UNIVERSITY OF CALIFORNIA, SAN DIEGO

Identification of Markers and Mechanisms in Infection and Immunity

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy

in

Bioengineering

by

Thomas Lafayette Hagan

Committee in charge:

Professor Shankar Subramaniam, Chair
Professor Gaurav Arya
Professor Todd Coleman
Professor Shyni Varghese
Professor Sheng Zhong

2015

The dissertation of Thomas Lafayette Hagan is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2015

DEDICATION

To Mom, Dad, and Daniel. Thanks for your constant love, encouragement, and support.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ANN – Artificial neural network

ASC – Antibody secreting cell

BTM – Blood transcriptional module

CAMK4 – Calcium-dependent protein kinase 4

DC – Dendritic cell

ELISA - Enzyme-linked immunosorbent assay

ELISPOT - Enzyme-linked immunospot assay

FACS – Fluorescence-activated cell sorting

GCN2 – General control nonderepressible 2 kinase

GSEA – Gene set enrichment analysis

HAI – Hemagluttinin inhibition assay

HDLSS – High dimension, low sample size

HLA – Human leukocyte antigen

IFN-γ – Interferon gamma

IL - Interleukin

LAIV – Live attenuated influenza vaccine

LPS - Lipopolysaccharide

LXR – Liver X receptor

miRNA – micro RNA

NFkB - Nuclear factor kappa-light-chain-enhancer of activated B cells

NK cell – Natural killer cell

PBMC – Peripheral blood mononuclear cell

RMA – Robust multi-array average

ssGSEA – Single sample gene set enrichment analysis

SVM – Support vector machine

T2D – Type 2 Diabetes

Th cell – T helper cell

TIV – Trivalent inactivated influenza vaccine

TLR - Toll-like receptor

TNF - Tumor necrosis factor

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge all of the people that I have worked with in the Subramaniam Lab over the last 5 years.

Chapter 1 is a modified presentation of material published as “Systems vaccinology: Enabling rational vaccine design with systems biological approaches” by Hagan T, Nakaya H, Subramaniam S, and Pulendran S in Vaccine 2015. The dissertation author was the primary author of this material.

Chapter 2 is a modified presentation of material that is being prepared for publication as “Systems analysis of immunity to influenza vaccination across multiple years and in diverse human populations” by Nakaya HI, Hagan T, Duraisingham SS, Lee EK, Kwissa M, Rouphael N, Frasca D, Gersten M, Mehta AK, Gaujoux R, Li GM, Gupta S, Ahmed R, Mulligan MJ, Shen-Orr S, Blomberg BB, Subramaniam S, and Pulendran B. The dissertation author was the co-primary investigator and author of this material.

Chapter 3 is a modified presentation of material that is being prepared for publication as “Systems analysis of immunity to influenza vaccination across multiple years and in diverse human populations” by Nakaya HI, Hagan T, Duraisingham SS, Lee EK, Kwissa M, Rouphael N, Frasca D, Gersten M, Mehta AK, Gaujoux R, Li GM, Gupta S, Ahmed R, Mulligan MJ, Shen-Orr S,

Blomberg BB, Subramaniam S, and Pulendran B. The dissertation author was the co-primary investigator and author of this material.

Chapter 4 is a modified presentation of material that is being prepared for publication as “Systems analysis of stimulated dendritic cells” by Hagan T, Bommakanti G, Li S, Bhargava V, Pulendran B, and Subramaniam S. The dissertation author was the co-primary investigator and author of this material.

Chapter 5 is a modified presentation of material published as “Systems vaccinology: Enabling rational vaccine design with systems biological approaches” by Hagan T, Nakaya H, Subramaniam S, and Pulendran S in Vaccine 2015. The dissertation author was the primary author of this material.

VITA

Education

2010-2015	University of California, San Diego, La Jolla, CA Doctor of Philosophy in Bioengineering
2006-2010	University of Notre Dame, Notre Dame, IN Bachelor of Science in Chemical Engineering, *summa cum laude*

ABSTRACT OF THE DISSERTATION

Identification of Markers and Mechanisms in Infection and Immunity

by

Thomas Lafayette Hagan

Doctor of Philosophy in Bioengineering

University of California, San Diego, 2015

Professor Shankar Subramaniam, Chair

Successful immune response to infection involves coordinated interactions among a myriad of cell types located across multiple tissues in the body. As a result of the highly complex nature of this biological system, many of the mechanisms involved in generating immune responses, particularly at a molecular level, remain unknown. Recent advances in high-throughput

technologies, along with the development of computational tools to integrate diverse data types and build comprehensive models of biological responses, are enabling insight into these molecular mechanisms at greater depth than ever before.

This dissertation details a systems biology approach to identifying reliable molecular signatures of immune response to vaccination and infection and utilizing these signatures to predict responsiveness to vaccination and improve understanding of the molecular processes underlying the immune system. Chapter 1 provides an overview of the field of systems vaccinology, which employs systems biology approaches to study vaccine response. Chapter 2 describes the development of a predictive classifier of influenza vaccine response, capable of discriminating between responders and nonresponders to vaccination through identification of distinguishing transcriptional signatures post-vaccination. In Chapter 3, the analysis of immune response to influenza vaccine is extended to examine age-associated changes in response, signatures of response longevity, and miRNA regulation of transcription post-vaccination. Chapter 4 presents an integrated transcriptional and metabolic analysis of dendritic cells, an important innate immune cell, stimulated *in vitro* with LPS, a bacterial endotoxin that stimulates immune responses to infection. Lastly, Chapter 5 discusses the potential impact of systems biology and predictive analysis on vaccine use and development, as well as future directions and challenges for this field.

Chapter 1: General Introduction

Introduction

Since Edward Jenner's discovery in the late 1700's that inoculation with the cowpox virus provided protection from smallpox infection, vaccines have emerged as one of the greatest public health tools in history. The last 60 years have established a golden age in the field of vaccinology, marked by events such as the eradication of smallpox by 1980 (Behbehani, 1983) and the development of polio vaccines in the 1950s, which have lead to near-eradication of the disease (Ehrenfeld et al., 2009). Despite the great success of these and other vaccines, there remain significant challenges for development of new vaccines against current global pandemics such as HIV and malaria. Among the many problems facing this field are: (i) most currently used vaccines were designed largely empirically. As a result there is little or no understanding of what the correlates and mechanisms of protection are for many vaccines. For example, although the two commercially available types of influenza vaccine, trivalent inactivated (TIV) and live attenuated (LAIV), provide similar levels of protection from infection (Sasaki et al., 2007), they generate significantly different immune responses. TIV induces higher levels of IgG antibody secreting cells (ASCs) in the blood as well as higher levels of serum antibodies than LAIV in adults. This is likely due to the different routes of administration, as LAIV, which is administered as an intranasal spray, is thought to produce a more local response in the nasal mucosa and upper respiratory tract, including IgA (mucosal) antibodies and cellular immune

responses. As a result, the correlate of protection for TIV is generally considered to be serum antibodies, while the correlate of protection for LAIV is less clear (Sasaki et al., 2007). (ii) The path to licensure of candidate vaccines involves very lengthy and expensive phase IIB and III clinical trials to assess their efficacy and safety. These trials involve thousands of subjects and can cost hundreds of millions of dollars to complete. As a result, very few vaccine concepts are tested in phase III trials. For example, during the past 30 years, only 4 HIV-1 vaccine concepts have been tested for clinical efficacy (Barouch and Michael, 2014), and despite repeated failures, the correlates and mechanisms of protective immunity against HIV remain poorly understood.

The conventional immunological methods, such as ELISA, ELISPOT, flow cytometry, etc., used to study vaccines have played a valuable role in the field of vaccinology, and will remain essential in evaluating responses to vaccination in the future. However these approaches are generally only able to analyze a single or small number of components of the immune system at a given time, and are insufficient to analyze the full complexity of the structure and dynamics of the human immune system as a whole. This represents a critical obstacle towards understanding the molecular mechanisms by which vaccines generate protective immune responses and identifying meaningful correlates of protection.

To address this issue, vaccinologists have turned to systems biology. By examining how coordinated interactions at a molecular level give rise to

immune responses, systems biology approaches enable a holistic view of the immune system and its many components. This developing field provides many promising tools to overcome the challenges facing current vaccine development. Enabling researchers to evaluate the immune responses of fewer subjects in a more in-depth and detailed fashion has the potential to dramatically improve our understanding of the mechanisms of protection of novel vaccines and decrease the length and costs of current clinical trials.

Systems vaccinology

Within the past 20 years, advances in high-throughput technologies have granted researchers the ability to interrogate the properties and abundances of entire classes of molecular components within the cell. For example, development of lower cost next-generation sequencing technology has facilitated the growth of transcriptomics, which seeks to measure the expression of all RNA transcripts within a cell or population of cells. By sequencing and mapping mRNA transcripts, RNA-sequencing enables the accurate quantification of gene expression as well as simultaneous identification of RNA structure such as transcription start site and exon usage/splice junctions, the regulation of which has been shown to play an important role in many biological processes, including within the immune system (Carpenter et al., 2014; Funderud et al., 2009). Simultaneously, in the growing domain of metabolomics, analytical chemistry techniques such as

liquid chromatography-mass spectrometry (LC-MS) have been harnessed to identify and quantify the set of metabolites within cells or tissues (Patti et al., 2012). Changes in metabolic activity are an important component of both innate and adaptive immune responses (Kominsky et al., 2010), such as the recognized role that lipid metabolism plays during inflammation (Bensinger and Tontonoz, 2008; Kominsky et al., 2010; Libby et al., 2011).

Systems vaccinology is an emerging field that applies such 'omics' technologies, in combination with bioinformatics tools such as transcriptional network analysis and predictive modeling, to study immune responses to vaccination (Nakaya et al., 2012; Pulendran, 2014; Pulendran et al., 2010). As a systems-based approach, it aims to use data generated through high-throughput measurements in the context of vaccination to characterize the interactions between individual components of the immune system in the interest of understanding and predicting behavior of the system as whole. This includes analysis of transcriptional, signaling, and metabolic pathways whose activity is perturbed in the various cells of the immune system in response to vaccination, as well as identification of molecular signatures that are predictive of various measurements of protection from infection. The knowledge obtained through these analyses can aid in the rational design of new vaccines that generate long-lasting protection and induce improved responses in populations with diminished immune function such as the elderly.

5 year historic perspective

The first examples of the use of such approaches to study responses to vaccination were performed on the yellow fever vaccine (Gaucher et al., 2008; Querec et al., 2009). This vaccine contains a live-attenuated strain (YF-17D) of the yellow fever virus, which induces potent and long-lived CD8+ T cell and neutralizing antibody responses (Pulendran, 2009; Pulendran et al., 2013). By combining high-throughput measurements such as microarray gene expression profiling and multiparameter flow cytometry with computational modeling, we were able to detect a regulated network of interferon and innate antiviral genes that were induced post-vaccination in peripheral blood mononuclear cells (PBMCs) (Querec et al., 2009). An independent YF-17D study by Gaucher et al. revealed induction of similar transcriptional responses to vaccination, including type I interferon and inflammatory pathways (Gaucher et al., 2008). In addition to examining innate immune pathways activated by vaccination, we successfully identified unique gene signatures that were capable of accurately predicting the CD8+ T cell and neutralizing antibody responses, respectively (Querec et al., 2009). The predictive CD8+ T cell signature contained complement protein C1qB and eukaryotic translation initiation factor 2 alpha kinase 4, which is an orchestrator of the integrated stress response. Meanwhile the B cell growth factor receptor TNFRSF17 was among the genes included in the antibody response signature. This work demonstrated for the first time that the immunogenicity of a vaccine could be

successfully predicted using early transcriptional measurements within 1 week of vaccination.

Following these initial studies, systems biology approaches have been used to examine immune responses to vaccines against a wide range of pathogens, including influenza (Bucasas et al., 2011; Nakaya et al., 2011), malaria (Vahey et al., 2010), smallpox (Reif et al., 2009), and HIV (Zak et al., 2012). In particular, as YF-17D is a live-attenuated vaccine that induces an acute viral infection, the study of influenza vaccination (TIV) enabled investigation into to whether or not similar methods could be used to identify molecular signatures predictive of response to an inactivated vaccine. We identified transcriptional signatures related to the expansion of plasmablasts and the unfolded protein response within B cells on day 7 post-vaccination that correlated with and were predictive of day 28 influenza-specific antibody responses (Nakaya et al., 2011). Indeed, these findings were consistent with studies by Bucasas et al. (Bucasas et al., 2011) and Obermoser et al. (Obermoser et al., 2013). Interestingly, TNFRSF17, which was predictive of antibody responses to YF-17D, also appeared in the signatures predictive of TIV response (Nakaya et al., 2011). Recently, Tsang et al. (Tsang et al., 2014), and Furman et al. (Furman et al., 2013) have extended this approach to search for baseline signatures capable of discriminating between high and low responders to vaccination. However, possibly due to limited sample sizes and weaker signal at baseline, neither study was able to successfully predict

antibody response using baseline transcriptional measurements alone. Instead, Tsang et al. utilized cell subset frequencies, while Furman et al. generated a model based on transcriptional modules, serum cytokines, cell subset frequencies, and pre-existing antibody titers. Additionally, as these studies were conducted using cohorts from an individual flu season, the effect of changes in influenza strains included in the TIV vaccine on the performance of these models remains to be examined. To address this question, we are performing a comprehensive analysis of over 400 adults vaccinated with seasonal TIV during 5 consecutive influenza seasons (Nakaya et al., manuscript in preparation). This analysis is an important step towards generating robust and clinically relevant signatures that can be used to predict the efficacy of vaccines in clinical trials.

Among the vaccines under investigation, malaria is one for which a human challenge model exists, allowing for identification of subjects who are protected from or susceptible to infection (Chulay et al., 1986). Vahey et al. used this model, in which subjects vaccinated with the RTS,S malaria vaccine were challenged using mosquitos infected with an antimalarial-sensitive strain of *Plasmodium*, to identify transcriptional signatures in PBMCs capable of discriminating between protected and nonprotected vaccinees (Vahey et al., 2010). Protected subjects had increased expression 2 weeks post-vaccination (but before challenge) of genes involved in the immunoproteasome degradation pathway, involved in MHC peptide processing, compared to

subjects who developed parasitemia (Vahey et al., 2010). These systems-based approaches can be used to identify early signatures of efficacy of other vaccines for which human challenge models have been developed, such as typhoid (Marwick, 1998; Waddington et al., 2014) and *Shigella* (Bodhidatta et al., 2012), enabling rapid evaluation of novel candidate vaccines in future clinical trials.

A critical question is the degree to which transcriptional responses to vaccination were conserved across vaccines, and whether or not there are universal signatures capable of predicting antibody responses to vaccination. To this end, we used a systems-based approach to compare signatures induced by different types of vaccines [YF-17D, LAIV, TIV, the carbohydrate meningococcal vaccine (Menimmune), and the conjugate meningococcal vaccine (Menectra)] (Li et al., 2014). We observed that while recall antibody responses to inactivated vaccines (e.g. seasonal influenza vaccine, diphtheria toxoid component of the conjugate meningococcal vaccine) were associated with transcriptional modules related to plasmablast differentiation, the antibody responses of live-attenuated vaccines (e.g. yellow fever vaccine) were highly correlated with modules involving innate immunity and type I interferon responses. Meanwhile, antibody responses to polysaccharide components of the meningococcal vaccines were associated with increased proinflammatory cytokines as well as activation of antigen-presenting cells (Li et al., 2014).

Thus, these preliminary results suggest that vaccine-induced signatures of immunity are significantly dependent on the class of vaccine in question.

Finally, systems approaches could also potentially be used to identify signatures of vaccine safety. Adverse reactions to vaccination pose major regulatory challenges to vaccine development. Relatively mild adverse reactions such as transient fever, local swelling at the injection site are fairly common and occur within a few hours or days after vaccination. The more serious adverse reactions such as the often fatal viscerotropic disease caused by yellow fever vaccination are very rare (1 in 250,000 cases (Pulendran, 2009)). The ability to identify signatures at baseline that would be predictive of such adverse reactions would thus represent an important advance, however a major obstacle to this is the rarity of such serious adverse events. In this context, detailed immunological characterization of a patient who developed viscerotropic disease after yellow fever vaccination, revealed a 200-fold expansion of CD14+CD16+ inflammatory monocytes during the period of disease, and a 20-fold elevation in the numbers of such cells even during the convalescent phase, long after the virus had been cleared (Pulendran et al., 2008). This raises the possibility that the persistently elevated levels of the CD14+CD16+ inflammatory monocytes may have predisposed this subject towards viscerotropic disease after vaccination with the yellow fever vaccine. The further evaluation of this possibility is stymied by the rarity of these adverse reactions, and the difficulty in obtaining samples to perform detailed

analysis. Future efforts should thus be aimed at creating a clinical infrastructure that would facilitate the acquisition of samples from vaccine adverse reactions at multiple time points, including the convalescent phase.

Drowning in a sea of big data

The revolution in high-throughput technologies has created a dramatic change in the landscape of biological research: with such large amounts of information generated with each experiment, we are now 'drowning in a sea of big data.' A cornerstone of successful systems vaccinology is the ability to extract meaningful knowledge from this wealth of data, and then to harness this knowledge to generate improved biological understanding of the immune system's response to vaccination. In the first transition, from data to knowledge, statistical approaches (Jeffery et al., 2006) are generally used to identify sets of genes, metabolites, etc. that undergo significant changes during system perturbation (e.g. before and after vaccination). However, a simple list of differentially expressed genes or metabolites is not highly informative, and may contain false positives. One approach to improve this type of analysis is to integrate expression measurements with *a priori* knowledge about the interactions and co-expression of genes within a given biological context. One such bioinformatics tool is gene-set enrichment analysis (Subramanian et al., 2005), which identifies enrichment of biologically relevant groups of genes (modules) within gene lists ranked by response to

perturbation. This can be performed using manually curated sets of biological pathways, such as the Kyoto Encyclopedia of Genes and Genomes (Kanehisa et al., 2006) and Reactome (Vastrik et al., 2009) databases, or through extraction of gene modules from existing data by searching for groups of co-expressed genes across experimental conditions or biological contexts (Chaussabel and Baldwin, 2014; Chaussabel et al., 2008; Li et al., 2014). MSigDB, a popular gene set database, contains collections of both types (Liberzon et al., 2011). Examining changes in activity of pathways or modules rather than individual genes reduces noise and appearance of false positives while simultaneously providing improved biological meaning and functional context to the data analysis.

Differential transcriptional pathways and networks can also be coupled with feature selection and machine learning algorithms (such as those mentioned in the previous section) to generate biologically informative predictive signatures of vaccine response (Fig. 1.1). These signatures can be used to assess vaccine efficacy and identify subjects at risk for infection. In addition, analysis of predictive pathways can yield insight into the mechanisms responsible for deficient immune responses. Once predictive signatures are generated using the data at hand, they should be validated in independent cohorts to ensure their robustness across diverse populations.

While the knowledge of biological pathways activated or suppressed in response to vaccination is useful in its own right, achieving deeper

understanding and control over the mechanisms by which vaccines induce protective immune responses requires human knowledge and insight coupled with traditional methods of experimental validation. Ideally, observations made through high-throughput data analysis can help generate biological hypothesis that are then tested in animal models such as knockout mice. Recently our group has used this approach to identify a novel role for the integrated stress response, mediated by EIF2AK4 [eukaryotic initiation factor 2 α-kinase 4, also known as general control nonderepressible 2 kinase (GCN2)], in stimulating autophagy and antigen presentation within dendritic cells during response to YF-17D (Ravindran et al., 2014). In our original systems biology analysis of the YF-17D vaccine, expression of GCN2 in PBMCs on day 7 post-vaccination was shown to be predictive of the later CD8+ T cell response (Querec et al., 2009). As a regulator of the integrated stress response, accumulation of uncharged tRNA during amino acid starvation activates GCN2, which results in phosphorylation of eukaryotic initiation factor 2 α (eIF2α). This process leads to reduced activity of the eIF2 complex, resulting in reduced rates of protein synthesis and stress granule formation (Anderson and Kedersha, 2002). However, GCN2's role in immune responses is not clear. The capability of GCN2 expression to predict CD8+ T cell responses to YF-17D led us to believe that this kinase may be important for priming of the adaptive immune response. Follow up work in mice revealed that YF-17D induced amino acid starvation in dendritic cells, inducing autophagy and

antigen cross-presentation in a GCN2-dependent manner (Ravindran et al., 2014). This reveals an unappreciated connection between the ancient nutrient sensing pathway and elicitation of adaptive immunity by dendritic cells.

Another mechanistic discovery enabled by a systems vaccinology is the role that the intestinal microbiota play in stimulating antibody responses to vaccination. This development was prompted by the finding in our systems-level analysis of human immune responses to TIV vaccination that expression of the toll-like receptor TLR5 on day 3 post-vaccination was significantly correlated with the day 28 antibody response (Nakaya et al., 2011). As TLR5 is a receptor specific for flagellin, the protein subunit of bacterial flagellum, its association with antibody response to a viral vaccine was unexpected. The importance of TLR5 sensing of intestinal bacteria during TIV vaccination was confirmed in experiments with antibiotic treated, germ-free, and $TLR5^{-/-}$ mice, all of which experienced significant reduction in antibody responses after TIV vaccination compared to controls (Oh et al., 2014). Further work demonstrated that sensing of flagellin by TLR5 promoted plasma cell differentiation directly and by stimulating lymph node macrophages to produce plasma cell growth factors. Interestingly, the loss of intestinal microbiota did not result in reduction of humoral responses to adjuvanted vaccines or the live-attenuated YF-17D vaccine, indicating that the microbiota may be acting as a natural adjuvant in the absence of other sources of immune stimulation. As the microbiome is affected by diet and varies widely across the world

(Arumugam et al., 2011), gut microbial composition adds another layer of diversity in the human population that should be considered in future studies of immune responses to vaccination. The experimental evaluation of the role of the microbiome in promoting immunity during vaccination in humans is ongoing.

Acknowledgements

Chapter 1 is a modified presentation of material published as “Systems vaccinology: Enabling rational vaccine design with systems biological approaches” by Hagan T, Nakaya H, Subramaniam S, and Pulendran S in Vaccine 2015. The dissertation author was the primary author of this material.

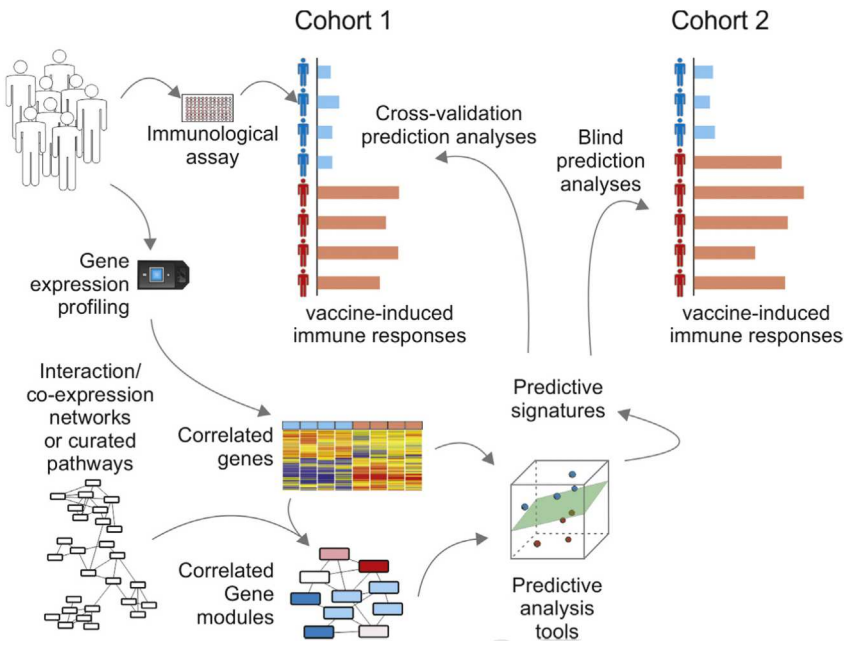

Figure 1.1 Predictive analysis of vaccine response.
The goal of predictive analyses is to identify early transcriptional signatures capable of predicting vaccine-induced immune responses. Integration of transcriptomic measurements with gene co-expression networks or biological pathways provides improved functional understanding of the immune processes activated in response to vaccination. These gene or module-level features can then be used as inputs to feature selection and machine learning algorithms. Such tools enable identification of molecular signatures that are predictive of responses to vaccination, such as antibody or T cell responses. Once predictive signatures are verified through blind prediction of independent datasets, they can be employed in a clinical setting to rapidly identify vaccinees with deficient responses and assess vaccine efficacy.

Chapter 2: Predictive analysis of response to influenza vaccination

Introduction

Influenza is a major public health problem, causing on average three to five million cases of severe illness and about 250,000 to 500,000 deaths per year worldwide (WHO, 2014). Annual vaccination has been established as one of the most effective methods for preventing influenza. However, vaccination does not always confer protection from infection, particularly in the elderly, who exhibit diminished responsiveness. For the trivalent inactivated influenza vaccine (TIV), serum antibodies have been established as a correlate of protection from infection (Sasaki et al., 2007). However, there is a significant delay of around 10 days between vaccination and peak antibody response (Hallilеy et al., 2010), which makes evaluation of vaccine response difficult in a clinical setting. A diagnostic test capable of identifying at early time points (or potentially even before vaccination) patients who will not respond to the vaccine would allow these patients to recieve alternate treatments to ensure their protection against infection, as well as provide researchers with a better understanding of the immune system's response to vaccination.

There have been several previous studies aimed at identifying markers of antibody response to the influenza vaccine. Nakaya et al. (Nakaya et al., 2011) used a SVM-based approach called 'discriminant analysis via mixed integer programming' (DAMIP) to identify several sets of 2-4 genes whose expression levels using day 3 or day 7 fold change data were predictive of

antibody response to the influenza vaccine. Not suprisingly, many of the genes identified on day 7 were associated with the adaptive immune response and are highly expressed in ASCs. However they did identify some novel markers, such as CAMK4, a calcium-dependent protein kinase that was negatively correlated on day 3 with the antibody response. CAMK4 had been previously implicated in T cell development and inflammatory responses, but not in B cell development. They were able to show through follow-up experiments in knockout mice that loss of CAMK4 does increase the antibody response to vaccination in mice by up to 6.5-fold. In another study, Furman et al. (Furman et al., 2013) examined features of baseline samples from 89 vaccinees using a logistic regression model in an attempt to identify pre-vaccination markers of antibody response. They characterized a variety of parameters from the baseline blood samples such as gene expression (using microarrays), serum cytokine levels, antibody levels against particular hemagglutinin peptides, and cell subset phenotypes. Among the markers that negatively correlated with the antibody response were the subject's age and pre-existing hemagglutinin antibodies, both of which have been previously implicated in reducing responsiveness to vaccination. In addition, they identified several gene modules related to apoptosis that positively correlated with antibody response. They hypothesized that apoptosis was required for turnover of old memory B cells that are not likely to be cross-reactive to new strains of influenza, which

creates room within the B cell compartment for expansion of ASCs which are reactive to the novel strains in the vaccine.

While both of these studies made interesting discoveries, they did suffer from certain limitations. The gene sets of Nakaya et al. were identified using fold change data, normalizing the day 3 or day 7 expression values to the baseline expression for each subject. If the conclusions of Furman et al. that there are gene expression markers within the blood at baseline which are predictive of antibody response are correct, using fold change gene expression values results in a loss of useful information. In addition, using fold change measurements is less ideal for diagnostic purposes as it requires two measurements, thereby doubling the cost and inconvenience for the patient. Meanwhile, Furman et al. used data obtained from a single flu season and location, meaning that their model could suffer from flu strain or location specific biases. Furthermore, although both studies claim to perform feature selection, neither SVMs nor logistic regression algorithms are methods for feature selection. In both cases it is left unclear how the dimension of the data was reduced in order to build their models (containing 2-4 inputs for Nakaya et al. and up to 16 inputs for Furman et al.). Depending on how these inputs were chosen, especially if large numbers of features were tested as inputs for the models, as well as which portion of the dataset was used during this procedure, it is quite possible that overfitting is occurring.

With this previous work in mind, the goal of this study is to identify robust early markers that are not strain or population specific from the RNA levels of isolated peripheral blood mononuclear cells (PBMCs) which are predictive of the later antibody response to the influenza vaccine, and to develop, using machine learning approaches, a classifier which is capable of accuractely preditcing antibody response to vaccination based on these early biomarkers. The identification of these markers using biologically informed methods of feature selection, as well as information gained by examining the output of the classifier (phenotypes and demographics of misclassified patients, effect of flu strain type on accuracy, etc.) will allow us to form biologically testable hypotheses about the immune response to vaccination and the mechanisms by which this response fails in certain patients, as depicted in Figure 2.1.

The dataset available to us consists of microarray data from isolated PBMCs on days 0, 3, and 7 post-vaccination and corresponding antibody titers from hemagglutination inhibition (HAI) assays 28 days post-vaccination of 212 total vaccinees from 5 separate flu seasons and 2 separate locations (Atlanta, GA and Miami, FL). Microarray data was normalized using the robust multi-array average algorithm (RMA) (Irizarry et al., 2003) and adjusted for batch effects between trials using ‘ComBat’ software (Johnson et al., 2007). Of the 212 subjects, 54 are elderly (65 or older) and 15 have type II diabetes. This large and heterogeneous dataset will provide us with the statistical power

required to detect signal from noise, as well as the ability to identify reliable markers which are not strain, age, or population specific. Based on the FDA's definition of seroconversion as an HAI titer of 1:40 or more and a minimum fourfold increase in antibody titer after vaccination (FDA, 2007), for preliminary analysis and comparison with previous work, the subjects have been classified as 'low responder' or 'high responder' according to whether or not they meet the criteria for seroconversion for at least 1 of the 3 strains included in the vaccine. The number of high and low responders in each of the 5 trials according to this criteria is listed in Table 2.1. It is noteworthy that in all trials except 2009 and 2011 there are significantly more high responders than low responders, which reflects the fact that flu vaccines are generally effective. However this bias should be kept in mind during the development of a classifier, as optimizing for overall accuracy will result in reduced ability to properly identify low responders (low specificity or high false positive rate). For initial analysis, the dataset was divided by randomly selecting 15% of the samples from each trial to create a testing set, and the remaining samples were used for training.

Feature selection

The goal of machine learning algorithms is to find the boundary or boundaries that optimally separate the samples within the feature space according to their labels. As such, a crucial step in building a well performing

classifier is selecting features to represent the samples such that samples are well separated according to their class label within the feature space. This step is particularly challenging with datasets such as this one because microarrays provide readouts for a very large number of probes (around 50,000 in this case) but are fairly costly to generate, resulting in a small number of samples. This results in a dataset with dimension p that is much larger than the sample size n, sometimes refered to as ‘high dimension, low sample size’ (HDLSS) data or ‘large p small n’ data. These types of data are challenging to work with because standard statisical methods are not well equipped to distinguish between true differences and randomness in this limit.

Using standard statistical tests in an undirected way ignores biological knowledge about the system that may increase our ability to detect predictive features. For example, interaction networks such as protein-protein interaction or transcription factor-target networks provide information about which groups of proteins (and their corresponding genes) are likely to be simultaneously expressed during a particular biological response. One way to include such information is to use these networks to find groups of genes whose aggregate expression distinguishes between two groups of samples. One such procedure, PinnacleZ, was developed by Chuang et al. (Chuang et al., 2007). In this algorithm, gene expression data is mapped to a given interaction network, and then a greedy search algorithm is used to find subnetworks of genes whose aggregate expression differentiates between the two classes of

samples. Significance testing is then performed either by generating a null distribution of random subnetworks, or by repeating the search using randomly permuted class labels. In their paper, Chuang et al. show that this method provides more accurate and robust prediction than simply using gene-level features on a metastatic versus non-metastatic breast cancer dataset. It makes sense that these features will have reduced variability because by using an aggregate expression value some of the inherent noise in the microarray measurements will be averaged out. Once significant modules are identified, they can be added as inputs to the classifier. In a technique called forward selection (a method which is applicable to any feature type), the identified features are added one at a time to the set of inputs, selecting the feature which increases the accuracy by the greatest amount at each iteration. This continues until the accuracy stops increasing significantly, such that near maximum accuracy is achieved while minimizing the complexity of the classifier (in terms of number of inputs).

This algorithm, when applied to the flu dataset, should help to identify networks of genes within pathways that are crucial to succesful antibody development in response to vaccination. However, initial analysis using selected modules as inputs to an ANN classifier in Figure 2.2 reveals that there is overfitting of the modules to the training data even when only using modules considered to be very significant by the PinnacleZ algorithm ($p<0.05$). As inputs are added to the classifier using forward selection, the training accuracy

continues to increase while the tesing accuracy does not, resulting in a training accuracy much higher than in testing. It is likely that this overfitting is generated as a result of the large number of possible combinations available to the algorithm during module generation. Because there are such a large number of interactions (□200,000 known interactions for the human network), the number of possible subnetworks is essentially infinite. The performance of a given module is supposedly compared to a distribution of random subnetworks as a signifigance test (Chuang et al., 2007). This will determine whether the perfomance of the module is better than those generated randomly, and should indicate that the selected modules contain genes involved in the biological response in question. However, because the algorithm is greedy (any gene which improves the t-test or mutual information score is included) and there are many options available due to the large number of interactions, it seems likely that although a selected module may perform better than random because some of the genes are truly relevant to the bioloigcal process, there will also be some genes included in the module that happened to increase the score on the training data by chance but are not bioloigcally involved. This type of overfitting was not addressed by the authors.

One possible approach to reduce the inclusion of overfitting genes into modules is to perform module generation on smaller portions of the training data, and then to look for common modules or sub-modules. In this particular case, the training data is easily separated into the 5 trials. As a first approach,

modules were generated using a leave-one-trial-out (LOTO) setup where the training data for each trial was left out once and modules were generated using the remaining training data. Modules generated from leaving the 2008 trial out were chosen as a basis, and were compared to the significant modules from the other LOTO cases. Those with a Jaccard index (size of intersection/size of union) of at least 0.20 with another module in each LOTO case were chosen as inputs to the ANN, and the results are also shown in Figure 2.2. Although the testing accuracy was still lower than training, it was encouraging that the accuracies followed the same trend as addtional inputs are added to the classifier. Unfortunately there were very few modules (5) that matched this criterion and the accuracy was still increasing once all 5 modules had been included.

Another approach to reducing overfitting during feature selection, in the context of transcriptional data, is to map gene expression values onto pre-defined biological pathways, rather than generating *ad hoc* modules. An additional benefit to this method is that it makes interpretation of results more straightforward, as each feature is established as relating to a particular biological function. To implement this mapping, we utilized Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005), an algorithm that generates enrichment scores for sets of genes based on their aggregate location in a ranked list (in this case using post-vaccination fold-change), to generate enrichment scores for a set of gene modules previously developed

by Li et al. (Li et al., 2014). Li et al. defined these 'Blood Transcriptional Modules' (BTMs) through co-expression analysis across many publicly available datasets containing gene expression from human blood samples, which contain many circulating immune cells. These modules were then manually annotated to identify the biological processes that the member genes reflected. Utilizing features generated from large amounts of external data (~500 data sets and ~30000 samples) allows us to identify highly robust signatures of response that are not context or dataset specific. The results of our predictive analysis using BTM features are described in Chapter 2.

Classification using artificial neural networks

Once features have been identified which represent as reliably as possible the differences in biological response, the next step is to train a machine learning algorithm using these features as inputs so that it can then make predictions on future samples whose antibody response is unkown. There are many different algorithms that are suitable for this task. Some of the more common algorithms applied to bioinformatics problems are logistic regression, classification trees, k-nearest neighbor algorithms (k-NNs), support vector machines (SVMs), and artificial neural networks (ANNs) (Larranaga et al., 2006). There is no consensus on which algorithm is best to use, even when restricted to microarray data, and the algorithms show varying orders of performance depending on the dataset, feature selection method used, and

author of the study. However, neural networks have advantages over linear methods such as logistic regression because they are nonlinear and are therefore able to capture nonlinear relationships between the features and the output, and they have been shown to outperform other methods such as SVMs and k-NNs at least on some datasets (Cho and Won, 2003). In addition, part of the success of a particular method comes from the author's familiarity with the algorithm and ability to make necessary adjustments. It was with this in mind that neural networks were chosen as the initial algorithm for classification on this dataset. Furthermore, the most important task in development of classifiers for high-dimensional data is to identify features that provide the best separation between classes, because if the samples are well separated in the feature space the classification will be successful independent of which method is used. Similarly, if there is very little or unreliable separation in the feature space all classification algorithms will perform poorly.

A standard two-layer neural network works by passing the input vectors into a set number of first layer neurons, with a weight assigned to each dimension of the input vector for each neuron. A transfer function (typically sigmoidal) is then applied to the result. The outputs of the first layer neurons are then passed into the second layer (output) neuron according to the weights assigned to each first layer neuron, as seen in Figure 2.3. When using a sigmoidal transfer function in the second layer, the output of the network will

range between 0 and 1. A hard limit function can be applied to this output, limiting the output to either 0 or 1. This effectively creates a boundary in the input space, such that samples which fall on opposite sides of the boundary will be classified into two separate groups. It can be shown that with this setup, it is possible to generate any desired classification boundary within the input space, with the complexity of the boundary being determined by the number of first layer neurons. During training, the algorithm uses training samples to determine how to adjust the weights of the network (using gradient descent or similar methods) such that the classification error is minimized. Overfitting is prevented by reserving a random portion of the training data for validation, and stopping training when the validation error begins to increase, a technique known as early stopping.

Because a random portion of the training data must be reserved for validation when using early stopping, repeated runs of the algorithm will use slightly different training sets and therefore have slightly different classification boundaries. In addition, different initial weights of the network (which are randomly generated) can cause the algorithm to reach a different final boundary. To take this variation into account, one can create a classifier composed of a committee of base classifiers (neural networks or other algorithms) using a technique known as ‘bagging’. Each committee member is created by training a single classifier on a random partitioning of the training data into training/validation. The output of the ensemble classifier is then

determined by a majority vote of the members of the committee. By averaging over the variation caused by initial weighting and training/validation partitioning, this method has been shown to increase the accuracy and reduce the variability of classification when using neural networks and other methods (Opitz and Maclin, 1999). In addition to these benefits, ensemble methods provide an intuitive way to estimate the confidence of prediction. If a sample receives unanimous voting by the committee into a particular class, that indicates that the prediction is independent of initial weighting or training data partitioning and that one can be fairly confident that the classification is accurate. However if a sample recieves a highly split vote, such as 60%/40% between the two classes, the sample must lie close to the classification boundary within the feature space such that noise in the system from the aforementioned sources can cause the sample to fall on either side of the boundary, and therefore one should be less confident in the classification of this sample. It is then possible to place a threshold on the votes such that samples which recieve less than the required number are labeled as 'low confidence.' This would be useful if the classifier is implemented as a diagnostic tool, because it would allow physicians to suggest further testing (such as performing later HAI assays to detect actual antibody titer) for patients who receive a low confidence prediction.

Effects of noise and bias on classification

In addition to binary classification (high/low response), we also explored the possibility of developing a continuous classifier able to predict quantitatively the magnitude of antibody fold-changes post-vaccination. Theoretically, the conversion of antibody responses into binary high/low categories results in a loss of information that may decrease classifier performance: while subjects with 4-fold and 512-fold increases in antibody titer will both be categorized as high responders, their phenotype and immune responses are likely quite different. However, classifier performance is also dependent on the reliability of measurements used as inputs and outputs. Unfortunately, HAI titers have been shown to have significant variability (cite), making accurate quantitative prediction difficult. Indeed, an initial comparison of ANN classifiers using probe-level features and forward selection on day 7 ComBat normalized data showed that using continuous outputs did not improve the overall accuracy of the classifier (Table 2.2). Visualization of the output of the classifier using 2 of the frequently selected probes shows the large degree of noise: subjects with similar gene expression show large differences in antibody titer fold-changes (Fig. 2.4).

However, training using continuous outputs did improve the specificity of the classifier, also known as the true negative rate (Table 2.2). In this case this represents the proportion of low responders who were correctly identified as such. As mentioned earlier, the dataset is biased in that there are significantly more high responders than low responders due to the efficacy of

the vaccine (Table 2.1). As a result, a binary classifier optimized for overall accuracy will have low specificity—i.e., it will have a tendency to classify subjects as high responders. Using antibody titer fold-changes as a continuous output removes this bias, and as a result there is a more balanced detection of high responders (sensitivity) and low responders (specificity).

Conclusions

The increase in HDLSS data associated with advances in high-throughput technologies poses significant challenges for development of robust and accurate predictive algorithms. In particular, the task of feature selection is critical to development of a successful classifier, and should be focused on in future research. Our work using pre-defined module-level features as a basis for predictive analysis of vaccine response is presented in Chapter 2.

In addition, measurement noise is an ever-present obstacle to reliable training of classification algorithms. In this context, current standards for antibody measurement, such as the HAI assay, present a limiting factor for predictive analysis. In order to develop clinically relevant diagnostic tools, improvements in the accuracy and reproducibility of these technologies are necessary.

Finally, important consideration should be given to determining the appropriate objective function in diagnostic applications. Although we chose

to evaluate performance based on overall accuracy in order to enable comparison with other previous work, it is apparent that this may not be the optimal marker of performance in a clinical setting. If the relative cost in misidentification of a high responder as a low responder (leading to administration of an additional unnecessary vaccination, for example) is small compared to the opposite case (a low responder misclassified as high, and therefore left untreated and potentially at risk for infection), then it would be reasonable to place more emphasis on specificity. Further analysis should be performed in order to evaluate the effects of optimizing alternate metrics of performance in this context.

Acknowledgements

Chapter 2 is a modified presentation of material that is being prepared for publication as “Systems analysis of immunity to influenza vaccination across multiple years and in diverse human populations” by Nakaya HI, Hagan T, Duraisingham SS, Lee EK, Kwissa M, Rouphael N, Frasca D, Gersten M, Mehta AK, Gaujoux R, Li GM, Gupta S, Ahmed R, Mulligan MJ, Shen-Orr S, Blomberg BB, Subramaniam S, and Pulendran B. The dissertation author was the co-primary investigator and author of this material.

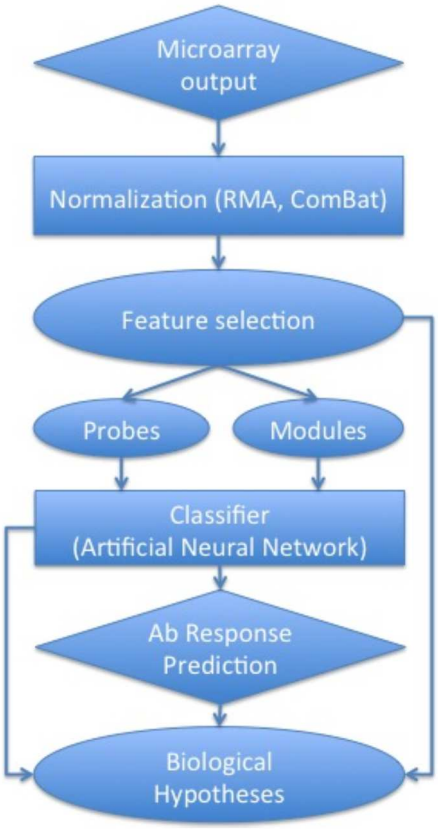

Figure 2.1 Flowchart for predictive analysis of vaccine response
Microarray expression data is first normalized to remove batch effects. Next, a feature selection algorithm such as forward selection is used to identify a small set of probes or gene modules (either pre-defined or generated *ad hoc*) that is capable of optimally distinguishing between classes (high or low responder). This set of probes or modules is then used as inputs into a classification algorithm, in this case an artificial neural network, which predicts as an output the antibody response. Analysis of classifier performance and misclassified subjects, along with the identity of selected probes or modules, can provide insight into the biological processes responsible for successful or impaired responses to vaccination. These discoveries generate hypotheses that can be further evaluated using mechanistic studies.

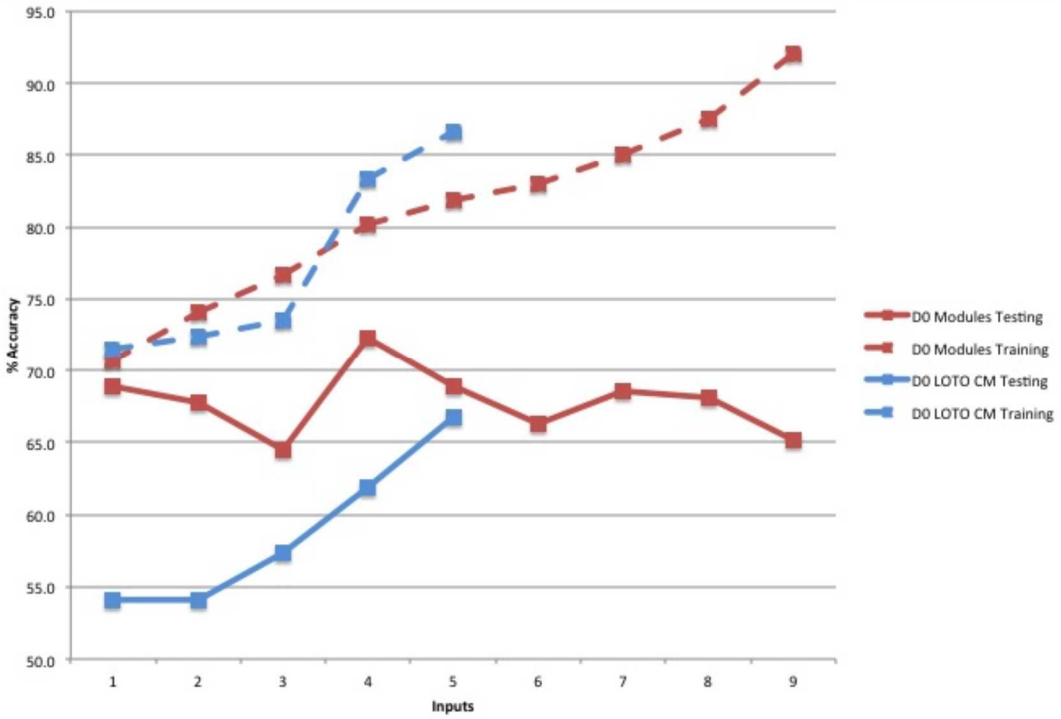

Figure 2.2 Training and testing accuracies of an ANN classifier with module-level features.

Module expression data was generated using day 0 ComBat normalized data as an input to the PinnacleZ algorithm. Significant modules ($p<0.05$) were used as potential inputs to a two-layer ANN with 5 first layer nodes. Inputs were selected using a forward selection algorithm as described in the text. Red lines represent results using all significant modules, blue lines represent results using only modules identified as common (Jaccard index ≥0.2) across all seasons in a leave-one-trial-out approach as described in the text. Dashed lines indicate training accuracies; solid lines indicate testing accuracies.

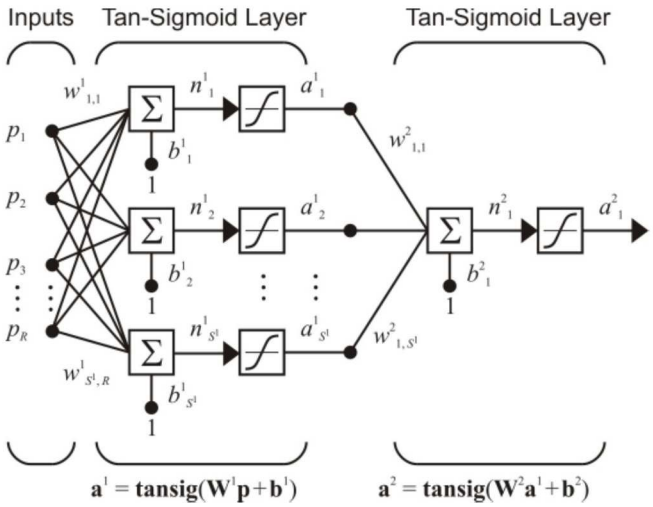

Figure 2.3 Two-layer Neural Network.

In the first layer, each node applies a vector of weights $w_{S^1}^1$ to the input vector p. A bias, b_S^1, is added to the result, creating a linear equation. A sigmoidal transfer function is applied to this equation, creating an output ranging between -1 and 1. This procedure is repeated in the second layer (which only has a single node), using the vector of first layer outputs a^1 as inputs. A hardlimit function (not shown) can be applied to the second layer output to create a binary classification scheme.

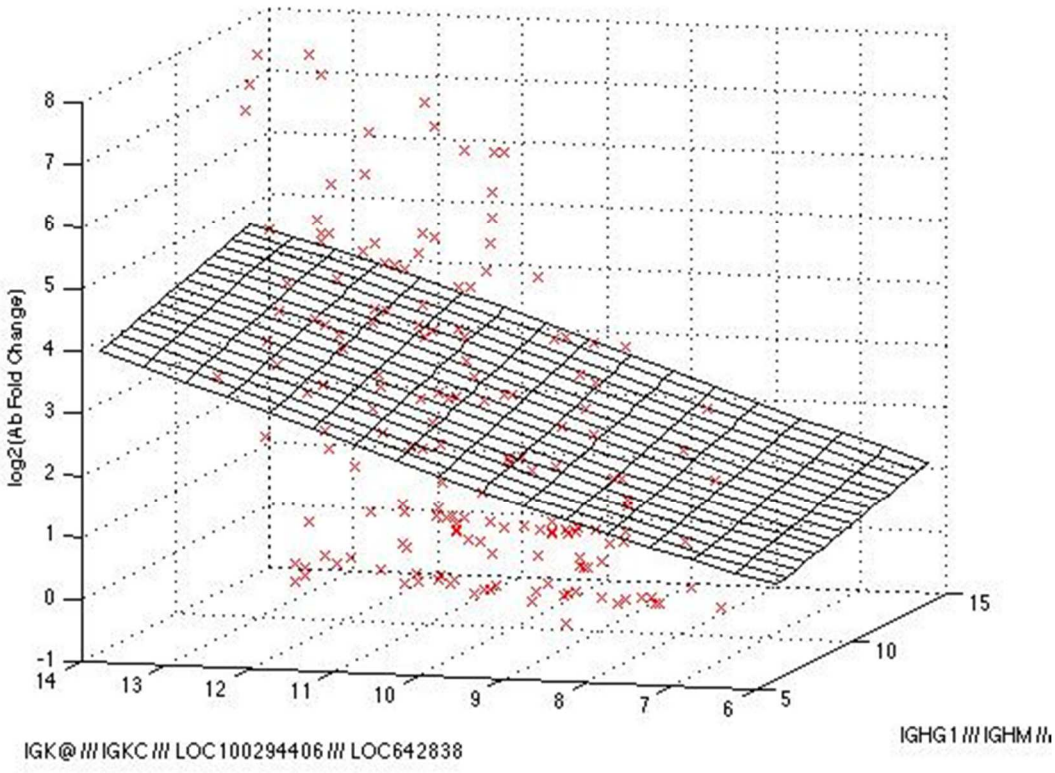

Figure 2.4 Quantitative prediction of antibody HAI titer fold-changes. An ANN with 5 first layer nodes was trained using the day 28/day 0 HAI titer fold change as a continuous output. In order to allow for visualization, 2 probes were used as inputs. Red crosses represent individual training subjects; plane represents output of trained network.

Table 2.1 Subject demographics by trial.

Trial	2007	2008	2009	2010	2011
Total Patients	9	28	28	75	72
High Responders	7	21	16	64	32
Low Responders	2	7	12	11	40
Young	9	28	28	36	57
Elderly	0	0	0	39	15

Table 2.2 Performance of binary and continuous ANN classifiers.

Classifier	Binary	Continuous
Accuracy (Training)	75.1	64.7
Accuracy (Testing)	81.9	74.1
Sensitivity (Training)	0.87	0.62
Sensitivity (Testing)	1	0.75
Specificity (Training)	0.53	0.7
Specificity (Testing)	0.55	0.73

Chapter 3: Systems analysis of immunity to influenza vaccination

Introduction

Seasonal influenza infection kills several hundred thousand people every year, with the majority of deaths occurring among elderly adults aged ≥65 years (Pica and Palese, 2013; Simonsen et al., 2005; Thompson et al., 2003). While vaccination is considered the most effective method for preventing influenza, it shows limited efficacy in the elderly (Sasaki et al., 2011). The vaccine recommended for this age group is the inactivated influenza vaccine that contains virus hemagglutinin (HA) proteins from 3 (trivalent, TIV) or 4 of the circulating influenza H1N1, H3N2 and B strains. The lower efficacy of influenza vaccine in elderly compared to young adults has been associated with the natural age-associated deterioration of immune functions (immunosenescence) (Boraschi et al., 2013; Duraisingham et al., 2013); which include quantitative defects in antibody-secreting cells (ASC) (Sasaki et al., 2011), loss of the less differentiated influenza-specific memory CD8+ T cells (Wagar et al., 2011) and lower frequency of effector memory CD4+ T cells (Kang et al., 2004). However, the molecular mechanisms responsible for the decrease in vaccine efficacy in the elderly remain unexplored (Oviedo-Orta et al., 2013).

Systems vaccinology, an emerging field that applies systems biology approaches and predictive modeling to vaccinology, provides a useful methodology for unraveling the molecular mechanisms of vaccine-induced immunity (Li et al., 2013a; Nakaya et al., 2012; Nakaya and Pulendran, 2012;

Pulendran, 2014; Pulendran et al., 2010). Recently, systems vaccinology has been successfully used to study the immune response to the influenza vaccine in young adults and children (Bucasas et al., 2011; Cao et al., 2014; Franco et al., 2013; Furman et al., 2013; Nakaya et al., 2011; Tsang et al., 2014) as well as to other vaccines such as the yellow fever (Gaucher et al., 2008; Querec et al., 2009) and meningococcus vaccines (Li et al., 2013a). These studies were able to identify innate gene signatures (Nakaya et al., 2011), as well as cell population frequencies and gene modules (Furman et al., 2013), that are predictive of the later antibody response at day 28 to influenza vaccination. In addition, such approaches provided important insights into the pathways driving both the innate and adaptive responses to vaccination (Oh et al., 2014; Ravindran et al., 2014).

However, several fundamental issues in the field remain unaddressed. First, there is still no comprehensive analysis of the similarity of signatures across multiple influenza seasons (Miller and Palese, 2014; Pica and Palese, 2013). This represents a critical issue as the virus strains in the seasonal influenza vaccine may change from year to year, and the impact of these variations on transcriptional signatures is unknown. Second, there is very limited information about the gene regulatory networks and cellular responses that underlie the sub-optimal immunity observed in the elderly population. Third, as all previous studies have focused on signatures that predict the antibody response at 4 weeks, the extent to which transcriptional signatures

associated with the longevity of the antibody response has not been examined. This analysis is important for understanding if and how innate immunity can generate long-lasting antibody responses to influenza vaccination, which is a critical issue, since waning immunity to influenza and other vaccines is a major challenge in vaccinology. Indeed, the mechanisms underlying durability of immune response to any vaccine are poorly understood. Finally, little is known about the role of micro-RNAs (miRNAs) in fine-tuning the transcriptional responses of immune cells after influenza vaccination.

To address these issues, we used systems vaccinology approaches to examine more than 400 young and elderly adults, including diabetics, vaccinated with seasonal TIV during 5 consecutive influenza seasons (2007-2011). With this very rich gene expression and immunological data spanning multiple seasons and studies, a systems-based approach allowed us to go beyond gene-level results and identify robust functional and pathway related changes in the young and elderly in response to influenza vaccination. Our analyses revealed that interferon and other innate immunity signaling networks were consistently activated in responders 1 to 3 days after vaccination, and that they, together with plasma cell and cell cycle gene modules on day 7, were associated with the day 28 antibody response in all 5 seasons. Importantly, molecular signatures involving these pathways were able to successfully predict day 28 antibody responses in both the young and elderly populations. Of note, such a signature was incapable of predicting the

longevity of the antibody response at day 180. Rather, a distinct signature, related to cell movement and adhesion, was associated with the longevity of the HAI titers at 180 days post-vaccination. Interestingly, signatures at baseline corresponding to B cells were highly correlated with vaccine induced HAI titers. Furthermore, molecular signatures associated with NK cells and monocytes were higher in the elderly, an observation later verified through flow cytometry analysis. Finally, the regulation of the response to TIV was examined through miRNA expression profiling at baseline and at days 1, 3, and 7 post-vaccination. Integrative analysis of the miRNA and mRNA data revealed miRNAs that are potential regulators of the interferon response on day 1 post-vaccination. Together these results represent the first systems analysis of several hundred subjects, to identify universal signatures of immunity to vaccination spanning across multiple years in diverse human populations including the young, elderly and diabetic.

Results

Antibody responses to influenza vaccination across 5 consecutive seasons and in diverse populations

We vaccinated with TIV a total of 212 subjects across 5 influenza vaccine seasons from 2007 to 2011, among whom 54 were elderly (65 or older in 2010 and 2011 cohorts) (Fig. 3.1a). The 2011 cohort also included 17 subjects diagnosed with type 2 diabetes. Eligible individuals were 20 years or older and with an established diagnosis of T2D for more than 6 months. In general the health of all was under control with no severe symptoms. Our healthy control population was derived from a Family Medicine cohort that is routinely checked for blood glucose and was not abnormal and thus would not be classified as diabetic. Presence of relevant co-morbidities was recorded and T2D patients did not have cancer, major infectious or autoimmune diseases for at least 12 months prior to enrollment. Pregnancy and documented current substance and/or alcohol abuse were also exclusion criteria, as well as use of medicines known to alter the immune response, such as high-dose corticosteroids, within six months prior to enrollment. As "co-morbidities", a few patients had hypertension, hyperlipidemia, pain (joint, back), or hypothyroidism. Individuals had to be free of influenza, or of any symptoms associated with respiratory infections at the time of enrollment. All T2D patients were under metformin treatment. Overall, there was a 65/35 ratio of

females to males, and the majority of subjects were white (Fig. 3.1b). Blood samples were collected at baseline and on days 1 (2010 cohort only), 3, 7, 14 (2010 cohort only), 28 and 180 days post-vaccination. We performed microarray transcriptome analyses on the peripheral blood mononuclear cells (PBMC) through day 7 post-TIV vaccination (Fig. 3.2a). For the 2008 and 2009 seasons, we included in our analyses published data from a previous study by Franco et al. (Franco et al., 2013) containing 218 additional subjects. For a subset of the 2010 season cohort, fluorescence-activated cell sorting (FACS) measurements as well as miRNA profiling were also performed (Fig. 3.2a).

We evaluated the plasma antibody responses of all vaccinees using hemagglutination-inhibition (HAI) antibody titer assays. The individual antibody response for each of the three influenza strains included in the vaccine was calculated as the fold change between the HAI titer at day 28 and the baseline titer. We then defined the magnitude of antibody responses (HAI response) as the maximum fold change among the three influenza strains (Nakaya et al., 2011) (Fig. 3.2b). Subjects were classified as “high responders” if (i) their HAI response had at least a fourfold increase 28 days after vaccination (Sullivan et al., 2010) and (ii) the day 28 antibody titer was 1:40 or more for at least one strain, and were classified as “low responders” otherwise. In the 2010 and 2011 seasons, in which elderly subjects were included, young adults (<65 years old) achieved higher antibody responses than elderly vaccinees (65 or older) (Fig. 3.1c and Fig. 3.2c), consistent with previous studies (Goodwin et

al., 2006; Seidman et al., 2012). Vaccines from different seasons induced varying levels of HAI response (Fig. 3.1d). Subjects vaccinated during the 2010 season, the year after the H1N1 2009 pandemic, generated the strongest HAI responses. There was no significant difference in antibody response based on gender (Fig. 3.1e), diabetic status (Fig. 3.1f), or race (data not shown).

Gene signatures of antibody responses across multiple seasons in the young and the elderly

In order to identify transcriptional pathways associated with the antibody response to vaccination, we performed Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005) on genes correlated with the HAI response in each season (Fig. 3.3a). For this, we used a set of blood transcriptional modules (BTMs) identified by our group through large-scale network integration of publicly available human blood transcriptomes (Tsang et al., 2014). BTMs related to the induction of interferons as well as activation of dendritic cells (Fig. 3.3a and 3.3b) were enriched on days 1 and 3 post-TIV vaccination, whereas modules related to T cells at these time points were negatively associated with the antibody response. These results were validated in both seasons of the Franco et al. dataset (Franco et al., 2013). On day 7, there was a robust enrichment of ASC and cell cycle-related modules, again consistent with our original study (Nakaya et al., 2011).

Additionally, we sought to compare transcriptional responses based on gender, age, race or diabetic status. We first performed differential expression analysis between these groups (male/female, type 2 diabetes positive/negative, and white/black/asian) using day 3 and day 7 fold change expression. There were very few differentially expressed genes based on gender (Fig. 3.4a) or race (data not shown), indicating similar responses post-vaccination. Comparison of gene expression between type 2 diabetes positive and negative subjects revealed a modest number of differentially expressed genes (Fig. 3.4b), however GSEA on genes ranked by correlation with the day 28 HAI response in these two groups showed a high degree of overlap in the modules associated with a high antibody response (Fig. 3.4c). Based on the similarity of these transcriptional responses, we decided not to segregate subjects based on gender, race, or diabetic status in subsequent analyses.

Based on the robustness of the module-level responses across different TIV seasons, we next wanted to investigate whether or not module-level features were capable of accurately predicting vaccine-induced immunity in a multi-year dataset. Therefore, we first generated BTM normalized enrichment scores for each subject using single sample GSEA (ssGSEA) (Barbie et al., 2009), and then used these scores as inputs to an artificial neural network classifier (Dreiseitl and Ohno-Machado, 2002) (see Experimental Procedures for details). In brief, young subjects from all seasons were randomly divided into training and testing groups in an 85%/15% ratio. Network training and

module selection were performed using the training group, and the trained network was then used to predict responses in the testing group. Using this approach, we identified BTMs that predicted HAI responses in training and testing groups with accuracies ranging between 79.0-80.2% and 64.7- 72.3%, respectively, using both day 3 and day 7 signatures (Fig. 3.3c). Examination of the modules most frequently selected by the algorithm in 100 randomized trials revealed pathways related to innate immune cell responses as well as leukocyte differentiation and antigen presentation on day 3, and B cell/immunoglobulin production and cell cycle pathways on day 7.

Importantly, by separating the elderly subjects into an independent testing set, we observed similar classification accuracy in both the young and elderly test sets when using the predictive signatures trained on the young dataset. This suggests that although the elderly have lower antibody responses to vaccination, the molecular signatures that are predictive of high HAI response remain the same. To further investigate the signatures associated with the antibody response in young and elderly subjects, we performed GSEA on genes ranked by correlation with HAI response (similar to the analysis of Fig. 3.3a) for young and elderly subjects in the 2010 cohort, which revealed significant overlap of enriched modules between the two groups (Fig. 3.5). Together these results demonstrate that module-level features are able to successfully predict vaccine immunogenicity in TIV seasons in both young and elderly vaccinees.

Baseline signatures associated with antibody responses

In addition to identifying robust signatures of the early immune response to vaccination, we also examined whether or not there were consistent transcriptional signatures pre-vaccination that were associated with either a high or low day 28 antibody response. While there has been previous work aimed at identifying baseline predictors of influenza vaccine response (Frasca et al., 2010; Frasca et al., 2012; Furman et al., 2013; Tsang et al., 2014), these studies only examined subjects within a single influenza season, and the robustness of these signatures in predicting immunity to influenza vaccination across multiple seasons has not been determined. As a first assessment, we performed GSEA (Subramanian et al., 2005) in each season using genes ranked by correlation between baseline expression and the HAI response, similar to the analysis presented in Fig. 3.3b. However, this approach resulted in little overlap between seasons, with no module consistently enriched in more than 3 out of 5 seasons (Fig. 3.6). This inability to identify consistent signatures across seasons at baseline may have been due to a reduced signal-to-noise ratio, with weaker transcriptional differences associated with the antibody response occurring pre-vaccination, as well as increased batch effects due to the inability to use fold-change expression at baseline. Although we attempted to adjust for batch effects in the day 0 expression data across seasons by using ComBat software (Johnson et al.,

2007), these effects may have not been completely removed.

To increase the power of this analysis to detect the more subtle transcriptional signatures at baseline, we repeated the procedure using all five seasons combined. To ensure the robustness of the results, publicly available data from previously published influenza vaccine studies by Franco et al. (Franco et al., 2013) and Furman et al. (Furman et al., 2013) were included as independent validation sets (see Experimental Procedures for details). This approach revealed several B cell and T cell-related modules whose expression pre-vaccination was positively correlated with an increased antibody response to vaccination in all studies (Fig. 3.7a). In contrast, modules related to monocytes were negatively correlated with antibody responses in most of the studies (Fig. 3.7b-g). Interestingly, contained within the monocyte related modules were genes such as TLR4, TLR8, NOD2, ASGR2, which encode for proteins involved in innate sensing of microbial stimuli, as well as genes encoding IFN□R, IL-13R, TIMP2, Lyn, Syk and other molecules involved in the inflammatory response. This supports the concept that inflammatory responses at baseline may be detrimental to the induction of vaccine-induced antibody responses (Haq and McElhaney, 2014; Pawelec et al., 2014). Investigation of the correlation with antibody response at a gene level (Fig. 3.7b-g) highlights the strength of pathway-level analyses: while the module enrichment was consistent across datasets, the individual genes contributing to this enrichment varied from study to study.

Signatures induced by vaccination with TIV in young and elderly

We next investigated the influence of age on the vaccine induced transcriptional signatures. To this end, we first identified differentially expressed genes post-vaccination in each age group (Fig. 3.8a). By far the largest difference in overall expression between the two groups occurred on day 1, where the young exhibited a much greater number of both up and down-regulated genes. Weighted correlation network analysis (WGCNA) (Langfelder and Horvath, 2008) was used to find clusters of highly correlated genes among the TIV regulated genes in young (clusters Y1-6) and elderly (clusters E1-5) (Fig. 3.8b). This method compares the correlation in expression patterns among genes across all young or elderly subjects, and utilizes hierarchical clustering with dynamic tree cut to define modules of genes that are temporally co-expressed within each group. Genes in common to clusters Y4 and E3 are associated with antibody-secreting cells (ASCs), whereas the overlap between clusters Y1 and E1 contains several interferon-related genes (Fig. 3.8c). Although the genes shared by clusters from young and elderly vaccinees have similar temporal expression profiles, the magnitude of the expression of interferon-related genes differs, being higher in young (Fig. 3.8c).

We then performed GSEA using genes ranked by their correlation with age to identify modules having enriched expression in either young or elderly

vaccinees (Fig. 3.8d). In order to ensure the robustness of our results, we performed this analysis separately on the two trials containing elderly subjects (2010 and 2011) and identified modules that were consistently enriched in both seasons. Natural killer (NK) cell related modules showed increased expression with age on both day 3 and day 7 post-vaccination, while several monocyte modules showed increased expression with age on day 7. In contrast, many B cell modules showed decreased expression in older subjects on both day 3 and day 7. These results indicate that elderly subjects may be mounting a significant innate response, but that that their adaptive B cell response is diminished. One of the modules showing increased expression in the elderly was BTM M61.0 (Fig. 3.8e), which contains many killer cell immunoglobulin-like receptor (KIR) and lectin-like receptor (KLR) genes. These are inhibitory receptors that suppress the cytotoxic activity of NK cells when bound to MHC I molecules (KIRs) and cadherins and other ligands (KLRs) (Long et al., 2013).

Next, we examined whether or not any of the age-related transcriptional differences were also associated with impaired antibody responses. By plotting the age-based enrichment score of each module against its antibody response-based enrichment score on day 7, we were able to identify five modules whose expression varied with age and were associated with either a high or low antibody response (Fig. 3.8f). Two B cell modules, M47.0 and M69, showed increased expression in younger subjects and in subjects with

high antibody responses, whereas three monocyte and myeloid cell related modules, S4, M4.3, and M11.0, were enriched in the elderly and were also associated with a decreased antibody response in the 2010 and 2011 seasons. Interestingly, two of these modules, M11.0 and M4.3, showed association with an increased antibody response in the 2007 and 2008 seasons (Fig. 3.3a), indicating that this association is not consistent across all 5 seasons on day 7. These results demonstrate that TIV induces distinct but overlapping transcriptional responses in the young versus elderly. The early innate response at day 1, comprised of antiviral and type IFN related genes seem to be impaired in the elderly. In contrast, several transcriptional modules related to NK cells and monocytes appear to have enhanced expression at baseline and after vaccination, in the elderly relative to the young.

Cellular responses induced by vaccination with TIV in the young and elderly

The aforementioned transcriptional signatures of NK cells and monocytes in the elderly subjects could have been caused by de novo transcriptional induction of genes expressed in these cell type, or by changing representations of these cell types within the PBMC compartment. In order to distinguish these two possibilities, we performed FACS analysis on PBMC in a subset of subjects from the 2010 cohort. Our analysis showed that proportions of total NK cells in elderly subjects were higher than those of

young at baseline as well as at all the time points studied (days 0-14). Of note vaccination induced an increase in the percentage of NK cells in the elderly subjects; in contrast, in the young there was a reduction in the percentage of NK cells on day 1 after vaccination (Fig. 3.9a). Similar trends were observed in the absolute numbers of NK cells (data not shown). These results confirm our findings at a molecular level (Fig. 3.8d) and show that the increase in NK cell-related expression in the elderly is due in part to enhanced representation of this cell population post-vaccination.

We also explored the distribution of the NK compartment into 3 subpopulations: CD56++ NK (CD56bright), CD56++CD16+ NK and CD56dim CD16++ NK (Fig. 3.9b). These NK subsets have distinct functions: CD56dim CD16++ cells, the most abundant population of NK cells in the blood, have significantly higher cytotoxic activity, while CD56++ cells are characterized by increased cytokine production and may exhibit immunoregulatory properties under certain conditions (Poli et al., 2009). The CD56++CD16+ subset is considered an intermediate population in transition from the CD56++ to the CD56dim CD16++ subset (Beziat et al., 2011). We observed a higher frequency of CD56dim CD16++ cells coincident with a lower frequency of CD56++ cells in the elderly at baseline, consistent with previous studies (Solana et al., 2014) (Fig. 3.9c). The distribution of these subpopulations was relatively stable during the response, with CD56dimCD16++ cells showing a small but significant increase in the elderly on day 1 post-vaccination and

CD56++ cells increasing in the young on day 7. Additionally, we examined the activation of these subsets through immunostaining for CD69, a marker associated with increased cytotoxicity and IFN-γ production (Clausen et al., 2003; Gorski et al., 2006) (Fig. 3.9d). Elderly subjects showed increasing activation post-vaccination in all 3 subsets, with highest expression on day 3. In contrast, the young only exhibited activation of the CD56dimCD16++ subset as measured by CD69 expression on day 1, and at a lower level than in the elderly.

In addition to NK cells, we also observed higher proportions of monocytes among the elderly at baseline and during the entire duration of the study compared to young subjects (Fig. 3.10a). In both groups, the percentage of total monocytes in PBMC increased substantially on day 1 post-vaccination, then returned to at or below baseline levels. We further examined the distribution of monocytes into 3 subsets: 'classical' (CD14+CD16-), 'intermediate' (CD14+CD16+), and 'non-classical' (CD14dimCD16++) (Fig. S6b). Classical as well as intermediate monocytes express proinflammatory cytokines such as IL-6 and TNF-α but also IL-10 and exhibit high phagocytic activity (Gordon and Taylor, 2005; Grage-Griebenow et al., 2001). In contrast, non-classical monocytes are weak phagocytes generally exhibiting an anti-inflammatory phenotype, but able to initiate proinflammatory responses, via TLR7 and TLR8 receptors, in response to viral infection (Cros et al., 2010). The distribution of monocyte subsets was similar in both groups, with the

elderly having a slight increase in intermediate monocytes compared to young (Fig. 3.10c). The young experienced a small decrease in classical monocytes and corresponding boost in non-classical monocytes on day 3 post-vaccination. This upregulation of CD16 post-vaccination is consistent with other recent studies of monocyte dynamics in response to influenza vaccination (Mohanty et al., 2015). The activation of these subsets was also analyzed by immunostaining for CCR5, a receptor for a number of inflammatory cytokines, and CD86, a costimulatory molecule involved in T cell activation (Fig. 3.10d). In both the young and elderly, CCR5 expression peaked in classical and intermediate monocytes on day 1 post-vaccination, corresponding with the inflammation associated with the innate immune response. Elderly subjects showed increased CCR5 expression in classical and intermediate monocytes compared to the young, whereas young subjects had higher CD86 expression in these subsets. These results are consistent with the transcriptional changes observed (Fig. 3.8), and support concept that persistent and excessive inflammatory responses may be detrimental to the induction of vaccine-induced antibody responses (Frasca et al., 2014; Haq and McElhaney, 2014; Pawelec et al., 2014).

Molecular signatures associated with the persistence of antibody responses

A desirable feature of a good vaccine is the ability to induce long lasting

protection from infection. To investigate how effectively the influenza vaccine generates a long-lived response and identify the mechanisms responsible for this response, we measured antibody titers in a subset of subjects at 180 days (D180) post-vaccination. The antibody responses peaked at day 28 (D28), with a significant decline in most subjects by day 180 (Fig. 3.11a). By applying the same FDA criteria for seroconversion used for the D28 high/low responder classification to the D180 antibody responses, we saw that just over half of D28 high responders maintained their responder status on day 180 ('persistent' responders), whereas the remainder no longer met the criteria for seroconversion on day 180 ('temporary' responders). Because the D180 antibody titer shows significant dependence on the original antibody response generated on day 28 (Fig. 3.11b), examining the molecular profile associated with the D180 response would identify many of the mechanisms responsible for generating the D28 response. In order to study the relative persistence of the response, we calculated residuals from a linear fit between the D28 and D180 responses, thereby removing the effect of the D28 response on the D180 response. We observed that subjects with a higher D28 titer had a greater decrease in HAI response between D28 and D180, indicating increased difficulty to maintain higher levels of antibodies. The linear fit therefore represents the expected D180 response given a particular initial D28 response, and the residual can be considered an 'adjusted' D180 response taking into account the original D28 response. Subjects with positive residuals

have a more persistent response, while a negative residual represents a waning response. As expected, correlation analysis of plasmablast module BTM S3 shows a positive association with the D180 antibody response due to the effect of D28 antibody responses, generated through plasmablast expansion (Fig. 3.11c). However, when we performed correlation between the plasmablast module activity and the D180/ D28 residual, we no longer saw an association with the plasmablast expression (Fig. 3.11d). This result suggests that factors other than the expansion of plasmablasts are responsible for maintaining a persistent antibody response.

To identify the pathways associated with a persistent or waning response, we performed GSEA on genes correlated with the D180/ D28 residual. Among the BTMs most enriched in subjects with a persistent response on both days 3 and 7 were several modules associated with cell movement and adhesion, such as BTM 51 (Fig. 3.11e). In fact, one of the genes in module M51 that was highly correlated with the response persistence was P-selectin (SELP), which plays an important role in the adhesion and extravasation of leukocytes out of the blood circulation and into organs or tissues (Wang et al., 2007). In addition, the day 7 expression of a number of T cell related modules was negatively associated with the D180/ D28 residual, indicating that increased T cell responses may potentially result in more transient antibody responses (Fig. 3.11f). Importantly, the plasmablast signature at day 7, which was predictive of the HAI titers at day 28, did not

correlate with the D180/ D28 residual (Fig. 3.11d).

Post-transcriptional gene regulation of the immune response to vaccination

MicroRNAs (miRNAs) have been identified as key regulators of gene expression at a post-transcriptional level (Filipowicz et al., 2008). While there is some knowledge about miRNA regulation of immune pathways (Lodish et al., 2008), their role in responses to vaccination remains unaddressed. To assess whether or not miRNA may contribute to the differences in response we saw between young and elderly vaccinees, we measured the miRNA expression profiles of 672 human miRNA on days 1, 3 and 7 post-vaccination from a subset of subjects in the 2010 cohort (Fig. 3.12a). We saw a significant difference in the miRNA profiles between these two groups. While a majority of the differentially expressed miRNA in the elderly were upregulated post-vaccination, particularly on days 1 and 7, the young exhibited predominantly downregulation of miRNA. We also identified miRNA whose expression correlated with the day 28 antibody response (Fig. 3.12b). We again saw differences between the young and elderly subjects, with the elderly having a larger number of miRNA negatively correlating with the antibody response when compared with the young. These results suggest that miRNA may indeed be important regulators of the immune response to influenza vaccination. As miRNA expression reduces the translation of its target genes

through mRNA degradation or silencing via the RNA-induced silencing complex, the increase in miRNA expression in the elderly post-vaccination suggests a possible mechanism for the impaired responses in this population that merits further exploration.

To identify the miRNA that regulate particular immune pathways during vaccine response, we implemented a strategy to integrate the mRNA and miRNA omics-level measurements (Fig. 3.12c). Because miRNA often have multiple related target genes, we first generated normalized enrichment scores for 252 blood transcription modules (BTMs) on a per subject basis by performing ssGSEA on the transcriptomic data. We then performed correlation analysis between all miRNA-BTM pairs, revealing groups of modules that appear to be regulated together by several miRNA (Fig 3.12d) on day 1 post-vaccination.

Finally, we used the TargetScan miRNA database (Garcia et al., 2011) to identify the genes in a given module predicted to be targets of negatively-correlating miRNAs. Of particular interest was the regulation of module M75-Antiviral Interferon Signature (Fig. 3.12e) on day 1 post-vaccination, as the interferon pathway plays a key role during the innate immune response. This miRNA-mRNA network suggests potential novel regulators of the interferon response after vaccination, such as miR-424. This miRNA was predicted to target OAS3, a 2'-5'-oligoadenylate synthase involved in viral RNA degradation (Samuel, 2001), as well as CXCL10, an important chemokine

induced by IFN-γ that serves as a chemoattractant for lymphocytes (Dufour et al., 2002).

Discussion

The use of systems-wide measurements combined with network and predictive modeling to study vaccine response, dubbed ‘systems vaccinology’, has advanced the field of vaccinology in recent years. Systems biology approaches have been used in several previous studies of the yellow fever (Gaucher et al., 2008; Querec et al., 2009) and the influenza (Franco et al., 2013; Furman et al., 2013; Nakaya et al., 2011; Tsang et al., 2014) vaccines to provide critical insights about the molecular mechanisms of action of these vaccines, and to successfully identify novel molecular signatures that are predictive of responsiveness to vaccination. However, there exist several poorly understood features of TIV and other vaccines; in particular the pathways and mechanisms responsible for the sub-optimal responses in the elderly population and for waning antibody responses post-vaccination remain unknown. Furthermore, the robustness of signatures identified in a given study and their ability to accurately predict immunity to vaccination in other studies performed in different seasons, and with diverse populations, remains unknown. We thus performed a series of clinical trials in a group of subjects diverse with respect to age, gender, race, and diabetic status across five consecutive influenza vaccine seasons, using genome-wide gene and miRNA

expression profiling and immunological measurements to address these unanswered questions. Interestingly, we found that although antibody levels and gene expression differed with age, other variables such as gender, race, and diabetic status did not have a noticeable effect on these responses. It is conceivable that examination of larger numbers of subjects may reveal putative differences in the immune response with regards to these variables.

In an unprecedented analysis of the transcriptional responses to TIV, we were able to identify robust signatures consistently enriched across five influenza vaccine seasons (Fig. 3.3a). The pathways positively associated with a successful antibody response on day 1 post-vaccination revealed a strong innate response marked by expression of interferons and activation of dendritic cells. This response was also observed on day 3, with enrichment of TLR signaling and antigen presentation as the viral particles are processed by dendritic cells and presented to the adaptive immune system. By day 7, the adaptive immune response was striking, with a strong signature from the expansion of plasmablasts present in all 5 seasons. These consistent signatures were only visible at a pathway level; when the same analysis was performed using differentially expressed genes the overlap across all seasons was low (data not shown). Using the same modules in a single sample approach as inputs to an artificial neural network classifier allowed us to successfully predict antibody responses to vaccination across all seasons included in our study. Using pathway-level features as predictors of immune

response is a promising approach to not only reduce the variability in gene expression across influenza seasons but also to provide improved biological context to the predictive signatures.

By extending this approach to examine the pre-vaccination time point, we were able to establish baseline signatures related to B and T cell expression that were associated with an increased antibody response at day 28 post-vaccination (Fig. 3.7). These signatures were validated in two independent datasets, confirming that they are not strain or study-specific artifacts. In addition, they agree in large part with previous findings by Tsang et al., who identified several B cell and T cell subsets whose frequency at baseline was predictive of day 28 antibody responses (Tsang et al., 2014). Therefore, the increased statistical power of our dataset allowed us to identify relatively weak baseline transcriptional differences that were only detectable using flow cytometry in previous studies. However, at this level it is not clear whether these signatures are due to increased naïve or memory subsets of these cells, which requires further investigation.

Our analysis of responses in young and elderly subjects revealed important aspects about the relationship between immunosenescence and vaccine response. We were able to show for the first time that although the elderly exhibit diminished B cell expression post-vaccination, they also mount a strong innate response, with increased NK cell and monocyte numbers compared to young vaccinees. Furthermore, the elderly displayed increased

frequency and activation of cytotoxic NK cells, as well as increased activation of proinflammatory monocytes via CCR5. This shift was associated with a decrease in the costimulatory molecule CD86, in the monocytes of elderly subjects. In addition, previous studies have shown that although the NK cell population increases with age, there are concurrent changes in the receptors of these cells, including decreases in the activating receptors NKp30 and NKp46 (Almeida-Oliveira et al., 2011). NKp30 is involved in crosstalk with dendritic cells (Walzer et al., 2005), and NKp46 has been shown to bind to influenza hemagglutinin to allow NK cell-mediated recognition of influenza-virus-infected cells (Mandelboim et al., 2001). Examination of the pathways associated with both age and HAI titer fold changes (Fig. 3.9f) found no effect of increased NK cell response on reduced antibody response in the elderly, but did identify three monocyte modules which may relate increased monocytes in the elderly to their diminished antibody response. Most importantly, monocytes were also increased in the elderly pre-vaccination (Fig. 3.10a), and our baseline analysis (Fig. 3.7) revealed a negative association between day 0 monocyte expression and magnitude of antibody response. This represents novel evidence of a potential connection between the baseline state of the immune system in the elderly and reduced responsiveness to vaccination. Possible pathways involved include changes in interferon activation, inflammatory response and antigen presentation, which merit further investigation. Together these results suggest potential mechanisms by

which changes to the innate response in the elderly may result in a diminished ability to prime the adaptive immune system, leading to impaired antibody responses.

In addition to examining the way in which the aging process affects vaccine response, we also investigated the mechanisms responsible for another challenging and poorly understood issue in the field of vaccinology, that of waning antibody responses. While the goal of all vaccines is to induce lasting protection from infection, most vaccines produce responses that diminish over time to varying degrees and require a booster dose after first immunization (Pichichero, 2009). Because the circulating strains are continually changing, the influenza vaccine must be administered every year for optimal protection. Even so, we observed a significant drop in antibody titers in a majority of subjects within 6 months of vaccination. By comparing the relative persistence of the day 180 antibody responses with BTM normalized enrichment scores (Fig. 3.11, we saw that the expansion of plasmablasts, which plays an important role in the development of the day 28 antibody response, had little effect on response longevity. Instead, we were able to show a potential role for cell movement and adhesion, in maintaining a persistent antibody response. This transcriptional signature may be an indicator of successful migration of plasmablasts into the bone marrow, a crucial step in the generation of long-lived plasma cells (Radbruch et al., 2006). This analysis demonstrates the role that initial immune processes, as early as

3 days post-vaccination, can have in shaping the antibody response as much as 6 months later.

Finally, we sought to explore how the previously examined transcriptional responses to vaccination might be regulated, in particular through the expression of miRNAs (Fig. 3.12). miRNAs have been shown to play an important role in modulating the signaling pathways of the immune system (Lodish et al., 2008), but their role in vaccine response has not been studied. By integrating transcriptomic and miRNA expression data, we were able to identify several potential miRNA regulators of the interferon response post-vaccination, such as miR-424. Previous studies have shown that absence of regulation of interferon signaling by miRNA can result in impaired CD8 T cell responses (Gracias et al., 2013). These results demonstrate how the balance between positive and negative signals controlling the innate response is necessary for a successful adaptive response and underscore the importance of understanding the way in which miRNA help achieve this balance. This additional layer of regulation adds to the complexity of the molecular mechanisms that mediate vaccine-induced immunity.

In summary, this longitudinal study across 5 influenza seasons provided the opportunity to identify elements consistently present in immune response to TIV vaccination. Our pathway-level systems biology analyses of this comprehensive dataset revealed previously unknown mechanisms that contribute to impaired responses to influenza vaccination in the elderly, and

allowed us to go beyond transcriptional signatures and identify potential miRNA regulators of the response to vaccination. However, our understanding of the biology of vaccine response remains incomplete. These systems approaches can be expanded to help identify and compare the molecular responses to vaccination in other immunocompromised populations, such as infants and HIV-positive individuals. They may also be helpful in understanding the pathways by which certain adjuvants improve vaccine response. By integrating this knowledge, we will be able to provide a more complete picture of how the immune system responds to vaccination and help guide the development of the next generation of vaccines that provide long-lasting immunogenicity and better protection of at-risk populations.

Experimental procedures

Human subjects and influenza vaccines

This study is comprised of subjects vaccinated with TIV during the 5 consecutive influenza seasons in, 2007-08 (9 subjects), 2008-09 (28 subjects), 2009-10 (28 subjects), 2010-11 (58 subjects), and 2011-12 (72 subjects). The age of vaccinees ranged from 20 to 86, with 39 and 15 subjects in the 2010-11 and 2011-12 seasons >65 years old, respectively. Subjects were vaccinated with one dose of TIV (Fluarix®, GlaxoSmithKline Biologicals, in 2007-08, 2008-09, and 2011-12 seasons, Fluvirin®, Novartis Vaccines and Diagnostics

Limited, in 2009-10 season, and Fluzone®, Sanofi Pasteur Inc., in 2010-11 season) following current guidelines for influenza vaccination. Written informed consent was obtained from each subject with institutional review and approval from the Emory University Institutional Review Board.

Cells, plasma and RNA isolation

Four blood samples were obtained from influenza vaccinees: at baseline, and at days 3, 7 and 28 (range, day 28 to 32) post-vaccination. Additional samples were collected at days 1, 14, and 180 for selected subjects. Peripheral blood mononuclear cells (PBMCs) and plasma were isolated from fresh blood (CPTs; Vacutainer® with Sodium Citrate; BD), following the manufacturer's protocol. PBMCs were frozen in DMSO with 10% FBS and stored at –80°C and then transferred on the next day to liquid nitrogen freezers (–210°C). Plasma samples from CPTs were stored at –80°C. Trizol (Invitrogen) was used to lyse fresh PBMCs (1 ml of Trizol to ~1.5 x 10^6 cells) and to protect RNA from degradation. Trizol samples were stored at –80°C.

Hemagglutination inhibition assays

Hemagglutination inhibition (HAI) titers were determined based on the standard WHO protocol, as previously described (Chen et al., 2010). Briefly, plasma samples were treated with receptor destroying enzyme (RDE; Denka Seiken Co.) by adding 1 part plasma to 3 parts RDE and incubating at 37 °C

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

Microarray experiments

Total RNA from fresh PBMCs (~1.5 x 10^6 cells) was purified using Trizol® (Invitrogen, Life Technologies Corporation) according to the manufacturer's instructions. All RNA samples were checked for purity using a ND-1000 spectrophotometer (NanoDrop Technologies) and for integrity by electrophoresis on a 2100 BioAnalyzer (Agilent Technologies). Two-round *in vitro* transcription amplification and labeling was performed starting with 50 ng intact, total RNA per sample, following the Affymetrix protocol. After hybridization on Human U133 Plus 2.0 Arrays (using GeneTitan platform, Affymetrix, or individual cartridges) for 16 h at 45 °C and 60 r.p.m. in a Hybridization Oven 640 (Affymetrix), slides were washed and stained with a

Fluidics Station 450 (Affymetrix). Scanning was performed on a seventh-generation GeneChip Scanner 3000 (Affymetrix), and Affymetrix GCOS software was used to perform image analysis and generate raw intensity data. Initial data quality was assessed by background level, 3' labeling bias, and pairwise correlation among samples. CEL files from outlier samples were excluded and the remaining CEL files of all the samples belonging to the same influenza season trial were grouped, and the microarray intensity data of probe sets (hereafter referred to "genes") were normalized by RMA, which includes global background adjustment and quantile normalization.

miRNA array experiments

Fresh PBMCs (~1.5 x 10^6 cells) were lysed using miRNA Lysis Buffer (2ml per subject, HTG Molecular Diagnostics). HTG Molecular Diagnostics core facility extracted and processed the total RNA, and performed the miRNA expression profiling according to the manufacturer's instructions. Data was normalized using 40% trimmed mean and log2 transformed.

Blood Transcriptional Module Enrichment

Probes were first ranked by Pearson correlation with either the day 28 antibody fold change, day 180/ day 28 residual, or subject age using fold change microarray data from days 0, 1, 3, and 7 post-vaccination. Separate rankings were generated for each flu season for the day 28 antibody and age

correlations. Probes were then collapsed to Entrez gene IDs, selecting the probe with the smallest correlation p value in cases where multiple probes mapped to a single Entrez ID. Gene set enrichment analysis software (GSEA) was then run in pre-ranked list mode with 1000 permutations to generate enrichment scores for 346 blood transcriptional modules (BTMs) based on the distribution of member genes of each module in the ranked list. Enrichment scores were visualized using the corrplot package in R (Wei, 2012). Individual modules and associated gene coexpression relationships were represented using Cytoscape software (Shannon et al., 2003).

Baseline Transcriptional Analysis

For analysis of baseline signatures of vaccine response (Fig. 3), ComBat software (Johnson et al., 2007) was applied to the RMA normalized baseline expression values to remove batch effects from each flu season within the study. Probes were then ranked by Pearson correlation between their ComBat normalized day 0 expression (across all 5 seasons combined) and the day 28 antibody fold change. This ranking was used with GSEA as described above to generate a list of enriched BTMs. As validation, this procedure was repeated separately in two independent datasets, from Franco et al. (Franco et al., 2013) and Furman et al. (Furman et al., 2013). The Franco et al. study included two separate cohorts that were analyzed separately. BTMs consistently enriched in 3 out of 4 of the datasets were

identified and represented in Fig. 3.

Flow Cytometry

PBMCs were thawed and stained with the live/dead marker Alexa Fluor® 430 (Lifetechnologies) to exclude dead cells. 2-3 x 10^6 cells were stained with an appropriate antibody cocktail. Cells were washed in PBS with 5% fetal bovine serum and fixed with the Cytofix buffer (BD) and analyzed on the LSRII flow cytometer (BD). All flow cytometry analysis was done using the FlowJo (Treestar). Blood NK cells were defined within the live singlets gate as $CD3^-CD4^-CD19^-CD14^-$ cells as the $CD56^{++}$ NK, $CD56^{++}CD16^+$ NK and CD56dim $CD16^{++}$ NK. Monocytes were defined within the live singlet $CD3^-$ $CD19^-$ PBMC as the $CD14^+CD16^-$, $CD14^+CD16^+$ and $CD14^{dim}CD16^{++}$ monocytes. Geometric mean fluorescence intensity (MFI) was reported for expression of CD69 on NK cell subsets and CCR5 and CD86 on monocytes.

Antibodies used for staining of human PBMCs: BD: CCR5 (2D7/CCR5), CD14 (MQp9 and M5E2), CD16 (3G8), CD56 (NCAM16.2), CD3 (SK7), CD19 (SJ25C1), CD69 (FN50). Biolegend: CD16 (3G8), CD56 (HCD56), CD86 (IT2.2). Molecular Probes: CD14 (TUK4).

Statistical analyses were performed with GraphPad Prism to identify significant change relevant to the baseline (day 0), mean ± SEM, t test. Areas Under Curve (AUC) were calculated to compare magnitudes of the cell subsets between cohorts throughout the study duration (days 0, 1, 3, 7 and

14) and compared by t-test.

Module-based responder classification via artificial neural network

In order to use an ANN to predict a high/low antibody response based on module-level features, blood transcription module enrichment scores were generated for each subject using single sample Gene Set Enrichment Analysis (ssGSEA) based on a ranking of the genes by fold change on days 3 and 7 post-vaccination. In order to reduce the computational complexity and reduce overfitting during feature selection, a pre-filter was applied by ranking the modules by independent two-sample t-test and selecting the top 25 modules with the smallest t-test p values for inclusion as possible features. A forward selection approach was then implemented for feature selection, in which the network is trained using each module separately as the sole input. The module that achieves the highest average accuracy across 10 trials is included in the final input set, and then modules are iteratively added to this set. Each iteration, all remaining modules are tested one at a time as an addition to the input set, and only the module that improves the average accuracy the most is kept. This process is stopped when the accuracy fails to increase in a given iteration.

Once the final input set was selected, the final classifier was generated using a 'bagging' approach (Breiman, 1996), in which a committee of networks is created by training single networks using randomized partitioning of the

training data. The output of the ensemble classifier is then determined by a majority vote of the output of the member networks of the committee. By averaging over the variation caused by initial weighting and training/validation partitioning, this method has been shown to increase the accuracy and reduce the variability of classification when using neural networks and other methods (Bauer and Kohavi, 1999). Based on preliminary analysis of the classifier performance with varying network and committee sizes, a committee of 20 two-layer neural networks with 5 first-layer neurons each was used in the final approach. All coding was done in MATLAB using the Neural Network Toolbox.

In order to evaluate the performance of the algorithm, a bootstrapping approach was implemented. The subjects across all seasons were randomly divided into one training and one testing group in an 85%/15% ratio, respectively. Elderly subjects (≥65 years old) were separated and included as an independent test set. To prevent overfitting, feature selection and training of the network weights were performed using only the training subjects, and then the performance of the trained network in blind prediction of the testing subjects was assessed. This process was repeated for 100 trials in order to ensure that the performance was robust across many random divisions of the data.

Differential expression analysis and WGCNA

For significance assessment of each season and time point, expression

differences between baseline and the day post-vaccination were assessed by pairwise Student's t-test (p-value < 0.01 and $\log_2$ fold-change > 0.2). Weighted correlation network analysis (WGCNA) was used to find clusters of highly correlated genes among the TIV regulated genes in young (power = 9) and elderly (power = 4).

Identification of signatures of antibody persistence

Day 180 HAI titers were measured for subjects in the 2007-2010 cohorts, using the same strain that achieved the maximum fold change on day 28 on a per subject basis. As the day 180 samples were collected and analyzed separately from the day 28 samples, the baseline HAI titer was also remeasured in order to provide accurate normalization during fold change calculation. Subjects whose baseline titer measurements differed by more than 2-fold were omitted from further analysis. Next, a simple linear regression was performed using the day 180 titer fold change as the dependent variable and the day 28 titer fold change as the independent variable. The distance to this regression line (residual) was computed for each subject and used as a measure of relative antibody response persistence, with a greater value representing a higher day 180 response relative to the original day 28 response. Blood transcriptional module enrichment analysis was then performed using genes ranked by correlation with the day 180/day 28 residual as described above.

Integrated mRNA-miRNA expression analysis

Differentially expressed miRNA were identified using pairwise Student's t-test (p-value <0.05) between baseline and post-vaccination expression values (days 1, 3, and 7). Correlated gene module-miRNA pairs were identified by performing Pearson correlation between normalized miRNA expression and single sample blood transcription module enrichment scores. Next, for modules of interest, each significantly negatively correlating ($p<0.05$) miRNA was mapped using Targetscan database to identify any predicted target genes (context score < 0) within the module. Pearson correlation was then used to validate the interactions of each miRNA with its predicted target genes within our dataset.

Acknowledgements

Chapter 3 is a modified presentation of material that is being prepared for publication as "Systems analysis of immunity to influenza vaccination across multiple years and in diverse human populations" by Nakaya HI, Hagan T, Duraisingham SS, Lee EK, Kwissa M, Rouphael N, Frasca D, Gersten M, Mehta AK, Gaujoux R, Li GM, Gupta S, Ahmed R, Mulligan MJ, Shen-Orr S, Blomberg BB, Subramaniam S, and Pulendran B. The dissertation author was the co-primary investigator and author of this material.

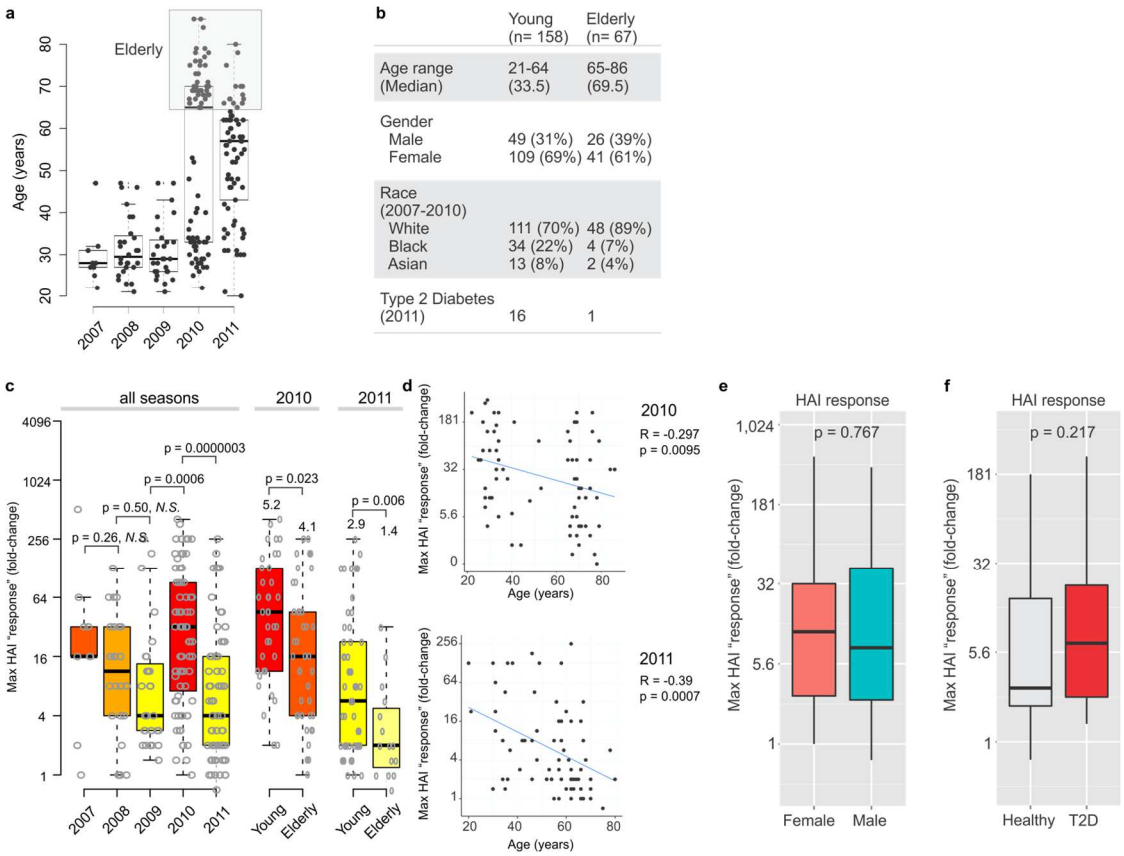

Figure 3.1. Study Demographics and HAI responses by phenotype.
(a) Age distribution of subjects in this study. (b) Demographics of young and elderly subjects in this study. (c) Antibody responses induced by TIV vaccination. The circos plot shows the $\log_2$ fold-induction of HAI titers on day 28 post-vaccination for each one of the three Influenza virus strains (heat maps) or for the highest fold-induction among all 3 strains (i.e. the HAI response; histogram). High antibody responses are shown as red bars in the histogram, while low antibody responses are blue. The y-axis of the histogram ranges from 0-9 in log2, which represents HAI fold changes from 1-512. Elderly antibody responses (in 2010 and 2011) are marked as pink in the inner circle of the plot. (c) HAI responses of young and elderly subjects. The box plots represent the median and first and third quartiles of HAI response in young (Y) (<65 years) and elderly (E) (≥65 years). *p < 0.03 (for t-test; two-tailed test). Data are from 58 subjects (2010) and 72 subjects (2011). (d) HAI responses by season. The maximum HAI response is shown for all 212 subjects separated by flu season, along with box plots indicating the first and third quartiles and median. For 2010 and 2011, subjects are also separated into young (<65 years old) and elderly (65 or older). P values represent results of independent two-sample t-test between responses of young and elderly. (e) HAI responses by gender (all seasons). (f) HAI responses by type 2 diabetes status (2011 season).

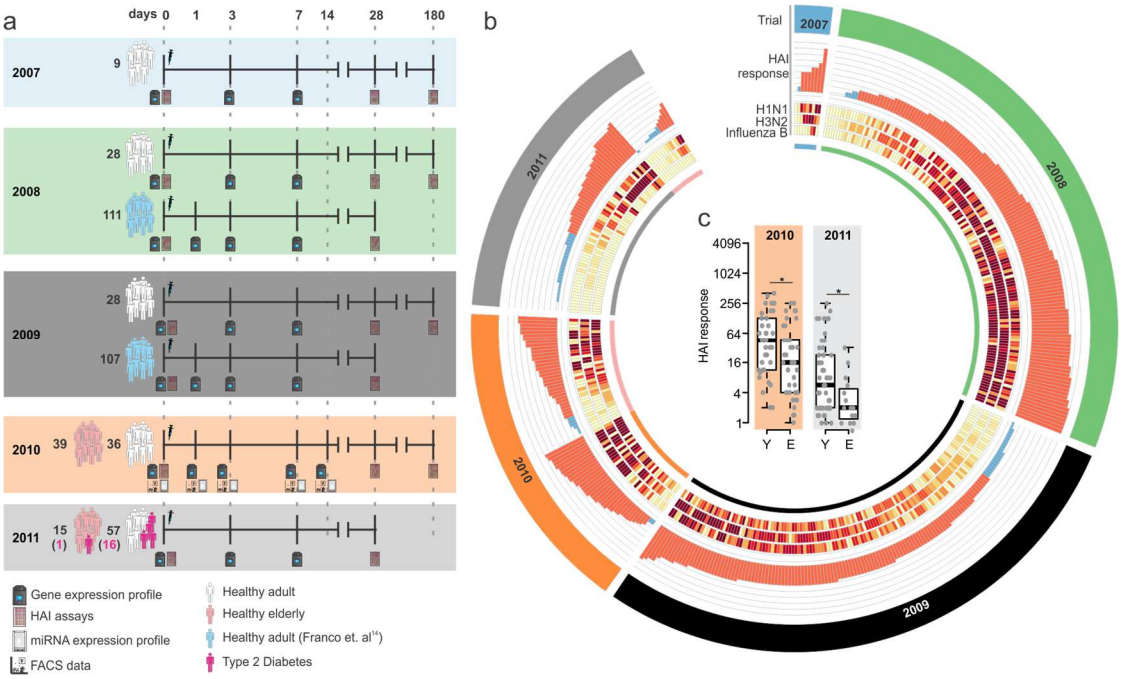

Figure 3.2. Experimental approach and humoral immunity to influenza vaccination in young and elderly subjects.

(a) Experimental approach used to study 5 consecutive influenza vaccination seasons. Microarray experiments and HAI assays were used to obtain, respectively, the gene expression profiles and antibody responses of 413 TIV vaccinees. Flow cytometry and miRNA expression data were obtained for vaccinees from 2010 season. For 2008 and 2009 seasons, publicly available data generated by Franco et al. (Franco et al., 2013) were included. (b) Antibody responses induced by TIV vaccination. The circos plot shows the $\log_2$ fold-induction of HAI titers on day 28 post-vaccination for each one of the three Influenza virus strains (heat maps) or for the highest fold-induction among all 3 strains (i.e. the HAI response; histogram). High antibody responses are shown as red bars in the histogram, while low antibody responses are blue. The y-axis of the histogram ranges from 0-9 in log2, which represents HAI fold changes from 1-512. Elderly antibody responses (in 2010 and 2011) are marked as pink in the inner circle of the plot. (c) HAI responses of young and elderly subjects. The box plots represent the median and first and third quartiles of HAI response in young (Y) (<65 years) and elderly (E) (≥65 years). *$p < 0.03$ (for t-test; two-tailed test). Data are from 58 subjects (2010) and 72 subjects (2011).

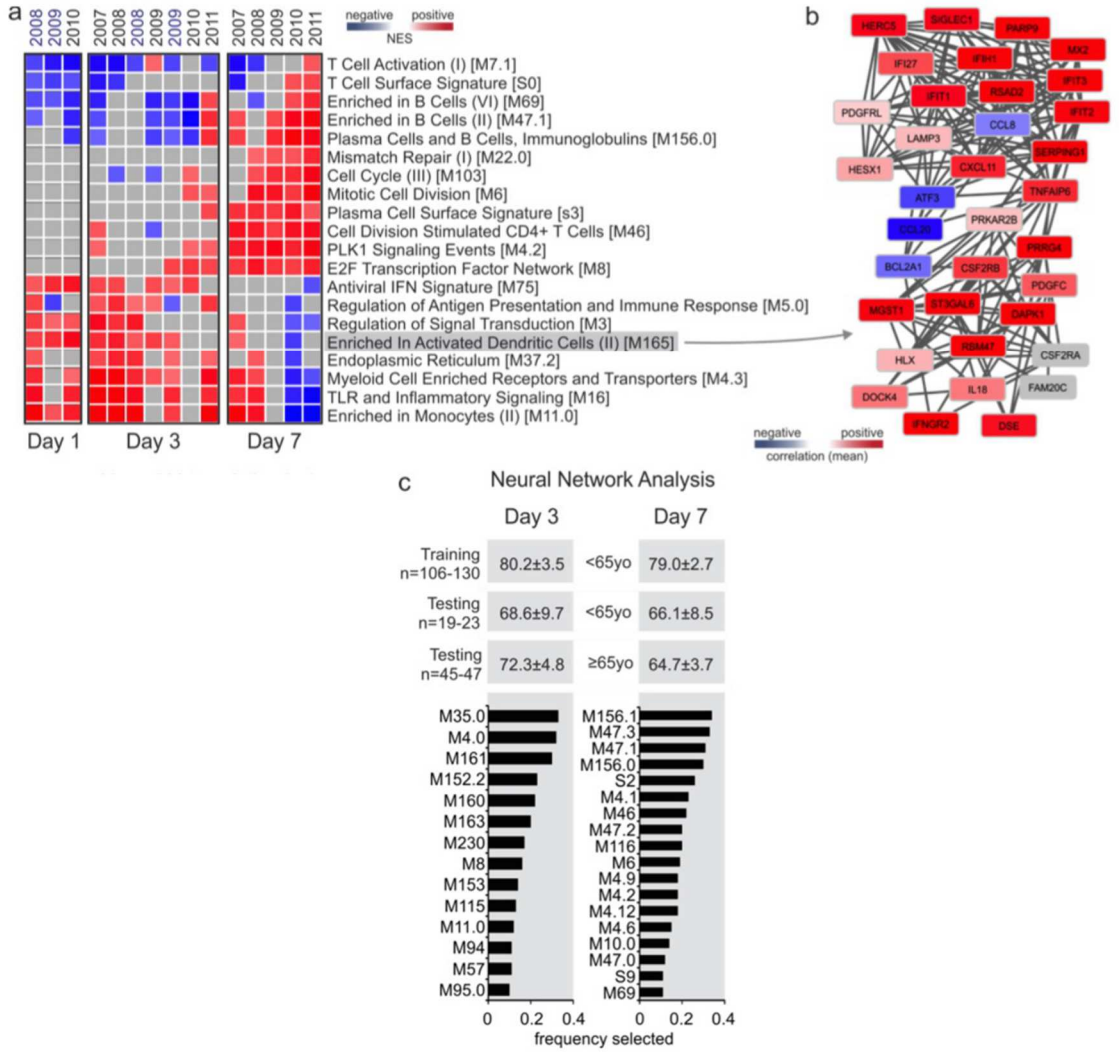

Figure 3.3. Signatures associated with the antibody response induced by TIV.

(a) Heat map of BTMs whose activity at days 1, 3 or 7 post-vaccination is associated with HAI response at day 28 after vaccination. GSEA was used to identify positive (red), negative (blue) or no (grey) enrichment of BTMs (gene sets) within pre-ranked gene lists, where genes were ranked according to their correlation between expression and HAI response. Seasons labeled in blue are from Franco et al. dataset. (b) Genes in BTM M165; each edge represents a coexpression relationship; colors represent the correlation between day 3 fold change gene expression and HAI response. (c) Identification of BTMs that predict antibody responses using ANNs. ssGSEA enrichment scores were generated for each BTM and used as inputs to an ANN classifier to predict the HAI response (see Experimental Procedures for details). Values represent results of 100 randomized trials.

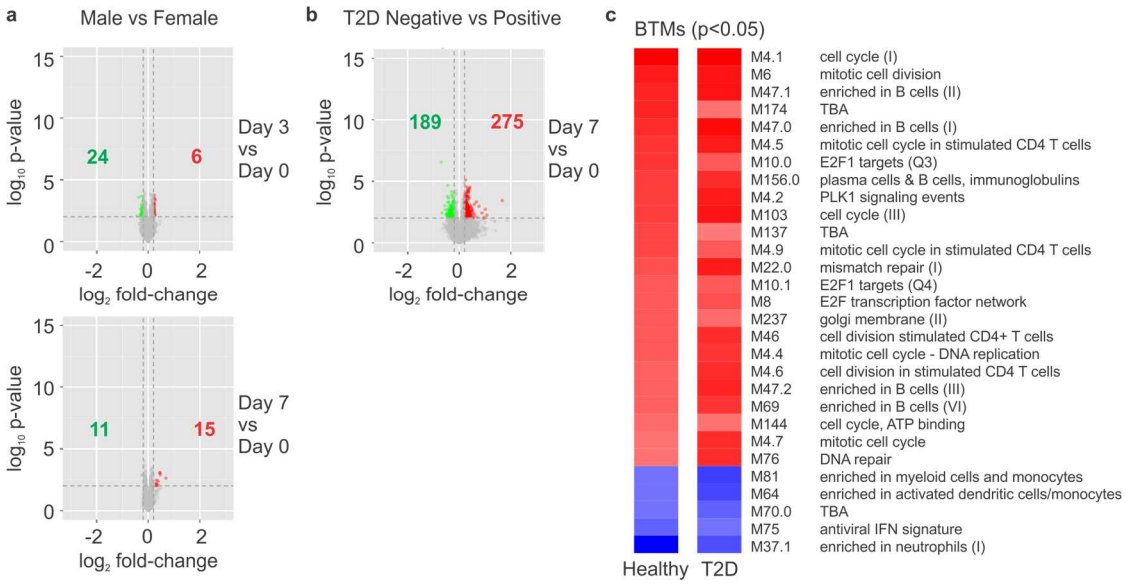

Figure 3.4. Comparison of transcriptional responses by phenotype.
(a) Volcano plot of differentially expressed genes ($p<0.01$ and $\log_2$ fold change>0.2) between male and female subjects on days 3 and 7 post-vaccination (using fold change expression compared to baseline). Green represents genes with increased expression in males, and red represents genes with increased expression in females. (b) Volcano plot of differentially expressed genes ($p<0.01$ and $\log_2$ fold change>0.2) between type 2 diabetes postitve and negative subjects on day 7 post-vaccination (2011 season, day 3 was not analyzed due to the limited number of type 2 diabetes positive subjects with available day 3 expression data (n=4)). Green represents genes with increased expression in type 2 diabetes negative subjects, and red represents genes with increased expression in type 2 diabetes positive subjects. (c) Heat map of Blood Transcription Modules (BTMs, rows) whose activity at day 7 post-vaccination is associated with HAI response at day 28 after vaccination in type 2 diabetes negative (left column) or positive (right column) subjects (2011 season). Gene Set Enrichment Analysis (GSEA, nominal $p<0.05$; 1,000 permutations) was used to identify positive (red) or negative (blue) enrichment of BTMs (gene sets) within pre-ranked gene lists, where genes were ranked according to their correlation between expression and HAI response. Intensity of color represents the normalized enrichment score (NES).

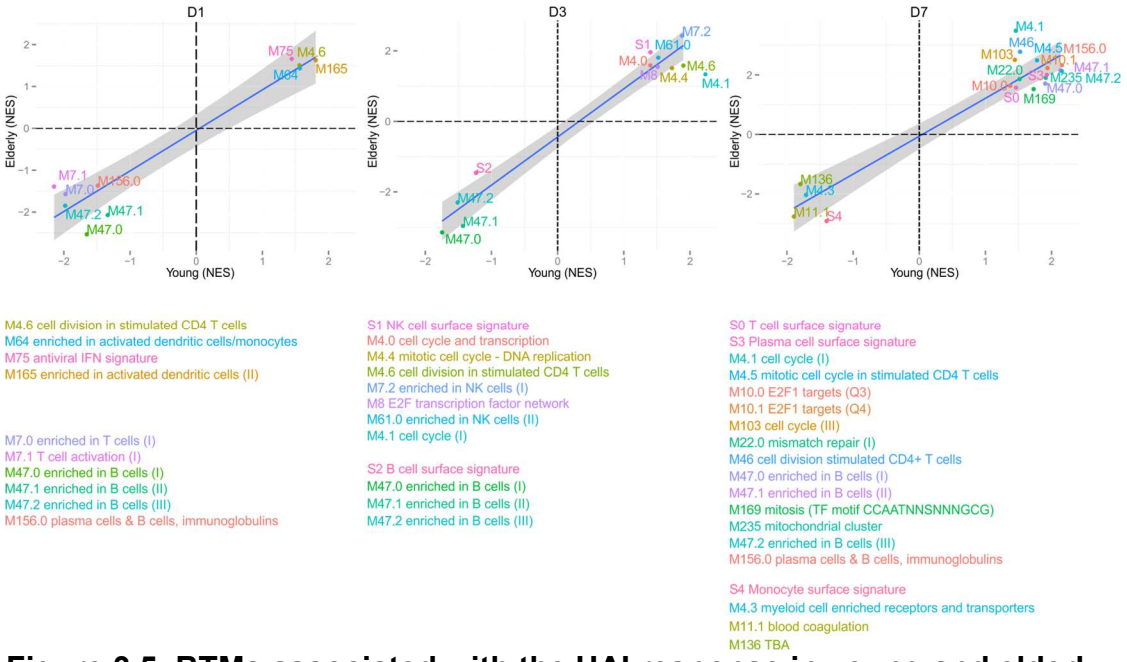

Figure 3.5. BTMs associated with the HAI response in young and elderly subjects from the 2010 season.

GSEA (nominal p<0.1; 1,000 permutations) was used to identify positive or negative enrichment of BTMs (gene sets) within pre-ranked gene lists, where genes were ranked according to their correlation between expression and HAI response in young and elderly subjects in the 2010 season. Modules represented are those consistently enriched for both young and elderly subjects, with Normalized Enrichment Score (NES) among Elderly subjects plotted on the y-axis and NES among young subjects plotted on the x-axis.

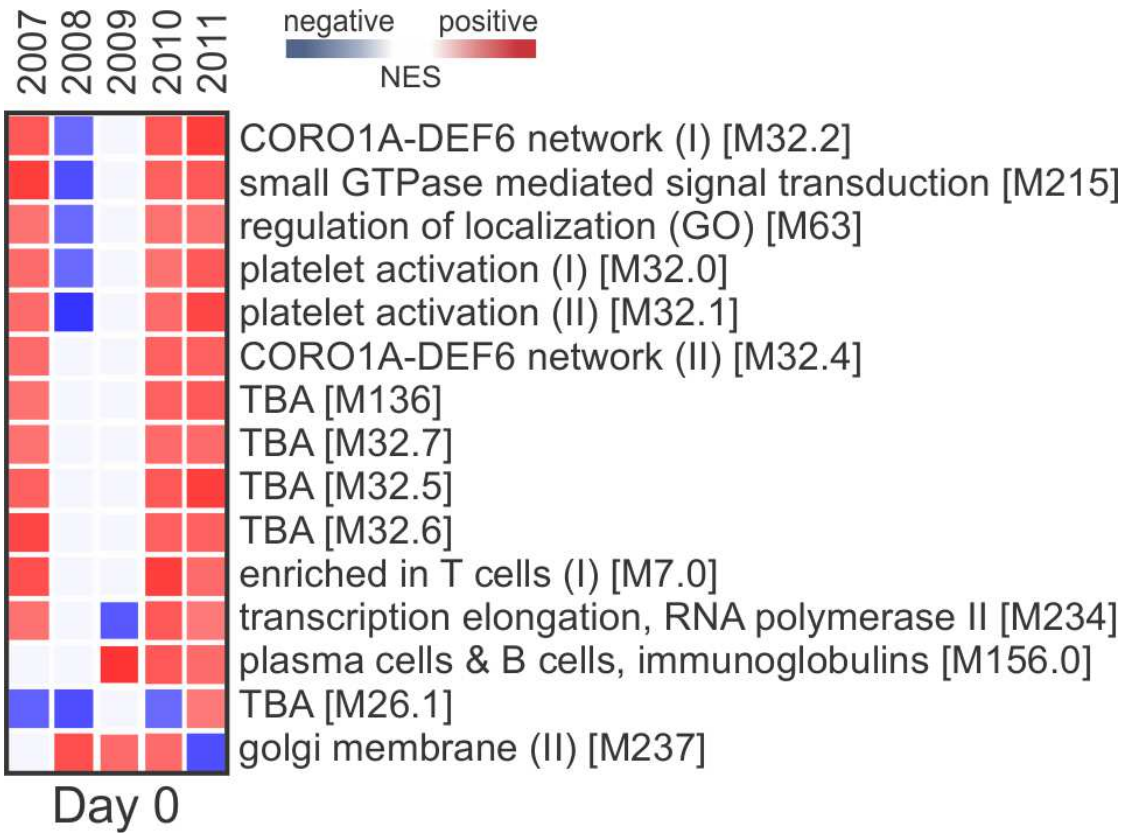

Figure 3.6. Baseline signatures associated with the antibody response in each season.

Heat map of Blood Transcription Modules (BTMs, rows) and TIV seasons (columns) whose activity at baseline is associated with HAI response at day 28 after vaccination. Gene Set Enrichment Analysis (GSEA, nominal $p<0.05$; 1,000 permutations) was used to identify positive (red), negative (blue) or no (grey) enrichment of BTMs (gene sets) within pre-ranked gene lists, where genes were ranked according to their correlation between expression and HAI response.

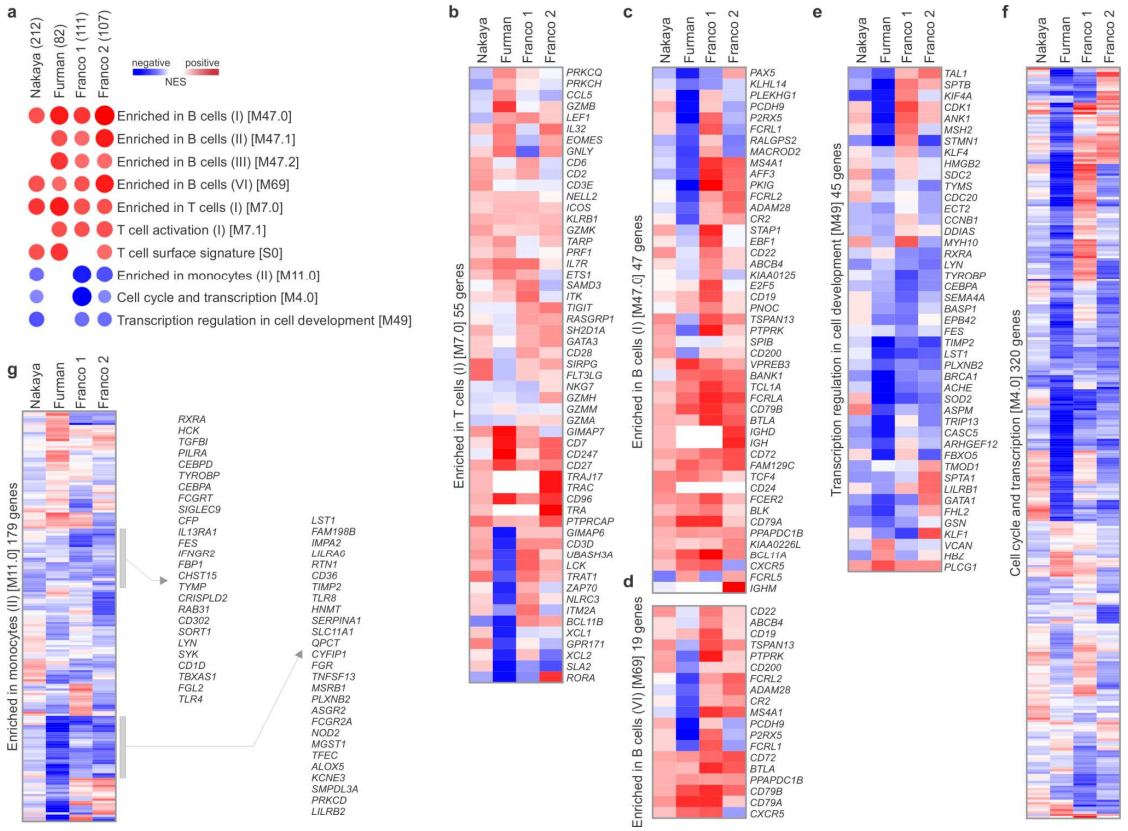

Figure 3.7. Baseline signatures associated with the antibody response. (a) Heat map of Blood Transcription Modules (BTMs, rows) and TIV studies (columns) whose activity pre-vaccination is associated with HAI response at day 28 after vaccination. Gene Set Enrichment Analysis (GSEA, nominal $p<0.05$; 1,000 permutations) was used to identify positive (red), or negative (blue) enrichment of BTMs (gene sets) within pre-ranked gene lists, where genes were ranked according to their correlation between expression and HAI response. Circle size is proportional to the normalized enrichment score (NES). Numbers in parenthesis next to each study represent number of subjects in the study. Modules shown are those consistently enriched in at least 3 out of 4 studies. (b-g) Heat maps of genes within BTMs from (a); colors represent the mean correlation in each study between baseline gene expression and HAI response at day 28 post-vaccination.

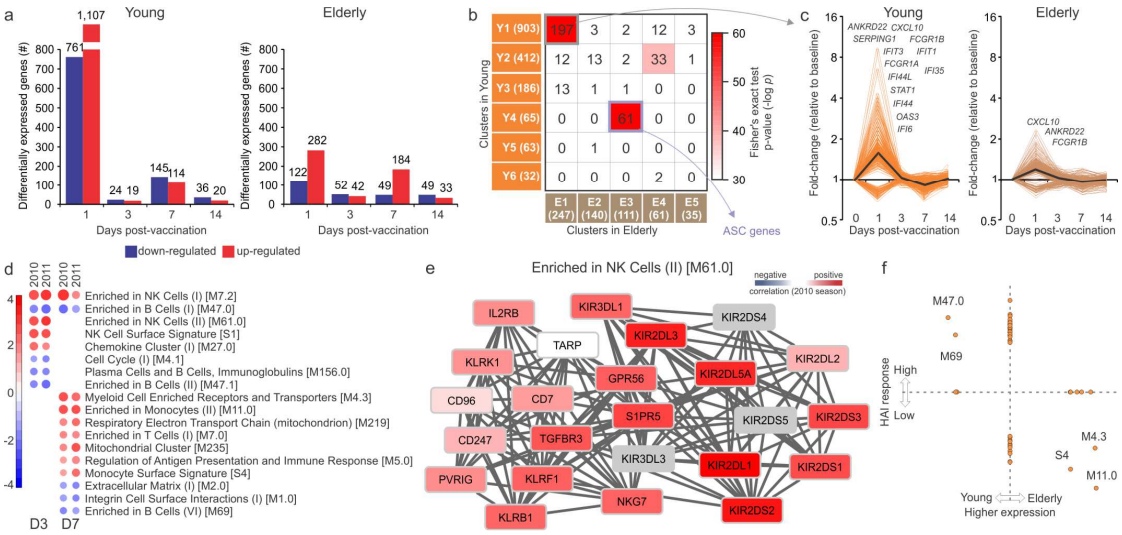

Figure 3.8. Molecular signatures induced by vaccination with TIV in young adults and in elderly.

(a) Number of genes differentially expressed (log_2 fold-change > 0.2 and t-test p value < 0.01) in young (<65 years) (left) and elderly (≥65 years) (right) vaccinees on days 1, 3, 7 and 14 post-vaccination (2010 season). (b) Heat map of highly correlated gene modules within the differentially expressed genes in (a) for young (rows, modules Y1 to Y6) and elderly (columns, modules E1 to E5), generated using Weighted Correlation Network Analysis (Langfelder and Horvath, 2008). The number of genes in each module is shown in parenthesis and the number of genes in common between 2 modules is shown inside the squares. Colors represent the Fisher's exact test p-value of the overlap between clusters. The 61 genes in common between Y4 and E3 are associated with antibody-secreting cells (ASCs). (c) Temporal expression patterns of 197 interferon-related genes in common between modules Y1 and E1 from (b). Black line represents the mean fold change of all genes (d) BTMs whose activity at days 3 or 7 post-vaccination is associated with the age of vaccinees from 2010 and 2011 seasons. GSEA was used to identify positive (red) or negative (blue) enrichment of BTMs (gene sets) within pre-ranked gene lists, where genes were ranked according to their correlation between expression and increasing age. The intensity of the color and the size of the circles represent the normalized enrichment score (NES) of GSEA. Modules shown are those consistently enriched in both seasons. (e) Genes in BTM M61.0; each 'edge' (gray line) represents a coexpression relationship, as described in Li et al. (Li et al., 2014); colors represent the correlation for 2010 season between day 3 fold change gene expression and the age of vaccinees. (f) BTMs whose activity at days 7 post-vaccination is correlated with the age of vaccinees (x-axis) and/or is correlated with HAI response (y-axis) in both 2010 and 2011 seasons. Values represent the mean of the NES obtained independently for each season. NES receives a value of zero if the BTM is not significantly associated with age or HAI response in either season.

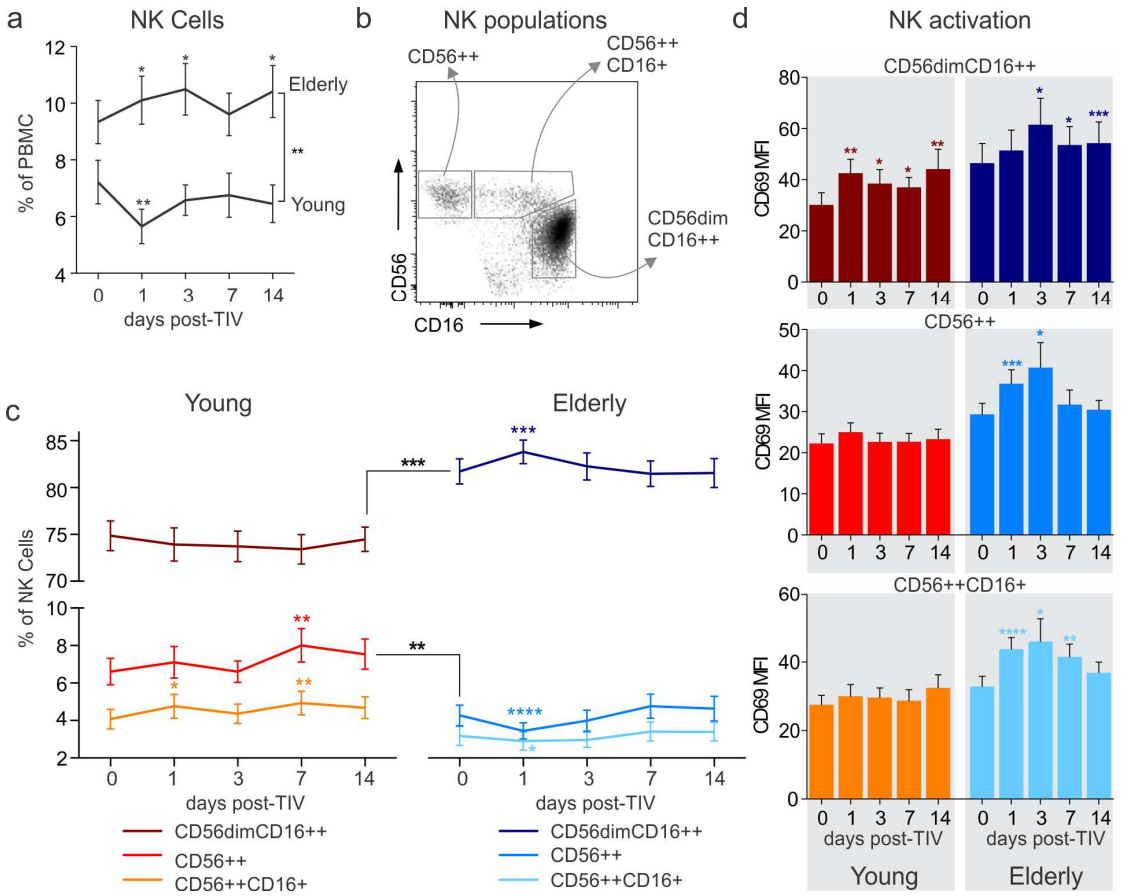

Figure 3.9. Flow cytometry analysis of NK cells in young and elderly after TIV vaccination.

(a) Changes in total NK cell population after vaccination represented as % within all PBMC for young and elderly. Mean +/-SEM. (b) Blood NK cells were defined within the CD3-CD4-CD19-CD14- PBMC. Dot-blot represents three distinct NK cell populations defined by CD56 and CD16 markers: CD56++ NK, CD56++CD16+ NK and CD56dim CD16++ NK. (c) Kinetics of magnitudes of CD56++ NK, CD56++CD16+ NK and CD56dim CD16++ NK cell subsets in young (left panel) and elderly (right panel) after vaccination. Mean +/-SEM. Areas Under Curve (AUC) in (a) and (c) were calculated to compare magnitudes of total NK cells and NK cells subsets between young and elderly cohorts throughout the study duration (days 0-14) and compared by t-test. Changes of each of the NK subset on the indicated time-points after vaccination were compared to the day 0 (baseline) timepoint by t-test. (d) Activation of each of the NK cell subsets were assessed by CD69 staining and compared with day 0 (baseline) by t-test. Data are represented as the geometric mean fluorescence intensity (MFI) for young (left panel) and elderly (right panel) at each time point, mean +/-SEM. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

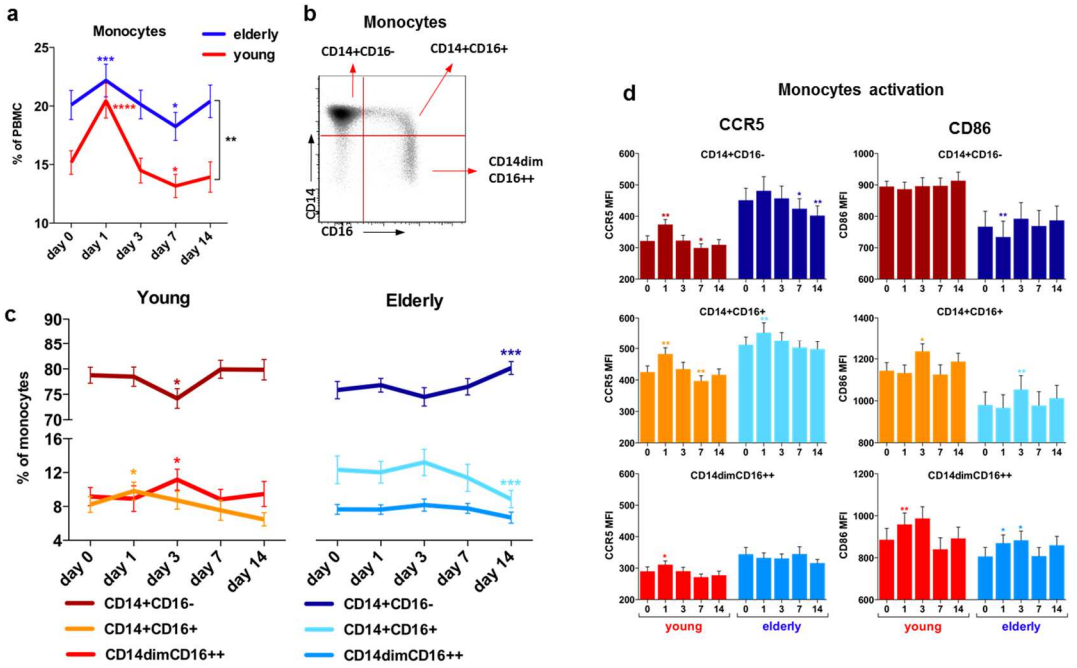

Figure 3.10. Flow cytometry analysis of monocytes in young and elderly after TIV vaccination.

(a) Changes in total monocytes population after vaccination represented as % within all PBMC for young (red) and elderly (blue). Mean +/-SEM. (b) Blood monocytes were defined within the CD3-CD19-CD14+ PBMC. Dot-blot represents three distinct populations defined by CD14 and CD16 markers: CD14+CD16- monocytes, CD14+CD16+monocytes and CD14dimCD16++ monocytes. (c) Kinetics of CD14+CD16- monocytes, CD14+CD16+monocytes and CD14dimCD16++ monocytes in young (left panel) and elderly (right panel) after vaccination. Mean +/-SEM. Areas Under Curve (AUC) in (a) and (c) were calculated to compare magnitudes of the monocyte subsets between young and elderly cohorts throughout the study duration (days 0-14) and compared by t-test. Changes of each of the monocytes subset on the indicated time-points after vaccination were compared to the day 0 (baseline) timepoint by t-test. (d) Activation of each of the monocyte subsets were assessed by CCR5 (left) and CD86 (right) staining and compared with day 0 (baseline) by t-test. Data are represented as the geometric mean fluorescence intensity (MFI) for young (left panel) and elderly (right panel) at each time point. Mean +/-SEM . *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

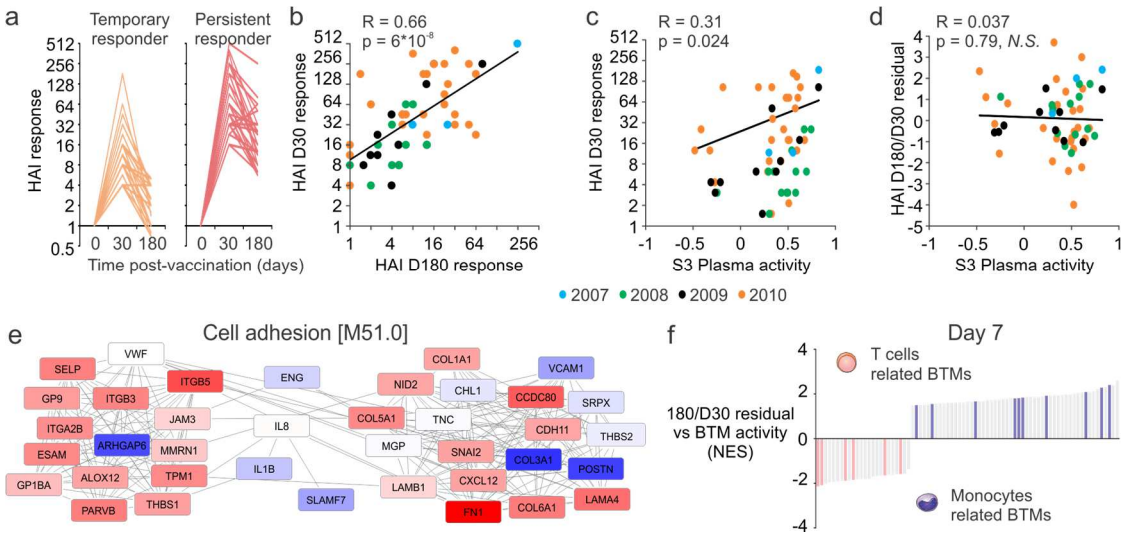

Figure 3.11. Signatures associated with the persistence of TIV-induced antibody response.

(a) HAI response (fold change of HAI titer relative to baseline) through 180 days for temporary (n=28) and persistent responders (n=34). Temporary responders met the FDA criteria for seroconversion (minimum 1:40 titer and fourfold increase post-vaccination) on day 28 but not on day 180, while persistent responders met the criteria on both days. (b) Comparison between day 28 and day 180 HAI responses. Each symbol represents a single vaccinee and the color represents the season that they were vaccinated (n=62). Black lines represent the regression line (Pearson) for all vaccinees combined. Day 180/day 28 residual is computed as the (vertical) distance from each sample to the regression line. (c) Comparison between "S3 Plasma" BTM activity and the HAI responses at day 28 post-vaccination. Each symbol represents a single vaccinee and the color represents the season that they were vaccinated (n=62). Black lines represent the regression line (Pearson) for all vaccinees combined. (d) Comparison between "S3 Plasma" BTM activity and the HAI D180/D28 residual. Each symbol represents a single vaccinee and the color represents the season that they were vaccinated (n=62). Black lines represent the regression line (Pearson) for all vaccinees combined. (e) Genes in BTM M51.0; each 'edge' (gray line) represents a coexpression relationship, as described in Li et al. (Li et al., 2014); colors represent the correlation between baseline-normalized gene expression at day 3 post-vaccination and the D180/D28 residual. (f) BTMs (bars) whose activity at day 7 post-vaccination is significantly associated with the HAI D180/D28 residual (GSEA; nominal p<0.05; 1,000 permutations). Vaccinees from 2007 to 2010 seasons were combined. BTMs related to T cell functions (pink bars) or monocyte functions (purple bars) are shown.

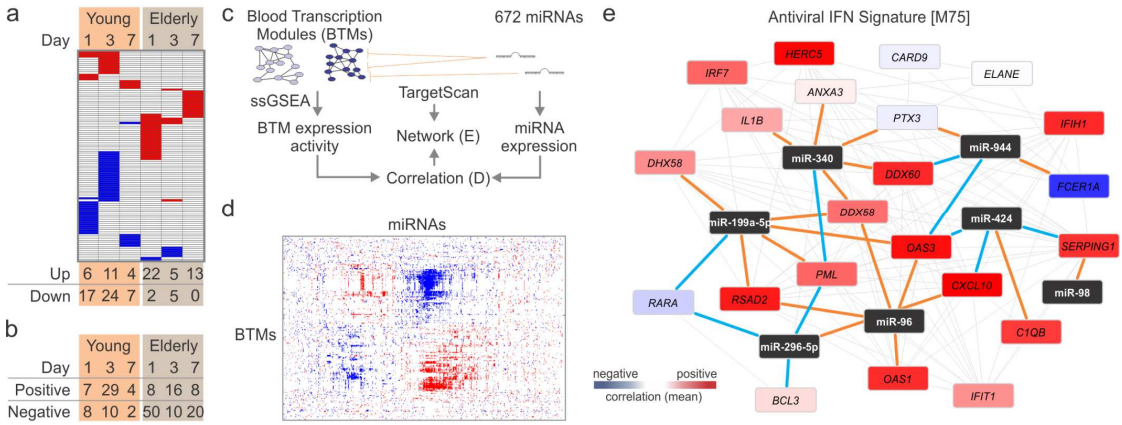

Figure 3.12. MicroRNA expression profiling of young and elderly TIV vaccinees.

(a) Heat map of miRNAs (rows) up- (red) or down-regulated (blue) at days 1, 3 and 7 post vaccination in young and elderly (columns); Paired t-test ($p<0.05$); total number of differentially expressed miRNAs are shown at the bottom. (b) MicroRNAs whose expression is positively or negatively correlated with HAI response in young and elderly; Pearson correlation ($p<0.05$). (c) Identification of networks potentially regulated by miRNAs. Activity of BTMs was determined by single sample GSEA (Barbie et al., 2009) and correlated with the expression of miRNAs. TargetScan database(Garcia et al., 2011) was used to identify the potential target genes of miRNAs. (d) Heat map of BTMs (rows) whose activity at day 1 post-vaccination correlated with the baseline-normalized expression of miRNAs (columns) at the same time point. Positive and negative correlations are shown in red and blue, respectively. (e) Genes in BTM M75; each gray line represents a coexpression relationship, as described in Li et al. (Li et al., 2014); each brown line connects a miRNA and its potential target gene; each blue line represents a negative correlation (Pearson, $p<0.15$) between the expression of miRNA and the expression of the potential target gene; colors represent the mean correlation between baseline-normalized gene expression at day 1 and HAI response at day 28 post-vaccination in the 2010 TIV season.

Chapter 4: Systems analysis of activated dendritic cells

Introduction

Successful immune response to pathogens in vertebrates requires complex coordination between many subsystems and cell types. As the primary antigen-presenting cells, dendritic cells (DCs) play a crucial role in this process by providing a bridge between the innate and adaptive immune responses. Among other functions, DCs help shape the adaptive immune response through control of T cell activation. By means of cytokines and co-stimulatory molecules expressed during antigen-presentation, DCs can help direct naïve T helper cells to differentiate into various subsets, such as IFN-γ producing Th1 cells or IL-4 producing Th2 cells (Banchereau and Steinman, 1998). Th1 cells are potent activators of macrophages and cytotoxic T cells, and as such are effective in response to intracellular bacteria and viruses. Meanwhile, Th2 cells are adept at controlling helminths through stimulation of B cell proliferation and IgE production and activation of eosinophils (Zhu et al., 2010). As such, proper DC control of Th cell differentiation according to the type of pathogen encountered is key to a successful immune response. As this process occurs in the lymph nodes, the contribution of dendritic cells to overall immunity and the molecular mechanisms by which this occurs cannot be evaluated in studies using peripheral blood cells, such as the work presented in Chapter 3. Instead, research in this area must often be performed *in vitro*.

Lipopolysaccharides (LPS), a major component of Gram-negative bacterial membranes, elicit strong immune responses and are often used to simulate the immune response to bacterial infections. LPS binds to Toll-like receptor 4 (TLR4) on innate immune cells, activating NFkB and interferon signaling through MyD88 and TRIF-dependent pathways, respectively (Lu et al., 2008; Takeda and Akira, 2004). While the transcriptional response in DCs to LPS is well understood (Amit et al., 2009; Patil et al., 2013), little work has been done to examine changes in lipid metabolism that occur during LPS treatment of DCs. As signaling molecules, various classes of lipids such as eicosanoids and sterols have been shown to influence a variety of functions within immune cells, including cytokine release, phagocytosis, migration, and antigen presentation (Harizi and Gualde, 2005; Spann and Glass, 2013). In addition, these molecules can directly induce transcriptional responses through interactions with transcription factors such as heat shock transcription factors and liver X receptors (Amici et al., 1992; Spann and Glass, 2013). Here we present an integrated systems biology analysis of the temporal transcriptional and metabolic changes in bone marrow-derived murine dendritic cells stimulated with LPS over a 24-hour time period.

Results

Transcriptional responses

To determine the temporal changes in gene expression that DCs undergo in response to LPS, we performed Designed Primer-based RNA sequencing (Bhargava et al., 2013) on bone marrow-derived DCs at 2, 6, 16, and 24 hours post-stimulation. By applying Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005) to gene lists ranked by z-score at each time point (see Materials and Methods), we identified transcriptional pathways progressively activated during the response (Fig. 4.1a). Among the processes enriched at early time points were B cell receptor and PI-3 kinase/Akt signaling. Although dendritic cells do not express B cell receptors, the B cell receptor signaling modules contain many proteasome components. Therefore enrichment of these modules likely reflects an increase in antigen processing and presentation, which rely on the function of the immunoproteasome (Ferrington and Gregerson, 2012). Meanwhile, PI-3 kinase and Akt activity has been shown to activate NFkB by initiating degradation of inhibitory IkB proteins associated with the NFkB complex (Andjelic et al., 2000; Pianetti et al., 2001), thus serving as an initial activator of the downstream inflammatory response. At later time points (16-24 hours), we observed enrichment of the oxidative stress response, as well as upregulation of aquaporins, particularly AQP3 and AQP4.

As expected, GSEA also identified enrichment of multiple modules related to interferon and interleukin signaling (Fig. 4.1b). Although these pathways were significantly enriched at all time points, this response peaked

at 6 hours. By examining the temporal expression of genes related to IFN-γ signaling, we observed significant upregulation of interferon regulatory factors Irf1, Irf7, and Irf9, but little to no change in Irf3, Irf4, or Irf8 (Fig. 4.1c). Stat1, a transcription factor activated by IFN-γ that interacts with several Irfs to induce transcriptional responses (Honda et al., 2006), was also modestly but consistently upregulated at later time points (6-24hr).

In order to compare transcriptional responses of dendritic cells to those of macrophages, a closely related innate immune cell, we repeated GSEA using microarray data available from the LIPID MAPS database (Fahy et al., 2007) (http://www.lipidmaps.org) generated from RAW264.7 cells treated with $(Kdo)_2$-lipid A (KLA), a component of LPS (Raetz et al., 2006), over 24 hours. In response to KLA, macrophages also induced significant proinflammatory signaling (Fig. 4.2), including upregulation of interferons, Nfkb signaling, and interleukins Il1a, Il1b, and Il6. The kinetics of this inflammatory response in macrophages appeared to be slightly faster than in dendritic cells, peaking at 2 hours instead of 6. In both cell types, induction of the inflammatory response was coincident with initiation of significant cell cycle arrest.

Transcription factor regulation of DC responses to LPS

In addition to describing the transcriptional responses that DCs undergo in response to LPS, we also sought to identify the key transcription factors that govern this response. To this end, we generated a transcription factor-target

interaction network by identifying differentially expressed transcription factors and their differentially expressed target genes at the same or later time points using the TRANSFAC database (Wingender, 2008). By analyzing the number of interactions for each species in the resulting network, we were able to identify important master regulators of the DC response to LPS (Fig. 4.3a). Sp1, a transcription factor involved in a wide variety of cellular functions, including cell proliferation, apoptosis, and immune response (Kaczynski et al., 2003; Park et al., 2002), had significantly more interactions than any other network member. A number of its downregulated targets (Fig. 4.3b), such as Rrm1, Srf, Dhfr, Hif1a, and Prim2, are involved in DNA replication or cell proliferation, suggesting that downregulation of Sp1 may contribute to the reduction in cell cycle-related gene expression post-stimulation as observed in Fig. 4.1b. In accordance with their role as key regulators of the inflammatory response, Nfkb and Rela were observed to jointly upregulate the response of many pro-inflammatory genes such as Nos2, Il12b, Ptgs2, and Cxcl10 (Fig. 4.3b).

Lipid metabolism in LPS stimulated DCs

To quantify changes in DC lipid metabolism induced by LPS, we performed liquid chromatography-mass spectrometry on cell extracts of DCs stimulated with LPS at the same time points as in the transcriptional analysis (2, 6, 16, and 24 hours), and mapped the annotated species onto pathways

from the LIPID MAPS database (Fahy et al., 2007) (http://www.lipidmaps.org). We first examined the synthesis of eicosanoids, in particular prostaglandins, which have been shown to regulate the inflammatory response, as well as the development and activity of a range of immune cells (Harris et al., 2002). Overall, there was an increase in prostaglandins post stimulation, however there was significant variation in across species (Fig. 4.4). Accompanying this increase in prostaglandin synthesis was a considerable upergulation in expression of Ptgs2, also known as COX2, an enzyme that converts arachidonic acid to prostaglandin H2 (PGH2) and has long been associated with inflammatory responses (Seibert and Masferrer, 1994). Concentration of prostaglandin E2 (PGE2), known to have diverse effects on B, T, and dendritic cells (Harris et al., 2002), was slightly downregulated at 2 hours, but upregulated at later time points, peaking at 16 hours. Meanwhile derivatives of prostaglandin D2 (PGD2), including 15-deoxy-PGJ2 and 15-deoxy-PGD2, exhibited the opposite trend, and were strongly downregulated at 16 hours.

We also sought to explore LPS-induced changes in sterol metabolism. Like prostaglandins, sterols and oxysterols are known to influence a variety of immune functions, including prevention of viral entry (Liu et al., 2013), immunoglobulin production (Bauman et al., 2009), and immune cell migration (Hannedouche et al., 2011). We observed an increase in the production of several oxysterol species, such as 25-hydroxy-cholesterol (Fig 4.5a). This increase peaked at early on in the response (2 hours) and was diminished by

24 hours, and was accompanied by upregulation in the transcripts of associated enzymes such as cholesterol 25-hyxroxylase (Ch25h).

As oxysterols are known to activate liver X receptor transcription factors LXRα and LXRβ (gene symbols Nr1h3 and Nr1h2, respectively) (Spann and Glass, 2013), we also examined changes in known LXR target genes post-LPS treatment (Fig. 4.5b). Although there was little increase in target LXR genes involved in lipid metabolism, inflammatory genes known to be suppressed by LXR activity peaked in expression at 6 hours, with diminished expression at later times.

Discussion

As the prototypical marker of infection with Gram-negative bacteria, the responses of immune cells to LPS have been a common subject of many previous studies. These responses involve changes in multiple cellular functions, including gene expression, metabolism, and cytokine signaling. The coordination of these pathways across multiple cell types and tissues is crucial for successful clearance of infection. In this current work, we sought to quantify the transcriptional and lipidomic dendritic cell responses to LPS, with the aim of describing the feedback between these two responses and comparing them to those of other closely related immune cells, such as macrophages.

In accordance with their role as the primary antigen-presenting cells of the body, many of the initial transcriptional changes we observed were involved with the induction of the immunoproteasome and cytokine signaling (Fig. 4.1). As expected, there was a large increase in pro-inflammatory cytokines (including IFNg, Il1a, Il1b, and Il6), as well as cytokines associated with the generation of Th1 cells, such as Il12b. Meanwhile our transcription factor-target analysis identified important regulators of this response (Fig. 4.3). Many of the highly connected transcription factors, such as Sp1, Sfpi1, Nfkb1, Rela, Jun, and Fos, were also identified as important hubs in a network of DC LPS response genes in recent work by Patil et al. (Patil et al., 2013). Interestingly, Relb, another member of the NFkB, family, has been shown to play an important role in DC function and priming of T cells (Shih et al., 2012; Zanetti et al., 2003). Our results, along with those of Patil et al., suggest that although Relb may be important for certain DC functions in response to antigen, Nfkb1 and Rela are still the dominant regulators of the inflammatory response to LPS in DCs.

The strong inflammatory response in DCs to LPS is similar to what is observed in other innate immune cells, in particular their close relative macrophages. Most DC and macrophage subsets develop from the same myeloid-committed hematopoietic stem cell precursors in the bone marrow (Geissmann et al., 2010). Comparing the transcriptional responses in this study (Fig. 4.1) with those of RAW264.7 macrophages stimulated with KLA

(Fig. 4.2), we observed enrichment of many common inflammatory pathways, including interferon and Nfkb signaling. Notably, IFNg, which was highly induced in LPS-treated DCs, was not upregulated in KLA-treated macrophages (data not shown). Overall, macrophages initiated a more rapid transcriptional response than DCs, with enrichments peaking at 2 hours post-treatment (as opposed to 6 hours in DCs).

In addition to inducing a strong transcriptional response, treatment with LPS caused significant changes to lipid metabolism in DCs. Analysis of eicosanoid metabolism (Fig. 4.4) revealed increases in pro-inflammatory prostaglandins, in particular PGH2, PGE2, and PGD2, along with significant upregulation of Ptgs2 (COX-2). We also observed production of oxysterols (Fig. 4.5a), including 25-, 27-, 4β-, and 7α-hydroxycholesterol. This lipidomic response in DCs is largely similar to that of macrophages (Dennis et al., 2010), which also generate increased prostaglandins and oxysterols in response to LPS/KLA. However, there are several notable differences between the two cell types. Both DCs and macrophages increased synthesis of prostaglandins post-stimulation, but the magnitude of production was considerably larger in macrophages (fold changes >10 compared with ~3 in DCs). Meanwhile, although both cell types generated similar levels of oxysterol production, the kinetics of this production was significantly different. Macrophages generated most oxysterols at later time points (12-24 hours), while DCs produced these compounds rapidly, peaking at 2 hours post-stimulation. This early production

of oxysterols may serve to help limit the duration of the inflammatory response in DCs, through LXR-mediated suppression of pro-inflammatory target gene expression (Fig. 4.5b).

In summary, stimulation with LPS induced coordinated pro-inflammatory transcriptional and lipidomic responses in DCs. These responses are largely similar to those of macrophages, with certain key differences, including the magnitude of prostaglandin synthesis and kinetics of oxysterol production. The impact of these variations in response is not clear and merits further study, in particular in the context of Th cell differentiation, a process predominantly regulated by DCs.

Experimental Procedures

Culture of bone marrow derived dendritic cells

Bone marrow cells were harvested from femurs and tibiae of C57BL/6J mice. Cells were cultured in RPMI supplemented with 10% FBS, 100 units/mL penicillin, 100 µg/mL streptomycin, 2mM L-Glutamine, 1 mM sodium pyruvate and 20ng/mL recombinant murine GM-CSF (Peprotech) at a density of 1X106 cells/mL in 10cm petri dishes at 37°C in a CO2 incubator (5% CO2). After 3 days of culture, 10ml of fresh culture medium was added to the plates. Cells were harvested after 7 days of culture and enriched for dendritic cells using CD11c MACS microbeads (Miltenyi).

Stimulation of BMDC and RNA isolation

Enriched BMDCs were plated at 1X106 cells/mL in a 6 well plate and either left unstimulated or treated with 100ng/mL of LPS (LPS-EB ultrapure, Invivogen). At 2hr, 6hr, 16hr and 24hr after treatment, cells were spun down and total RNA was isolated using the Qiagen miRNeasy kit as per the manufacturer's protocol.

Liquid chromatography-mass spectrometry

Cell pellets were extracted by chilled acetonitrile. The samples were kept on ice for 30 minutes and then centrifuged for 10 minutes at 13,400 x rpm at 4°C. The supernatant was removed and placed into autosampler vials. Mass spectral data were collected with a 10-minute gradient on a Thermo LTQ-Velos Orbitrap mass spectrometer (Thermo Fisher, San Diego, CA) set to collect data from m/z 85 to 2000 in positive ionization mode with a resolution of 60,000. Three technical replicates were run for each sample, using reverse phase C18 chromatography, with positive electrospray ionization (ESI). Following LC-MS, the data from the mouse and human samples were independently collected as .raw files and converted to CDF format using the XCalibur file converter software (Thermo Fisher, San Diego, CA). Peak detection, noise filtering, m/z and retention time alignment, feature

quantification, and data quality filtering were performed using apLCMS (Yu et al., 2009).

Acknowledgements

Chapter 4 is a modified presentation of material that is being prepared for publication as “Systems analysis of stimulated dendritic cells” by Hagan T, Bommakanti G, Li S, Bhargava V, Pulendran B, and Subramaniam S. The dissertation author was the co-primary investigator and author of this material.

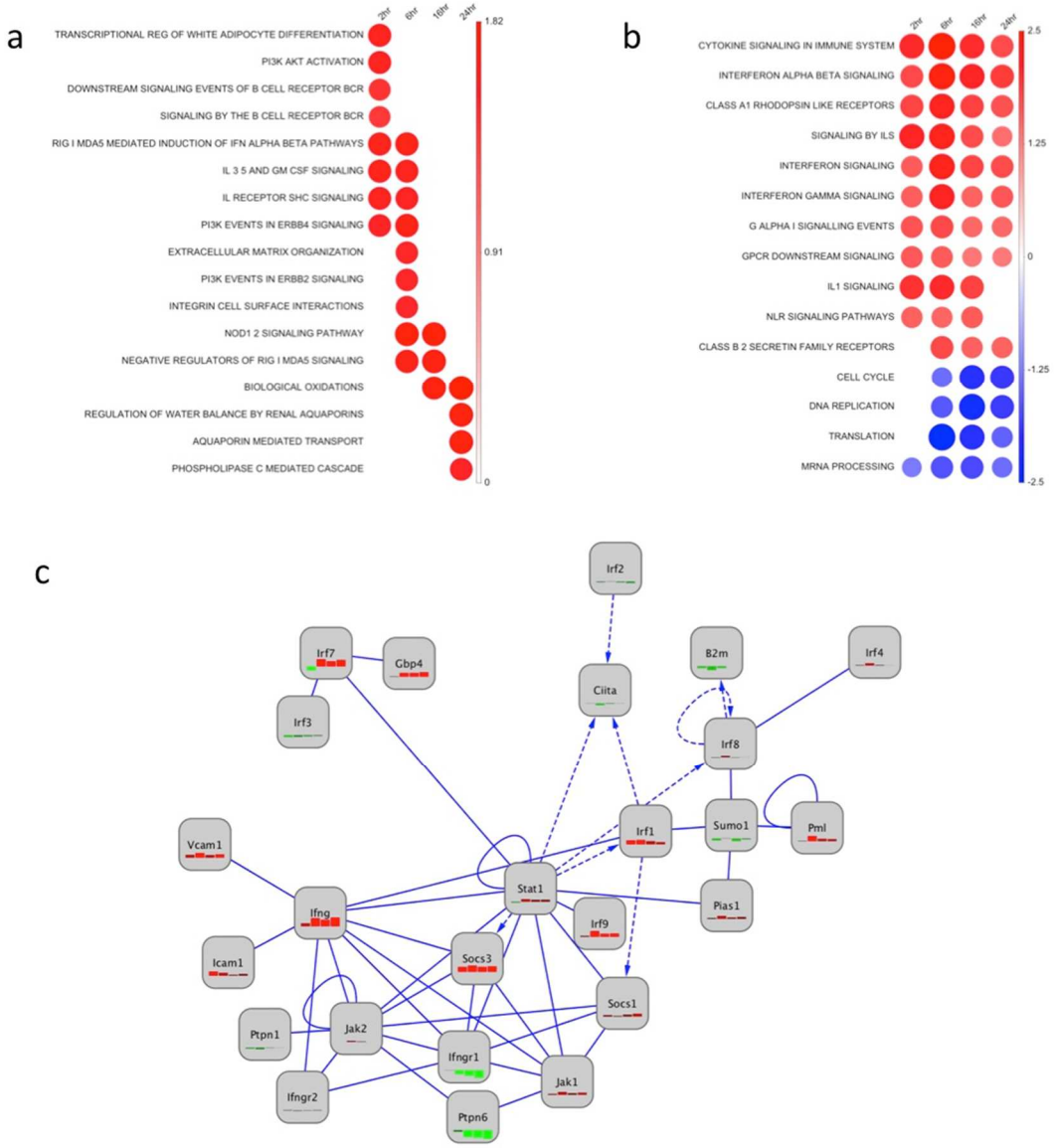

Figure 4.1. Transcriptional processes activated in DCs by LPS.
a) Heat map of reactome pathways that showed significant activation in two or fewer time points. GSEA was used to identify enrichment of reactome pathways (gene sets) within pre-ranked gene lists, where genes were ranked according to their normalized fold-change between the LPS-stimulated and unstimulated sample at each time point. b) Heat map of reactome pathways that showed significant positive (red) or negative (blue) enrichment in three or more time points. c) Expression of genes within reactome pathway 'Interferon Gamma Signaling'. The fold-change expression of each gene at all 4 time points (2 hr, 6hr, 16hr, 24hr left to right) is plotted within the corresponding node (red – positive fold change, green – negative fold change). Solid edges represent a protein-protein interaction; dashed edges represent a transcription factor-target interaction (arrow points to the target gene).

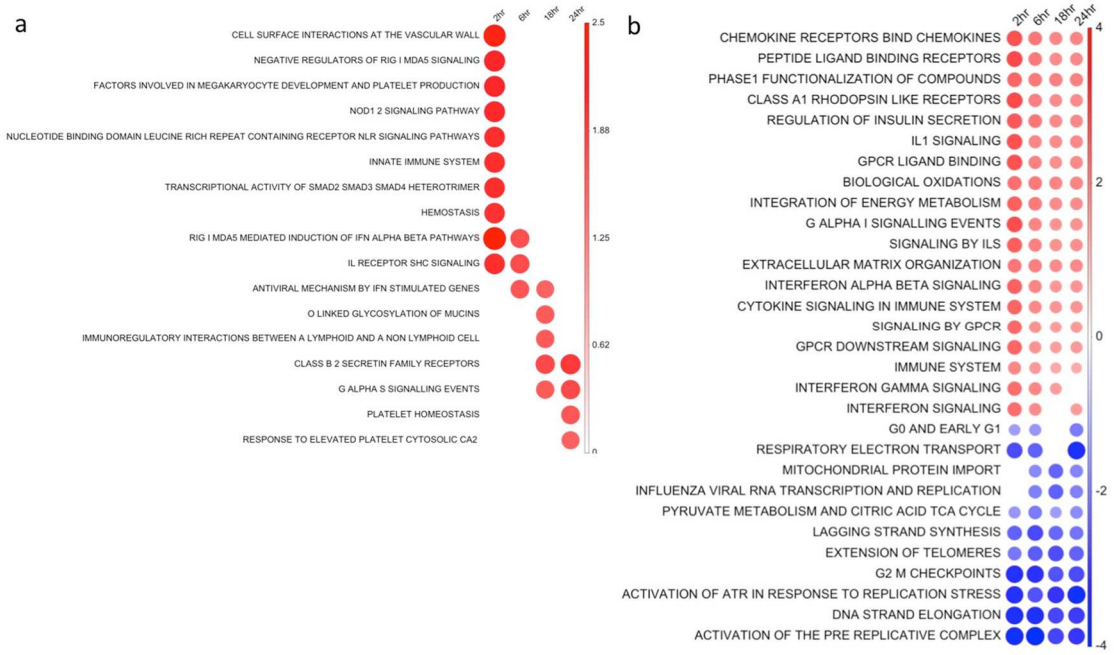

Figure 4.2. Transcriptional processes activated in RAW 264.7 macrophages by KLA.

a) Heat map of reactome pathways that showed significant activation in two or fewer time points. Gene Set Enrichment Analysis (GSEA, nominal $p<0.05$; 1,000 permutations) was used to identify enrichment of reactome pathways (gene sets) within pre-ranked gene lists, where genes were ranked according to their fold-change between the KLA-stimulated and unstimulated sample at each time point. (NES – normalized enrichment score) b) Heat map of reactome pathways that showed significant positive (red) or negative (blue) enrichment in three or more time points.

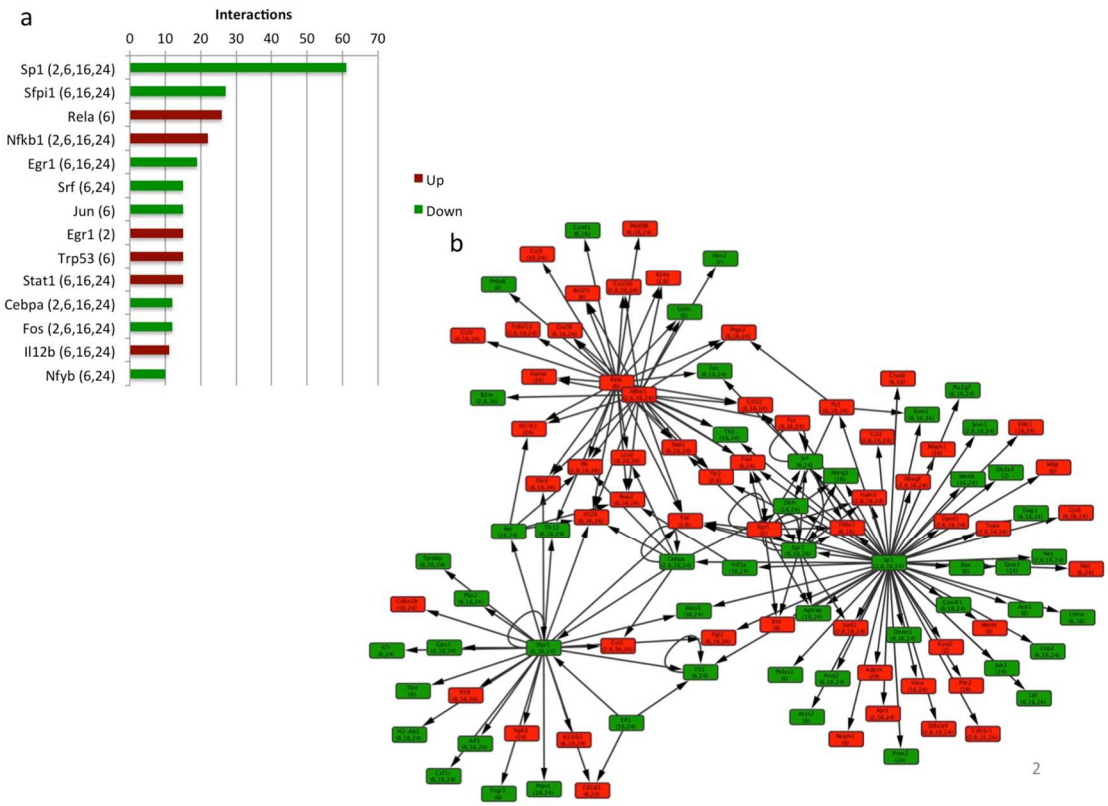

Figure 4.3. Transcription factor regulation of DC responses to LPS.
A transcription factor-target interaction network was generated by identifying differentially expressed transcription factors that had targets which were also differentially expressed at the same or later time point. a) Number of interactions of the most connected nodes. Numbers represent time points at which the gene is differentially expressed. b) Interaction network of the top 4 most connected transcription factors and their associated targets. Red – upregulated, green – downregulated.

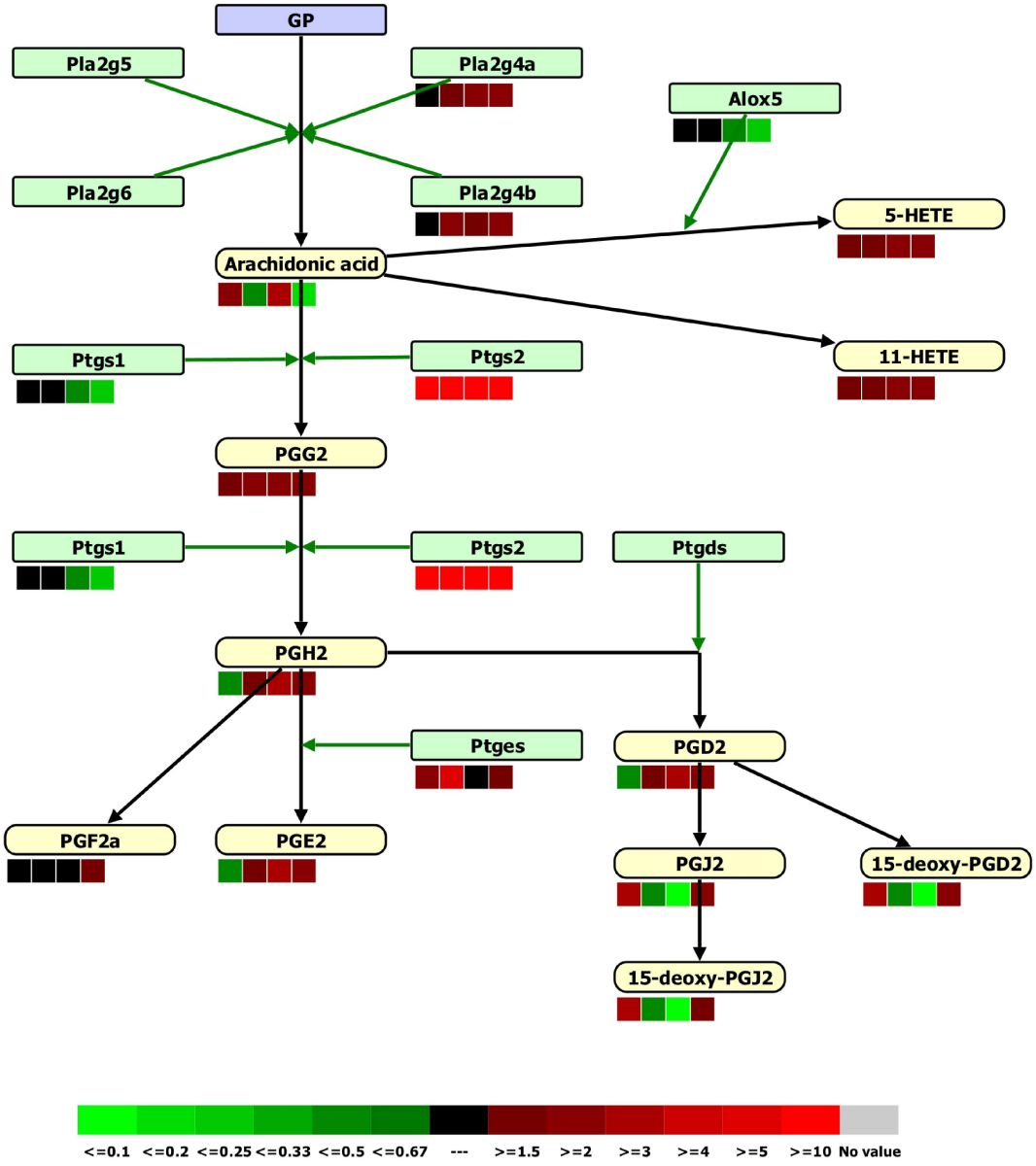

Figure 4.4. Prostaglandin synthesis in LPS-treated DCs.
Green boxes represent genes coding for enzymes that catalyze the indicated conversions, whereas yellow boxes represent lipid metabolites. The heat maps (see color key) below each lipid and gene represent the LPS/control fold-change at 2, 6, 16, and 24 h of treatment.

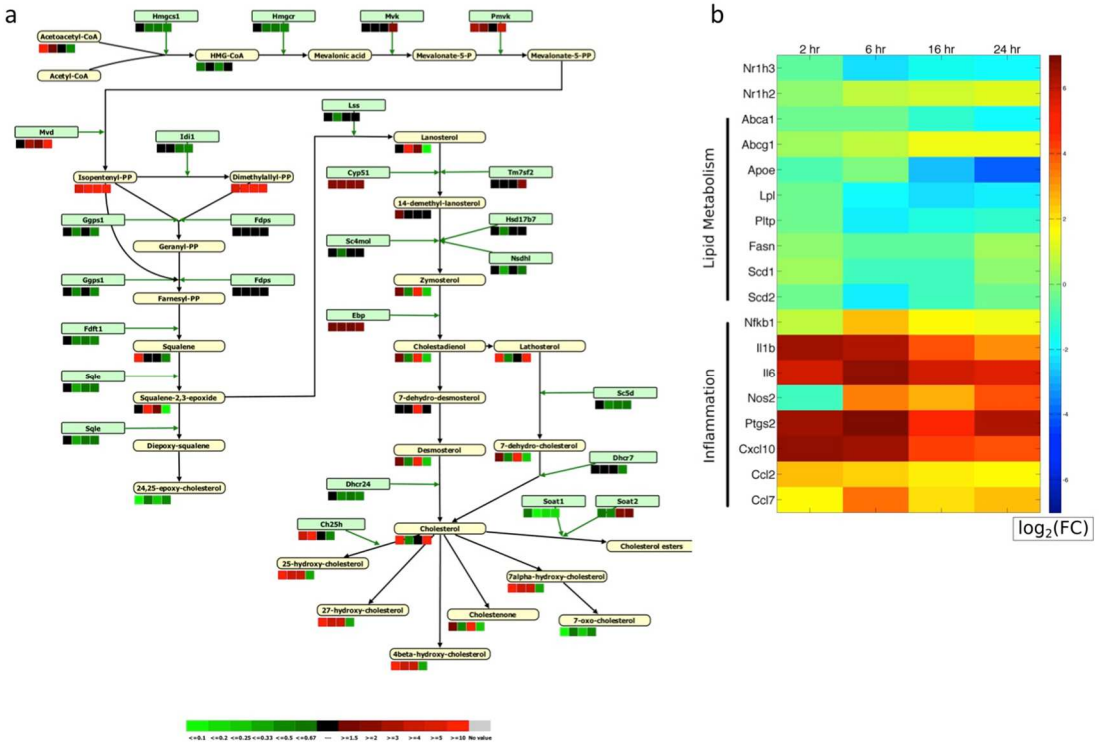

Figure 4.5. Sterol synthesis and LXR activity in LPS-treated DCs.
a) Changes in sterols during LPS treatment. Green boxes represent genes coding for enzymes that catalyze the indicated conversions, whereas yellow boxes represent lipid metabolites. The heat maps (see color key) below each lipid and gene represent the LPS/control fold-change at 2, 6, 16, and 24 h of treatment. b) Changes in LXR and associated target genes during LPS treatment. Colors represent the log2 fold-change between LPS and control.

Chapter 5: Concluding discussion and future directions

In this thesis, we have presented novel systems biology approaches to understanding and predicting immune responses. As research in this area continues, these tools will hopefully lead to diagnostic platforms capable of identifying, potentially pre-vaccination, patients at risk for poor vaccine response. These technologies, along with the improved knowledge of molecular mechanisms driving successful immune responses derived through their development, have the capability to greatly improve the development and application of future vaccines. However, there are several obstacles that must be overcome if this field is to reach its full potential.

Challenges

Systems vaccinology analyses face challenges from both the biological and technological perspectives. These problems must be tackled simultaneously in order to achieve a deeper understanding of how the immune system as a whole responds to vaccination. This is why it is important that systems vaccinology be carried out through collaboration between vaccinologists, immunologists and bioinformaticians. As biological knowledge and insight as well as rigorous and sound data analysis are both required for successful discovery, these two groups should work in close contact and coordinate throughout each phase of the scientific process, including hypothesis generation, experimental design, execution, and data analysis.

Biological challenges

A major factor complicating the analysis of responses to vaccination is the sheer diversity within the human population. The state of the immune system is a function of a host of variables, including genetics (Orru et al., 2013), previous infections/vaccinations, age (Duraisingham et al., 2013), chronic diseases, and environmental factors such as diet (Kau et al., 2011), gut microbial composition (Hooper et al., 2012), stress (Padgett and Glaser, 2003), and physical activity (Woods et al., 2006). While the role of genetics in creating susceptibility to cancer and many heritable diseases has been well established, its effect on vaccine response is less clear. A recent study of immune function in twins by Brodin et al. found that while a majority of immune variation is due to non-heritable factors, certain responses, such as IL-2 and IL-7 induced STAT5 phosphorylation in T cells, were highly heritable (Brodin et al., 2015). In addition, candidate gene studies as well as genome-wide association studies are beginning to be used to identify polymorphisms in genes associated with improved or diminished vaccine responses (O'Connor and Pollard, 2013). Another source of genetic immune variation in humans is that of HLA alleles. As a key step in generating adaptive immune responses involves antigen processing and presentation to T and B cells, the capability of the HLA molecule to bind peptides derived from processed infectious agents or vaccines affects the ability of the immune system to successfully respond to infection or vaccination. Therefore, recent studies have examined the impact

that variation in HLA alleles has on antibody responses to vaccines against influenza (Gelder et al., 2002; Moss et al., 2013), measles (Poland, 1999), smallpox (Ovsyannikova et al., 2014; Ovsyannikova et al., 2011), and hepatitis B (Li et al., 2013b).

Although human diversity is a challenging aspect of immunology and vaccinology, systems biology approaches are particularly well suited to characterize and account for the complex inter-individual differences within a population (Pulendran, 2014). Regression models and machine learning algorithms are able to analyze the effect of many phenotypic variables on responsiveness to vaccination simultaneously. As our understanding of how factors such as genetics, age, and gut microbiome affect immune responses improves, it is reasonable to believe that we will see development of specific vaccine formulations tailored to specific populations such as the elderly or immunocompromised that are designed to stimulate the deficient components of the response unique to each group.

In particular, effect of age on vaccine response is a critical issue. While vaccines are largely successful in preventing infection among the young and in healthy adults, many vaccines, such as the influenza vaccine (Gardner et al., 2006), zoster vaccine, and pneumococcal vaccine (Jackson and Janoff, 2008) show reduced effectiveness in the elderly. This represents a pressing problem for vaccine development, as the world population is rapidly aging, with an expected doubling in the proportion of the population over 60 from 10% in

2000 to 21.8% in 2050 (Lutz et al., 2008). The mechanisms responsible for the decrease in vaccine efficacy in the elderly are just beginning to be examined. For example, the lower efficacy of TIV in elderly compared to young adults has been associated with the some of the changes resulting from natural age-associated deterioration of immune functions (immunosenescence) (Duraisingham et al., 2013). At a cellular level, this includes a reduction in number of antibody-secreting cells (ASC) (Sasaki et al., 2011), loss of the less differentiated influenza-specific memory CD8+ T cells (Wagar et al., 2011) and lower frequency of effector memory CD4+ T cells (Kang et al., 2004). Our analysis of young and elderly vaccinees in Chapter 2 identified several novel age-associated changes in immune response to influenza vaccination. Elderly subjects showed increases in NK cell frequencies post-vaccination, along with increased NK cell activation. In contrast, young subjects exhibited decreased NK cell frequencies post-vaccination, as well as a decreased proportion of cytotoxic CD56dimCD16++ NK cells. Interestingly, transcriptional analysis of pathways associated with the antibody titer fold-changes did not find any significant association of this increased NK cell response with decreased antibody responses. Instead, we identified several monocyte modules that may link deficits in antibody responses with increased monocyte frequencies in the elderly. In particular, our baseline analysis of signatures associated with antibody responses revealed novel evidence of potential connections between age-associated immune changes at baseline and decreased vaccine

responsiveness through identification of negative correlations between pre-vaccination monocyte signatures and antibody fold-changes.

Meanwhile, for vaccines stimulating recall responses such as the influenza and zoster vaccines, history of exposure to the pathogen is very important, as pre-existing antibodies are associated with a decreased response to vaccination (Lauterbach et al., 2005; Sasaki et al., 2008). As these antibodies may be neutralizing viral particles from the vaccine and preventing an otherwise capable immune system from generating a response, this factor should be considered when examining the mechanisms of response. While several groups have attempted various methods to normalize the measure of antibody response to account for variations in baseline antibody levels (Bucasas et al., 2011; Tsang et al., 2014), there is not yet a consensus approach.

Finally, a considerable challenge for developing successful vaccines is to identify reliable correlates of protection from infection. While antigen-specific antibody titers are established as the primary correlate of protection for most vaccines (Plotkin, 2008), emerging research suggests that in many cases other measures may be equally if not more indicative of protected status. For example, there has been recent work suggesting that serum antibody titers may not be predictive of risk of influenza infection in the elderly (McElhaney et al., 2006). Instead, McElhaney et al. found that strong influenza-specific T cell responses (identified by high IFN-γ:IL-10 ratios

following *ex vivo* stimulation of PBMCs with influenza virus) were indicative of a reduction in risk of illness in elderly subjects (McElhaney et al., 2006). In addition, there is evidence that T cell responses may play an important role in preventing infection for many of the vaccines under current development such as HIV (Pantaleo and Koup, 2004), tuberculosis (Hoft, 2008), and malaria (Reyes-Sandoval et al., 2009).

Even when antigen-specific antibody titers are a robust correlate of protection, the techniques used to measure them may not be reliable. Two of the most widely used assays to measure influenza-specific antibodies, hemagglutination inhibition (HAI) and virus neutralization (VN) have been shown to have significant reproducibility issues. In a comparison of antibody titer measurements on identical samples across several laboratories, HAI and VN assays had geometric coefficients of variation ranging from 138-261% and 256-369%, which corresponded to fold differences of 16-128 and 91-724 respectively (Stephenson et al., 2007). For the VN assay, 21% of replicate measurements within the same laboratory differed by more than 2-fold (Stephenson et al., 2007). Many studies of influenza vaccination use the FDA criteria of seroconversion (Frey et al., 2015; Greenberg et al., 2013; Nakaya et al., 2011) (>4 fold increase in HAI titer post-vaccination) to categorize subjects into ‘responders’ and ‘nonresponders’ to vaccination; these results demonstrate that this classification may vary greatly between laboratories. As the accuracy of the endpoint measurement is crucial for obtaining meaningful

analysis in any study, improvements in standardization of these assays will be of great benefit to the field.

Alternatively, systems biology approaches can be harnessed to establish novel correlates of protection. Currently, nested case-control studies are used in clinical trials to retrospectively compare subjects who developed infection with subjects that remained uninfected and identify distinguishing biomarkers (Haynes et al., 2012; Olotu et al., 2014). The development of experimental challenge models (such as those available for malaria, typhoid, *Shigella*, and dengue fever) provides a complementary approach to enable detection of subjects protected from or vulnerable to infection. Systems level profiling of these subjects can then be used to reveal reliable molecular markers of vaccine-induced protection.

Technological challenges

A distinct feature of a true systems analysis of vaccine response is the integration of diverse measurement types. Due to the high coverage and relative ease of measurement provided by microarrays and RNA-sequencing, most of the initial work in systems vaccinology has focused on transcriptional responses to vaccination. However, modulation of gene expression represents only one of many methods by which biological systems respond to perturbation. Among other factors, changes in protein, metabolite, and signaling molecule abundances, post-translational modifications, and shifts in

cell subset populations all contribute to the immune response post-vaccination. As the technologies capable of detecting these additional layers of biological regulation improve, future systems vaccinology analyses should incorporate these measurements in order to develop a more complete picture of the biological mechanisms of vaccine-induced immunity.

As with any type of assay, measurements made using high-throughput technologies suffer from varying degrees of noise. This creates difficulty in reliable detection of weak signals, such as those coming from low-expressed transcripts or metabolites present in limited amounts. Unfortunately, the abundance of a particular molecule is not necessarily an indicator of its importance, and these difficult to detect components often have significant biological impact. Primary examples of this are transcription factors, which are often present in very low numbers but are important regulators of many biological processes (Spitz and Furlong, 2012). Measurement noise, along with batch effects (particularly in the case of microarrays) (Leek et al., 2010), can also reduce the reproducibility of results. This is an important obstacle that needs to be overcome in order for laboratory findings such as predictive signatures of vaccine efficacy to be translated into tools with clinical diagnostic value.

Another challenge common to many areas of medical research is the limited number of samples and time points available in a given dataset due to ethical and logistical constraints. As sample collection and processing is both

taxing on trial participants and costly, studies are often restricted to a small group of subjects who are evaluated at a handful of time points. While small sample sizes create obvious problems in statistical hypothesis testing, the limited availability of time course data also poses a significant obstacle to successful kinetic modeling of immune responses to vaccination. As immune responses involve coordinated interactions between diverse cell types across multiple tissues, understanding of cellular and transcriptional dynamics post-vaccination is an important goal of systems vaccinology. However, while transcriptional responses can take place within minutes to hours (Amit et al., 2009), in clinical studies sample collection is often only feasible on the time scale of days to weeks. This is complicated by the fact that while immune cells migrate between the blood, the site of infection/vaccination, and the lymph nodes, analysis is often performed only on immune cells within peripheral blood. These limitations, combined with the inter-individual variability in responses, make it extremely difficult to capture the entirety of transcriptional and cellular events that occur *in vivo* post-vaccination.

In addition to limited number of samples, in some cases (e.g. pediatric studies) researchers are also constrained by the small sample volume that can be collected from subjects. Typical guidelines for blood sample volume limits range from 1-5% of total blood volume, in children this equates to 1-3 mL/kg of body weight (Kauffman, 2000). The available blood drawn from neonates may therefore be <5 mL. As the concentration of PBMCs is typically around 1

million per mL of blood, a neonatal blood sample may only contain a few million PBMCs. This limits the potential analyses that can be performed, as assays such as flow cytometry and ELISPOT normally utilize larger numbers of cells. Development of more sensitive technologies that require fewer cells or less starting material, such as RNA-sequencing in place of microarrays, will enable improved analysis of vaccine response in these populations.

Unlike conventional techniques, which generate a limited number of measurements, high-throughput technologies produce tens or hundreds of thousands of measurements per sample. These developments have resulted in an explosion of datasets where the number of features (or dimension, p) is much larger than the sample size (n), known as ‘large p, small n’ or ‘high dimension, low sample size’ (HDLSS) data. Under these conditions, traditional statistical and computational algorithms perform poorly due to the resulting sparsity of the dataset; this is known as the ‘curse of dimensionality’. As a result, an entire branch of statistics has emerged to handle this type of data and reduce spurious results (Johnstone and Titterington, 2009). In order to avoid being swamped by false positives, suitable corrections for multiple hypothesis testing must be used to control false discovery rates (Benjamini and Hochberg, 1995; Storey and Tibshirani, 2003). Machine learning techniques (Kotsiantis et al., 2006) such as the artificial neural networks (ANNs) developed in Chapter 1 of this thesis, are being used to predict antibody responses to vaccination based on expression values of sets of

genes (or other measurements such as cell population frequencies) made shortly after vaccination (or even potentially at baseline). However, a large majority of the measurements made using high-throughput technologies are not relevant to the particular biological processes being examined, and these algorithms are not equipped to identify the subset of features (genes, metabolites, etc.) that are most predictive of responsiveness to vaccination. Therefore, an additional step known as feature selection must be performed, in order to reduce the computational complexity of the problem and to prevent overfitting (in which the algorithm performs well on training data but poorly on independent testing data). Various approaches to this task have been developed, such as random forests (Breiman, 2001), filter and wrapper approaches (Blum and Langley, 1997), and Bayesian methods (Inza et al., 2000). In Chapter 1, we applied modified forward selection approaches to identify combinations of transcriptional modules capable of accurately predicting antibody responses to influenza vaccination. Feature selection, particularly in the context of biological data, is an area of ongoing research, with no clear optimal approach and variability in performance across datasets and biological contexts. This is a critical step, because the identification of a small number of predictive features is necessary for clinical diagnostic applications, and the identity of these features provides insight into the underlying biology of the immune response to vaccination.

Finally, the generation of such large amounts of information requires appropriate computational resources as well as data management and storage. Creation of relevant and centralized databases such as the Immunology Database and Analysis Portal (ImmPort) (Bhattacharya et al., 2014) (http://immport.niaid.nih.gov/) will promote data sharing between groups and facilitate integration of diverse data types. Standardization of measurement techniques as well as data formatting should be encouraged, as this will empower the meta-analysis of multiple related studies, enabling discovery of novel associations which were previously undetectable.

Conclusion

Developments in high-throughput technologies are enabling vaccinologists to investigate immune responses at a greater depth than ever before. More importantly, these advances are facilitating the identification of robust molecular and cellular signatures of protective immunity, which can help to generate diagnostic tools that reduce the length and cost of current clinical trials. In this developing field, vaccinologists, immunologists, bioinformaticians, and systems biologists must work hand in hand to overcome the unique challenges associated with systems vaccinology and advance our understanding of the molecular mechanisms by which vaccines induce protective immunity. These critical insights will help to drive development of the next generation of vaccines.

Acknowledgements

Chapter 5 is a modified presentation of material published as “Systems vaccinology: Enabling rational vaccine design with systems biological approaches” by Hagan T, Nakaya H, Subramaniam S, and Pulendran S in Vaccine 2015. The dissertation author was the primary author of this material.

REFERENCES

Almeida-Oliveira, A., Smith-Carvalho, M., Porto, L.C., Cardoso-Oliveira, J., Ribeiro, A.D., Falcao, R.R., Abdelhay, E., Bouzas, L.F., Thuler, L.C.S., Ornellas, M.H., and Diamond, H.R. (2011). Age-related changes in natural killer cell receptors from childhood through old age. Human Immunology *72*, 319-329.

Amici, C., Sistonen, L., Santoro, M.G., and Morimoto, R.I. (1992). Antiproliferative prostaglandins activate heat shock transcription factor. Proc Natl Acad Sci U S A *89*, 6227-6231.

Amit, I., Garber, M., Chevrier, N., Leite, A.P., Donner, Y., Eisenhaure, T., Guttman, M., Grenier, J.K., Li, W., Zuk, O., Schubert, L.A., Birditt, B., Shay, T., Goren, A., Zhang, X., Smith, Z., Deering, R., McDonald, R.C., Cabili, M., Bernstein, B.E., Rinn, J.L., Meissner, A., Root, D.E., Hacohen, N., and Regev, A. (2009). Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. Science *326*, 257-263.

Anderson, P., and Kedersha, N. (2002). Visibly stressed: the role of eIF2, TIA-1, and stress granules in protein translation. Cell Stress Chaperones *7*, 213-221.

Andjelic, S., Hsia, C., Suzuki, H., Kadowaki, T., Koyasu, S., and Liou, H.C. (2000). Phosphatidylinositol 3-kinase and NF-kappa B/Rel are at the divergence of CD40-mediated proliferation and survival pathways. J Immunol *165*, 3860-3867.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R., Fernandes, G.R., Tap, J., Bruls, T., Batto, J.M., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., Kurokawa, K., Leclerc, M., Levenez, F., Manichanh, C., Nielsen, H.B., Nielsen, T., Pons, N., Poulain, J., Qin, J.J., Sicheritz-Ponten, T., Tims, S., Torrents, D., Ugarte, E., Zoetendal, E.G., Wang, J., Guarner, F., Pedersen, O., de Vos, W.M., Brunak, S., Dore, J., Weissenbach, J., Ehrlich, S.D., Bork, P., and Consortium, M. (2011). Enterotypes of the human gut microbiome. Nature *473*, 174-180.

Banchereau, J., and Steinman, R.M. (1998). Dendritic cells and the control of immunity. Nature *392*, 245-252.

Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C., Sandy, P., Meylan, E., Scholl, C., Frohling, S., Chan, E.M., Sos, M.L., Michel, K., Mermel, C., Silver, S.J., Weir, B.A., Reiling, J.H., Sheng, Q., Gupta, P.B., Wadlow, R.C., Le, H., Hoersch, S., Wittner, B.S., Ramaswamy, S., Livingston, D.M., Sabatini, D.M., Meyerson, M., Thomas, R.K., Lander, E.S., Mesirov, J.P., Root, D.E., Gilliland, D.G., Jacks, T., and Hahn, W.C. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature *462*, 108-112.

Barouch, D.H., and Michael, N.L. (2014). Accelerating HIV-1 Vaccine Efficacy Trials. Cell *159*, 969-972.

Bauer, E., and Kohavi, R. (1999). An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. Machine Learning *36*, 105-139.

Bauman, D.R., Bitmansour, A.D., McDonald, J.G., Thompson, B.M., Liang, G., and Russell, D.W. (2009). 25-Hydroxycholesterol secreted by macrophages in response to Toll-like receptor activation suppresses immunoglobulin A production. Proc Natl Acad Sci U S A *106*, 16764-16769.

Behbehani, A.M. (1983). The smallpox story: life and death of an old disease. Microbiol Rev *47*, 455-509.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B Met *57*, 289-300.

Bensinger, S.J., and Tontonoz, P. (2008). Integration of metabolism and inflammation by lipid-activated nuclear receptors. Nature *454*, 470-477.

Beziat, V., Duffy, D., Quoc, S.N., Le Garff-Tavernier, M., Decocq, J., Combadiere, B., Debre, P., and Vieillard, V. (2011). CD56brightCD16+ NK cells: a functional intermediate stage of NK cell differentiation. J Immunol *186*, 6753-6761.

Bhargava, V., Ko, P., Willems, E., Mercola, M., and Subramaniam, S. (2013). Quantitative transcriptomics using designed primer-based amplification. Sci Rep *3*, 1740.

Bhattacharya, S., Andorf, S., Gomes, L., Dunn, P., Schaefer, H., Pontius, J., Berger, P., Desborough, V., Smith, T., Campbell, J., Thomson, E., Monteiro, R., Guimaraes, P., Walters, B., Wiser, J., and Butte, A.J. (2014). ImmPort: disseminating data to the public for the future of immunology. Immunol Res *58*, 234-239.

Blum, A.L., and Langley, P. (1997). Selection of relevant features and examples in machine learning. Artif Intell *97*, 245-271.

Bodhidatta, L., Pitisuttithum, P., Chamnanchanant, S., Chang, K.T., Islam, D., Bussaratid, V., Venkatesan, M.M., Hale, T.L., and Mason, C.J. (2012). Establishment of a Shigella sonnei human challenge model in Thailand. Vaccine *30*, 7040-7045.

Boraschi, D., Aguado, M.T., Dutel, C., Goronzy, J., Louis, J., Grubeck-Loebenstein, B., Rappuoli, R., and Del Giudice, G. (2013). The gracefully aging immune system. Science translational medicine *5*, 185ps188.

Breiman, L. (1996). Bagging predictors. Machine Learning *24*, 123-140.

Breiman, L. (2001). Random forests. Mach Learn *45*, 5-32.

Brodin, P., Jojic, V., Gao, T., Bhattacharya, S., Angel, C.J., Furman, D., Shen-Orr, S., Dekker, C.L., Swan, G.E., Butte, A.J., Maecker, H.T., and Davis, M.M. (2015). Variation in the human immune system is largely driven by non-heritable influences. Cell *160*, 37-47.

Bucasas, K.L., Franco, L.M., Shaw, C.A., Bray, M.S., Wells, J.M., Nino, D., Arden, N., Quarles, J.M., Couch, R.B., and Belmont, J.W. (2011). Early Patterns of Gene Expression Correlate With the Humoral Immune Response to Influenza Vaccination in Humans. J Infect Dis *203*, 921-929.

Cao, R.G., Suarez, N.M., Obermoser, G., Lopez, S.M., Flano, E., Mertz, S.E., Albrecht, R.A., Garcia-Sastre, A., Mejias, A., Xu, H., Qin, H., Blankenship, D., Palucka, K., Pascual, V., and Ramilo, O. (2014). Differences in antibody responses between trivalent inactivated influenza vaccine and live attenuated influenza vaccine correlate with the kinetics and magnitude of interferon signaling in children. Journal of Infectious Diseases *210*, 224-233.

Carpenter, S., Ricci, E.P., Mercier, B.C., Moore, M.J., and Fitzgerald, K.A. (2014). Post-transcriptional regulation of gene expression in innate immunity. Nat Rev Immunol *14*, 361-376.

Chaussabel, D., and Baldwin, N. (2014). Democratizing systems immunology with modular transcriptional repertoire analyses. Nat Rev Immunol *14*, 271-280.

Chaussabel, D., Quinn, C., Shen, J., Patel, P., Glaser, C., Baldwin, N., Stichweh, D., Blankenship, D., Li, L., Munagala, I., Bennett, L., Allantaz, F., Mejias, A., Ardura, M., Kaizer, E., Monnet, L., Allman, W., Randall, H., Johnson, D., Lanier, A., Punaro, M., Wittkowski, K.M., White, P., Fay, J., Klintmalm, G., Ramilo, O., Palucka, A.K., Banchereau, J., and Pascual, V. (2008). A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. Immunity *29*, 150-164.

Chen, G.L., Lamirande, E.W., Jin, H., Kemble, G., and Subbarao, K. (2010). Safety, immunogencity, and efficacy of a cold-adapted A/Ann Arbor/6/60 (H2N2) vaccine in mice and ferrets. Virology *398*, 109-114.

Cho, S.-B., and Won, H. (2003). Machine Learning in DNA Microarray Analysis for Cancer Classification. Proceedings of the First Asia-Pacific Bioinformatics Conference on Bioinformatics.

Chuang, H.Y., Lee, E., Liu, Y.T., Lee, D., and Ideker, T. (2007). Network-based classification of breast cancer metastasis. Mol Syst Biol *3*, 140.

Chulay, J.D., Schneider, I., Cosgriff, T.M., Hoffman, S.L., Ballou, W.R., Quakyi, I.A., Carter, R., Trosper, J.H., and Hockmeyer, W.T. (1986). Malaria transmitted to humans by mosquitoes infected from cultured Plasmodium falciparum. Am J Trop Med Hyg *35*, 66-68.

Clausen, J., Vergeiner, B., Enk, M., Petzer, A.L., Gastl, G., and Gunsilius, E. (2003). Functional significance of the activation-associated receptors CD25 and CD69 on human NK-cells and NK-like T-cells. Immunobiology *207*, 85-93.

Cros, J., Cagnard, N., Woollard, K., Patey, N., Zhang, S.Y., Senechal, B., Puel, A., Biswas, S.K., Moshous, D., Picard, C., Jais, J.P., D'Cruz, D., Casanova, J.L., Trouillet, C., and Geissmann, F. (2010). Human CD14dim monocytes patrol and sense nucleic acids and viruses via TLR7 and TLR8 receptors. Immunity *33*, 375-386.

Dennis, E.A., Deems, R.A., Harkewicz, R., Quehenberger, O., Brown, H.A., Milne, S.B., Myers, D.S., Glass, C.K., Hardiman, G., Reichart, D., Merrill, A.H., Jr., Sullards, M.C., Wang, E., Murphy, R.C., Raetz, C.R., Garrett, T.A., Guan, Z., Ryan, A.C., Russell, D.W., McDonald, J.G., Thompson, B.M., Shaw, W.A., Sud, M., Zhao, Y., Gupta, S., Maurya, M.R., Fahy, E., and Subramaniam, S. (2010). A mouse macrophage lipidome. J Biol Chem *285*, 39976-39985.

Dreiseitl, S., and Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: a methodology review. Journal of Biomedical Informatics *35*, 352-359.

Dufour, J.H., Dziejman, M., Liu, M.T., Leung, J.H., Lane, T.E., and Luster, A.D. (2002). IFN-gamma-inducible protein 10 (IP-10; CXCL10)-deficient mice reveal a role for IP-10 in effector T cell generation and trafficking. J Immunol *168*, 3195-3204.

Duraisingham, S.S., Rouphael, N., Cavanagh, M.M., Nakaya, H.I., Goronzy, J.J., and Pulendran, B. (2013). Systems biology of vaccination in the elderly. Curr Top Microbiol Immunol *363*, 117-142.

Ehrenfeld, E., Modlin, J., and Chumakov, K. (2009). Future of polio vaccines. Expert Rev Vaccines *8*, 899-905.

Fahy, E., Sud, M., Cotter, D., and Subramaniam, S. (2007). LIPID MAPS online tools for lipid research. Nucleic Acids Res *35*, W606-612.

FDA (2007). Guidance for Industry: Clinical Data Needed to Support the Licensure of Pandemic Influenza Vaccines. (Rockville, Maryland).

Ferrington, D.A., and Gregerson, D.S. (2012). Immunoproteasomes: structure, function, and antigen presentation. Prog Mol Biol Transl Sci *109*, 75-112.

Filipowicz, W., Bhattacharyya, S.N., and Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nature Reviews: Genetics *9*, 102-114.

Franco, L.M., Bucasas, K.L., Wells, J.M., Nino, D., Wang, X., Zapata, G.E., Arden, N., Renwick, A., Yu, P., Quarles, J.M., Bray, M.S., Couch, R.B., Belmont, J.W., and Shaw, C.A. (2013). Integrative genomic analysis of the human immune response to influenza vaccination. Elife *2*, e00299.

Frasca, D., Diaz, A., Romero, M., Landin, A.M., and Blomberg, B.B. (2014). High TNF-alpha levels in resting B cells negatively correlate with their response. Exp Gerontol *54*, 116-122.

Frasca, D., Diaz, A., Romero, M., Landin, A.M., Phillips, M., Lechner, S.C., Ryan, J.G., and Blomberg, B.B. (2010). Intrinsic defects in B cell response to seasonal influenza vaccination in elderly humans. Vaccine *28*, 8077-8084.

Frasca, D., Diaz, A., Romero, M., Phillips, M., Mendez, N.V., Landin, A.M., and Blomberg, B.B. (2012). Unique biomarkers for B-cell function predict the serum response to pandemic H1N1 influenza vaccine. International Immunology *24*, 175-182.

Frey, S.E., Bernstein, D.I., Brady, R.C., Keitel, W.A., Sahly, H.E., Rouphael, N.G., Mulligan, M.J., Atmar, R.L., Edupuganti, S., Patel, S.M., Dickey, M., Graham, I., Anderson, E.L., Noah, D.L., Hill, H., Wolff, M., and Belshe, R.B. (2015). Phase II trial in adults of concurrent or sequential 2009 pandemic H1N1 and 2009-2010 seasonal trivalent influenza vaccinations. Vaccine *33*, 163-173.

Funderud, A., Aas-Hanssen, K., Aksaas, A.K., Hafte, T.T., Corthay, A., Munthe, L.A., Orstavik, S., and Skalhegg, B.S. (2009). Isoform-specific regulation of immune cell reactivity by the catalytic subunit of protein kinase A (PKA). Cell Signal *21*, 274-281.

Furman, D., Jojic, V., Kidd, B., Shen-Orr, S., Price, J., Jarrell, J., Tse, T., Huang, H., Lund, P., Maecker, H.T., Utz, P.J., Dekker, C.L., Koller, D., and Davis, M.M. (2013). Apoptosis and other immune biomarkers predict influenza vaccine responsiveness. Mol Syst Biol *9*, 659.

Garcia, D.M., Baek, D., Shin, C., Bell, G.W., Grimson, A., and Bartel, D.P. (2011). Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. Nature structural & molecular biology *18*, 1139-1146.

Gardner, E.M., Gonzalez, E.W., Nogusa, S., and Murasko, D.M. (2006). Age-related changes in the immune response to influenza vaccination in a racially diverse, healthy elderly population. Vaccine *24*, 1609-1614.

Gaucher, D., Therrien, R., Kettaf, N., Angermann, B.R., Boucher, G., Filali-Mouhim, A., Moser, J.M., Mehta, R.S., Drake, D.R., 3rd, Castro, E., Akondy, R., Rinfret, A., Yassine-Diab, B., Said, E.A., Chouikh, Y., Cameron, M.J., Clum, R., Kelvin, D., Somogyi, R., Greller, L.D., Balderas, R.S., Wilkinson, P., Pantaleo, G., Tartaglia, J., Haddad, E.K., and Sekaly, R.P. (2008). Yellow fever vaccine induces integrated multilineage and polyfunctional immune responses. J Exp Med *205*, 3119-3131.

Geissmann, F., Manz, M.G., Jung, S., Sieweke, M.H., Merad, M., and Ley, K. (2010). Development of monocytes, macrophages, and dendritic cells. Science *327*, 656-661.

Gelder, C.M., Lambkin, R., Hart, K.W., Fleming, D., Williams, O.M., Bunce, M., Welsh, K.I., Marshall, S.E., and Oxford, J. (2002). Associations between human leukocyte antigens and nonresponsiveness to influenza vaccine. J Infect Dis *185*, 114-117.

Goodwin, K., Viboud, C., and Simonsen, L. (2006). Antibody response to influenza vaccination in the elderly: A quantitative review. Vaccine *24*, 1159-1169.

Gordon, S., and Taylor, P.R. (2005). Monocyte and macrophage heterogeneity. Nature Reviews: Immunology *5*, 953-964.

Gorski, K.S., Waller, E.L., Bjornton-Severson, J., Hanten, J.A., Riter, C.L., Kieper, W.C., Gorden, K.B., Miller, J.S., Vasilakos, J.P., Tomai, M.A., and Alkan, S.S. (2006). Distinct indirect pathways govern human NK-cell activation by TLR-7 and TLR-8 agonists. International Immunology *18*, 1115-1126.

Gracias, D.T., Stelekati, E., Hope, J.L., Boesteanu, A.C., Doering, T.A., Norton, J., Mueller, Y.M., Fraietta, J.A., Wherry, E.J., Turner, M., and Katsikis, P.D. (2013). The microRNA miR-155 controls CD8(+) T cell responses by regulating interferon signaling. Nat Immunol *14*, 593-602.

Grage-Griebenow, E., Flad, H.D., and Ernst, M. (2001). Heterogeneity of human peripheral blood monocyte subsets. Journal of leukocyte biology *69*, 11-20.

Greenberg, D.P., Robertson, C.A., Noss, M.J., Blatter, M.M., Biedenbender, R., and Decker, M.D. (2013). Safety and immunogenicity of a quadrivalent inactivated influenza vaccine compared to licensed trivalent inactivated influenza vaccines in adults. Vaccine *31*, 770-776.

Halliley, J.L., Kyu, S., Kobie, J.J., Walsh, E.E., Falsey, A.R., Randall, T.D., Treanor, J., Feng, C., Sanz, I., and Lee, F.E. (2010). Peak frequencies of circulating human influenza-specific antibody secreting cells correlate with serum antibody response after immunization. Vaccine *28*, 3582-3587.

Hannedouche, S., Zhang, J., Yi, T., Shen, W., Nguyen, D., Pereira, J.P., Guerini, D., Baumgarten, B.U., Roggo, S., Wen, B., Knochenmuss, R., Noel, S., Gessier, F., Kelly, L.M., Vanek, M., Laurent, S., Preuss, I., Miault, C., Christen, I., Karuna, R., Li, W., Koo, D.I., Suply, T., Schmedt, C., Peters, E.C., Falchetto, R., Katopodis, A., Spanka, C., Roy, M.O., Detheux, M., Chen, Y.A., Schultz, P.G., Cho, C.Y., Seuwen, K., Cyster, J.G., and Sailer, A.W. (2011). Oxysterols direct immune cell migration via EBI2. Nature *475*, 524-527.

Haq, K., and McElhaney, J.E. (2014). Immunosenescence: Influenza vaccination and the elderly. Current Opinion in Immunology *29*, 38-42.

Harizi, H., and Gualde, N. (2005). The impact of eicosanoids on the crosstalk between innate and adaptive immunity: the key roles of dendritic cells. Tissue Antigens *65*, 507-514.

Harris, S.G., Padilla, J., Koumas, L., Ray, D., and Phipps, R.P. (2002). Prostaglandins as modulators of immunity. Trends Immunol *23*, 144-150.

Haynes, B.F., Gilbert, P.B., McElrath, M.J., Zolla-Pazner, S., Tomaras, G.D., Alam, S.M., Evans, D.T., Montefiori, D.C., Karnasuta, C., Sutthent, R., Liao, H.X., DeVico, A.L., Lewis, G.K., Williams, C., Pinter, A., Fong, Y., Janes, H., DeCamp, A., Huang, Y.D., Rao, M., Billings, E., Karasavvas, N., Robb, M.L., Ngauy, V., de Souza, M.S., Paris, R., Ferrari, G., Bailer, R.T., Soderberg, K.A., Andrews, C., Berman, P.W., Frahm, N., De Rosa, S.C., Alpert, M.D., Yates, N.L., Shen, X.Y., Koup, R.A., Pitisuttithum, P., Kaewkungwal, J., Nitayaphan, S., Rerks-Ngarm, S., Michael, N.L., and Kim, J.H. (2012). Immune-Correlates Analysis of an HIV-1 Vaccine Efficacy Trial. New Engl J Med *366*, 1275-1286.

Hoft, D.F. (2008). Tuberculosis vaccine development: goals, immunological design, and evaluation. Lancet *372*, 164-175.

Honda, K., Takaoka, A., and Taniguchi, T. (2006). Type I interferon [corrected] gene induction by the interferon regulatory factor family of transcription factors. Immunity *25*, 349-360.

Hooper, L.V., Littman, D.R., and Macpherson, A.J. (2012). Interactions between the microbiota and the immune system. Science *336*, 1268-1273.

Inza, I., Larranaga, P., Etxeberria, R., and Sierra, B. (2000). Feature Subset Selection by Bayesian network-based optimization. Artif Intell *123*, 157-184.

Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics *4*, 249-264.

Jackson, L.A., and Janoff, E.N. (2008). Pneumococcal vaccination of elderly adults: new paradigms for protection. Clin Infect Dis *47*, 1328-1338.

Jeffery, I.B., Higgins, D.G., and Culhane, A.C. (2006). Comparison and evaluation of methods for generating differentially expressed gene lists from microarray data. BMC Bioinform *7*, 359.

Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*, 118-127.

Johnstone, I.M., and Titterington, D.M. (2009). Statistical challenges of high-dimensional data. Philos Trans A Math Phys Eng Sci *367*, 4237-4253.

Kaczynski, J., Cook, T., and Urrutia, R. (2003). Sp1- and Kruppel-like transcription factors. Genome Biol *4*, 206.

Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., and Hirakawa, M. (2006). From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res *34*, D354-357.

Kang, I.S., Hong, M.S., Nolasco, H., Park, S.H., Dan, J.M., Choi, J.Y., and Craft, J. (2004). Age-associated change in the frequency of memory CD4(+) T cells impairs long term CD4(+) T cell responses to influenza vaccine. J Immunol *173*, 673-681.

Kau, A.L., Ahern, P.P., Griffin, N.W., Goodman, A.L., and Gordon, J.I. (2011). Human nutrition, the gut microbiome and the immune system. Nature *474*, 327-336.

Kauffman, R.E. (2000). Clinical trials in children: problems and pitfalls. Paediatr Drugs *2*, 411-418.

Kominsky, D.J., Campbell, E.L., and Colgan, S.P. (2010). Metabolic shifts in immunity and inflammation. J Immunol *184*, 4062-4068.

Kotsiantis, S.B., Zaharakis, I.D., and Pintelas, P.E. (2006). Machine learning: a review of classification and combining techniques. Artif Intell Rev *26*, 159-190.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.

Larranaga, P., Calvo, B., Santana, R., Bielza, C., Galdiano, J., Inza, I., Lozano, J.A., Armananzas, R., Santafe, G., Perez, A., and Robles, V. (2006). Machine learning in bioinformatics. Brief Bioinform *7*, 86-112.

Lauterbach, H., Ried, C., Epstein, A.L., Marconi, P., and Brocker, T. (2005). Reduced immune responses after vaccination with a recombinant herpes simplex virus type 1 vector in the presence of antiviral immunity. J Gen Virol *86*, 2401-2410.

Leek, J.T., Scharpf, R.B., Bravo, H.C., Simcha, D., Langmead, B., Johnson, W.E., Geman, D., Baggerly, K., and Irizarry, R.A. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. Nat Rev Genet *11*, 733-739.

Li, S., Rouphael, N., Duraisingham, S., Romero-Steiner, S., Presnell, S., Davis, C., Schmidt, D.S., Johnson, S.E., Milton, A., Rajam, G., Kasturi, S., Carlone, G.M., Quinn, C., Chaussabel, D., Palucka, A.K., Mulligan, M.J., Ahmed, R., Stephens, D.S., Nakaya, H.I., and Pulendran, B. (2014). Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. Nat Immunol *15*, 195-204.

Li, S.Z., Nakaya, H.I., Kazmin, D.A., Oh, J.Z., and Pulendran, B. (2013a). Systems biological approaches to measure and understand vaccine immunity in humans. Seminars in Immunology *25*, 209-218.

Li, Z.K., Nie, J.J., Li, J., and Zhuang, H. (2013b). The effect of HLA on immunological response to hepatitis B vaccine in healthy people: A meta-analysis. Vaccine *31*, 4355-4361.

Libby, P., Ridker, P.M., and Hansson, G.K. (2011). Progress and challenges in translating the biology of atherosclerosis. Nature *473*, 317-325.

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdottir, H., Tamayo, P., and Mesirov, J.P. (2011). Molecular signatures database (MSigDB) 3.0. Bioinformatics *27*, 1739-1740.

Liu, S.Y., Aliyari, R., Chikere, K., Li, G., Marsden, M.D., Smith, J.K., Pernet, O., Guo, H., Nusbaum, R., Zack, J.A., Freiberg, A.N., Su, L., Lee, B., and Cheng,

G. (2013). Interferon-inducible cholesterol-25-hydroxylase broadly inhibits viral entry by production of 25-hydroxycholesterol. Immunity *38*, 92-105.

Lodish, H.F., Zhou, B., Liu, G., and Chen, C.Z. (2008). Micromanagement of the immune system by microRNAs. Nature Reviews: Immunology *8*, 120-130.

Long, E.O., Kim, H.S., Liu, D., Peterson, M.E., and Rajagopalan, S. (2013). Controlling natural killer cell responses: integration of signals for activation and inhibition. Annual review of immunology *31*, 227-258.

Lu, Y.C., Yeh, W.C., and Ohashi, P.S. (2008). LPS/TLR4 signal transduction pathway. Cytokine *42*, 145-151.

Lutz, W., Sanderson, W., and Scherbov, S. (2008). The coming acceleration of global population ageing. Nature *451*, 716-719.

Mandelboim, O., Lieberman, N., Lev, M., Paul, L., Arnon, T.I., Bushkin, Y., Davis, D.M., Strominger, J.L., Yewdell, J.W., and Porgador, A. (2001). Recognition of haemagglutinins on virus-infected cells by NKp46 activates lysis by human NK cells. Nature *409*, 1055-1060.

Marwick, C. (1998). Volunteers in typhoid infection study will aid future vaccine development. JAMA *279*, 1423-1424.

McElhaney, J.E., Xie, D., Hager, W.D., Barry, M.B., Wang, Y., Kleppinger, A., Ewen, C., Kane, K.P., and Bleackley, R.C. (2006). T cell responses are better correlates of vaccine protection in the elderly. J Immunol *176*, 6333-6339.

Miller, M.S., and Palese, P. (2014). Peering into the crystal ball: influenza pandemics and vaccine efficacy. Cell *157*, 294-299.

Mohanty, S., Joshi, S.R., Ueda, I., Wilson, J., Blevins, T.P., Siconolfi, B., Meng, H., Devine, L., Raddassi, K., Tsang, S., Belshe, R.B., Hafler, D.A., Kaech, S.M., Kleinstein, S.H., Trentalange, M., Allore, H.G., and Shaw, A.C. (2015). Prolonged proinflammatory cytokine production in monocytes modulated by interleukin 10 after influenza vaccination in older adults. Journal of Infectious Diseases *211*, 1174-1184.

Moss, A.J., Gaughran, F.P., Karasu, A., Gilbert, A.S., Mann, A.J., Gelder, C.M., Oxford, J.S., Stephens, H.A., and Lambkin-Williams, R. (2013). Correlation between Human Leukocyte Antigen Class II Alleles and HAI Titers Detected Post-Influenza Vaccination. PLOS ONE *8*.

Nakaya, H.I., Li, S., and Pulendran, B. (2012). Systems vaccinology: learning to compute the behavior of vaccine induced immunity. Wiley Interdiscip Rev Syst Biol Med *4*, 193-205.

Nakaya, H.I., and Pulendran, B. (2012). Systems vaccinology: its promise and challenge for HIV vaccine development. Curr Opin HIV AIDS *7*, 24-31.

Nakaya, H.I., Wrammert, J., Lee, E.K., Racioppi, L., Marie-Kunze, S., Haining, W.N., Means, A.R., Kasturi, S.P., Khan, N., Li, G.M., McCausland, M., Kanchan, V., Kokko, K.E., Li, S., Elbein, R., Mehta, A.K., Aderem, A., Subbarao, K., Ahmed, R., and Pulendran, B. (2011). Systems biology of vaccination for seasonal influenza in humans. Nat Immunol *12*, 786-795.

O'Connor, D., and Pollard, A.J. (2013). Characterizing Vaccine Responses Using Host Genomic and Transcriptomic Analysis. Clin Infect Dis *57*, 860-869.

Obermoser, G., Presnell, S., Domico, K., Xu, H., Wang, Y., Anguiano, E., Thompson-Snipes, L., Ranganathan, R., Zeitner, B., Bjork, A., Anderson, D., Speake, C., Ruchaud, E., Skinner, J., Alsina, L., Sharma, M., Dutartre, H., Cepika, A., Israelsson, E., Nguyen, P., Nguyen, Q.A., Harrod, A.C., Zurawski, S.M., Pascual, V., Ueno, H., Nepom, G.T., Quinn, C., Blankenship, D., Palucka, K., Banchereau, J., and Chaussabel, D. (2013). Systems scale interactive exploration reveals quantitative and qualitative differences in response to influenza and pneumococcal vaccines. Immunity *38*, 831-844.

Oh, J.Z., Ravindran, R., Chassaing, B., Carvalho, F.A., Maddur, M.S., Bower, M., Hakimpour, P., Gill, K.P., Nakaya, H.I., Yarovinsky, F., Sartor, R.B., Gewirtz, A.T., and Pulendran, B. (2014). TLR5-mediated sensing of gut microbiota is necessary for antibody responses to seasonal influenza vaccination. Immunity *41*, 478-492.

Olotu, A., Clement, F., Jongert, E., Vekemans, J., Njuguna, P., Ndungu, F.M., Marsh, K., Leroux-Roels, G., and Bejon, P. (2014). Avidity of Anti-

Circumsporozoite Antibodies following Vaccination with RTS,S/AS01(E) in Young Children. PLOS ONE *9*.

Opitz, D., and Maclin, R. (1999). Popular Ensemble Methods: An Empirical Study. J Artif Intell Res *11*, 169-189.

Orru, V., Steri, M., Sole, G., Sidore, C., Virdis, F., Dei, M., Lai, S., Zoledziewska, M., Busonero, F., Mulas, A., Floris, M., Mentzen, W.I., Urru, S.A., Olla, S., Marongiu, M., Piras, M.G., Lobina, M., Maschio, A., Pitzalis, M., Urru, M.F., Marcelli, M., Cusano, R., Deidda, F., Serra, V., Oppo, M., Pilu, R., Reinier, F., Berutti, R., Pireddu, L., Zara, I., Porcu, E., Kwong, A., Brennan, C., Tarrier, B., Lyons, R., Kang, H.M., Uzzau, S., Atzeni, R., Valentini, M., Firinu, D., Leoni, L., Rotta, G., Naitza, S., Angius, A., Congia, M., Whalen, M.B., Jones, C.M., Schlessinger, D., Abecasis, G.R., Fiorillo, E., Sanna, S., and Cucca, F. (2013). Genetic variants regulating immune cell levels in health and disease. Cell *155*, 242-256.

Oviedo-Orta, E., Li, C.K., and Rappuoli, R. (2013). Perspectives on vaccine development for the elderly. Current opinion in immunology *25*, 529-534.

Ovsyannikova, I.G., Pankratz, V.S., Salk, H.M., Kennedy, R.B., and Poland, G.A. (2014). HLA alleles associated with the adaptive immune response to smallpox vaccine: a replication study. Hum Genet *133*, 1083-1092.

Ovsyannikova, I.G., Vierkant, R.A., Pankratz, V.S., Jacobson, R.M., and Poland, G.A. (2011). Human Leukocyte Antigen Genotypes in the Genetic Control of Adaptive Immune Responses to Smallpox Vaccine. J Infect Dis *203*, 1546-1555.

Padgett, D.A., and Glaser, R. (2003). How stress influences the immune response. Trends Immunol *24*, 444-448.

Pantaleo, G., and Koup, R.A. (2004). Correlates of immune protection in HIV-1 infection: what we know, what we don't know, what we should know. Nat Med *10*, 806-810.

Park, S.Y., Shin, H.M., and Han, T.H. (2002). Synergistic interaction of MEF2D and Sp1 in activation of the CD14 promoter. Mol Immunol *39*, 25-30.

Patil, A., Kumagai, Y., Liang, K.C., Suzuki, Y., and Nakai, K. (2013). Linking transcriptional changes over time in stimulated dendritic cells to identify gene networks activated during the innate immune response. PLoS Comput Biol *9*, e1003323.

Patti, G.J., Yanes, O., and Siuzdak, G. (2012). Innovation: Metabolomics: the apogee of the omics trilogy. Nat Rev Mol Cell Biol *13*, 263-269.

Pawelec, G., Goldeck, D., and Derhovanessian, E. (2014). Inflammation, ageing and chronic disease. Current Opinion in Immunology *29*, 23-28.

Pianetti, S., Arsura, M., Romieu-Mourez, R., Coffey, R.J., and Sonenshein, G.E. (2001). Her-2/neu overexpression induces NF-kappaB via a PI3-kinase/Akt pathway involving calpain-mediated degradation of IkappaB-alpha that can be inhibited by the tumor suppressor PTEN. Oncogene *20*, 1287-1299.

Pica, N., and Palese, P. (2013). Toward a universal influenza virus vaccine: prospects and challenges. Annual review of medicine *64*, 189-202.

Pichichero, M.E. (2009). Booster Vaccinations: Can Immunologic Memory Outpace Disease Pathogenesis? Pediatrics *124*, 1633-1641.

Plotkin, S.A. (2008). Vaccines: correlates of vaccine-induced immunity. Clin Infect Dis *47*, 401-409.

Poland, G.A. (1999). Immunogenetic mechanisms of antibody response to measles vaccine: the role of the HLA genes. Vaccine *17*, 1719-1725.

Poli, A., Michel, T., Theresine, M., Andres, E., Hentges, F., and Zimmer, J. (2009). CD56bright natural killer (NK) cells: an important NK cell subset. Immunology *126*, 458-465.

Pulendran, B. (2009). Learning immunology from the yellow fever vaccine: innate immunity to systems vaccinology. Nat Rev Immunol *9*, 741-747.

Pulendran, B. (2014). Systems vaccinology: probing humanity's diverse immune systems with vaccines. Proc Natl Acad Sci U S A *111*, 12300-12306.

Pulendran, B., Li, S., and Nakaya, H.I. (2010). Systems vaccinology. Immunity *33*, 516-529.

Pulendran, B., Miller, J., Querec, T.D., Akondy, R., Moseley, N., Laur, O., Glidewell, J., Monson, N., Zhu, T., Zhu, H., Staprans, S., Lee, D., Brinton, M.A., Perelygin, A.A., Vellozzi, C., Brachman, P., Jr., Lalor, S., Teuwen, D., Eidex, R.B., Cetron, M., Priddy, F., del Rio, C., Altman, J., and Ahmed, R. (2008). Case of yellow fever vaccine--associated viscerotropic disease with prolonged viremia, robust adaptive immune responses, and polymorphisms in CCR5 and RANTES genes. J Infect Dis *198*, 500-507.

Pulendran, B., Oh, J.Z., Nakaya, H.I., Ravindran, R., and Kazmin, D.A. (2013). Immunity to viruses: learning from successful human vaccines. Immunol Rev *255*, 243-255.

Querec, T.D., Akondy, R.S., Lee, E.K., Cao, W., Nakaya, H.I., Teuwen, D., Pirani, A., Gernert, K., Deng, J., Marzolf, B., Kennedy, K., Wu, H., Bennouna, S., Oluoch, H., Miller, J., Vencio, R.Z., Mulligan, M., Aderem, A., Ahmed, R., and Pulendran, B. (2009). Systems biology approach predicts immunogenicity of the yellow fever vaccine in humans. Nat Immunol *10*, 116-125.

Radbruch, A., Muehlinghaus, G., Luger, E.O., Inamine, A., Smith, K.G., Dorner, T., and Hiepe, F. (2006). Competence and competition: the challenge of becoming a long-lived plasma cell. Nat Rev Immunol *6*, 741-750.

Raetz, C.R., Garrett, T.A., Reynolds, C.M., Shaw, W.A., Moore, J.D., Smith, D.C., Jr., Ribeiro, A.A., Murphy, R.C., Ulevitch, R.J., Fearns, C., Reichart, D., Glass, C.K., Benner, C., Subramaniam, S., Harkewicz, R., Bowers-Gentry, R.C., Buczynski, M.W., Cooper, J.A., Deems, R.A., and Dennis, E.A. (2006). Kdo2-Lipid A of Escherichia coli, a defined endotoxin that activates macrophages via TLR-4. J Lipid Res *47*, 1097-1111.

Ravindran, R., Khan, N., Nakaya, H.I., Li, S., Loebbermann, J., Maddur, M.S., Park, Y., Jones, D.P., Chappert, P., Davoust, J., Weiss, D.S., Virgin, H.W., Ron, D., and Pulendran, B. (2014). Vaccine activation of the nutrient sensor GCN2 in dendritic cells enhances antigen presentation. Science *343*, 313-317.

Reif, D.M., Motsinger-Reif, A.A., McKinney, B.A., Rock, M.T., Crowe, J.E., Jr., and Moore, J.H. (2009). Integrated analysis of genetic and proteomic data identifies biomarkers associated with adverse events following smallpox vaccination. Genes Immun *10*, 112-119.

Reyes-Sandoval, A., Pearson, F.E., Todryk, S., and Ewer, K. (2009). Potency assays for novel T-cell-inducing vaccines against malaria. Curr Opin Mol Ther *11*, 72-80.

Samuel, C.E. (2001). Antiviral actions of interferons. Clinical microbiology reviews *14*, 778-809.

Sasaki, S., He, X.S., Holmes, T.H., Dekker, C.L., Kemble, G.W., Arvin, A.M., and Greenberg, H.B. (2008). Influence of prior influenza vaccination on antibody and B-cell responses. PLOS ONE *3*, e2975.

Sasaki, S., Jaimes, M.C., Holmes, T.H., Dekker, C.L., Mahmood, K., Kemble, G.W., Arvin, A.M., and Greenberg, H.B. (2007). Comparison of the influenza virus-specific effector and memory B-cell responses to immunization of children and adults with live attenuated or inactivated influenza virus vaccines. J Virol *81*, 215-228.

Sasaki, S., Sullivan, M., Narvaez, C.F., Holmes, T.H., Furman, D., Zheng, N.Y., Nishtala, M., Wrammert, J., Smith, K., James, J.A., Dekker, C.L., Davis, M.M., Wilson, P.C., Greenberg, H.B., and He, X.S. (2011). Limited efficacy of inactivated influenza vaccine in elderly individuals is associated with decreased production of vaccine-specific antibodies. J Clin Invest *121*, 3109-3119.

Seibert, K., and Masferrer, J.L. (1994). Role of inducible cyclooxygenase (COX-2) in inflammation. Receptor *4*, 17-23.

Seidman, J.C., Richard, S.A., Viboud, C., and Miller, M.A. (2012). Quantitative review of antibody response to inactivated seasonal influenza vaccines. Influenza and other respiratory viruses *6*, 52-62.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software

environment for integrated models of biomolecular interaction networks. Genome Res *13*, 2498-2504.

Shih, V.F., Davis-Turak, J., Macal, M., Huang, J.Q., Ponomarenko, J., Kearns, J.D., Yu, T., Fagerlund, R., Asagiri, M., Zuniga, E.I., and Hoffmann, A. (2012). Control of RelB during dendritic cell activation integrates canonical and noncanonical NF-kappaB pathways. Nat Immunol *13*, 1162-1170.

Simonsen, L., Reichert, T.A., Viboud, C., Blackwelder, W.C., Taylor, R.J., and Miller, M.A. (2005). Impact of influenza vaccination on seasonal mortality in the US elderly population. Archives of Internal Medicine *165*, 265-272.

Solana, R., Campos, C., Pera, A., and Tarazona, R. (2014). Shaping of NK cell subsets by aging. Current opinion in immunology *29*, 56-61.

Spann, N.J., and Glass, C.K. (2013). Sterols and oxysterols in immune cell function. Nat Immunol *14*, 893-900.

Spitz, F., and Furlong, E.E. (2012). Transcription factors: from enhancer binding to developmental control. Nat Rev Genet *13*, 613-626.

Stephenson, I., Das, R.G., Wood, J.M., and Katz, J.M. (2007). Comparison of neutralising antibody assays for detection of antibody to influenza A/H3N2 viruses: An international collaborative study. Vaccine *25*, 4056-4063.

Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. P Natl Acad Sci USA *100*, 9440-9445.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A *102*, 15545-15550.

Sullivan, S.J., Jacobson, R., and Poland, G.A. (2010). Advances in the vaccination of the elderly against influenza: role of a high-dose vaccine. Expert Review of Vaccines *9*, 1127-1133.

Takeda, K., and Akira, S. (2004). TLR signaling pathways. Semin Immunol *16*, 3-9.

Thompson, W.W., Shay, D.K., Weintraub, E., Brammer, L., Cox, N., Anderson, L.J., and Fukuda, K. (2003). Mortality associated with influenza and respiratory syncytial virus in the United States. Journal of the American Medical Association *289*, 179-186.

Tsang, J.S., Schwartzberg, P.L., Kotliarov, Y., Biancotto, A., Xie, Z., Germain, R.N., Wang, E., Olnes, M.J., Narayanan, M., Golding, H., Moir, S., Dickler, H.B., Perl, S., Cheung, F., Baylor, H.C., and Consortium, C.H.I. (2014). Global analyses of human immune variation reveal baseline predictors of postvaccination responses. Cell *157*, 499-513.

Vahey, M.T., Wang, Z., Kester, K.E., Cummings, J., Heppner, D.G., Jr., Nau, M.E., Ofori-Anyinam, O., Cohen, J., Coche, T., Ballou, W.R., and Ockenhouse, C.F. (2010). Expression of genes associated with immunoproteasome processing of major histocompatibility complex peptides is indicative of protection with adjuvanted RTS,S malaria vaccine. J Infect Dis *201*, 580-589.

Vastrik, I., D'Eustachio, P., Schmidt, E., Gopinath, G., Croft, D., de Bono, B., Gillespie, M., Jassal, B., Lewis, S., Matthews, L., Wu, G., Birney, E., and Stein, L. (2009). Reactome: a knowledge base of biologic pathways and processes (vol 8, pg 39, 2007). Genome Biol *10*.

Waddington, C.S., Darton, T.C., Jones, C., Haworth, K., Peters, A., John, T., Thompson, B.A.V., Kerridge, S.A., Kingsley, R.A., Zhou, L.Q., Holt, K.E., Yu, L.M., Lockhart, S., Farrar, J.J., Sztein, M.B., Dougan, G., Angus, B., Levine, M.M., and Pollard, A.J. (2014). An Outpatient, Ambulant-Design, Controlled Human Infection Model Using Escalating Doses of Salmonella Typhi Challenge Delivered in Sodium Bicarbonate Solution. Clin Infect Dis *58*, 1230-1240.

Wagar, L.E., Gentleman, B., Pircher, H., McElhaney, J.E., and Watts, T.H. (2011). 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*.

Walzer, T., Dalod, M., Robbins, S.H., Zitvogel, L., and Vivier, E. (2005). Natural-killer cells and dendritic cells: "l'union fait la force". Blood *106*, 2252-2258.

Wang, H.B., Wang, J.T., Zhang, L., Geng, Z.H., Xu, W.L., Xu, T., Huo, Y., Zhu, X., Plow, E.F., Chen, M., and Geng, J.G. (2007). P-selectin primes leukocyte integrin activation during inflammation. Nat Immunol *8*, 882-892.

Wei, T. (2012). Package 'corrplot' – Visualization of a correlation matrix. (cran.r-project.org.).

WHO (2014). WHO Influenza Fact Sheet.

Wingender, E. (2008). The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation. Brief Bioinform *9*, 326-332.

Woods, J.A., Vieira, V.J., and Keylock, K.T. (2006). Exercise, inflammation, and innate immunity. Neurol Clin *24*, 585-599.

Yu, T., Park, Y., Johnson, J.M., and Jones, D.P. (2009). apLCMS--adaptive processing of high-resolution LC/MS data. Bioinformatics *25*, 1930-1936.

Zak, D.E., Andersen-Nissen, E., Peterson, E.R., Sato, A., Hamilton, M.K., Borgerding, J., Krishnamurty, A.T., Chang, J.T., Adams, D.J., Hensley, T.R., Salter, A.I., Morgan, C.A., Duerr, A.C., De Rosa, S.C., Aderem, A., and McElrath, M.J. (2012). Merck Ad5/HIV induces broad innate immune activation that predicts CD8(+) T-cell responses but is attenuated by preexisting Ad5 immunity. Proc Natl Acad Sci U S A *109*, E3503-3512.

Zanetti, M., Castiglioni, P., Schoenberger, S., and Gerloni, M. (2003). The role of relB in regulating the adaptive immune response. Ann N Y Acad Sci *987*, 249-257.

Zhu, J., Yamane, H., and Paul, W.E. (2010). Differentiation of effector CD4 T cell populations (*). Annu Rev Immunol *28*, 445-489.