Genes and Immunity (2016) **17**, 371–379

www.nature.com/gene

ORIGINAL ARTICLE

Gene signatures associated with adaptive humoral immunity following seasonal influenza A/H1N1 vaccination

IG Ovsyannikova[1], HM Salk[1], RB Kennedy[1], IH Haralambieva[1], MT Zimmermann[2], DE Grill[2], AL Oberg[2] and GA Poland[1]

This study aimed to identify gene expression markers shared between both influenza hemagglutination inhibition (HAI) and virus-neutralization antibody (VNA) responses. We enrolled 158 older subjects who received the 2010–2011 trivalent inactivated influenza vaccine. Influenza-specific HAI and VNA titers and mRNA-sequencing were performed using blood samples obtained at Days 0, 3 and 28 post vaccination. For antibody response at Day 28 versus Day 0, several gene sets were identified as significant in predictive models for HAI ($n = 7$) and VNA ($n = 35$) responses. Five gene sets (comprising the genes *MAZ*, *TTF*, *GSTM*, *RABGGTA*, *SMS*, *CA*, *IFNG* and *DOPEY*) were in common for both HAI and VNA. For response at Day 28 versus Day 3, many gene sets were identified in predictive models for HAI ($n = 13$) and VNA ($n = 41$). Ten gene sets (comprising biologically related genes, such as *MAN1B1*, *POLL*, *CEBPG*, *FOXP3*, *IL12A*, *TLR3*, *TLR7* and others) were shared between HAI and VNA. These identified gene sets demonstrated a high degree of network interactions and likelihood for functional relationships. Influenza-specific HAI and VNA responses demonstrated a remarkable degree of similarity. Although unique gene set signatures were identified for each humoral outcome, several gene sets were determined to be in common with both HAI and VNA response to influenza vaccine.

Genes and Immunity (2016) **17,** 371–379; doi:10.1038/gene.2016.34; published online 18 August 2016

INTRODUCTION

Each year, seasonal influenza infection places a substantial burden on society, both medically and socioeconomically.[1] In the 2014–2015 season alone, nearly 18 000 hospitalizations and 5.0–9.3% of all deaths were directly attributed to influenza infection and related complications, including pneumonia in the United States.[2] Meanwhile, additional research suggests that medical expenses and loss of productivity associated with influenza infection place a financial burden of over $80 billion on the US economy annually.[1] Yearly influenza vaccination is currently the most effective method available to reduce this morbidity and mortality.

Currently, the hemagglutination inhibition (HAI) assay is the most widely accepted measure to assess the protection provided by influenza vaccine. An HAI titer of 1:40 is commonly accepted as a correlate of protection associated with ~50% protection against subsequent influenza virus challenge.[3–6] However, owing to the inability of HAI titers to completely predict susceptibility to future infection, it is clear that HAI antibodies (Abs) provide only a partial explanation of the immunity induced by influenza vaccination. For instance, a recent study noted the ability of patients with little or no HAI Abs to resist influenza infection, suggesting that non-HAI neutralizing activity, such as broader virus neutralization Abs (VNA), may contribute to protection.[7] Interestingly, HAI and VNA titers are often observed to be correlated, although the literature in this area is limited.[8] Another report suggests that VNA titers are a better correlate of protection against influenza infection than HAI titers after immunization with live-attenuated influenza vaccine.[9] Thus, it is important to consider both HAI and VNA titers when determining correlates of protection after influenza immunization.

In order to further explore this issue, it is critical to understand the genomic mechanisms underlying HAI and VNA responses to influenza vaccination. Currently, few biological or genetic markers[10–13] have been identified that characterize the adaptive immune response to influenza vaccine or predict vaccine failure, and additional research in this area is needed. In this study we used a systems biology approach to explore and identify such potential markers. Our aim for this study was to identify genomic markers (both individual genes and gene sets) in common between both influenza A/H1N1-specific HAI and VNA responses following seasonal trivalent inactivated influenza vaccine (TIV) in older adults. Application of a systems biology approach will advance our knowledge by identifying novel mechanisms and generating hypotheses for variations in adaptive immune responses to vaccines.

RESULTS

Subject demographics

In total, 159 subjects participated in this study[14] and the median (interquartile range) age of the cohort was 59.5 (55.3, 66.3) years. Female participants represented 61.6% of the cohort, whereas males represented 38.4%. Overall, a majority of the cohort self-identified as Caucasian (98.7%); the remaining 1.3% reported a race of 'Other.' One subject's sample failed mRNA-sequencing quality control; therefore, data from158 subjects were carried forward for analyses.

Ab responses to influenza A/H1N1 in study subjects

Influenza A/H1N1-specific Ab responses were measured in pre-vaccination serum samples (Day 0) and at Days 3 and 28 post

[1]General Internal Medicine, Mayo Clinic Vaccine Research Group, Mayo Clinic, Rochester, MN, USA and [2]Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA. Correspondence: Dr GA Poland, General Internal Medicine, Mayo Clinic Vaccine Research Group, Mayo Clinic, Guggenheim 611C, 200 First Street SW, Rochester, MN 55905, USA.
E-mail: poland.gregory@mayo.edu
Received 15 December 2015; revised 7 June 2016; accepted 9 June 2016; published online 18 August 2016

vaccination. Day 0 influenza-specific median HAI and VNA titers (median 1:80; interquartile range 1:40–1:320) demonstrated the presence of pre-existing Abs in all subjects. Both HAI and VNA titers increased by Day 28 (1:320; interquartile range 1:160–1:640, $P < 0.001$ for both outcomes). As expected, we found a strong positive linear relationship between HAI titer and VNA titer for Days 0 ($r = 0.924$, $P = 7.07 \times 10^{-67}$), 3 ($r = 0.934$, $P = 1.65 \times 10^{-71}$) and 28 ($r = 0.940$, $P = 8.62 \times 10^{-75}$), which suggests that Ab responses measured by the HAI and VNA assays are similar (Figure 1). The intra-class correlation coefficient was 0.91 for the HAI assay (Day 0) and 0.83 for the VNA assay (Day 0). There were no significant differences in HAI ($P = 0.22$, $P = 0.17$ and $P = 0.79$, for Days 0, 3 and 28, respectively) or VNA ($P = 0.37$, $P = 0.31$ and $P = 0.41$, for Days 0, 3 and 28, respectively) titers between male and female subjects at any timepoint (Figure 2a). Figure 2b illustrates that there were no significant differences at any time point in HAI ($P = 0.56$, $P = 0.44$ and $P = 0.15$, for Days 0, 3 and 28, respectively) or VNA ($P = 0.86$, $P = 0.51$ and $P = 0.32$, for Days 0, 3 and 28, respectively) responses.

Common gene sets (Log_2 Day 28 vs Day 0) associated with HAI and VNA responses

Externally defined gene sets with significant mRNA expression changes from Day 0 to Day 28 ($P < 0.005$) were used in cross-validated penalized regression models, to predict HAI or VNA change from Day 0 to Day 28 (adaptive immune response). We identified 7 and 35 gene sets containing genes with changes in expression that demonstrated as association with variability in the response and had at least one gene entering the model for the HAI and VNA responses, respectively. In contrast, no gene sets entered the models to predict early HAI or VNA change from Day 0 to Day 3. The genes associated with HAI are RNA transcription factors (*TTF2* and *MAZ*), chemokine/cytokine/receptor (*CCR9*, *IL10RA* and *IFNG*), transferase activity markers (*GSTM1*, *2*, *SMS* and *RABGGTA*), carbonic-anhydrases (*CA2*, *6*, *8*, *11* and *14*), oxidoreductases (*NDUFS3*, *CYB5R2*, *3*, *CYB561*, *NQO1* and *2*) and the endosome transport dopey family marker (*DOPEY2*) (Supplementary Table 1). Among gene sets associated with VNA response are tumor necrosis factor ligand TNFSF11, cytokines/receptors (*IFNG*, *IL7* and *IL27*), interferon (IFN)-inducible transcription factors (*IRF7* and *IRF9*) and other genes with unknown role in the regulation of humoral immunity (Supplementary Table 2). Five gene sets were in common with both HAI and VNA (Table 1). These genes/gene sets were involved in cellular protein processes (*MAZ* and *TTF2*), transferase (*GSTM1*, *GSTM2*, *RABGGTA* and *SMS*) and carbonate dehydratase (*CA2*, *CA6*, *CA8*, *CA11* and *CA14*) activities, KEGG pathway (*IFNG*) and endosome transport (*DOPEY2*).

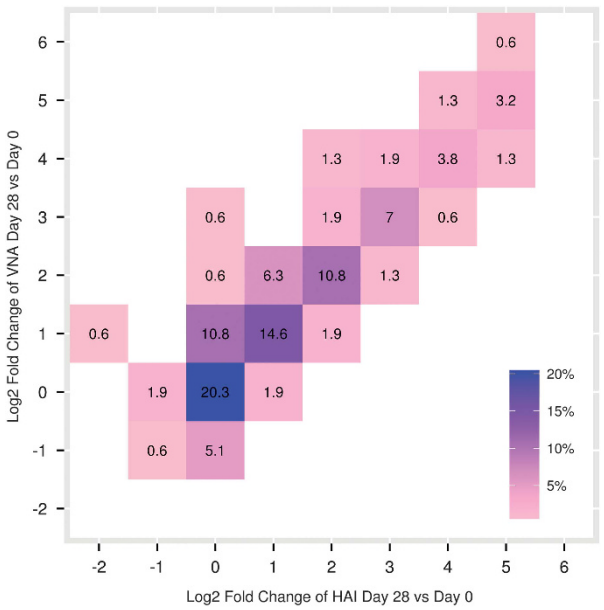

Figure 1. Comparison of HAI and VNA responses (Day 28 vs Day 0) ($n = 158$ subjects). Heatmap of the overlap between the log_2 fold change of HAI (rows) and VNA (columns) for Day 28 relative to Day 0. The color scheme is determined by the percent of the total subjects in each cell with white indicating no overlap, pink indicating a small percentage of overlap and blue indicating the largest overlap. The majority of the results either fall on the diagonal or off-diagonal, indicating strong concordance in these assays.

Common gene sets (Log_2 Day 28 vs Day 3) associated with HAI and VNA responses

Using the approach detailed above for gene sets having statistically significant changes between Days 3 (innate immune response) and 28 post influenza immunization, we identified 13 gene sets associated with HAI titers. In addition, 41 gene sets were associated with variation in VNA response to vaccine. Genes in the HAI regression model include endoplasmic reticulum-associated enzyme (*MAN1B1*), DNA polymerase (*POLL*) and many other genes with unknown immune function (Supplementary Table 3). Genes in the VNA regression model include signal transducers (*STAT1* and *STAT3*), tyrosine kinase (*TYK2*), Golgi vesicular transport protein (*GOSR1*) and others. For additional genes and gene sets, see Supplementary Table 4. Overall, 10 gene sets and related genes were associated with both HAI and VNA responses. Included are immune-related genes *MAN1B1*, *POLL*, *CEBPG*, *FOXP3*, *IL12A*, *TLR3* and *TLR7* (Table 2).

Biologic functions implicated by genes associating with HAI and VNA responses

Network interactions among all genes within significant gene sets were extracted, revealing a number of highly interconnected modules. Our statistically prioritized genes were typically representatives from these modules (Supplementary Figure S2). Within this network, we calculated the extent of interconnectivity between genes prioritized by our statistical approach (Figure 3) and identified 51% as directly connected to at least one other prioritized gene. Randomly selected gene sets (see Methods) exhibited a mean interconnectivity of 29 ± 4%, indicating a significant level of interrelationships among our prioritized genes; none of the randomly generated sets of genes showed this extent of interconnectivity. Focusing on statistically prioritized genes, many of those associated with both HAI and VNA (including *INFG* and *TLR* genes) were found to have known interactions with genes that were uniquely associated with HAI or VNA. Further, a high degree of network interactions were identified between the gene sets prioritized at different time points, indicating a high likelihood for functional relationships between the genes. Thus, the distinct genes/gene sets prioritized for each outcome and time point tend to directly interact with one another, indicating probable participation in common biologic mechanisms.

To add further annotation-based evidence for functional relationships between these genes, we performed Gene Ontology (GO) term enrichment. The most significantly enriched VNA-specific terms were purine salvage and metabolism. The most significantly enriched terms shared by both outcomes are the biosynthetic processes 'long chain fatty-acyl-CoA,' 'positive regulation of *IFNG*,' and 'triglyceride.' Although many of the statistically enriched terms matched prior expectation (for example, *IFNG* regulation),[10] some present more novel hypotheses for future study. See Supplementary Figure S1 for the full list of significantly enriched GO terms, by outcome and time point.

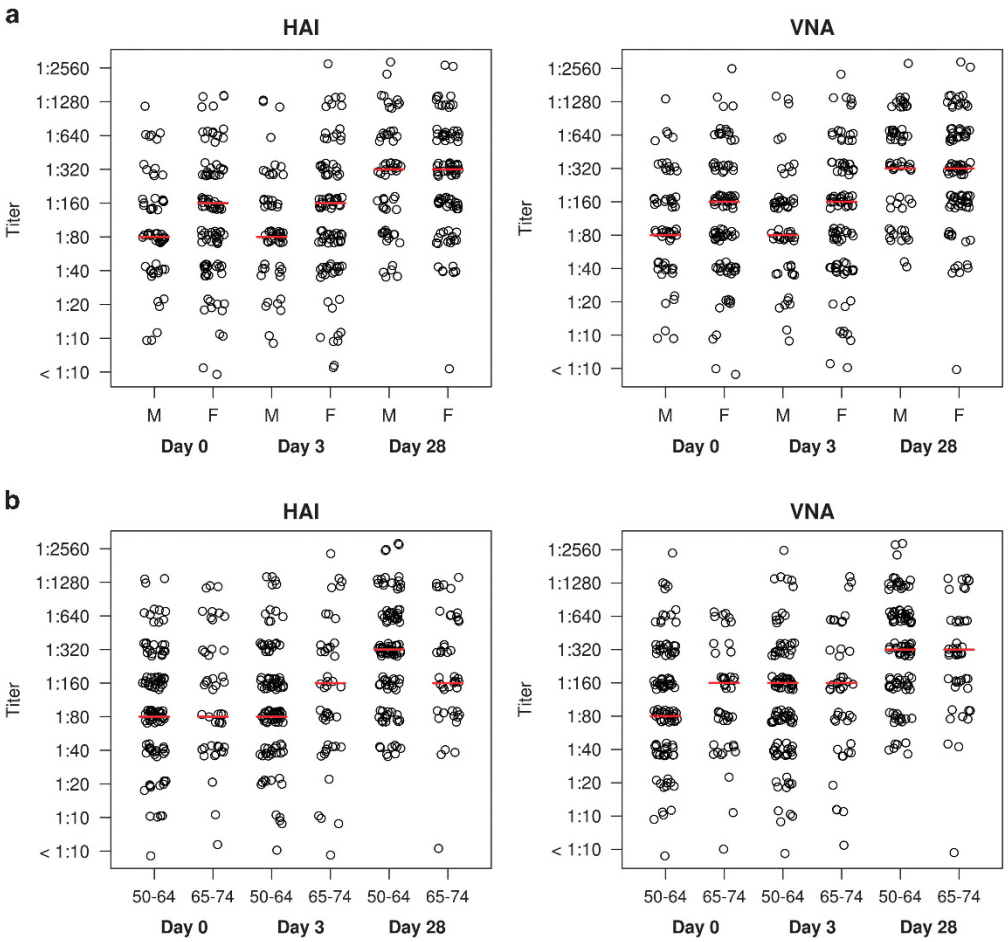

Figure 2. Distribution of HAI and VNA responses (Day 0, Day 3 and Day 28) by sex and age group. (**a**) Scatterplot of HAI and VNA titers by male (M) and female (F) at Day 0, Day 3 and Day 28. (**b**) Scatterplot of HAI and VNA titers by age: 50–64 years old and 65 years and older.

Table 1. Common gene sets with genes entering regression models

Gene set[64]	*HAI*		*VNA*		*Coefficient*[a]	*Median log_2 fold change*
	MSE[b]	*Genes*	*MSE*[b]	*Genes*		
CPCD: (CELLULAR_PROTEIN_COMPLEX_DISASSEMBLY)	2.245	MAZ	2.290	MAZ	–	−0.005
				SUPT16H	–	0.026
		TTF2		TTF2	–	−0.009
				UPF1	+	0.037
ALKYL: (TRANSFERASE_ACTIVITY_TRANSFERRING_ALKYL_OR_ARYLOTHER_THAN_METHYLGROUPS)	2.279	GSTM1	2.334	GSTM1	–	−0.029
		GSTM2		GSTM2	–	−0.039
				MAT2A	+	0.070
		RABGGTA		RABGGTA	+	0.004
		SMS		SMS	–	−0.005
CDA: (CARBONATE_DEHYDRATASE_ACTIVITY)	2.286	CA11	2.346	CA11	+	−0.080
		CA14		CA14	–	−0.101
		CA2		CA2	–	−0.038
		CA6		CA6	–	−0.100
		CA8		CA8	–	0.197
RA: (KEGG_REGULATION_OF_AUTOPHAGY)	2.290	IFNG	2.403	IFNG	+	0.113
ET: (ENDOSOME_TRANSPORT)	2.330	DOPEY2	2.357	DOPEY2	+	0.023

Abbreviations: HAI, hemagglutination inhibition; MSE, mean squared error; VNA, virus-neutralization antibody. Common gene sets with genes entering regression models for HAI and VNA responses with the log_2 Day 28 versus Day 0 fold-change in gene expression as the explanatory variables. [a]A positive (+) coefficient from the regression models indicates that as the log_2 fold change for the gene increases from Day 0 to Day 28, then the estimated response increases (upregulated with respect to the change). If the coefficient is negative (−) as the log_2 fold change for the gene increases, the estimate response decreases. [b]Cross-validated MSE. Gene sets presented had genes remain in the penalized regression models for both HAI and VNA. The gene set name provides the abbreviation that is used for simplicity in the text, a brief description and gene set name from the MSigDB[64] and the actual gene.

Table 2. Common gene sets with genes entering regression models

Gene set[64]	*HAI*		*VNA*		*Coefficient*[a]	*Median log_2 Fold- Change*
	MSE[b]	*Genes*	*MSE*[b]	*Genes*		
ERCC: (REACTOME_N_GLYCAN_TRIMMING_IN_THE_ER_AND_CALNEXIN_CALRETICULIN_CYCLE)	2.222	MAN1B1	2.333	MAN1B1	–	−0.022
NHEJ: (KEGG_NON_HOMOLOGOUS_END_JOINING)	2.288	POLL	2.374	POLL	–	−0.026
IFNG: (INTERFERON_GAMMA_PRODUCTION)	2.302	CEBPG	2.419	CEBPG	–	−0.022
		EBI3			–	0.004
		FOXP3		FOXP3	–	−0.041
		IL12A		IL12A	+	−0.026
		INHBA		INHBA	+	0.005
		TLR3		TLR3	+	0.151
		TLR7		TLR7	+	0.058
		TLR9			+	0.147
BER: (KEGG_BASE_EXCISION_REPAIR)	2.302		2.313	LIG3	+	0.047
				NEIL1	+	−0.016
		PARP2		PARP2	–	−0.001
		POLL		POLL	–	−0.026
				SMUG1	+	0.015
FAB: (REACTOME_FATTY_ACYL_COA_BIOSYNTHESIS)	2.306	ACSL1	2.381	ACSL1	+	0.063
		ACSL5		ACSL5	+	0.020
				ELOVL7	+	−0.012
		FASN		FASN	+	0.013
		HSD17B12			–	0.034
		SLC25A1		SLC25A1	+	0.009
SRA: (SECRETIN_LIKE_RECEPTOR_ACTIVITY)	2.309	BAI1	2.398	BAI1	+	−0.029
		TAPT1		TAPT1	–	−0.057
KREB: (BIOCARTA_KREB_PATHWAY)	2.310		2.360	ACO2	+	0.017
		CS		CS	–	−0.002
				FH	–	0.016
				IDH2	–	0.002
		OGDH		OGDH	–	0.055
				SDHA	+	0.036
		SUCLA2		SUCLA2	+	0.043
PNS: (PERIPHERAL_NERVOUS_SYSTEM_DEVELOPMENT)	2.318	BAI1	2.422	BAI1	+	−0.029
		GSTM3		GSTM3	–	−0.067
MCC: (MEIOTIC_CELL_CYCLE)	2.325	LIG3	2.363	LIG3	+	0.047
				RAD51B	+	−0.032
		RAD54B		RAD54B	–	−0.060
		REC8		REC8	+	−0.018
		STAG3		STAG3	–	−0.013
		TUBG1		TUBG1	–	0.030
LM: (LEARNING_AND_OR_MEMORY)	2.326	ACSL4	2.354	ACSL4	+	0.015
		CRHBP		CRHBP	+	−0.025
		DLG4		DLG4	–	−0.011
				FYN	–	0.024
		GALR2		GALR2	–	0.059
				NF1	+	0.069
		S100B		S100B	–	0.030
				VLDLR	–	0.071

Abbreviations: HAI, hemagglutination inhibition; MSE, mean squared error; VNA, virus-neutralization antibody. Common gene sets with genes entering regression models for HAI and VNA Responses, with the log_2 Day 28 vs Day 3 fold-change in gene expression as the explanatory variables. [a]A positive (+) coefficient from the regression models indicates that as the log_2 fold change for the gene increases from Day 3 to Day 28, then the estimated response increases (upregulated with respect to the change). If the coefficient is negative (−) as the log_2 fold change for the gene increases, the estimate response decreases. [b]Cross-validated MSE. Gene sets presented had genes remain in the penalized regression models for both HAI and VNA. The gene set name provides the abbreviation that is used for simplicity in the text, a brief description and gene set name from the MSigDB[64] and the actual gene.

DISCUSSION

The goal of this study was to identify genomic markers in common with both HAI and VNA responses following influenza A/H1N1 vaccination. Although the magnitude of HAI and VNA titers to influenza vaccine are often correlated, as they are in this study (Figure 1), the two measures of humoral immunity are different; therefore, it was expected that gene expression models would vary somewhat between the two immune outcomes. Thus, we chose to identify common gene sets/genes based on similarities in HAI and VNA responses.

Through our modeling, we identified five gene sets containing genes with expression changes from Day 0 to Day 28 that demonstrated association with both HAI and VNA titers (Table 1). Similarly, we identified 10 gene sets containing genes with expression pattern changes from Day 3 to Day 28 (Table 2) that were associated with variation in both HAI and VNA responses. One of the gene sets associated with both HAI and VNA for Day 28 versus Day 0 expression was CPCD. Genes of particular interest within this gene set include *MAZ*, a transcription factor, which was prioritized in both models, and *TTF2*, which is a transcription

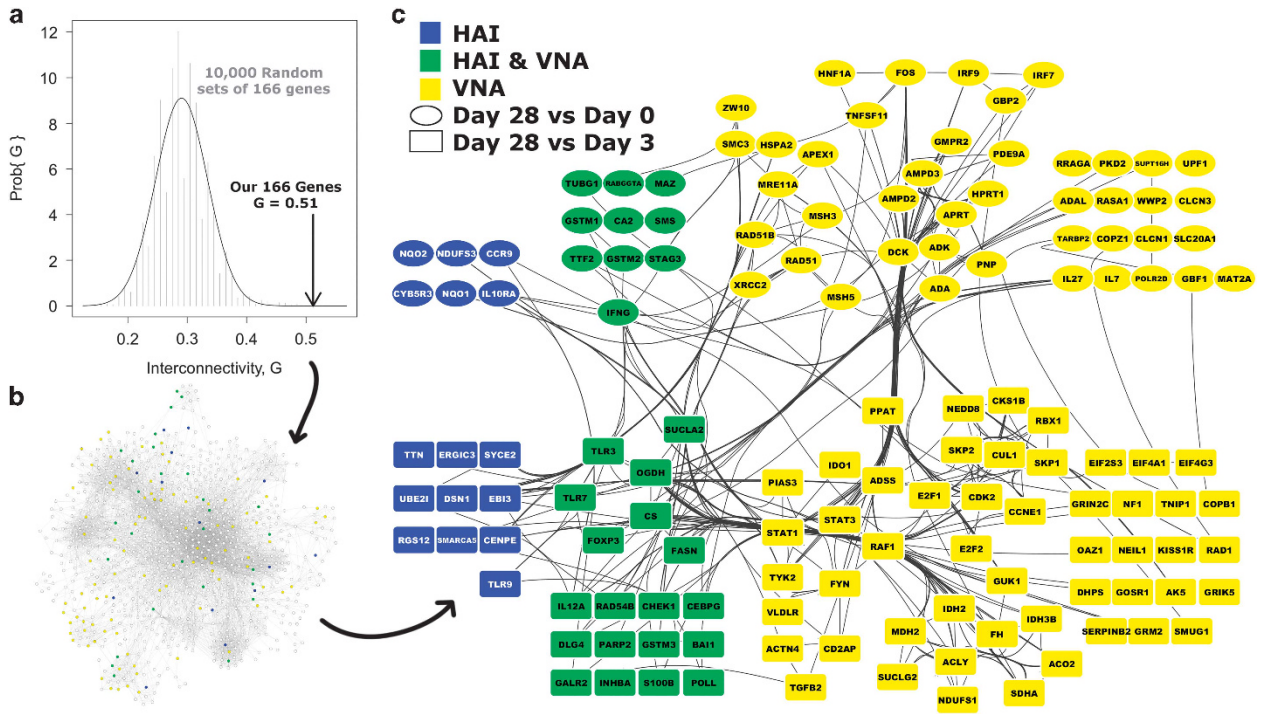

Figure 3. Statistically prioritized genes exhibit a high degree of network interactions. (**a**) Comparing with randomly selected gene sets, our prioritized genes have a significant level of direct interactions. (**b**) Visualizing the full network of all genes within prioritized gene sets reveals the presence of network modules (Supplementary Figure S2). (**c**) The network interactions between our statistically prioritized genes demonstrate significant interactions across time points and outcomes. Edges are bundled to increase legibility.

termination factor for RNA polymerase II.[15] Often referred to as Pur1, *MAZ* encodes for a protein that regulates inflammation-induced expression of serum amyloid A proteins.[16,17] Data suggest that serum amyloid A is involved in at least two aspects of immune regulation. Serum amyloid A proteins are released from hepatocytes during acute inflammation where they typically collaborate with high-density lipoprotein.[18] This newly formed complex prompts the synthesis of several cytokines and acts as a chemotactic agent for both neutrophils and mast cells. In addition, serum amyloid A is capable of binding to, and activating, TLR4 on B lymphocytes to initiate maturation.[18,19]

We also identified the RA gene set, with significant change between Day 28 versus Day 0. This gene set includes the immune gene *IFNG*. Studies have shown that IFNγ is capable of inhibiting proliferation of pre-activated B lymphocytes or stimulating activated B-lymphocyte proliferation and isotype switching.[20] Thus, *MAZ* or *IFNG* gene expression signature may contribute to differential influenza-specific Ab expression post immunization.

On analysis of changes in gene expression between Day 28 versus Day 3 post influenza vaccination, we identified 10 gene sets associated with both HAI and VNA titers (Table 2). There was one gene in the ERCC gene set (*MAN1B1*) that entered the regression model. *MAN1B1* encodes the enzyme endoplasmic reticulum mannosyl-oligosaccharide 1,2-α-mannosidase (ER mannosidase I, Mnl1). *MAN1B1* is commonly referenced in disorders of lysosomal storage[21] due to the enzyme's role in cleaving mannose monomers from newly synthesized peptides within the ER. As Mnl1 is an essential component of quality control for proper protein folding, secretion and function,[22,23] polymorphisms in *MAN1B1* could potentially result in an alteration of the protein profile of immune response to TIV antigens. Recent studies using kifunensine (Mnl1 inhibitor) suggest that Mnl1 affects immune regulation by preventing naive CD4+ T-cell activation,[24] leading to a lack of B-cell co-stimulation and Ab production. Although this phenomenon has been more widely studied for *MAN1A1*-encoded Mnl1, the observed actions could potentially be generalized to *MAN1B1*-encoded Mnl1.

When evaluating the change in early-activation gene expression from Day 3 and Day 28 as predictors in the model, the NHEJ gene set had genes enter the regression model for both HAI and VNA. Within this gene set is the *POLL* gene, which codes for DNA polymerase λ, an enzyme with a critical role in both DNA replication and repair.[25] Previous studies suggest that polymerase λ is involved in DNA base excision repair in the wake of oxidative damage.[26,27] One study shows that polymerase λ-deficient mice experience a decreased rate of germinal center B-cell receptor somatic hypermutation.[27] This observed decline in somatic hypermutation is thought to be caused by a reduction in the prevalence of B cells due to their inability to successfully repair oxidative DNA damage. It can be proposed that mutations in the *POLL* gene may affect the ability of germinal center B cells to diversify Ab populations, thereby altering the humoral immune response to influenza.

The immune-associated gene set IFNG also had genes that entered the model for both HAI and VNA when the change in gene expression for Day 28 versus Day 3 was evaluated. Genes of interest in this gene set are *FOXP3*, *INHBA*, *TLR3* and *TLR7*; others include *CEBPG*, *EBI3*, *IL12A* and *TLR9*. However, *TLR9* entered models only for the HAI and not for the VNA response.

Studies have shown that *FOXP3* serves as the master regulator of T-regulatory cell (Treg) production. Tregs, which are critical for dampening the immune response, have also been shown to secrete cytokines, such as IFNγ, which have important immunoregulatory functions.[28] IFNγ is essential for the induction of *FOXP3* in CD4+ T cells.[29,30] Less research has been conducted on the influence of Tregs on humoral immunity; however, recent studies suggest that Tregs also inhibit B-cell responses to antigens through IFNγ-related mechanisms. It has been suggested that

alterations in the ratio of Tregs to T-follicular helper cells within germinal centers exert control over Ab responses.[31]

Of further importance in the IFNG gene set is *INHBA*. This gene encodes the peptide inhibin βA, a subunit of both activin and inhibin; these proteins operate antagonistically within the immune system.[32] Activin A is produced by activated B cells before it directly stimulates naive B cells to increase production of IgG and indirectly stimulates activated B cells to increase IgG and IgE secretion.[33]

The Toll-like receptor (TLR) genes, *TLR3* and *TLR7*, also entered the models for both the HAI and VNA response to influenza antigen models when using Day 28 versus Day 3 post vaccination expression. TLRs are well known for their involvement in the innate immune response; however, it has been shown that B cells also express TLRs that, when bound to ligand, enhance cell survival and provoke increased Ab production.[34] Other genes in this gene set include *CEBPG*, a B-cell transcription factor;[35] *EBI3*, a gene encoding a subunit of both interleukin (IL)-27 and IL-35;[36] and *IL12A*, a gene encoding a subunit of the Treg inducer IL-35.[37]

We have demonstrated that the genes prioritized by our statistical methods are involved in diverse cellular functions that are important biologically. Our statistical approach filters out genes with highly correlated expression (that is, members of protein complexes or proteins mediating similar biologic functions). Thus, the somewhat diverse set of genes prioritized by our methods are likely to be representatives of broader classes of genes or of specific protein complexes (see Supplementary Figure S2).

To interpret the coordinated activity of prioritized genes, we used two types of prior information: network resources and GO term enrichment. These two types of resources provide complementary information: evidence of physical or molecular interactions and participation in the same molecular or biochemical processes, respectively. It is noteworthy that statistically significant GO term enrichment in this context is a descriptive measure about the gene's functions: the genes were prioritized by our statistical approach and we use term enrichment as a means to describe what biologic activities to which they contribute. Through mapping all of the genes identified by our statistical approach onto biologic networks and inspecting the degree of known relationships among them, we find that many share direct interactions. The set of common genes, those prioritized by models of both HAI and VNA, have many direct connections with genes that were prioritized by one outcome, indicating participation in similar functions. This is also true across time points. In order to further identify the functions shared by these genes, we performed GO term enrichment for the entire set and also for each subset (by time point and outcome). Some functions are identified across the individual outcome comparisons, but some are only identified when considering genes across our outcome-associated results. Specifically, multiple lipid metabolism terms including phosphatidylinositol signaling (including PIP3), lipid (cholesterol, triglyceride and phosphatidylinositol) metabolism and lipid transport are implicated by genes across our comparisons. This broad network-based summary highlights the multifaceted nature of immune response.

Our study design used rigorous randomization approaches in order to minimize the influence of experimental artifacts. Our data reduction steps were agnostic to the outcome being modeled and used externally available immune system knowledge. All modeling involving the outcome was cross-validated to ensure reproducibility. The novelty of this study is evident in that it is the first of its kind to identify gene sets and genes whose changes in expression influence variation in both HAI and VNA responses to influenza A/H1N1 immunization. However, there are limitations to this study. First, the utilization of a larger cohort would enhance statistical power. The examination of gene expression underlying the innate immune response would be beneficial to this study; in order to do this, sample collection on Day 1 post immunization would have been ideal. Lastly, the scope of this study is limited to A/H1N1 influenza and may or may not be generalizable to other components of the TIV.

The unique data and results generated from this study could potentially be used to predict humoral response to influenza vaccine in the future and inform the development of an individualized vaccine schedule paradigm.[38–40] The identification of gene signatures associated with humoral immunity may provide a better understanding of the genetic markers of immune response and may assist with the design of better vaccines and adjuvants.

MATERIALS AND METHODS

Study subjects

The methodology used for the selection and recruitment of study subjects has been previously reported elsewhere.[14,41–43] Briefly, the study included 159 healthy adults, ranging in age from 50 to 74 years old, who were immunized with an intramascular single dose (0.5 ml) of the 2010–2011 seasonal TIV Fluarix (GlaxoSmithKline, Research Triangle Park, NC, USA), containing A/California/7/2009 (H1N1), A/Perth/16/2009 (H3N2) and B/Brisbane/60/2008 viral strains.[14,41–43] TIV is prepared from split-virion influenza viruses propagated in embryonated chicken eggs and contains a ratio of 15 μg of the hemagglutinin antigen for each of the three influenza strains.[44]

Before the onset of the study, each participant provided written, informed consent. All subjects reported stable health and provided detailed vaccination histories. Subjects were excluded from the study if they already received the 2010–2011 TIV. Exclusion criteria included the display of flu-like signs or symptoms at any point throughout the duration of the study. Subjects were also excluded from enrollment if they were diagnosed with influenza, or exhibited symptoms consistent with influenza, at any time from the beginning of the influenza season in Minnesota (as defined by the first reported cases over the period of the individual subject's participation).[45] In addition, one participant was excluded due to cDNA library preparation malfunction. Blood samples (90 ml) from each subject were obtained at three separate time points: pre-vaccination (Day 0) and Days 3 (innate immune response) and 28 (adaptive immune response) post vaccination.[14] This study was approved by the Mayo Clinic Institutional Review Board.

HAI and VNA assays

We have previously described the HAI and VNA assays.[14,41,43] As published elsewhere,[46–48] the standard WHO protocol[49] was used to determine influenza H1N1-specific (virus strain A/California/07/2009) Ab titers from each subject's serum at all three time points. The HAI titer was defined as the highest dilution of serum that inhibits red blood cell (0.5%) hemagglutination. The VNA titer was defined as the reciprocal of the highest dilution of serum that neutralizes 200 plaque-forming units of influenza A/H1N1 virus.

Seroconversion to influenza vaccine antigens was defined as a fourfold increase in serum Ab titers between Day 0 (before vaccination) and Day 28, or as an increase in serum Ab titers from < 10 to $\geqslant 40$ from Day 0 to Day 28.[50] The average coefficients of variation for the assays performed in this study were 2.9 and 4.7% for HAI and VNA, respectively.

Separation of PBMCs

Methodology used for peripheral blood mononuclear cell (PBMC) preparation is identical to what we have previously published.[42] PBMCs were isolated from blood samples (100 ml) from each subject at each timepoint pre- and post-vaccination using cell preparation tubes with sodium citrate (CPT; BD Biosciences, San Jose, CA, USA), as previously published.[42] Purified PBMCs were resuspended at a concentration of 1×10^7 per ml in RPMI-1640 medium containing L-glutamine (Invitrogen; Thermo Fisher Scientific, Carlsbad, CA, USA), supplemented with 10% dimethyl sulfoxide and 20% fetal calf serum (Hyclone; GE Healthcare Life Sciences, Logan, UT, USA). Cells were frozen overnight at −80 °C in freezing containers (Thermo Fisher Scientific) to reach an optimal rate of cooling and then transferred for storage to liquid nitrogen.

mRNA sequencing

To carry out mRNA next-generation sequencing, we used protocols identical to those used and published for our transcriptomic studies.[51,52]

To summarize, the RNeasy Plus mini Kit and the RNAprotect reagent from Qiagen (Valencia, CA, USA) were used to extract total RNA from 1×10^6 PBMCs. Manufacturer protocols were used to create full-length cDNA libraries using the mRNA-Seq 8 Sample Prep Kit by Illumina (San Diego, CA, USA). The DNA 1000 Nano Chip kits were run on an Agilent 2100 Bioanalyzer (Agilent, Palo Alto, CA, USA) for library validation and quantification, and cDNA libraries (5–7 pM) were loaded onto individual flow cell lanes. The Illumin HiSeq 2000 (Illumina), in addition to Illumina's Single Read Cluster Generation kit (v2) and 50 Cycle Illumina Sequencing Kit (v3), was used to perform single-end read sequencing and the sequencing reads were aligned to the human genome build 37.1 using TopHat (1.3.3, Baltimore, MD, USA) and Bowtie (0.12.7, Baltimore, MD, USA). Gene counting was performed using HTSeq (0.5.3p3, EMBL Heidelberg, Germany) and the reads mapping to individual exons were counted using BEDTools (2.7.1, Salt Lake City, UT, USA).[53–55]

Statistical methods

Demographic results are presented as a percent of the total sample for discrete variables or as the median and interquartile range for continuous variables. Spearman's correlation was used to calculate associations between Ab titers. The Wilcoxon rank-sum test was used to test for difference in HAI and VNA titers between male and female subjects, and between subjects who were 50–64 years of age and those who were 65 years and older.

Conditional quantile normalization, which adjusts for gene length and GC content, was used to calculate a normalization offset for use in subsequent analyses of the counts from the mRNA-sequencing data.[56] Moderated dispersion estimates were estimated using edgeR, assuming variance was a nonlinear function of the mean, with the tagwise dispersion calculated from the trended dispersion.[57,58] To identify genes that changed significantly over time, we used generalized linear models, assuming the negative binomial distribution with generalized estimating equations, in order to account for correlation between multiple observations within a subject to obtain per gene *P*-values.[59–62] Self-contained gene set tests were performed using the gamma method with soft truncation threshold of 0.15, an extension of Fisher's method of combining per-gene *P*-values.[63] The gene sets tested were externally defined gene sets from the Molecular Signatures Database[64,65] that were in the Biocarta, GO, KEGG, Reactome, Sigma-Aldrich, Signaling Gateway Signal Transduction KE pathways or had 'virus,' 'infect,' 'pathogen,' or 'innate' as a keyword in the gene set description.

Penalized regression methods were used to build regression models with the goal of understanding the biological processes that may explain variation in vaccine response.[66] Two response variables were evaluated: (1) $\log_2$ fold change of the Day 28 HAI titer relative to the Day 0 HAI titer; and (2) $\log_2$ fold change for Day 28 VNA titer relative to Day 0 VNA titer. Predictor variables were the individual genes from gene sets having statistically significant changes over time ($P < 0.005$, $n = 339$). Specifically, gene sets with significant changes in Day 28 versus Day 0 and Day 28 versus Day 3 were used in the models (log2 difference in normalized expression between time points). Models were fit to genes from one gene set at a time. The predictor variables were first filtered using redundancy analysis,[67] with an $R^2 < 0.75$ as the cutoff. The remaining genes were included as independent variables in the 'glmnet' function in R, which was used for model selection.[68] Gender was also included as a covariate in the models. Tenfold cross validation was used to select the lambda parameter (based on the minimum cross-validated mean squared error), governing the selection of genes to be incorporated in the final model. The tuning parameter α was set to 0.9, reflective of the elastic net penalty (a combination of the L1 LASSO and L2 ridge penalties). The R statistical software version 3.0.2 was used for all analyses (www.r-project.org).

Genes prioritized by the above statistical models were used in network analysis. In order to be comprehensive and also focus on high-confidence interaction data, multiple network resources were combined including HPRD,[69] CCSB,[70] the Pathway Interaction Database[71] and the subset (7.8%) of STRING,[72] where all interactions had a confidence score of at least 70%. We evaluated the significance level of network connections using random sampling. To do this, we generated a network interconnectivity metric defined as the fraction of genes that are first neighbors of each other, using all statistically significant genes that map to the network ($n = 166$). We then randomly selected 10 000 sets of the same size from the network, limited to genes that were detected by RNA-Seq in our study ($n = 15{,}708$). Network operations were performed using the igraph R package, version 0.7.1. Networks were visualized using Cytoscape[73] version 3.2.1 and layouts refined using AllegroLayout v.2.2.1.[74] The biologic functions of gene sets were evaluated using GO term enrichment.[75,76] GO terms were extracted from the human GO Annotation database[77] and hypergeometric tests were used to determine enrichment. We reported FDR-corrected *P*-values (*q*-values) for terms significant at the $q < 0.01$ level.

CONFLICT OF INTEREST

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., CSL Biotherapies, Avianax, Dynavax, Novartis Vaccines and Therapeutics, Emergent Biosolutions, Adjuvance, Microdermis, Seqirus, NewLink, Protein Sciences, GSK Vaccines and Sanofi Pasteur. Drs Poland and Ovsyannikova hold two patents related to vaccinia and measles peptide research. 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. This research has been reviewed by the Mayo Clinic Conflict of Interest Review Board and was conducted in compliance with Mayo Clinic Conflict of Interest policies. All other authors declare no conlict of interest.

ACKNOWLEDGEMENTS

We thank the Mayo Clinic Vaccine Research Group and the subjects who participated in our studies. We thank Randy A Albrecht and Adolfo García-Sastre of the Mount Sinai School of Medicine (New York, NY) for performing the HAI and VNA assays. We thank Krista M Goergen, who served as our statistical programmer analyst in this effort. We thank Caroline L Vitse for her editorial assistance. We acknowledge support from NIH grant U01AI089859 for this work. This publication was supported by Grant Number UL1 TR000135 from the National Center for Advancing Translational Sciences (NCATS). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

1 Molinari NA, Ortega-Sanchez IR, Messonnier ML, Thompson WW, Wortley PM, Weintraub E *et al.* The annual impact of seasonal influenza in the US: measuring disease burden and costs. *Vaccine* 2007; **25**: 5086–5096.
2 Appiah GD, Blanton L, D'Mello T, Kniss K, Smith S, Mustaquim D *et al.* Influenza activity - United States, 2014-15 season and composition of the 2015-16 influenza vaccine. *MMWR* 2015; **64**: 583–590.
3 Coudeville L, Bailleux F, Riche B, Megas F, Andre P, Ecochard R. Relationship between haemagglutination-inhibiting antibody titres and clinical protection against influenza: development and application of a bayesian random-effects model. *BMC Med Res Methodol* 2010; **10**: 18.
4 Hobson D, Curry RL, Beare AS, Ward-Gardner A. The role of serum haemagglutination-inhibiting antibody in protection against challenge infection with influenza A2 and B viruses. *J Hyg (Lond)* 1972; **70**: 767–777.
5 Fox JP, Cooney MK, Hall CE, Foy HM. Influenza virus infections in Seattle families, 1975-1979. II. Pattern of infection in invaded households and relation of age and prior antibody to occurrence of infection and related illness. *Am J Epidemiol* 1982; **116**: 228–242.
6 Ng S, Fang VJ, Ip DK, Chan KH, Leung GM, Peiris JS *et al.* Estimation of the association between antibody titers and protection against confirmed influenza virus infection in children. *J Infect Dis* 2013; **208**: 1320–1324.
7 Fox A, Mai le Q, Thanh le T, Wolbers M, Le Khanh Hang N, Thai PQ *et al.* Hemagglutination inhibiting antibodies and protection against seasonal and pandemic influenza infection. *J Infect* 2015; **70**: 187–196.
8 Benne CA, Kroon FP, Harmsen M, Tavares L, Kraaijeveld CA, de Jong JC. Comparison of neutralizing and hemagglutination-inhibiting antibody responses to influenza A virus vaccination of human immunodeficiency virus-infected individuals. *Clin Diagn Lab Immunol* 1998; **5**: 114–117.
9 Lee MS, Mahmood K, Adhikary L, August MJ, Cordova J, Cho I *et al.* Measuring antibody responses to a live attenuated influenza vaccine in children. *Pediatr Infect Dis J* 2004; **23**: 852–856.
10 Bucasas KL, Franco LM, Shaw CA, Bray MS, Wells JM, Nino D *et al.* Early patterns of gene expression correlate with the humoral immune response to influenza vaccination in humans. *J Infect Dis* 2011; **203**: 921–929.
11 Franco LM, Bucasas KL, Wells JM, Nino D, Wang X, Zapata GE *et al.* Integrative genomic analysis of the human immune response to influenza vaccination. *Elife* 2013; **2**: e00299.

12 Nakaya HI, Wrammert J, Lee EK, Racioppi L, Marie-Kunze S, Haining WN *et al.* Systems biology of seasonal influenza vaccination in humans. *Nat Immunol* 2011; **12**: 786–795.
13 Furman D, Jojic V, Kidd B, Shen-Orr S, Price J, Jarrell J *et al.* Apoptosis and other immune biomarkers predict influenza vaccine responsiveness. *Mol Syst Biol* 2013; **9**: 659.
14 Ovsyannikova IG, White SJ, Larrabee BR, Grill DE, Jacobson RM, Poland GA. Leptin and leptin-related gene polymorphisms, obesity, and influenza A/H1N1 vaccine-induced immune responses in older individuals. *Vaccine* 2014; **32**: 881–887.
15 Online Mendelian Inheritance in Man, OMIM. 2000. Johns Hopkins University, Baltimore, MD. MIM Number: *604718. Available from http://omim.org/entry/604718?search=TTF2&highlight=ttf2.
16 Online Mendelian Inheritance in Man, OMIM. 1996. Johns Hopkins University, Baltimore, MD. MIM Number: *600999. Available from http://omim.org/entry/600999.
17 UniProt. MAZ. Available from http://www.uniprot.org/uniprot/P56270.
18 Eklund KK, Niemi K, Kovanen PT. Immune functions of serum amyloid A. *Crit Rev Immunol* 2012; **32**: 335–348.
19 Hayashi EA, Akira S, Nobrega A. Role of TLR in B cell development: signaling through TLR4 promotes B cell maturation and is inhibited by TLR2. *J Immunol* 2005; **174**: 6639–6647.
20 Vazquez MI, Catalan-Dibene J, Zlotnik A. B cells responses and cytokine production are regulated by their immune microenvironment. *Cytokine* 2015; **74**: 318–326.
21 Online Mendelian Inheritance in Man, OMIM. 1999. Johns Hopkins University, Baltimore, MD. MIM Number: *604346. Available from http://www.omim.org/entry/604346?search=man1b1&highlight=man1b1.
22 Fagioli C, Sitia R. Glycoprotein quality control in the endoplasmic reticulum. Mannose trimming by endoplasmic reticulum mannosidase I times the proteasomal degradation of unassembled immunoglobulin subunits. *J Biol Chem* 2001; **276**: 12885–12892.
23 Hammond C, Braakman I, Helenius A. Role of N-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality control. *Proc Natl Acad Sci USA* 1994; **91**: 913–917.
24 Gebuhr I, Keeren K, Vogt K, Hoflich C, Appelt C, Schlieer U *et al.* Differential expression and function of alpha-mannosidase I in stimulated naive and memory CD4+ T cells. *J Immunother* 2011; **34**: 428–437.
25 Online Mendelian Inheritance in Man, OMIM. 2001. Johns Hopkins University, Baltimore, MD. MIM Number: *606343. Available from http://www.omim.org/entry/606343?search=POLL&highlight=poll.
26 Bertocci B, De Smet A, Flatter E, Dahan A, Bories JC, Landreau C *et al.* Cutting edge: DNA polymerases mu and lambda are dispensable for Ig gene hypermutation. *J Immunol* 2002; **168**: 3702–3706.
27 Schrader CE, Linehan EK, Ucher AJ, Bertocci B, Stavnezer J. DNA polymerases beta and lambda do not directly affect Ig variable region somatic hypermutation although their absence reduces the frequency of mutations. *DNA Repair (Amst)* 2013; **12**: 1087–1093.
28 Li Z, Li D, Tsun A, Li B. FOXP3 regulatory T cells and their functional regulation. *Cell Mol Immunol* 2015; **12**: 558–565.
29 Stewart R, Ohta Y, Minter RR, Gibbons T, Horton TL, Ritchie P *et al.* Cloning and characterization of Xenopus beta2-microglobulin. *Dev Comp Immunol* 2005; **29**: 723–732.
30 Online Mendelian Inheritance in Man, OMIM. 2001. Johns Hopkins University, Baltimore, MD. MIM Number: *300292. Available from http://www.omim.org/entry/300292?search=FOXP3&highlight=foxp3.
31 Sage PT, Sharpe AH. T follicular regulatory cells in the regulation of B cell responses. *Trends Immunol* 2015; **36**: 410–418.
32 Online Mendelian Inheritance in Man, OMIM. 1987. Johns Hopkins University, Baltimore, MD. MIM Number: *147290. Available from http://www.omim.org/entry/147290?search=INHBA&highlight=inhba.
33 Ogawa K, Funaba M, Tsujimoto M. A dual role of activin A in regulating immunoglobulin production of B cells. *J Leukocyte Biol* 2008; **83**: 1451–1458.
34 Winston DJ, Vikram HR, Rabe IB, Dhillon G, Mulligan D, Hong JC *et al.* Donor-derived West Nile virus infection in solid organ transplant recipients: report of four additional cases and review of clinical, diagnostic, and therapeutic features. *Transplantation* 2014; **97**: 881–889.
35 Online Mendelian Inheritance in Man, OMIM. 1991. Johns Hopkins University, Baltimore, MD. MIM Number: *138972. Available from http://www.omim.org/entry/138972?search=CEBPG&highlight=cebpg.
36 Online Mendelian Inheritance in Man, OMIM. 2001. Johns Hopkins University, Baltimore, MD. MIM Number: *605816. Available from http://www.omim.org/*entry/605816?search=EBI3&highlight=ebi3.
37 Online Mendelian Inheritance in Man, OMIM. 1992. Johns Hopkins University, Baltimore, MD. MIM Number: *161560. Available from http://www.omim.org/entry/161560?search=IL12A&highlight=il12a.
38 Lambert ND, Ovsyannikova IG, Pankratz VS, Jacobson RM, Poland GA. Understanding the immune response to seasonal influenza vaccination in older adults: a systems biology approach. *Expert Rev Vaccines* 2012; **11**: 985–994.
39 Poland GA, Ovsyannikova IG, Kennedy RB, Haralambieva IH, Jacobson RM. Vaccinomics and a new paradigm for the development of preventive vaccines against viral infections. *Omics* 2011; **15**: 625–636.
40 Poland GA, Ovsyannikova IG, Kennedy RB, Lambert ND, Kirkland JL. A systems biology approach to the effect of aging, immunosenescence and vaccine response. *Curr Opin Immunol* 2014; **29**: 62–68.
41 Ovsyannikova IG, White SJ, Albrecht RA, Garcia-Sastre A, Poland GA. Turkey versus guinea pig red blood cells: hemagglutination differences alter hemagglutination inhibition responses against influenza A/H1N1. *Viral Immunol* 2014; **27**: 174–178.
42 Salk HM, Haralambieva IH, Ovsyannikova IG, Goergen KM, Poland GA, Granzyme B. ELISPOT assay to measure influenza-specific cellular immunity. *J Immunol Methods* 2013; **398-399**: 44–50.
43 Haralambieva IH, Painter SD, Kennedy RB, Ovsyannikova IG, Lambert ND, Goergen KM *et al.* The Impact of immunosenescence on humoral immune response variation after influenza A/H1N1 vaccination in older subjects. *PLoS ONE* 2015; **10**: e0122282.
44 U.S. Food and Drug Administration. Influenza Virus Vaccine for the 2010-2011 Season. Silver Spring, MD 2015. Available from http://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Post-MarketActivities/LotReleases/ucm202750.htm.
45 Jacobson RM, Grill DE, Oberg AL, Tosh PK, Ovsyannikova IG, Poland GA. Profiles of influenza A/H1N1 vaccine response using hemagglutination-inhibition titers. *Hum Vaccin Immunother* 2015; **11**: 961–969.
46 Bentebibel SE, Lopez S, Obermoser G, Schmitt N, Mueller C, Harrod C *et al.* Induction of ICOS+CXCR3+CXCR5+ TH cells correlates with antibody responses to influenza vaccination. *Sci Transl Med* 2013; **5**: 176ra32.
47 Kasturi SP, Skountzou I, Albrecht RA, Koutsonanos D, Hua T, Nakaya HI *et al.* Programming the magnitude and persistence of antibody responses with innate immunity. *Nature* 2011; **470**: 543–547.
48 Wang S, Taaffe J, Parker C, Solorzano A, Cao H, Garcia-Sastre A *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. *J Virol* 2006; **80**: 11628–11637.
49 World Health Organization. *WHO Global Influenza Surveillance Network: Manual for the Laboratory Diagnosis and Virological Surveillance of Influenza* 2011. Available from http://apps.who.int/iris/bitstream/10665/44518/1/9789241548090_eng.pdf.
50 Brady RC, Treanor JJ, Atmar RL, Keitel WA, Edelman R, Chen WH *et al.* Safety and immunogenicity of a subvirion inactivated influenza A/H5N1 vaccine with or without aluminum hydroxide among healthy elderly adults. *Vaccine* 2009; **27**: 5091–5095.
51 Kennedy RB, Oberg AL, Ovsyannikova IG, Haralambieva IH, Grill DE, Poland GA. Transcriptomic profiles of high and low antibody responders to smallpox vaccine. *Genes Immun* 2013; **14**: 277–285.
52 Haralambieva IH, Oberg AL, Ovsyannikova IG, Kennedy RB, Grill DE, Middha S *et al.* Genome-wide characterization of transcriptional patterns in high and low antibody responders to rubella vaccination. *PLoS ONE* 2013; **8**: e62149.
53 Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009; **10**: R25.
54 Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010; **26**: 841–842.
55 Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009; **25**: 1105–1111.
56 Hansen KD, Irizarry RA, Wu Z. Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics* 2012; **13**: 204–216.
57 Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010; **26**: 139–140.
58 Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* 2010; **11**: R25.
59 Hardin JW, Hilbe JM. *Generalized Estimating Equations*. Chapman and Hall/CRC: London, 2003.
60 McCullagh P, Nelder JA. *Generalized Linear Models*. Chapman and Hall: London, 1983.
61 McCulloch CE, Searle SR, Neuhaus JM. *Generalized, Linear, and Mixed Models*2nd ednWiley: Hoboken, NJ, 2008.
62 Oberg AL, Bot BM, Grill DE, Poland GA, Therneau TM. Technical and biological variance structure in mRNA-Seq data: life in the real world. *BMC Genomics* 2012; **13**: 304.

63 Fridley BL, Jenkins GD, Grill DE, Kennedy RB, Poland GA, Oberg AL. Soft truncation thresholding for gene set analysis of RNA-seq data: application to a vaccine study. *Sci Rep* 2013; **3**: 2898.
64 Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005; **102**: 15545–15550.
65 Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 2011; **27**: 1739–1740.
66 Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Ser B* 2005; **67** (Part 2): 301–320.
67 Harrell FE. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer: New York, 2001.
68 Friedman J, Hastie T, Tibshirani R. regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 2010; **33**: 1–22.
69 Prasad TS, Kandasamy K, Pandey A. Human protein reference database and human proteinpedia as discovery tools for systems biology. *Methods Mol Biol* 2009; **577**: 67–79.
70 Rolland T, Tasan M, Charloteaux B, Pevzner SJ, Zhong Q, Sahni N *et al.* A proteome-scale map of the human interactome network. *Cell* 2014; **159**: 1212–1226.
71 Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, Hannay T *et al.* PID: the pathway interaction database. *Nucleic Acids Res* 2009; **37** (Database issue): D674–D679.
72 Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P *et al.* The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res* 2011; **39** (Database issue): D561–D568.
73 Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003; **13**: 2498–2504.
74 AllegroViva. AllegroLayout 2015. Available from http://allegroviva.com/allegrolayout2.
75 Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; **25**: 25–29.
76 Rhee SY, Wood V, Dolinski K, Draghici S. Use and misuse of the gene ontology annotations. *Nat Rev Genet* 2008; **9**: 509–515.
77 Gene Ontologly Annotations. GO Terms from GOA 2015. Available from http://geneontology.org/gene-associations/gene_association.goa_human.gz.

Supplementary Information accompanies this paper on Genes and Immunity website (http://www.nature.com/gene)