle (I) (M4.1)	$<10^{-25}$	
Greenyellow	135	plasma cells & B cells, immunoglobulins (M156.0)	$<10^{-25}$	B cell activity
		enriched in B cells (I), (II), (VI) (M47.0, M47.1, M69)	$<10^{-25}$	
		B cell surface signature (S2)	$10^{-23.3}$	
		enriched in B cells (III)	$10^{-13.2}$	
Yellow	1780	enriched in monocytes (II) (M11.0)	$<10^{-25}$	Monocytes and inflammation
		cell cycle and transcription (M4.0)	$<10^{-25}$	
		Monocyte surface signature (S4)	$10^{-11.7}$	
		TLR and inflammatory signaling (M16)/lysosome (M209)	$10^{-10.5}$	
Magenta	269	Platelet activation - actin binding (M196)	$10^{-22.3}$	Platelets and monocytes
		platelet activation and blood coagulation (M199)	$10^{-14.5}$	
		enriched in myeloid cells and monocytes (M81)	$10^{-14.9}$	
Salmon	96	NK cells surface signature (S1)	$10^{-21.6}$	NK cell activity
		enriched in NK cells (I) (M7.2)	$10^{-17.2}$	
		enriched in NK cells (II) (M61.0)	$10^{-9.9}$	
Purple	225	enriched in NK cells (I) (M7.2)	$<10^{-25}$	NK and T cell activity
		enriched in NK cells (II) (M61.0)	$10^{-5.8}$	
		enriched in T cells (I) (M7.0)	$10^{-5.8}$	
Pink	289	immune activation - generic cluster (M37.0)	$10^{-6.7}$	Monocyte activation
		enriched in monocytes (II) (M11.0)	$10^{-6.6}$	
		CCR1, γ and cell signaling (M59)	$10^{-4.9}$	
Green	1144	enriched in T cells (I) (M7.0)	$10^{-5.6}$	T cell activation
		T cell activation (I) (M7.1)	$10^{-4.7}$	
		T cell activation (III) (M7.4)	$10^{-2.8}$	
Black	629	T cell activation (I) (M7.1)	$10^{-4.5}$	T cell activation
		enriched in T cells (I) (M7.0)	$10^{-4.4}$	
		T cell activation (III) (M7.4)/T cell differentiation (M14)	$10^{-2.6}$	
Blue	2504	TBA (M177.0)	$10^{-2.5}$	none
		No further enrichment found	—	
Brown	1988	TBA (M177.0)	$10^{-0.9}$	none
		No further enrichment found	—	
Grey	1112	No enrichment found	—	none
Turquoise	2790	No enrichment found	—	none
Red	1067	No enrichment found	—	none
Cyan	54	No enrichment found	—	none

Table 1. WGCNA-determined gene clusters are strongly enriched for genes linked to particular cell types and functions. The top 100 genes in each gene cluster most concordant with the cluster eigengene were tested for enrichment for BTMs. The top 3 enriched-for BTMs are listed for each cluster.

WGCNA cluster gene membership allows for identification of biologically-relevant gene markers of vaccine-induced immunity. To see if WGCNA clustering of gene expression data can be used to detect biologically relevant gene markers of vaccine-induced immunity, we examined the gene membership in the WGCNA clusters that correlate with subject immune outcomes. The top five genes in each cluster—with behavior most closely mirroring each cluster’s “average” eigengene behavior—are listed in Table 2.

We found that the top genes in each cluster typically display clear biological association with the summary class (See Table 1) of that cluster, and logically link to the immune outcomes with which they correlate. Some genes are better correlated with immune outcomes than others, and may therefore be better gene markers of these immune outcomes. WGCNA allows creation of gene signatures of different aspects of vaccine-induced immunity with clear biological meaning, using data-driven methods alone, unbiased by current knowledge.

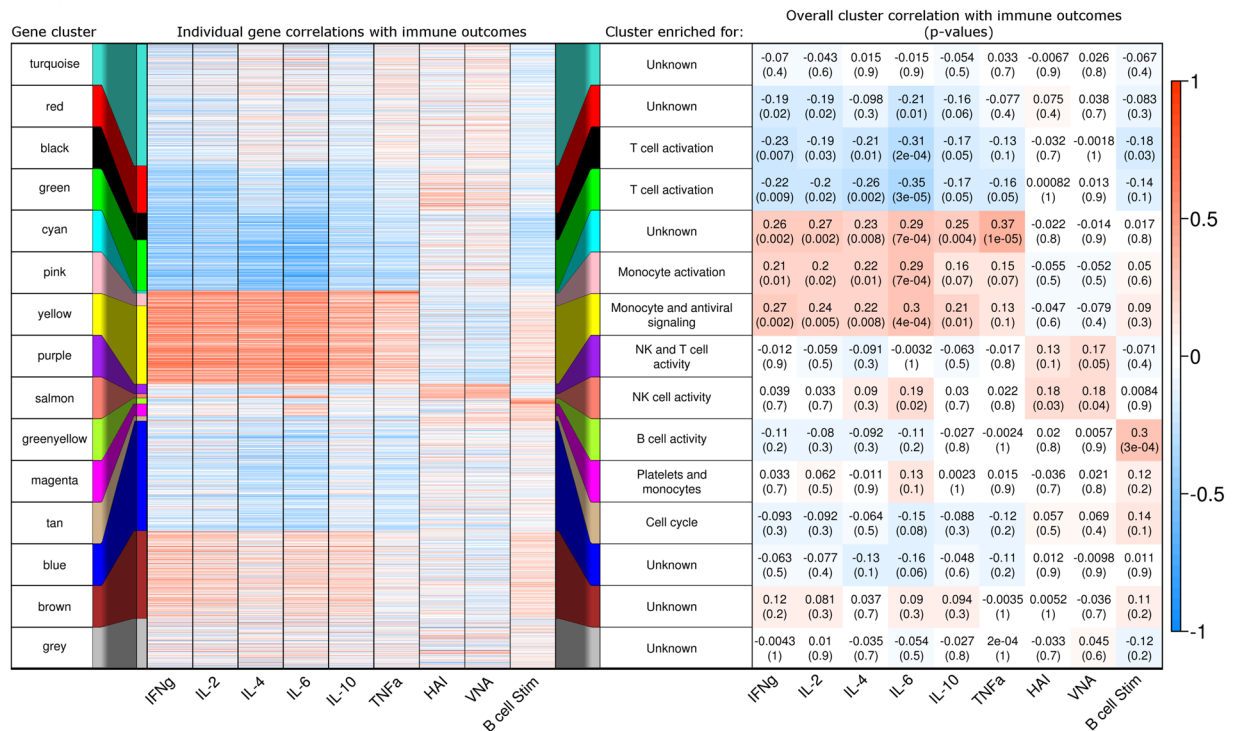


Figure 3. Gene clusters (transcriptomic data 28 days post-vaccination) correlate meaningfully with immune response outcomes after vaccination. Gene expression (mRNA-Seq) data from PBMC samples of 138 older adults after influenza vaccination was clustered using WGCNA. Cluster membership is indicated by color. Left panel: heatmap of the correlation between each gene (row) and subjects' immune response outcomes – cytokine recall responses (IFN- γ , IL-2, IL-4, IL-6, IL-10, TNF- α), H1N1 antibody responses (HAI/VNA), and memory B-cell responses (B-cell ELISPOT). Center column: cluster enrichment labels from Table 1. Right panel: Pearson's correlation coefficient and p-value for relationships between cluster eigenvector (representing overall behavior) and subject immune outcomes. Heatmap shading indicates strength of correlations in both panels; see scale bar at right.

WGCNA-derived gene clusters represent distinct but interacting immune processes. To characterize the relationships of these non-overlapping WGCNA-derived gene clusters with one another, we examined the interaction of genes in each cluster with the genes in each of the other gene clusters. For each pair of gene clusters, we calculated the fraction of genes in one cluster that have known interactions with genes in the second cluster, and used this to calculate a cluster-cluster interaction score indicating whether the two gene clusters are enriched (or depleted) for interacting genes. The matrix of cluster-cluster interconnectivity scores is displayed in Fig. 4. Groups of highly interacting gene-expression clusters can be seen along the diagonal.

The interconnectivity analysis resulted in five major groups of gene clusters, as illustrated by the boxed clusters in Fig. 4. Highly interacting gene clusters tended to correlate with similar immune response outcomes. For example, the expression of genes in the Yellow, Cyan, and Pink cluster grouping all correlate positively with cytokine responses (Fig. 3). These three clusters represent three sets of genes with different expression patterns from one another, yet interact heavily with one another and correlate with the same immune outcomes. Conversely, the highly interacting Red-Black-Green gene clusters correlate with similar biological responses (negatively correlated with cytokine recall responses from PBMCs), yet the Black cluster is not particularly interconnected with the Red cluster. This illustrates that different gene sets may contribute to the same immune outcomes, yet they do not display high levels of gene-gene interactions. The Green cluster is highly interconnected with all other clusters, possibly through the inclusion of a few particularly well-linked genes (i.e., transcription factors, ribosomal proteins).

The Salmon, Purple, and Pink clusters are enriched for interacting genes. The antibody-correlated Salmon and Purple clusters interact with Pink cluster genes, which appear to link these influenza antibody-response related genes with those involved in cytokine recall responses (i.e., the Cyan, Yellow, Magenta, and Black clusters). While the Salmon and Purple cluster genes interact highly with one another, and both of these clusters correlate with antibody titers, we find evidence that they do have distinct behavior from one another: the Salmon cluster is also correlated with IL-6 responses, while the Purple cluster is not. Finally, the Greenyellow cluster interacts solely with the Black cluster, which is the only other gene cluster correlated with B-cell ELISPOT outcomes. This suggests that these genes are exclusively involved in driving ELISPOT responses.

Time development of correlations between PBMC cytokine recall responses with gene expression clusters reveals signatures of pre-existing immunity. To look for pre-existing gene expression signatures that predict influenza immune outcomes, we examined the relationship of baseline (Day 0) gene

Immune Outcome	Genes from Cluster:	Cluster enrichment summary	Top 5 genes	Correlation with B cell ELISPOT	Correlation p-value	Gene name
B-cell ELISPOT	Greenyellow	B cell activity	<i>CD79A</i>	0.31	2.1E-04	B-cell antigen receptor complex-associated protein alpha chain
			<i>RALGPS2</i>	0.30	2.9E-04	Ras-specific guanine nucleotide-releasing factor
			<i>CD22</i>	0.28	9.1E-04	B-cell receptor
			<i>CD19</i>	0.33	8.1E-05	B-lymphocyte antigen
			<i>FCRLA</i>	0.37	7.6E-06	Fc receptor-like A
Immune Outcome	Genes from Cluster:	Cluster enrichment summary	Top 5 genes	Correlation with HAI titers	Correlation p-value	Gene name
Antibody titers	Salmon	NK cell activity	<i>SPON2</i>	0.24	4.0E-03	Spondin-2
			<i>KLRF1</i>	0.19	2.5E-02	Killer cell lectin-like receptor subfamily F member 1
			<i>AKRIC3</i>	0.23	6.3E-03	Aldo-keto reductase family 1 member C3
			<i>NCR1</i>	0.26	2.4E-03	Natural cytotoxicity triggering receptor 1
			<i>PRF1</i>	0.20	1.8E-02	Perforin-1
	Purple	NK and T cell activity	<i>FCRL6</i>	0.11	1.9E-01	Fc receptor-like protein 6
			<i>GZMA</i>	0.15	8.3E-02	Granzyme A
			<i>MATK</i>	0.16	6.6E-02	Megakaryocyte-associated tyrosine-protein kinase
			<i>CST7</i>	0.12	1.6E-01	Cystatin-F
			<i>NKG7</i>	0.15	7.5E-02	G-CSF-induced gene 1 protein
Immune Outcome	Genes from Cluster:	Cluster enrichment summary	Top 5 genes	Correlation with <i>in vitro</i> IL-6 secretion	Correlation p-value	Gene name
PBMC cytokine recall	Yellow	Monocytes and inflammation	<i>ANKRD50</i>	0.22	8.2E-03	Ankyrin repeat domain-containing protein 50
			<i>RAB31</i>	0.22	1.1E-02	Ras-related protein Rab-31
			<i>CORO1C</i>	0.29	6.6E-04	Coronin-1C
			<i>C8orf83</i>	0.21	1.4E-02	Triple QxxK/R motif-containing protein
			<i>CD68</i>	0.22	9.7E-03	Macrosialin
	Pink	Monocyte activation	<i>USP32</i>	0.23	6.2E-03	Nucleolar transcription factor 1
			<i>VPS37C</i>	0.26	2.3E-03	Ubiquitin carboxyl-terminal hydrolase 32
			<i>DIRC2</i>	0.23	7.9E-03	Vacuolar protein sorting-associated
			<i>NHS</i>	0.13	1.2E-01	Disrupted in renal carcinoma protein 2
			<i>STX11</i>	0.25	2.8E-03	26S protease regulatory subunit 8
	Cyan	None	<i>PTPRS</i>	0.19	2.9E-02	Receptor-type tyrosine-protein phosphatase E
			<i>KCNK10</i>	0.11	1.8E-01	Potassium channel subfamily K member 10
			<i>CLEC4C</i>	0.24	5.1E-03	C-type lectin domain family 4 member C
			<i>CUX2</i>	0.19	2.8E-02	Homeobox protein cut-like 2
	Green	T cell activation	<i>LILRA4</i>	0.21	1.4E-02	Leukocyte immunoglobulin-like receptor subfamily A member 4
			<i>RPL13A</i>	-0.32	1.0E-04	60S ribosomal protein L13a
			<i>RPL3</i>	-0.33	9.5E-05	60S ribosomal protein L3
			<i>OCIAD2</i>	-0.30	4.3E-04	OCIA domain-containing protein
			<i>RPL10</i>	-0.32	1.1E-04	60S ribosomal protein L10
	Black	T cell activation	<i>SNRPN</i>	-0.28	1.1E-03	small nuclear ribonucleoprotein-associated protein N
<i>CDR2</i>			-0.32	1.3E-04	Cerebellar degeneration related protein 2	
<i>FLT3LG</i>			-0.34	4.7E-05	Fms-related tyrosine kinase 3 ligand	
<i>MLLT3</i>			-0.22	8.1E-03	Protein AF-9	
<i>CAMK4</i>			-0.36	1.2E-05	Calcium/calmodulin-dependent protein kinase type IV	
Red	None	<i>SFMBT1</i>	-0.27	1.1E-03	Scm-like with four MBT domains protein 1	
		<i>PRPF39</i>	-0.17	4.7E-02	Pre-mRNA processing factor 39	
		<i>ZNF37BP</i>	-0.21	1.3E-02	Zinc finger protein 37B, pseudogene	
		<i>TUBGCP6</i>	-0.17	4.8E-02	Gamma-tubulin complex component 6	
			<i>CUL9</i>	-0.17	5.2E-02	Cullin-9
			<i>NEURL4</i>	-0.12	1.4E-01	Neuralized-like protein 4

Table 2. Top five genes in immune-related WGCNA gene clusters. The five genes in each WGCNA cluster that are most closely correlated with the cluster eigengene are listed, along with their correlation with the relevant immune outcomes.

expression data with subject immune outcomes 28 days after vaccination. To do so, we clustered Day 0 gene expression data according to WGCNA clusters described above. The correlation of each gene cluster's eigengene (median expression behavior) at Day 0 with PBMC cytokine recall responses measured at Day 28 is shown in Fig. 5.

Genes in clusters related to PBMC cytokine recall responses showed baseline (Day 0) expression levels that correlated with post-vaccination PBMC cytokine responses. This correlation suggests the presence of significant pre-existing cellular immunity to influenza virus. This method has identified transcriptomic signatures of pre-existing cellular immunity embedded in PBMC transcriptomic profiles prior to vaccination. We found transcriptional signatures of preexisting immunity for PBMC cytokine responses specific to influenza A/H1N1, but did not find equivalents for antibody HAI/VNA or memory B-cell ELISPOT responses to the same virus.

Time development of correlations between humoral and cellular immunity measures and gene expression clusters reveals stages of humoral and cellular vaccine response. To further examine the development of our gene expression correlations with antibody and B-cell ELISPOT, for which we did not find transcriptomic signatures of pre-existing immunity, we clustered baseline and early responding gene expression data (Day 0 and Day 3) according to the WGCNA clusters developed using the Day 28 gene expression data. The correlation of the eigengene (overall behavior) of each gene cluster at each time point with final antibody (VNA) and memory B-cell (ELISPOT) responses was also calculated. Significant correlations are shown in Fig. 6, along with correlations between each of the top 50 genes in each cluster and the corresponding Day 28 immune outcome (inset panels).

Development of the transcriptomic clusters' correlation with immune outcomes reflects the progression of immune responses to vaccination. Initially (Day 0), the Yellow "monocytes and inflammation" gene cluster shows significant negative correlation with peak antibody (VNA) titers, which disappears after vaccination. Late in response, the Purple "NK and T cell activity," and Salmon "NK cell activity" gene clusters show correlation with antibody titer (Fig. 6A).

While baseline expression of genes in the Greenyellow gene cluster does not correlate with B-cell ELISPOT responses to vaccination, transcriptomic activity of these genes three days after vaccination is related to ultimate ELISPOT responses, and this relationship is maintained through Day 28 (Fig. 6B). Expression of genes in the Greenyellow cluster appears to be important for development of B-cell immune memory to the influenza vaccine strains, both early and late in responses.

Discussion

Development of humoral and cellular immunity to influenza after seasonal influenza vaccination is a complex process integrating the activity of multiple immune cells. Transcriptomic data from human PBMCs sampled post-vaccination contains a wealth of valuable information about human immune responses to vaccination; however, such data is complex and difficult to meaningfully analyze. We used WGCNA, a hierarchical clustering technique, to examine PBMC gene expression data from 138 subjects before and after vaccination and more fully examine relationships between gene expression and subject immune outcomes.

When applied to post-vaccination PBMC transcriptomic data, the data-driven WGCNA algorithm clustered the 14,197 genes into 15 gene clusters, which ranged in size from 2,790 to 54 genes. Gene expression in nine of these clusters (5,419 genes) demonstrated significant correlation with immune responses measures. Six gene clusters (4,963 genes) correlated with PBMC cytokine recall responses, and two clusters (321 genes) correlated significantly with influenza A/H1N1 neutralizing antibody measures (Purple, Salmon). A single small 135-gene cluster (Greenyellow) correlated strongly with B-cell ELISPOT measures.

Gene expression relationships with post-vaccination PBMC cytokine recall responses. Cross-reactive T-cell and other cellular memory responses are important for long-term protection against influenza in older adults. While not yet well worked out, *ex vivo* cellular immune response measures have been associated with vaccine-induced protection in the elderly, while serum antibody response may be a poorer measure of vaccine efficacy in older people when used alone^{6,68,69}.

Our data identified gene clusters related to cellular immunity after vaccination, potentially elucidating the processes underlying the development of this immunity. Of the six gene clusters that correlated with our *ex vivo* measures of PBMC cytokine recall responses, we identified two negatively correlated clusters (Black, Green) enriched for T cell genes. The genes whose behavior is most representative of these clusters included ribosomal proteins (RPL13A, RPL3, RPL10, SNRPN), and FLT3L, which is a key gene acting to support dendritic cell action and enhance systemic T cell and humoral immunity after vaccination^{70,71}. CAMK4, a kinase previously implicated in transcriptional regulation in lymphocytes⁷², and MLLT3 (aka AF9), a transcriptional regulator known to be involved in hematopoiesis and involved in monocyte-macrophage maturation^{73,74}, were both prominent members of the Black gene cluster, suggesting that these genes have significant roles in immune responses to vaccination. CAMK4 has been previously identified as a key negative regulator of antibody responses to trivalent seasonal influenza vaccines²⁵, thus validating the ability of WGCNA to identify key genes regulating immune responses after vaccination. The role of MLLT3 in vaccine responses has not been previously described. MLLT3 is a component of the super elongation complex that functions to increase the rate of RNA polymerase II transcription. MLLT3 is a regulator of human hematopoiesis, where mutations are associated with various leukemias⁷⁵, and increased expression promotes the output of erythroid and megakaryocytic progenitors⁷³. It is possible that lower expression of this protein instead encourages the formation of monocytes, thus affecting immune responses to vaccination. Such hypotheses should be further explored in future studies.

We also detected transcriptomic signatures of pre-existing cellular immunity to influenza (Day 0 gene expression that correlated with PBMC cytokine secretion) in these Red, Black, Green, Cyan, Pink, and Yellow clusters that

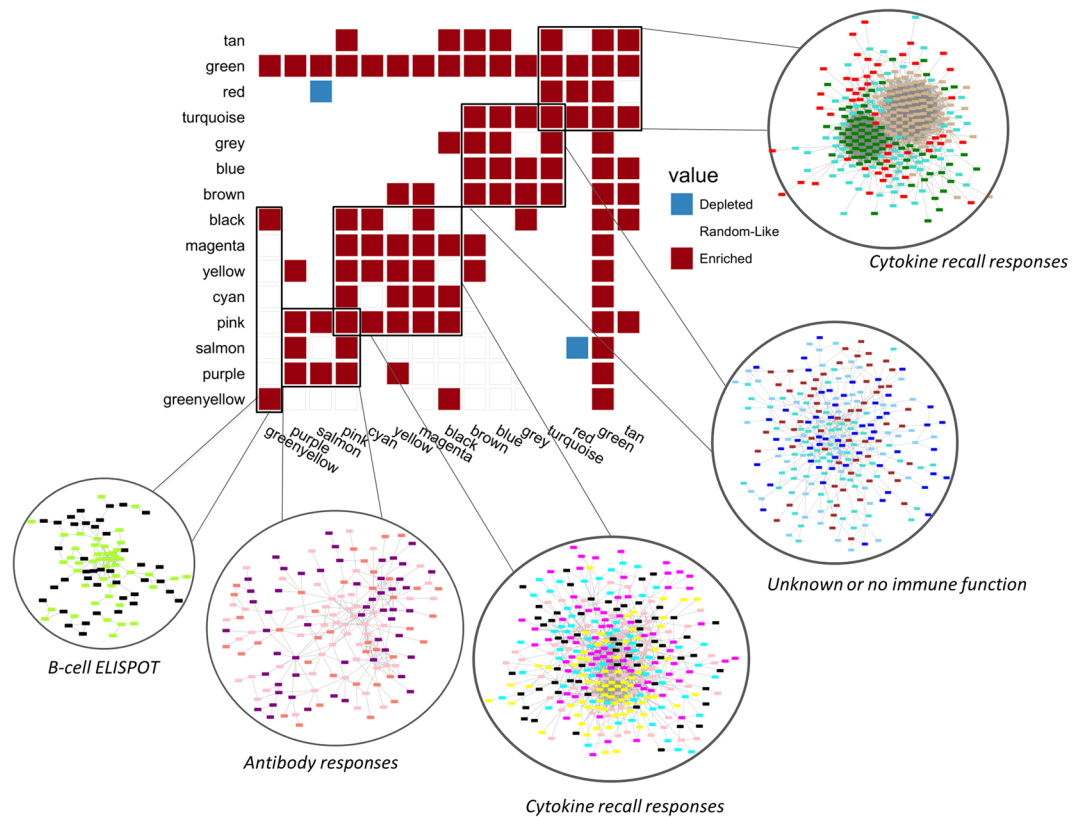


Figure 4. Cluster interconnectivity matrix. The cluster interconnectivity matrix shows gene cluster pairs found to be enriched or depleted for cross-cluster gene interactions. Red squares indicate that two gene clusters are enriched for cross-cluster gene interactions, blue squares indicate two clusters depleted for cross-cluster gene interactions, and a lack of either enrichment or depletion is denoted in white. Groups of highly interconnected gene clusters are identified (black boxes), and network diagrams of these meta-clusters, displaying known gene-gene interactions between the top 100 individual genes from each involved cluster, are displayed in the inset circles.

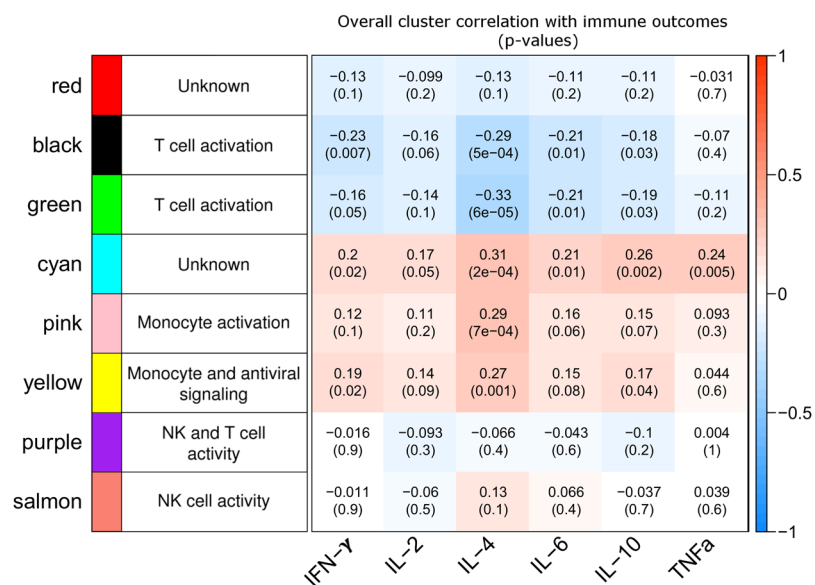


Figure 5. PBMC gene expression prior to vaccination (Day 0) correlates with post-vaccination (Day 28) PBMC cytokine recall responses. Gene expression data at baseline (Day 0) was clustered according to the Day 28 WGCNA gene clusters, and the correlation coefficient of each cluster's Day 0 eigengene with PBMC cytokine responses is shown, colored as in previous figures.

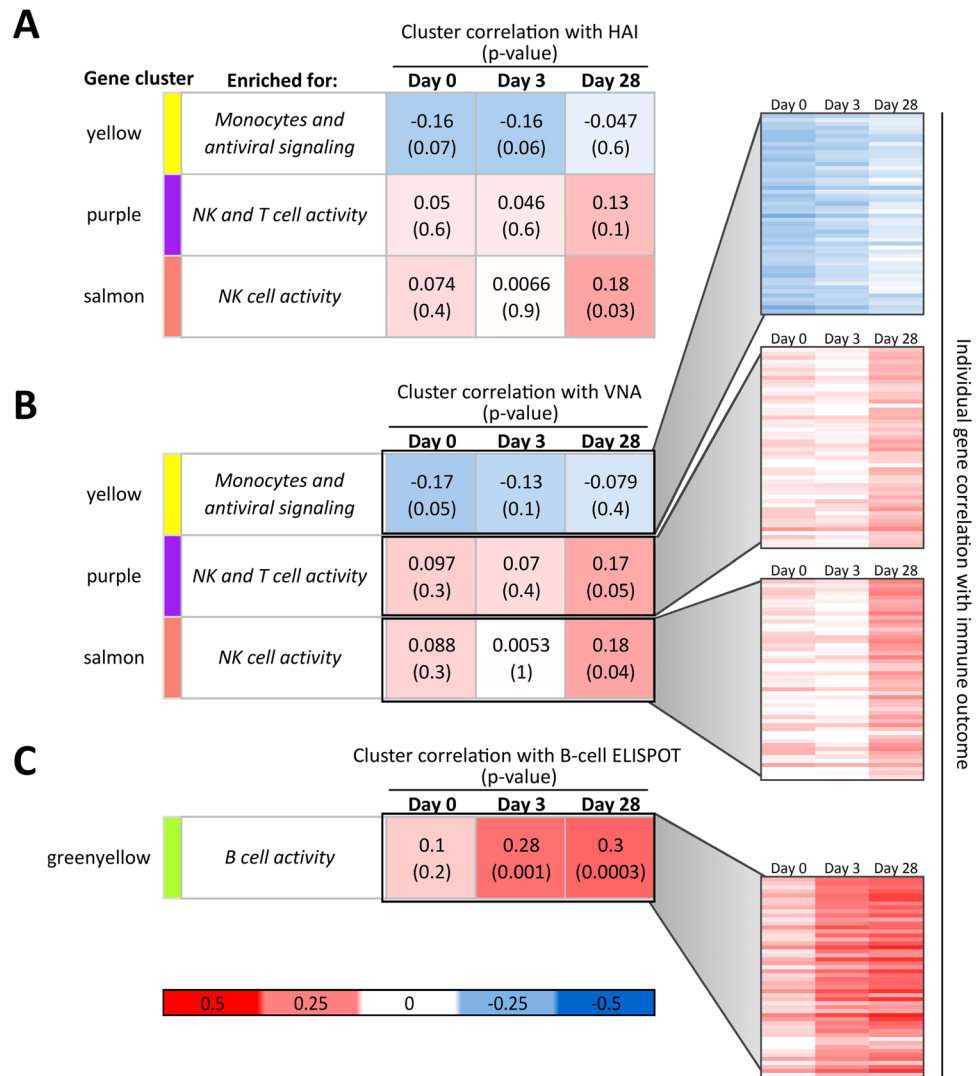


Figure 6. Time development of gene expression correlation with antibody and B-cell ELISPOT immune outcomes reveals temporal patterns leading to vaccine responses. Gene expression data at baseline (Day 0), Day 3, and Day 28 after vaccination was clustered according to the Day 28 WGCNA gene clusters, and the correlation of each cluster's eigengene with final antibody (panel A) and B-cell ELISPOT (panel B) immune outcomes was calculated at each time point. Inset panels: time development of individual genes' correlation with indicated immune outcomes for the top 50 genes in each displayed cluster.

correlated with cytokine recall responses. We note that no equivalent transcriptional signatures for pre-existing antibody or B-cell ELISPOT responses were found. This may indicate that pre-existing B cell responses require little maintenance on a transcriptional level, and/or may reflect that influenza-specific memory B cells reside predominantly outside of the blood and their transcriptional signatures may be too weak to detect in PBMCs.

Interestingly, when the WGCNA algorithm was applied to fold-change gene expression datasets, the resulting gene clusters were larger in size, fewer in number, and did not correspond to biological functions (via enrichment analysis). We believe this is due to pre-existing immunity, as seen in Fig. 5. For subjects with high levels of pre-existing immunity, subtracting the baseline gene expression data, which already captures this pre-existing immunity, from the Day 28 post-vaccination responses (once again in immune homeostasis) may obscure the existing signature of immunity.

Accounting for pre-existing immunity to influenza virus is a considerable challenge in studies of influenza vaccine-induced immune responses, yet maintenance of cellular immunity via annual revaccination is important for protection in older adults^{76,77}. Our study provides insight into the genes involved in this pre-existing cellular immunity.

Gene expression relationships with humoral immune responses. Baseline gene expression of the Yellow gene cluster enriched for genes involved in monocyte induction, cell cycle & transcription, and inflammatory/antiviral signaling was found to be negatively correlated with antibody titers. This correlation disappears post-vaccination. Curiously, in addition to genes related to monocytes and inflammation, the Yellow gene cluster

includes multiple toll-like receptors (TLRs) known to be involved in innate responses (*TLR1*, 4, 5, 6, 7, and 8) that have been previously identified as part of early (Day 3) gene signatures *positively* predicting antibody responses to seasonal influenza vaccine^{25,51}. This finding is consistent with findings of Nakaya *et al.*²³: in a meta-analysis of multiple influenza vaccine studies, the group found baseline expression in monocyte- and cell cycle-related gene modules that included many TLRs (BTMs M11.0 and M4.0, corresponding to many genes in our Yellow cluster) that also correlate negatively with antibody titers²³. Our findings lend further support to the hypothesis that baseline inflammation may inhibit the induction of vaccine-induced antibody responses, as proposed by others^{23,78,79}. Interestingly, in a comparison of baseline expression in young vs. old subjects, Nakaya *et al.*²³ also found increased levels of monocyte cell populations and corresponding monocyte-related gene expression, suggesting that monocyte activity may be a key mediator of inflammation and a correspondingly reduced ability of the elderly to develop antibody responses to influenza vaccination.

Of the five genes whose behavior most strongly mirrored that of the Yellow gene cluster as a whole, two—*CORO1C* and *C8orf83*—are less well-studied and have not been previously associated with influenza vaccine responses. *CORO1C* encodes the coronin-1C protein, a member of the WD repeat protein family that is involved in a variety of cellular processes, including signal transduction and gene regulation, and may be involved in fibroblast migration⁸⁰. *C8orf83* encodes TRIQK, which is a small protein highly conserved across vertebrates with largely unknown function and may have undiscovered roles in immunity⁸¹. These genes may play heretofore undescribed roles in the development of immunity in response to influenza vaccine.

Day 28 gene expression of two WGCNA gene clusters (Purple and Salmon; 321 total genes) correlated significantly with HAI and VNA antibody measures; however, in these clusters, gene expression in Days 0 and 3 demonstrated no correlation with antibody titers. The genes in these clusters were found to be enriched for genes involved in the control of NK and cytotoxic T cells rather than B cells. While NK and dendritic cells are involved early in the induction of immune responses leading to antibody secretion⁸², the fact that no correlation was observed between baseline/Day 3 expression of these NK-related genes and antibody measures leads us to believe that the genes in the Purple and Salmon gene clusters may not be directly involved in the *induction* of antibody responses, but rather in processes that parallel the development of antibody-secreting B cells. In the Purple and Salmon NK-related gene clusters, eight and four genes, respectively, are KIR and KLR (killer cell immunoglobulin-like and lectin-like receptor) genes, which are known to be inhibitory receptors expressed by NK and CD8⁺ T cells and inhibit their cytotoxic activity⁸³. This leads us to believe that higher expression in these clusters may lead to lower cytotoxic activity, subsequent lower levels of inflammation, and culminate in high levels of antibody production. Others have noted that the frequency, type, and activity of NK cells tends to change with age^{23,24,84–86}, with increased frequencies of NK and CD8⁺ T cells, and corresponding gene expression changes. Thus, these Purple and Salmon gene clusters may serve as post-vaccination indicators of successful humoral responses in older populations, and may also account for immunosenescence.

Additional genes found in the Salmon clusters of note include *SPON2* and *AKRIC3* genes. *AKRIC3* encodes a protein involved in sex steroid metabolism⁸⁷. Expression of *AKRIC3* is regulated by IL-6⁸⁸, appears to change in PBMCs during aging, and may be an endocrine link with immunosenescence⁸⁹. Further investigation of this gene may help us better understand the poor responses of the elderly to influenza vaccine. The *SPON2* gene encodes Spondin-2, also known as mindin. Mindin is an essential component of the innate immune response, as it is a pattern recognition molecule that binds to macrophage-presented receptors during responses to pathogens⁹⁰. Mindin has previously been associated with the intranasal clearance of influenza viruses in mice⁹¹, by both binding directly to influenza virus particles and also activating macrophages. Mice lacking mindin have been demonstrated to have impaired responses to bacterial and influenza virus infections^{90,91}, indicating a potentially critical role in influenza vaccine responses that has not previously been reported and should be further studied.

Additional prominent genes in the Purple gene cluster include *MATK* (encoding megakaryocyte-associated tyrosine-protein kinase), which is an essential regulator of immune cells in mice⁹². *MATK* may play a role in signal transduction during megakaryocytopoiesis and may enhance MAPK activation in RAS-mediated signaling⁹³. *CST7* encodes cystatin-F, an immune cell-specific inhibitor that plays a role in immune responses through regulation of specific enzyme targets such as cathepsin C, which in turn regulate the cytotoxicity of NK and T cells^{94–96}. Thus, these two genes not previously connected with influenza vaccine responses may well influence immune responses to influenza vaccine and are candidates for future studies.

B-cell ELISPOT measures correlated strongly with one small gene cluster (Greenyellow, 135 genes) whose Day 0, Day 3, and Day 28 expression patterns were found to be largely unassociated with antibody titers. The Greenyellow cluster is dominated by B-cell genes, with four of the top five genes involved in the B-cell antigen receptor complex (CD19, CD22, CD79A, FCRLA). The final top gene, RALGPS2, is a RAS-specific GTPase implicated in cell survival and cell cycle control⁹⁷. Interestingly this gene's expression patterns grouped it with B-cell related genes, rather than with other cell cycle-related genes in the Brown or Yellow gene clusters. This suggests a potential role of RALGPS2 in B-cell activity beyond that of other RAS-specific GTPases.

Many of the B cell-related genes found in our Greenyellow cluster were found by others²⁵ to correlate with antibody responses; however, these correlations were measured at Day 7 post-vaccination and reflect peak antibody-secreting cell (ASC) populations and activity²⁴, a timepoint our study did not include. If a future study were conducted with a Day 7 timepoint, we would expect to find such correlations between Day 7 ASC gene expression and antibody titer.

A previous study examining a subset of this dataset used per-gene logistic regression models to identify individual genes whose expression correlated with HAI response across 94 subjects⁵¹. Of the 30 top-associated genes in that study, nine are members of WGCNA gene clusters associated with immunity, of which just six are found in our HAI-associated Salmon or Purple clusters. Individual genes from this transcriptomic dataset that correlated with B-cell ELISPOT responses were also previously identified⁶⁴. Notably, just a single gene from this previous study, *USP6NL*, was clustered into our B-cell ELISPOT-associated Greenyellow cluster. These results indicate that

logistic regression models and WGCNA analysis highlight significantly different genes and genesets. Langfelder *et al.* demonstrated that, in biological systems where marginal associations are weak or noisy, WGCNA gene cluster membership is more reproducible than single-gene selection by marginal³⁷. Thus, the use of WGCNA as an alternative or additional method for analysis of large-scale post-vaccination gene expression data may prove invaluable for future studies. Additionally, WGCNA identification of genes whose expression follows patterns similar to other immune-relevant genes may allow for discovery of genes with important but previously undescribed immunological roles.

Strengths and limitations of the study. The transcriptome of human PBMCs after influenza vaccination reflects contributions from many cell subsets and cellular processes leading to protective immunity, yet such data is highly complex and difficult to interpret. Studying individual immune cell subsets is a common strategy to simplify gene expression datasets; however, this reductionist approach leaves much to be desired. While transcriptomic profiles of each cell subset differ significantly from the profiles of PBMCs²⁸, measurement of gene expression in all possible or even all relevant cell subsets is a difficult and costly process. Additionally, as immune cell subsets signal to and affect one another, reductionist study of individual cell subsets necessarily disrupts or ignores the behavior of the immune system as a whole. The WGCNA algorithm appears to be capable of separating genes involved in different T- and B-cell functions (i.e., cytokine secretion vs. cytotoxic responses, and cytokine-producing vs. antibody-producing, respectively) from whole-PBMC transcriptomic data. We propose the use of WGCNA as a potential tool with which to study transcriptomic responses after vaccination without the disadvantages of a reductionist approach.

Multiple hypothesis testing is always a consideration in high-dimensional studies. The Bonferroni correction, which assumes independence, is inappropriately strict in systems with strong correlations between statistical tests. We present the actual p-values to allow the reader to draw their own conclusions about the significance of the data, as others in the field have also done. Future validation work will be conducted to confirm our results.

Use of a trivalent seasonal influenza vaccine that contains influenza A/H1N1, influenza A/H3N2, and influenza B is a study limitation. While our gene expression data (mRNA-seq) reflects responses to vaccination with all three influenza strains, due to practical limitations we conducted studies of immune development (PBMC cytokine recall responses after *ex vivo* influenza stimulation, B-cell ELISPOT, antibody assays) specific to just influenza A/H1N1. Despite this, relationships between gene expression and immune outcomes are expected, as the development of immunity to the three virus strains occurs concurrently. Further studies may now be conducted on stored PBMC samples to examine the development of cellular and humoral immunity to the other two influenza strains.

Conclusions

Through WGCNA clustering of PBMC gene expression data collected after trivalent seasonal influenza vaccination, we identified individual gene clusters whose expression correlated with subject cellular, memory B-cell and antibody responses. This clustering procedure allowed us to identify key marker genes that reflect development of vaccine-induced immunity. Our findings validated gene signatures of influenza A/H1N1 vaccination responses previously identified by others, and expanded upon these by identification of previously uncharacterized proteins with possible immunological roles. Further analysis suggested the involvement of different genes and immune cell subsets at key points in the development of long-term immunity after influenza virus vaccination. These results may help elucidate the roles that genes and immune cell subsets play during human immune responses to seasonal influenza vaccination, and suggest further utility of the WGCNA gene clustering algorithm in vaccinology.

Further work may use WGCNA clustering to examine the effects of subject characteristics such as age, sex, and obesity on pre- and post-vaccination gene expression, and the relationship of these differentially expressed genes with vaccine-induced immunity. Additional, future studies may examine the behavior of the cell subsets identified in this report as likely contributors to particular aspects of vaccine-induced immunity to influenza, allowing us to identify the mechanisms by which these cells contribute to immune response. This information may be used in the future to reverse-engineer novel influenza vaccine candidates that trigger directed immune responses to enhance development of protection.

References

1. World Health Organization. *Influenza (seasonal)*. 2016. <http://www.who.int/mediacentre/factsheets/fs211/en/>. Date accessed October 30, 2017.
2. Centers for Disease Control and Prevention. 2014–2015 Influenza Season Week 15 ending April 18, 2015. <http://www.cdc.gov/flu/weekly/>. Date accessed: October 30, 2017.
3. D’Mello, T. *et al.* Update: Influenza activity—United States, September 28, 2014–February 21, 2015. *Morbidity and Mortality Weekly Report* **64**, 206–212 (2015).
4. Thompson, W. W. *et al.* Mortality associated with influenza and respiratory syncytial virus in the United States. *Journal of the American Medical Association* **289**, 179–186 (2003).
5. Poland, G. A., Ovsyannikova, I. G. & Jacobson, R. M. Immunogenetics of seasonal influenza vaccine response. *Vaccine* **26S**, D35–D40 (2008).
6. McElhane, J. E. Influenza vaccine responses in older adults. *Ageing Research Reviews* **10**, 379–388 (2011).
7. Poland, G. A., Ovsyannikova, I. G., Kennedy, R. B., Lambert, N. D. & Kirkland, J. L. A systems biology approach to the effect of aging, immunosenescence and vaccine response. *Current Opinion in Immunology* **29C**, 62–68. <https://doi.org/10.1016/j.coi.2014.04.005> (2014).
8. Gardner, E. M., Gonzalez, E. W., Nogusa, S. & Murasko, D. M. Age-related changes in the immune response to influenza vaccination in a racially diverse, healthy elderly population. *Vaccine* **24**, 1609–1614 (2006).
9. Lambert, N. D., Ovsyannikova, I. G., Pankratz, V. S., Jacobson, R. M. & Poland, G. A. Understanding the immune response to seasonal influenza vaccination in older adults: a systems biology approach. *Expert Review of Vaccines* **11**, 985–994 (2012).

10. Grubeck-Loebenstein, B. *et al.* Immunosenescence and vaccine failure in the elderly. *Aging Clinical and Experimental Research* **21**, 201–209 (2009).
11. Reber, A. J. *et al.* Immunosenescence and Challenges of Vaccination against Influenza in the Aging Population. *Aging and Disease* **3**, 68–90 (2012).
12. Lang, P. O. *et al.* Effectiveness of influenza vaccine in aging and older adults: comprehensive analysis of the evidence. *Clinical Interventions in Aging* **7**, 55–64, <https://doi.org/10.2147/CI.A.S25215> (2012).
13. Nichol, K. L., Margolis, K. L., Wuorenma, J. & Von Sternberg, T. The efficacy and cost effectiveness of vaccination against influenza among elderly persons living in the community. *New England Journal of Medicine* **331**, 778–784 (1994).
14. Vu, T., Farish, S., Jenkins, M. & Kelly, H. A meta-analysis of effectiveness of influenza vaccine in persons aged 65 years and over living in the community. *Vaccine* **20**, 1831–1836 (2002).
15. Hill, T., Platzer, A. & Reyes, C. Influenza deaths in spite of immunization and prophylaxis. *Clinical Infectious Diseases* **40**, 492–493 (2005).
16. McElhaney, J. E. The unmet need in the elderly: designing new influenza vaccines for older adults. *Vaccine* **23**(Suppl 1), S10–S25 (2005).
17. Poland, G. A. & Mulligan, M. J. The imperative of influenza vaccines for elderly individuals—an evolving story. *Journal of Infectious Diseases* **200**, 161–163 (2009).
18. Haralambieva, I. H. & Poland, G. A. Vaccinomics, predictive vaccinology and the future of vaccine development. *Future Microbiol* **5**, 1757–1760 (2010).
19. Bernstein, A., Pulendran, B. & Rappuoli, R. Systems vaccinomics: the road ahead for vaccinology. *Omics* **15**, 529–531 (2011).
20. Querec, T. D. *et al.* Systems biology approach predicts immunogenicity of the yellow fever vaccine in humans. *Nat Immunol* **10**, 116–125 (2009).
21. Davis, M. M., Tato, C. M. & Furman, D. Systems immunology: just getting started. *Nature immunology* **18**, 725–732, <https://doi.org/10.1038/ni.3768> (2017).
22. Furman, D. *et al.* Apoptosis and other immune biomarkers predict influenza vaccine responsiveness. *Mol Syst Biol* **9**, 659, <https://doi.org/10.1038/msb.2013.15> (2013).
23. Nakaya, H. I. *et al.* Systems Analysis of Immunity to Influenza Vaccination across Multiple Years and in Diverse Populations Reveals Shared Molecular Signatures. *Immunity* **43**, 1186–1198, <https://doi.org/10.1016/j.immuni.2015.11.012> (2015).
24. Tsang, J. S. *et al.* Global analyses of human immune variation reveal baseline predictors of postvaccination responses. *Cell* **157**, 499–513, <https://doi.org/10.1016/j.cell.2014.03.031> (2014).
25. Nakaya, H. I. *et al.* Systems biology of seasonal influenza vaccination in humans. *Nature Immunology* **12**, 786–795 (2011).
26. Franco, L. M. *et al.* Integrative genomic analysis of the human immune response to influenza vaccination. *Elife* **2**, e00299, <https://doi.org/10.7554/eLife.00299> (2013).
27. Furman, D. *et al.* Systems analysis of sex differences reveals an immunosuppressive role for testosterone in the response to influenza vaccination. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 869–874, <https://doi.org/10.1073/pnas.1321060111> (2014).
28. Hoek, K. L. *et al.* A cell-based systems biology assessment of human blood to monitor immune responses after influenza vaccination. *PLoS ONE* **10**, e0118528, <https://doi.org/10.1371/journal.pone.0118528> (2015).
29. Walhout, M., Vidal, M. & Dekker, J. *Handbook of Systems Biology*. (Elsevier, 2013).
30. Krawetz, Stephen (Ed.) *Bioinformatics for Systems Biology*. (Humana Press, 2009).
31. Stuart, J. M., Segal, E., Koller, D. & Kim, S. K. A gene-coexpression network for global discovery of conserved genetic modules. *Science* **302**, 249–255, <https://doi.org/10.1126/science.1087447> (2003).
32. Presson, A. P. *et al.* Integrated weighted gene co-expression network analysis with an application to chronic fatigue syndrome. *BMC Systems Biology* **2**, 95, <https://doi.org/10.1186/1752-0509-2-95> (2008).
33. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559, <https://doi.org/10.1186/1471-2105-9-559> (2008).
34. Saris, C. G. *et al.* Weighted gene co-expression network analysis of the peripheral blood from Amyotrophic Lateral Sclerosis patients. *BMC Genomics* **10**, 405, <https://doi.org/10.1186/1471-2164-10-405> (2009).
35. Levine, A. J. *et al.* Systems analysis of human brain gene expression: mechanisms for HIV-associated neurocognitive impairment and common pathways with Alzheimer's disease. *BMC Medical Genomics* **6**, 4, <https://doi.org/10.1186/1755-8794-6-4> (2013).
36. Timm, C. M. *et al.* Two Poplar-Associated Bacterial Isolates Induce Additive Favorable Responses in a Constructed Plant-Microbiome System. *Frontiers in Plant Science* **7**, 497, <https://doi.org/10.3389/fpls.2016.00497> (2016).
37. Langfelder, P., Mischel, P. S. & Horvath, S. When is hub gene selection better than standard meta-analysis? *PLoS ONE* **8**, e61505, <https://doi.org/10.1371/journal.pone.0061505> (2013).
38. Fuller, T. F. *et al.* Weighted gene coexpression network analysis strategies applied to mouse weight. *Mammalian Genome* **18**, 463–472, <https://doi.org/10.1007/s00335-007-9043-3> (2007).
39. MacLennan, N. K. *et al.* Weighted gene co-expression network analysis identifies biomarkers in glycerol kinase deficient mice. *Molecular Genetics and Metabolism* **98**, 203–214, <https://doi.org/10.1016/j.yjmgme.2009.05.004> (2009).
40. Mason, M. J., Fan, G., Plath, K., Zhou, Q. & Horvath, S. Signed weighted gene co-expression network analysis of transcriptional regulation in murine embryonic stem cells. *BMC Genomics* **10**, 327, <https://doi.org/10.1186/1471-2164-10-327> (2009).
41. Kennedy, R. B. *et al.* Statistical modeling using early markers of innate immunity to explain variation in humoral responses to influenza vaccine in older adults. *Vaccine* **33**, 3682–3688, <https://doi.org/10.1016/j.vaccine.2015.06.031> (2015).
42. Salk, H. M., Haralambieva, I. H., Ovsyannikova, I. G., Goergen, K. M. & Poland, G. A. Granzyme B ELISPOT assay to measure influenza-specific cellular immunity. *Journal of Immunological Methods* **398–399**, 44–50, <https://doi.org/10.1016/j.jim.2013.09.007> (2013).
43. Umlauf, B. J. *et al.* Associations between demographic variables and multiple measles-specific innate and cell-mediated immune responses after measles vaccination. *Viral Immunology* **25**, 29–36, <https://doi.org/10.1089/vim.2011.0051> (2012).
44. Painter, S. D., Haralambieva, I. H., Ovsyannikova, I. G., Grill, D. E. & Poland, G. A. Detection of Influenza A/H1N1-Specific Human IgG-Secreting B Cells in Older Adults by ELISPOT Assay. *Viral Immunology* **27**, 32–38, <https://doi.org/10.1089/vim.2013.0099> (2014).
45. Haralambieva, I. H. *et al.* The Impact of Immunosenescence on Humoral Immune Response Variation after Influenza A/H1N1 Vaccination in Older Subjects. *PLoS ONE* **10**, e0122282, <https://doi.org/10.1371/journal.pone.0122282> (2015).
46. Jacobson, R. M. *et al.* Profiles of influenza A/H1N1 vaccine response using hemagglutination-inhibition titers. *Human Vaccines and Immunotherapeutics* **11**, 961–969 (2015).
47. FLUARIX. Highlights of Prescribing Information. Package insert. <http://id-ea.org/wp-content/uploads/2012/05/Fluarix-Package-Insert.pdf> (2010). Date accessed: October 30, 2017.
48. Chaudhry, M., Rashid, H. B., Thrusfield, M., Welburn, S. & Bronsvoort, B. M. A case-control study to identify risk factors associated with avian influenza subtype H9N2 on commercial poultry farms in Pakistan. *PLoS ONE* **10**, e0119019, <https://doi.org/10.1371/journal.pone.0119019> (2015).
49. Wang, S. *et al.* Hemagglutinin (HA) proteins from H1 and H3 serotypes of influenza A viruses require different antigen designs for the induction of optimal protective antibody responses as studied by codon-optimized HA DNA vaccines. *Journal of Virology* **80**, 11628–11637 (2006).

50. World Health Organization. WHO Global Influenza Surveillance Network: Manual for the laboratory diagnosis and virological surveillance of influenza. 1–139 (2011).
51. Ovsyannikova, I. G. *et al.* Gene signatures related to HAI response following influenza A/H1N1 vaccine in older individuals. *Heliyon* **2**, e00098, <https://doi.org/10.1016/j.heliyon.2016.e00098> (2016).
52. Hansen, K. D., Irizarry, R. A. & Wu, Z. Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics* **13**, 204–216 (2012).
53. Langfelder, P. & Horvath, S. Fast R Functions for Robust Correlations and Hierarchical Clustering. *Journal of Statistical Software* **46**, <https://www.jstatsoft.org/article/view/v046i011> (2012).
54. R: a language for statistical computing (R Foundation for Statistical Computing, Vienna, Austria, 2009).
55. Rapid Integration of Term Annotation and Network resources. <https://github.com/MTZimmer/RITAN> (2017).
56. Li, S. *et al.* Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. *Nature Immunology* **15**, 195–204, <https://doi.org/10.1038/ni.2789> (2014).
57. Schaefer, C. F. *et al.* PID: the Pathway Interaction Database. *Nucleic Acids Research* **37**, D674–679, <https://doi.org/10.1093/nar/gkn653> (2009).
58. Yusuf, D. *et al.* The transcription factor encyclopedia. *Genome Biology* **13**, R24, <https://doi.org/10.1186/gb-2012-13-3-r24> (2012).
59. Vinayagam, A. *et al.* A directed protein interaction network for investigating intracellular signal transduction. *Science Signaling* **4**, rs8, <https://doi.org/10.1126/scisignal.2001699> (2011).
60. Prasad, T. S., Kandasamy, K. & Pandey, A. Human Protein Reference Database and Human Proteinpedia as discovery tools for systems biology. *Methods in Molecular Biology* **577**, 67–79, https://doi.org/10.1007/978-1-60761-232-2_6 (2009).
61. Rolland, T. *et al.* A proteome-scale map of the human interactome network. *Cell* **159**, 1212–1226, <https://doi.org/10.1016/j.cell.2014.10.050> (2014).
62. Lee, I., Blom, U. M., Wang, P. I., Shim, J. E. & Marcotte, E. M. Prioritizing candidate disease genes by network-based boosting of genome-wide association data. *Genome Research* **21**, 1109–1121, <https://doi.org/10.1101/gr.118992.110> (2011).
63. Szklarczyk, D. *et al.* The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Research* **39**, D561–568, <https://doi.org/10.1093/nar/gkq973> (2011).
64. Haralambieva, I. H. *et al.* Transcriptional signatures of influenza A/H1N1-specific IgG memory-like B cell response in older individuals. *Vaccine* **34**, 3993–4002, <https://doi.org/10.1016/j.vaccine.2016.06.034> (2016).
65. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**, 44–57, <https://doi.org/10.1038/nprot.2008.211> (2009).
66. Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research* **37**, 1–13, <https://doi.org/10.1093/nar/gkn923> (2009).
67. Gene Ontology Consortium: going forward. *Nucleic Acids Research* **43**, D1049–1056, <https://doi.org/10.1093/nar/gku1179> (2015).
68. McElhaney, J. E., Kuchel, G. A., Zhou, X., Swain, S. L. & Haynes, L. T-Cell Immunity to Influenza in Older Adults: A Pathophysiological Framework for Development of More Effective Vaccines. *Frontiers in Immunology* **7**, 41, <https://doi.org/10.3389/fimmu.2016.00041> (2016).
69. Murasko, D. M. *et al.* Role of humoral and cell-mediated immunity in protection from influenza disease after immunization of healthy elderly. *Experimental Gerontology* **37**, 427–439 (2002).
70. Anandasabapathy, N. *et al.* Classical Flt3L-dependent dendritic cells control immunity to protein vaccine. *The Journal of Experimental Medicine* **211**, 1875–1891, <https://doi.org/10.1084/jem.20131397> (2014).
71. King, G. D. *et al.* Combined Flt3L/TK gene therapy induces immunological surveillance which mediates an immune response against a surrogate brain tumor neoantigen. *Molecular Therapy* **19**, 1793–1801, <https://doi.org/10.1038/mt.2011.77> (2011).
72. Racioppi, L. & Means, A. R. Calcium/calmodulin-dependent kinase IV in immune and inflammatory responses: novel routes for an ancient traveller. *Trends in Immunology* **29**, 600–607, <https://doi.org/10.1016/j.it.2008.08.005> (2008).
73. Pina, C., May, G., Soneji, S., Hong, D. & Enver, T. MLLT3 regulates early human erythroid and megakaryocytic cell fate. *Cell Stem Cell* **2**, 264–273, <https://doi.org/10.1016/j.stem.2008.01.013> (2008).
74. Martino, V. *et al.* Down-regulation of MLL-AF9, MLL and MYC expression is not obligatory for monocyte-macrophage maturation in AML-M5 cell lines carrying t(9;11)(p22;q23). *Oncology Reports* **15**, 207–211 (2006).
75. Lin, C. *et al.* AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia. *Molecular Cell* **37**, 429–437, <https://doi.org/10.1016/j.molcel.2010.01.026> (2010).
76. Voordouw, A. C. *et al.* Annual revaccination against influenza and mortality risk in community-dwelling elderly persons. *Journal of the American Medical Association* **292**, 2089–2095 (2004).
77. de Bruijn, I. A. *et al.* Annually repeated influenza vaccination improves humoral responses to several influenza virus strains in healthy elderly. *Vaccine* **15**, 1323–1329 (1997).
78. Haq, K. & McElhaney, J. E. Immunosenescence: influenza vaccination and the elderly. *Current Opinion in Immunology* **29C**, 38–42, <https://doi.org/10.1016/j.coi.2014.03.008> (2014).
79. Pawelec, G., Goldeck, D. & Derhovanessian, E. Inflammation, ageing and chronic disease. *Current Opinion in Immunology* **29**, 23–28, <https://doi.org/10.1016/j.coi.2014.03.007> (2014).
80. Williamson, R. C. *et al.* Coronin-1C Protein and Caveolin Protein Provide Constitutive and Inducible Mechanisms of Rac1 Protein Trafficking. *The Journal of Biological Chemistry* **290**, 15437–15449, <https://doi.org/10.1074/jbc.M115.640367> (2015).
81. Onuma, Y., Watanabe, A., Aburatani, H., Asashima, M. & Whitman, M. TRIQK, a novel family of small proteins localized to the endoplasmic reticulum membrane, is conserved across vertebrates. *Zoolog Sci* **25**, 706–713, <https://doi.org/10.2108/zsj.25.706> (2008).
82. Walzer, T., Dalod, M., Vivier, E. & Zitvogel, L. Natural killer cell-dendritic cell crosstalk in the initiation of immune responses. *Expert Opinion in Biological Therapy* **5**(Suppl 1), S49–59, <https://doi.org/10.1517/14712598.5.1.S49> (2005).
83. Long, E. O., Kim, H. S., Liu, D., Peterson, M. E. & Rajagopalan, S. Controlling natural killer cell responses: integration of signals for activation and inhibition. *Annual Review of Immunology* **31**, 227–258, <https://doi.org/10.1146/annurev-immunol-020711-075005> (2013).
84. Solana, R., Campos, C., Pera, A. & Tarazona, R. Shaping of NK cell subsets by aging. *Current Opinion in Immunology* **29**, 56–61, <https://doi.org/10.1016/j.coi.2014.04.002> (2014).
85. Wagar, L. E., Gentleman, B., Pircher, H., McElhaney, J. E. & Watts, T. H. Influenza-specific T cells from older people are enriched in the late effector subset and their presence inversely correlates with vaccine response. *PLoS ONE* **6**, e23698 (2011).
86. Ibegbu, C. C. *et al.* Expression of killer cell lectin-like receptor G1 on antigen-specific human CD8+ T lymphocytes during active, latent, and resolved infection and its relation with CD57. *Journal of Immunology* **174**, 6088–6094 (2005).
87. Azzarello, J., Fung, K. M. & Lin, H. K. Tissue distribution of human AKR1C3 and rat homolog in the adult genitourinary system. *Journal of Histochemistry & Cytochemistry* **56**, 853–861, <https://doi.org/10.1369/jhc.2008.951384> (2008).
88. Chun, J. Y. *et al.* Interleukin-6 regulates androgen synthesis in prostate cancer cells. *Clinical Cancer Research* **15**, 4815–4822, <https://doi.org/10.1158/1078-0432.CCR-09-0640> (2009).
89. Hammer, F. *et al.* Sex steroid metabolism in human peripheral blood mononuclear cells changes with aging. *The Journal of Clinical Endocrinology & Metabolism* **90**, 6283–6289, <https://doi.org/10.1210/jc.2005-0915> (2005).
90. He, Y. W. *et al.* The extracellular matrix protein mindin is a pattern-recognition molecule for microbial pathogens. *Nature Immunology* **5**, 88–97, <https://doi.org/10.1038/ni1021> (2004).

91. Jia, W., Li, H. & He, Y. W. Pattern recognition molecule mindin promotes intranasal clearance of influenza viruses. *Journal of Immunology* **180**, 6255–6261 (2008).
92. Lee, B. C., Avraham, S., Imamoto, A. & Avraham, H. K. Identification of the nonreceptor tyrosine kinase MATK/CHK as an essential regulator of immune cells using Matk/CHK-deficient mice. *Blood* **108**, 904–907, <https://doi.org/10.1182/blood-2005-12-4885> (2006).
93. Fernandez-Duenas, J. *et al.* [Severe mitral regurgitation in Libman-Sacks endocarditis. Conservative surgery]. *Revista Espanola de Cardiologia* **58**, 1118–1120 (2005).
94. Magister, S. & Kos, J. Cystatins in immune system. *Journal of Cancer* **4**, 45–56, <https://doi.org/10.7150/jca.5044> (2013).
95. Perisic Nanut, M., Sabotic, J., Jewett, A. & Kos, J. Cysteine cathepsins as regulators of the cytotoxicity of NK and T cells. *Front Immunol* **5**, 616, <https://doi.org/10.3389/fimmu.2014.00616> (2014).
96. Hamilton, G., Colbert, J. D., Schuettelkopf, A. W. & Watts, C. Cystatin F is a cathepsin C-directed protease inhibitor regulated by proteolysis. *The EMBO Journal* **27**, 499–508, <https://doi.org/10.1038/sj.emboj.7601979> (2008).
97. A. O. S., Parrini, M. C. & Camonis, J. RalGPS2 Is Essential for Survival and Cell Cycle Progression of Lung Cancer Cells Independently of Its Established Substrates Ral GTPases. *PLoS ONE* **11**, e0154840, <https://doi.org/10.1371/journal.pone.0154840> (2016).

Acknowledgements

We thank the Mayo Clinic Vaccine Research Group staff and the study participants. We also thank Krista M. Goergen for her work in programming and statistical analysis, and Caroline L. Vitse for her editorial assistance with this manuscript. Research reported in this publication was supported by the National Institute of Allergy And Infectious Diseases of the National Institutes of Health under award number U01AI089859. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author Contributions

All authors were involved in study conceptualization and reviewed/edited the manuscript. E.A.V., W.L.S., M.T.Z. and D.E.G. prepared the original draft of the manuscript. R.B.K., I.G.O. and W.L.S. supervised and conducted the laboratory work. E.A.V., D.E.G. and M.T.Z. worked on the methodology used for analysis, as well as performed the formal analysis. M.T.Z. and E.A.V. prepared the figures.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-017-17735-x>.

Competing Interests: Dr. Poland is the chair of a Safety Evaluation Committee for novel investigational vaccine trials being conducted by Merck Research Laboratories. Dr. Poland offers consultative advice on vaccine development to Merck & Co. Inc., Avianax, Dynavax, Novartis Vaccines and Therapeutics, Emergent Biosolutions, Adjuvance Technologies, Seqirus, and Protein Sciences. Drs. Poland and Ovsyannikova hold three patents related to vaccinia and measles peptide research. Dr. Kennedy has received funding from Merck Research Laboratories to study waning immunity to mumps vaccine. These activities have been reviewed by the Mayo Clinic Conflict of Interest Review Board and are conducted in compliance with Mayo Clinic Conflict of Interest policies.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018