Transcriptional atlas of the human immune response to 13 vaccines reveals a common predictor of vaccine-induced antibody responses

Thomas Hagan [1,2], Bram Gerritsen [3], Lewis E. Tomalin [4], Slim Fourati [5], Matthew P. Mulè[6,7], Daniel G. Chawla [3], Dmitri Rychkov[8], Evan Henrich[9], Helen E. R. Miller[9], Joann Diray-Arce[10,11], Patrick Dunn [12], Audrey Lee[13], The Human Immunology Project Consortium (HIPC)*, Ofer Levy [10,11,14], Raphael Gottardo [9,15,16], Minne M. Sarwal[8], John S. Tsang[6], Mayte Suárez-Fariñas[4], Rafick-Pierre Sékaly[5], Steven H. Kleinstein [3,25] ✉ and Bali Pulendran [13,25] ✉

Systems vaccinology has defined molecular signatures and mechanisms of immunity to vaccination. However, comparative analysis of immunity to different vaccines is lacking. We integrated transcriptional data of over 3,000 samples, from 820 adults across 28 studies of 13 vaccines and analyzed vaccination-induced signatures of antibody responses. Most vaccines induced signatures of innate immunity and plasmablasts at days 1 and 7, respectively, after vaccination. However, the yellow fever vaccine induced an early transient signature of T and B cell activation at day 1, followed by delayed antiviral/interferon and plasmablast signatures that peaked at days 7 and 14–21, respectively. Thus, there was no evidence for a 'universal signature' that predicted antibody response to all vaccines. However, accounting for the asynchronous nature of responses, we defined a time-adjusted signature that predicted antibody responses across vaccines. These results provide a transcriptional atlas of immunity to vaccination and define a common, time-adjusted signature of antibody responses.

Systems vaccinology uses high-throughput omics measurements together with systems-based analysis approaches to better understand immune responses to vaccination[1,2]. The recent growth of this field, which began with initial studies of yellow fever[3,4] and seasonal influenza[5,6] vaccines, has rapidly expanded to include studies profiling responses to a range of vaccines and vaccine platforms, including those targeting diverse pathogens and age groups[7–20]. These studies have led to important discoveries such as the role for the nutrient sensor general control nonderepressible 2 (GCN2) in enhancing antigen presentation

during responses to yellow fever vaccination[21], as well as the impact of the gut microbiota in promoting antibody responses to inactivated influenza vaccination[22,23]. However, apart from a few studies[8–10], thus far the vast majority have examined immune responses to a single vaccine, hindering the ability to contextualize the findings and understand how differences in vaccine formulations can impact immunogenicity.

Another important outcome of such studies has been the identification of early transcriptional signatures predictive of immune response quality such as subsequent antibody[3,5,18] or antigen-specific
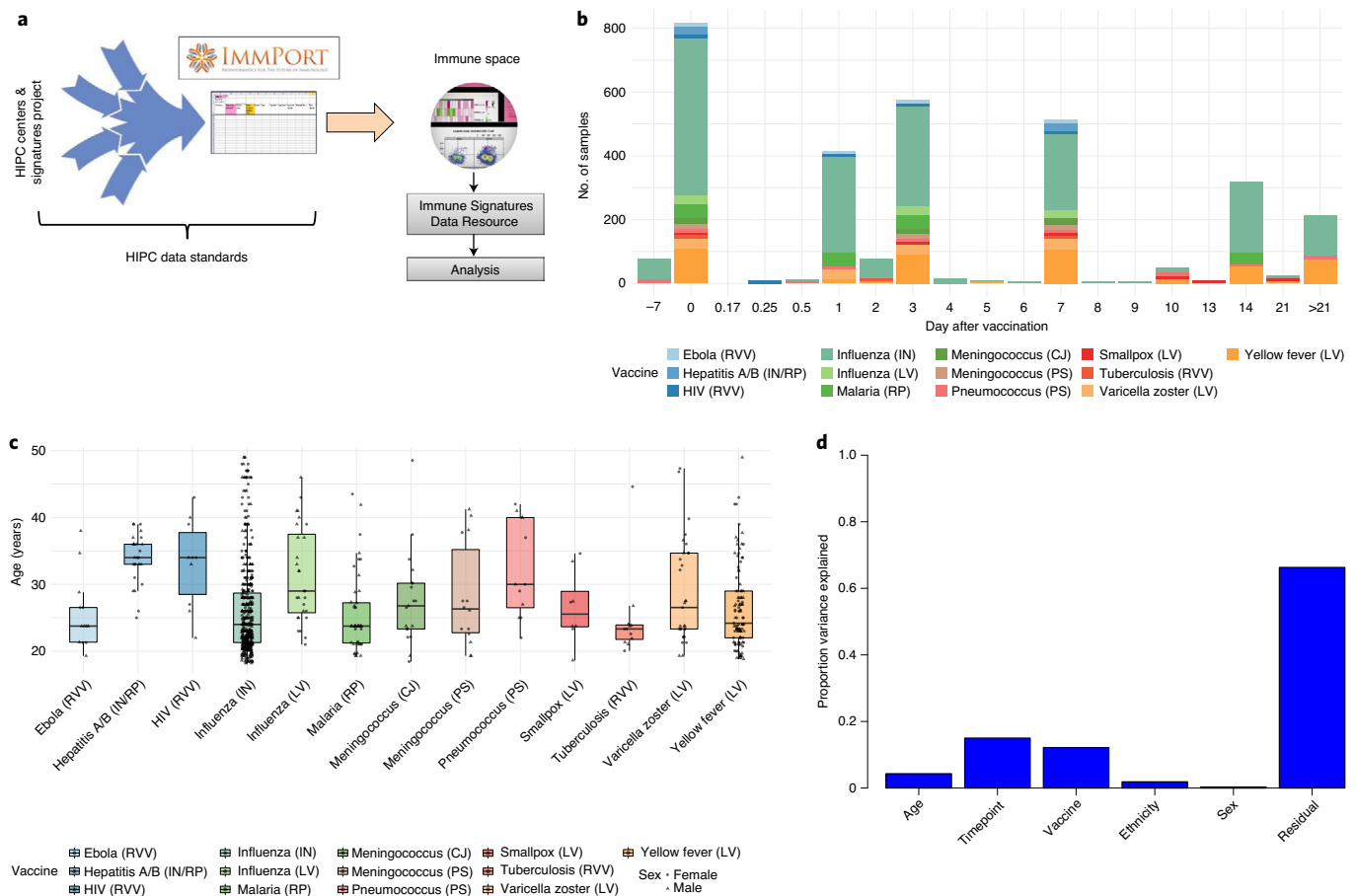
Fig. 1 | An integrated database of transcriptional responses to vaccination. **a**, Workflow for collection, curation and standardization of datasets in the Immune Signatures Data Resource. **b**, Histogram of the number of samples included per vaccine at each time point. Day 0 represents day of vaccination. **c**, Age distribution of participants in the Immune Signatures Data Resource by vaccine. The shape of points denotes the participant's inferred sex based on Y-chromosome-specific gene expression. For participants with missing age data, ages were estimated using RAPToR[43]. For the boxplot, the center line represents the median, box limits represent upper and lower quartiles, and whiskers indicate

1.5 times the interquartile range. Ebola (RVV), n = 13; hepatitis A/B (IN/RP), n = 26; HIV (RVV), n = 10; influenza (IN), n = 496; influenza (LA), n = 28; malaria (RP), n = 44; meningococcus (CJ), n = 19; meningococcus (PS), n = 14; pneumococcus (PS), n = 12; smallpox (LA), n = 8; tuberculosis (RVV), n = 12; varicella zoster (LA), n = 31; yellow fever (LA), n = 107. **d**, Bar plot representing the proportion of variance in post-vaccination transcriptional responses that can be attributed to clinical (age, sex, ethnicity) and experimental (time after vaccination, vaccine) variables via principal variance component analysis. The residual represents the proportion of the variance that could not be explained by any of the included variables.

T cell[3,7,14] responses. These findings may enable more rapid and personalized evaluation of vaccine efficacy and development of improved next-generation vaccines. Yet again, a current limitation is that the identified predictive signatures thus far have been described in the context of responses to a single vaccine, and the extent to which predictors of immune response quality are conserved across vaccines is unclear[24,25]. We previously sought to address the question of whether there was a 'universal signature' that could be used to predict antibody responses to any vaccine by analyzing the transcriptional response to five different human vaccines. Our analysis revealed distinct transcriptional signatures of antibody responses to different classes of vaccines, and provided key insights into primary viral, protein recall and anti-polysaccharide responses[8], yet failed to identify a universal signature of vaccination.

Here we leverage the Immune Signatures Data Resource[26], a curated database of publicly available datasets containing transcriptional and immune response profiling of peripheral blood following vaccination in humans, to perform a comparative analysis of transcriptional responses from 820 healthy young adults across 13 different vaccines. We find that while a common transcriptional program is shared across many vaccines, there is considerable heterogeneity especially

in the kinetics of immune responses. In particular, the live-attenuated yellow fever vaccine induces a unique transcriptional response, with a surprisingly early upregulation of B and T cell modules within a day of vaccination, and a delayed induction of innate responses, including antiviral and interferon signaling, peaking at 10–14 d following vaccination. Furthermore, in an analysis of predictive signatures of antibody responses across vaccines, adjusting for time of peak expression enabled a gene module associated with plasma cells and immunoglobulins to consistently predict antibody responses across vaccines, demonstrating the importance of accounting for immune response kinetics in the development of universal predictors of response quality. Together, these findings highlight the spectrum of immune responses across vaccines and serve as a basis for future studies to understand the mechanisms underlying variation in immune responses across vaccines and inform future vaccine development.

Results

An integrated database of transcriptional responses to vaccination

As part of an effort to enable comparative studies and benchmarking of human vaccine responses, we curated a database of transcriptomic

Table 1 | Table of included vaccines

Vaccine	Pathogen	Vaccine type	Adjuvant/vector
rVSV-ZEBOV	Ebola	RVV	VSV
Hepatitis A/B vaccine[a]	Hepatitis A/B	IN/RP	None
MRKAd5/HIV	HIV	RVV	Adenovirus
Influenza vaccine	Influenza	IN	None
Influenza vaccine live	Influenza	LV	Live-attenuated influenza virus
RTS,S	Malaria	RP	AS01/AS02
MCV4	Meningococcus	CJ	None
MPSV4	Meningococcus	PS	None
PPSV-23	Pneumococcus	PS	None
Smallpox vaccine	Smallpox	LV	Vaccinia virus
MVA85A	Tuberculosis	RVV	Vaccinia virus
Zoster vaccine live	Varicella zoster	LV	Varicella zoster virus
Yellow fever vaccine	Yellow fever	LV	Yellow fever 17D

CJ, conjugate; IN, inactivated; LV, live virus; PS, polysaccharide; RP, recombinant protein; RVV, recombinant viral vector; VSV, vesicular stomatitis virus. [a]Participants receiving hepatitis A/B vaccine also received tetanus/diphtheria and cholera vaccines at the same time.

responses of 820 healthy adults (18–50 years old) across 13 different vaccines from previously published datasets. These datasets were compiled into ImmPort, a National Institutes of Health (NIH)-funded repository for immunological data[27], and uploaded to ImmuneSpace (https://www.immunespace.org/) for centralized quality control (QC) and processing (Fig. 1a). This combined database, named the Immune Signatures Data Resource[26], includes responses to a broad range of vaccines, including live viruses (for example, yellow fever, smallpox and influenza vaccines), recombinant viral vectors (for example, Ebola and HIV vaccines), inactivated viruses (for example, seasonal influenza vaccine), glycoconjugate vaccines (for example, pneumococcal and meningococcal vaccines; Table 1 and Supplementary Table 1). It also contains samples spanning multiple response time points, ranging from a few hours to more than 3 weeks after vaccination (Fig. 1b). Included participants in our initial analysis were restricted to 18–50 years old, and there were similar age and sex distributions across vaccines (Fig. 1c). For analysis, all post-vaccination samples were normalized by pairwise fold-change calculation with their matched pre-vaccination samples. Principal variance component analysis[28] revealed that demographic features such as age and sex had relatively small contributions to variation in responses. In contrast, the post-vaccination time point of the sample explained 15% of the variance in the data, suggesting that there are shared kinetics of immune responses across vaccines (Fig. 1d).

Common and unique transcriptional responses across vaccines

To examine the overlap in responses across vaccines, we identified differentially expressed genes after vaccination relative to the pre-vaccination baseline as well as differential expression of blood transcriptional modules (BTMs), a set of gene modules developed through large-scale network integration of publicly available human blood transcriptomes[10]. There was much less overlap at a gene level (Extended Data Fig. 1a) than at a module level (Extended Data Fig. 1b), where a majority of differentially expressed modules were shared across four or more vaccines. Based on temporal expression patterns, the most commonly induced modules clustered into four groups (Fig. 2a). Cluster 1, upregulated at days 1 and 3 after vaccination, represented BTMs related to innate responses and included modules associated

with Toll-like receptor (TLR) and inflammatory signaling, antigen presentation and monocyte signatures. Cluster 2 contained multiple natural killer cell modules and was significantly downregulated on day 1 (Extended Data Fig. 1c). Finally, clusters 3 and 4 generally peaked in activity on day 7 and reflected plasma cell and cell cycle signatures, respectively, corresponding with expansion of antibody-producing plasmablasts. The 'innate' cluster 1 was most prominently induced in vaccines containing a live viral vector (Ebola, HIV), or an adjuvant (malaria; Fig. 2b). Meanwhile, the plasma cell signature in cluster 3 was strongly increased in the polysaccharide pneumococcal vaccine and the conjugate meningococcal vaccine (Fig. 2c).

We next analyzed how differentially expressed modules were shared across vaccine responses (Fig. 3). In agreement with the previous analysis of common responses (Fig. 2), the response to most vaccines on days 1 (Fig. 3a) and 7 (Fig. 3c) reflected innate and plasma cell/cell cycle responses, respectively, while the day 3 response (Fig. 3b) appeared as an intermediate between these states, with both innate and cell cycle signatures present. However, such responses were not universally shared across all of the vaccines. In particular, the early innate and antiviral responses common to most vaccines on day 1 were not observed in the varicella zoster and yellow fever vaccine responses. While these signatures appeared at later time points (days 3 and 7) in yellow fever vaccine responses, they were not observed at all following varicella zoster viral vaccine. Additionally, the day 7 cell cycle signature was not observed following smallpox, varicella zoster and polysaccharide meningococcal vaccines. Notably, this signature was observed in the case of the meningococcal conjugate vaccine, where the bacterial polysaccharides have been conjugated to a diphtheria toxoid protein to induce memory and helper T cell responses[29]. Since diphtheria toxoid protein is used in other vaccines such as the *Haemophilus influenzae* type B vaccine[30], the cell cycle signature observed at day 7 likely reflects the plasmablast response of the recall response to diphtheria toxoid, consistent with our previous study[10].

Early adaptive and delayed innate signatures of yellow fever vaccine

At the gene level, responses were highly correlated across most vaccines on day 1 after vaccination (Fig. 4a) but became more divergent at later time points (Extended Data Fig. 2a,b). On day 1, Ebola, inactivated influenza, HIV and malaria vaccines exhibited the strongest similarity (Fig. 4a and Extended Data Fig. 2c). However, the yellow fever vaccine YF-17D induced a very distinct response that had little or even negative correlation with responses to all other vaccines, including other live viral vaccines such as varicella zoster virus, HIV and Ebola (Fig. 4a and Extended Data Fig. 2d). The innate pathways that were upregulated in other vaccine responses were, in fact, downregulated in response to yellow fever vaccine on day 1 (Fig. 4b). Instead, YF-17D induced early expression of multiple B and T cell modules. Analysis using the xCell deconvolution algorithm[31] detected an increase in estimated frequencies of total B cells (Extended Data Fig. 2e,f), suggesting that this signature may reflect an induction of adaptive cells into the periphery.

Another surprising feature of the yellow fever vaccine response was the relatively late expression of antiviral and interferon pathways, whose expression started to be observed on day 3 and peaks on day 7 (Fig. 4c,d). While these modules were also upregulated at this time point in Ebola vaccine responses, their expression waned rapidly following a robust early induction at day 1. Some of the genes in these pathways that were strongly upregulated on day 1 in response to most vaccines, such as *CXCL10* and *OAS1*, were upregulated as late as 21 d after vaccination with YF-17D (Fig. 4e). Importantly, both the early adaptive and delayed innate responses were consistent across multiple studies from diverse geographical locations (Extended Data Fig. 2g,h). Together, these results highlight the unique kinetics of transcriptional responses to yellow fever vaccine relative to other vaccines.
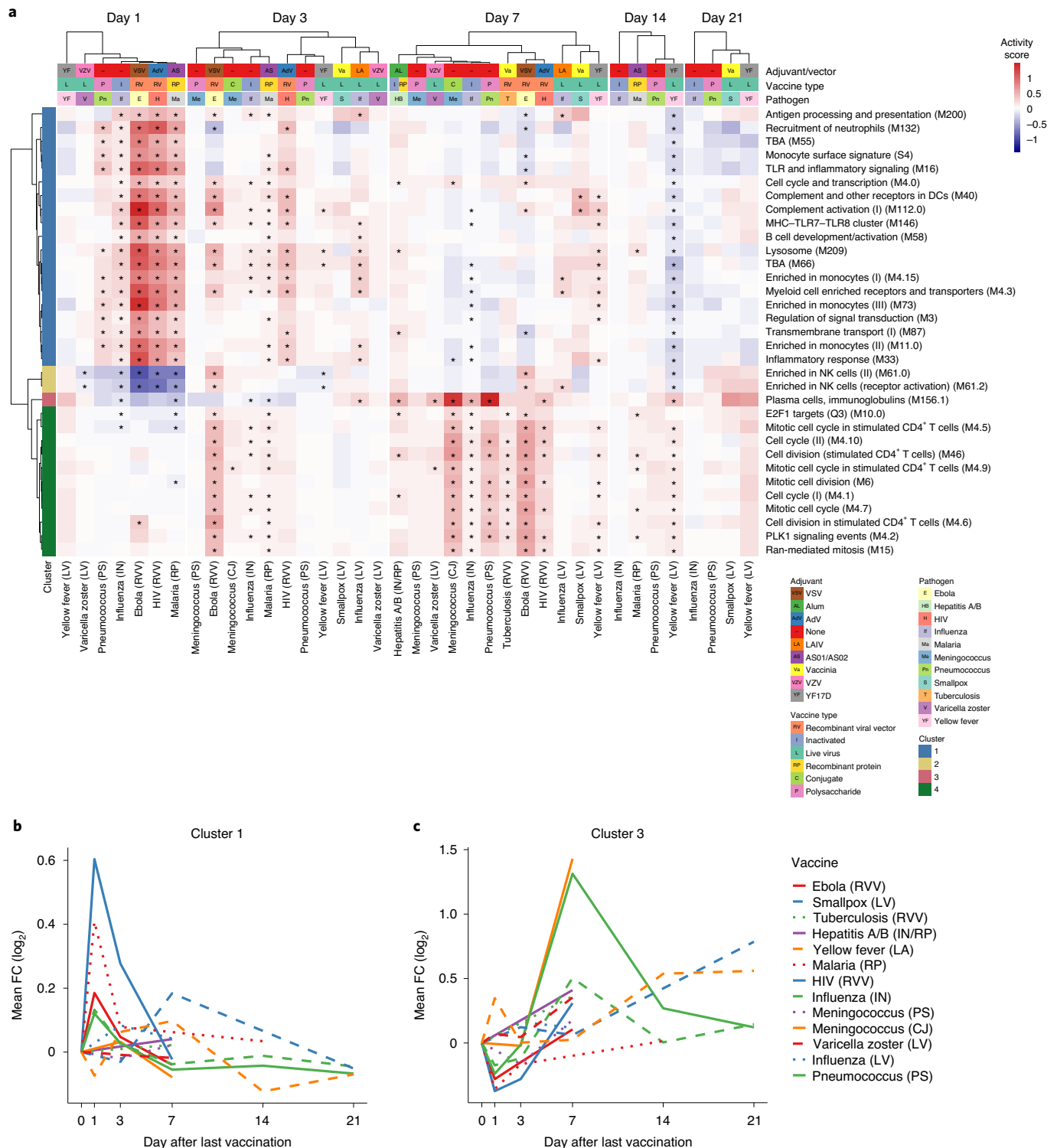
Fig. 2 | Common and unique transcriptional responses across different vaccines. a, Heat map of common differentially expressed modules (regulated in 7 or more vaccines) over time (asterisks denote false discovery rate (FDR) < 0.05). Color represents the QuSAGE activity score. Clustering on columns was performed separately for days 1, 3, 7, 14 and 21 after vaccination. Clusters 1, 2, 3 and 4 are indicated by the blue, yellow, pink and green vertical bars to the left of the heat map. TBA, to be annotated. **b**, Kinetics of the mean FC of cluster 1 modules across vaccines. **c**, Kinetics of the mean FC of cluster 3 modules across vaccines. DCs, dendritic cells; FC, fold change; MHC, major histocompatibility complex; NK, natural killer.

Impact of preexisting immunity on transcriptional responses to vaccination

One possible contributor to the unique kinetics of responses to the yellow fever vaccine is that of the studies included in the data resource; all participants were naïve to yellow fever and the vaccine was inducing a primary response. This is in contrast to other vaccines such as influenza and varicella zoster that elicit a recall response in the adult populations studied. However, it is difficult to evaluate the impact of preexisting immunity and compare primary and recall responses across vaccines because of intrinsic differences in the nature of these vaccines as well

as other factors such as the vaccine platform and type of pathogen that also affect the nature and kinetics of immune response and confound such an analysis.

To address this, we therefore performed a vaccine-specific analysis using all inactivated influenza studies. The inactivated influenza vaccine was ideal for this analysis as there are a large number of such studies in the data resource, and there is substantial heterogeneity in the amount of preexisting immunity and previous exposure to influenza across the population. We used pre-vaccination antibody levels as a marker of preexisting immunity and defined 'high' and 'low' baseline antibody participants as the top and bottom 30% of participants in each study based on baseline geometric mean antibody titer. Increased levels of preexisting antibodies resulted in diminished induction of interferon (Extended Data Fig. 3a,c) and plasma cell (Extended Data Fig. 3b,d) signatures on days 1 and 7, respectively, but did not appear to impact the kinetics of these responses (Extended Data Fig. 3e,f). One possible explanation for this effect is that increased amounts of preexisting antibodies may bind antigen from the vaccine, thereby reducing the amount available to be processed and presented by innate cells and lowering the vaccine 'take'.

Time-adjusted transcriptional predictors of antibody responses

A key goal of systems vaccinology is to identify early signatures predictive of subsequent protection from infection. Antibody titers have been established as a reliable correlate of protection against many pathogens[32] and previous studies have identified transcriptional signatures predictive of antibody responses to several vaccines, including inactivated influenza[6,11,18,33,34], yellow fever[3] and hepatitis B[12]. However, these signatures have thus far been developed for single vaccines, and it remains to be seen whether a 'universal signature' exists that can predict antibody responses across vaccines. Our curated data resource is uniquely suited to address this question, as ten of the encompassed vaccines had at least one dataset with antibody titer measurements before and ~1 month after vaccination (Extended Data Fig. 4a). As there was substantial variability in antibody responses across vaccines, we defined 'high' and 'low' responders on a per dataset basis as the top and bottom 30% of participants according to antibody titer fold changes. We then used an elastic-net machine learning algorithm to develop classifiers capable of distinguishing between high and low responders based on early transcriptional signatures (Methods).

As an initial approach, we wanted to examine whether a model trained using responses to a single vaccine could reliably predict responses to other vaccines. We therefore first built models using all 15 inactivated influenza vaccine datasets (the vaccine for which there was the largest number of samples) in a leave-one-study-out training/testing configuration. As validation that our model could predict responses within the same vaccine, classifiers trained using fold-change expression data at day 7 were able to predict high-versus-low antibody response in the left-out influenza dataset, with areas under the receiver operating characteristic (ROC) curve (AUCs), a measure of classification performance taking into account both true positive and false positive rates, ranging between 0.55 and 0.9 (Fig. 5a). The modules in these classifiers were primarily associated with cell cycle and plasma cell modules (Extended Data Fig. 4b). These results are consistent with

previous work showing that classifiers built using similar pathways are predictive of antibody responses to influenza vaccination across multiple seasons[18]. However, when we examined their performance in other vaccines, they were not reliably predictive (Extended Data
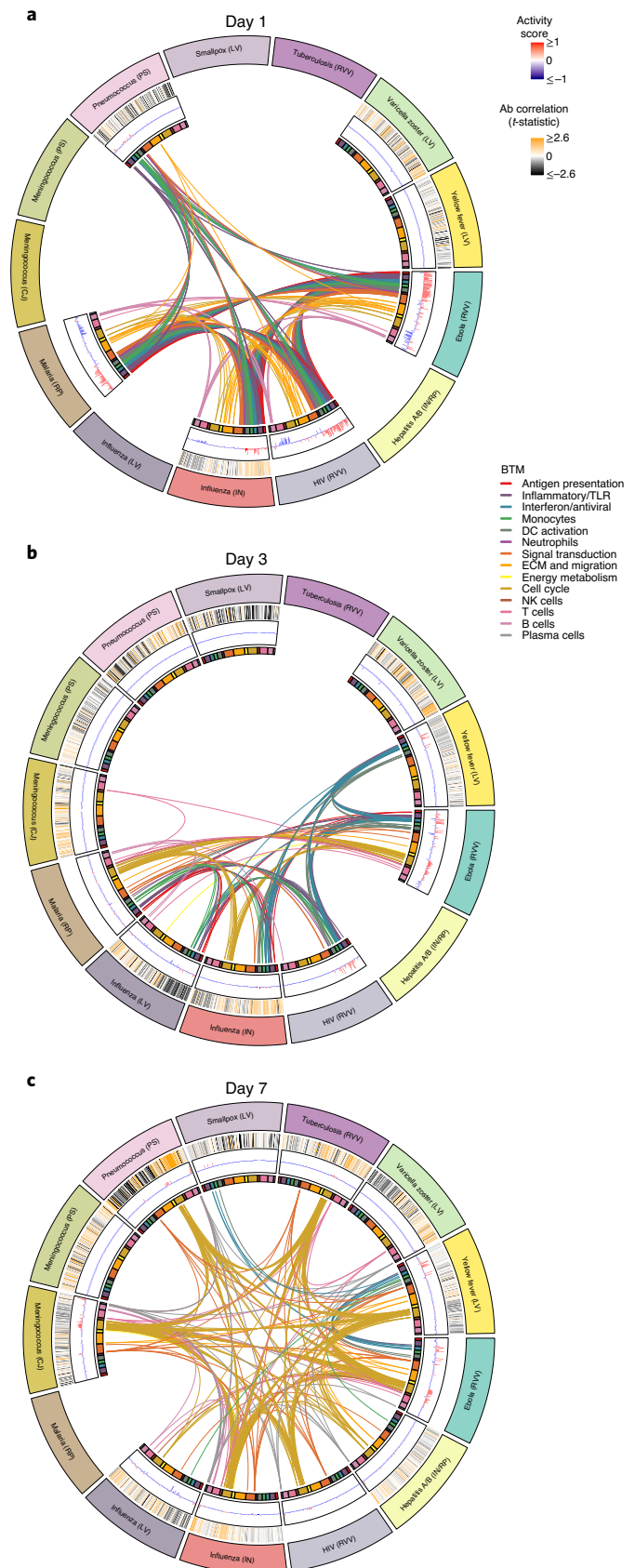
Fig. 3 | Overlap in transcriptional responses across vaccines. a–c, Circos plots of the overlap in differentially expressed modules (FDR < 0.05) across vaccines on days 1 (**a**), 3 (**b**) and 7 (**c**). Each segment of the circle represents one vaccine, and each point in a segment represents a single module. Heat maps in the outermost ring represent the correlation with day 28 antibody responses, and bars in the middle ring represent the activity score of differentially expressed modules. Lines connect modules with a significant positive score shared between vaccines. Boxes and line colors in the innermost ring represent the functional groups of the modules. ECM, extracellular matrix.
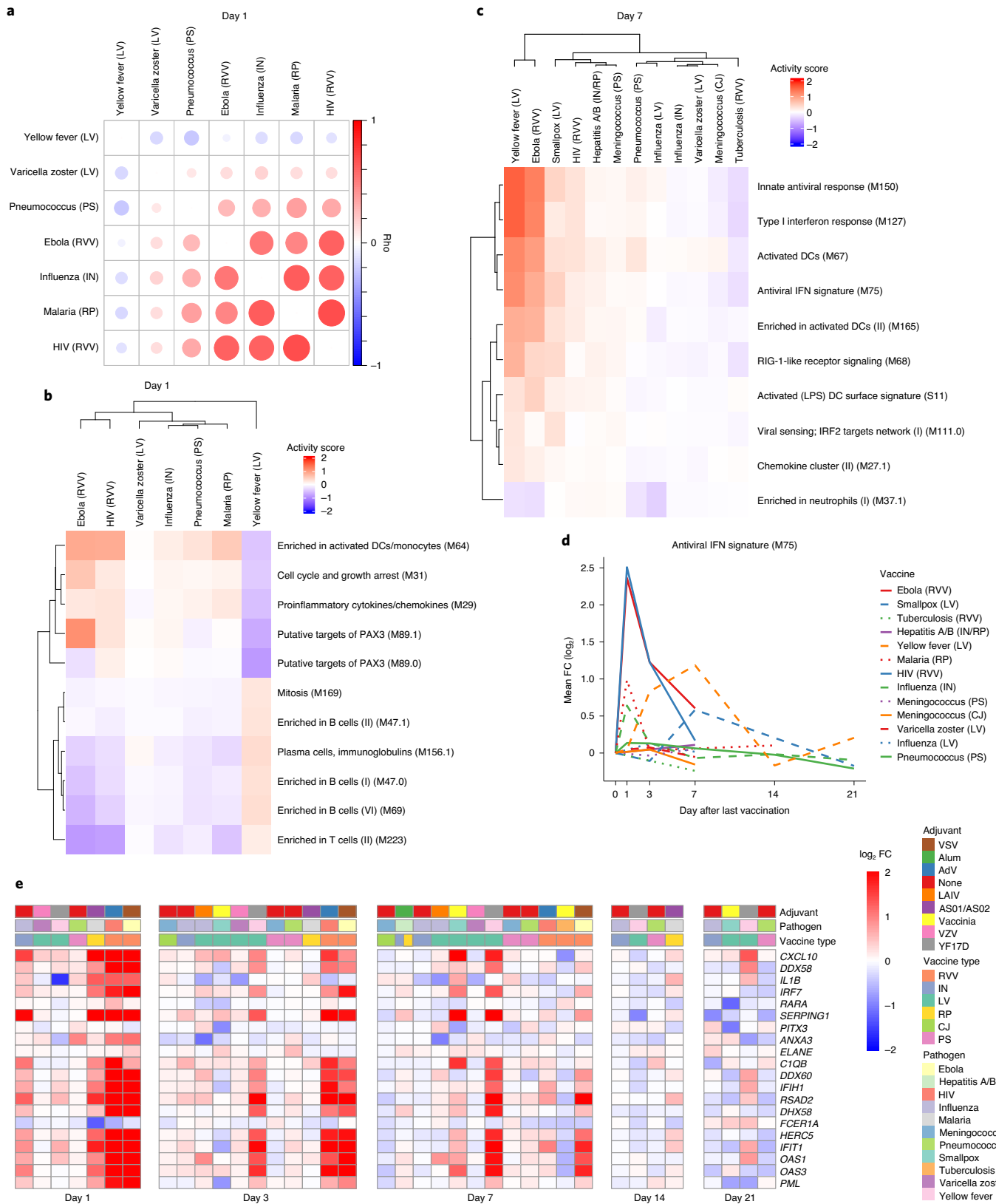
Fig. 4 | Early adaptive and delayed innate transcriptional signatures of yellow fever vaccine. a, Correlation matrix of pairwise Spearman correlations of day 1 gene-level FC values between vaccines. **b**, Heat map of day 1 activity scores of modules differentially expressed in response to yellow fever vaccination (QuSAGE FDR < 0.2). **c**, Heat map of day 7 activity scores of modules differentially expressed in response to yellow fever vaccination (QuSAGE FDR < 0.05, activity score > 0.2). **d**, Kinetics of the mean FC of module M75 across vaccines. **e**, Heat map of the post-vaccination FC of genes in module M75. LPS, lipopolysaccharides.

Fig. 4c). Moreover, the expression of modules associated with antibody response to inactivated influenza vaccination at day 7 was not generally correlated with antibody responses across vaccines (Fig. 5b).

We then asked whether training across multiple vaccines would improve the universality of the identified signatures. Neither a leave-one-vaccine-out approach, nor a tenfold cross-validation approach combining all datasets, was able to identify signatures on day 3 or day 7 after vaccination that could accurately discriminate high versus low responders across all vaccines (Extended Data Fig. 4d,e). However, analysis of the predictive power of specific modules over time, such as M156.1, one of the plasma cell modules associated with response in influenza vaccination on day 7, revealed that this module was predictive of response across many vaccines but at different time points (Fig. 5c). While many vaccines saw a strong association between M156.1 on day 7 and subsequent antibody response, in certain vaccines such as yellow fever and smallpox, expression of the module was not associated with response until much later, at days 10–14 and 21, respectively, consistent with the delayed kinetics of this BTM with these vaccines (Fig. 2).

These results suggest that differential kinetics of immune responses across vaccines pose a confounding variable in the identification of universal predictive signatures of response at a single time point, but that using vaccine-specific time points dictated by the particular kinetics of immune responses for identification of predictive biomarkers of vaccine responses may improve the universality of such signatures. To test this hypothesis in the context of the plasma cell signature, we identified the time point at which expression of the plasma cell module M156.1 peaked in response to each vaccine (Fig. 2c). We then trained a logistic regression classifier with M156.1 expression as an input in a tenfold cross-validation approach using fold-change data at the peak M156.1 expression time point for each vaccine. Indeed, using M156.1 peak expression time points improved the overall performance of the classifier compared with using a single time point (day 7) for all of the vaccines (Fig. 5d). This improvement was driven by increases in response prediction among vaccines in which the plasma cell signature peaked at time points other than day 7, such as the yellow fever and smallpox vaccines (Fig. 5e). Thus, expression of the plasma cell module M156.1 acts as a time-variable universal signature of antibody responses to vaccination.

Impact of aging on transcriptional responses to vaccination

The impairment of vaccine efficacy with age is a major challenge for vaccine development and public health. Although declining vaccine efficacy can broadly be attributed to effects of immunosenescence such as loss and dysfunction of naïve T cells[35], diminished class-switch capability of B cells[36] and decreased TLR function among innate cells[37,38], the molecular mechanisms responsible for impaired vaccine responses among older adults are not yet fully understood. While most of the curated datasets in the Human Immunology Project Consortium (HIPC) resource contained only young adult participants, some studies, including those of inactivated influenza[18], varicella zoster[13] and hepatitis B[12] vaccines, profiled responses of both young (≤50 years) and older (≥60 years) vaccinees. As expected, post-vaccination antibody responses were diminished in older compared to younger participants across all three vaccines (Fig. 6a).

We sought to examine for the effect of aging on immune responses across vaccines by comparing BTM activity scores of the most commonly induced BTMs (Fig. 2a) between young and older participants across all three vaccines at each time point. Broadly, transcriptional responses to the three vaccines were similar between the two age groups (Extended Data Fig. 5a). However, there were significant age-associated differences in several pathways in response to inactivated influenza vaccination, including decreased expression of interferon and other innate immune modules in older compared to young participants early after vaccination (Fig. 6b,c), consistent with previous findings[18]. Despite these differences, the power of the plasma cell signature to predict the antibody response was similar in both young and older individuals (Fig. 6d). These results suggest conservation in the pathways responsible for successful antibody production after vaccination, consistent with earlier findings for influenza vaccination[20].

Discussion

The high degree of homology in the vaccine-induced signatures demonstrates that diverse vaccines that differ widely in target pathogens and composition stimulate conserved immunological networks. Despite this homology, there was still substantial heterogeneity in both the magnitude and kinetics of the induced responses across vaccines. The most distinct in this regard were responses to the yellow fever vaccine YF-17D, which displayed several unique features: (1) a delayed innate and antiviral response that did not peak until days 3–7 after vaccination (Fig. 4d), (2) an early upregulation of B and T cell signatures at day 1 (Fig. 4b and Extended Data Fig. 2e) not observed in other vaccines until much later, and (3) a delay in cell cycle and plasma cell signatures typically associated with the expansion of antigen-specific antibody-secreting cells (Fig. 2a,b).

The observed temporal differences in the plasmablast signature likely reflects differences in the kinetics of the plasmablast response. This in turn will depend on a number of factors, such as immune memory caused by previous exposure to the vaccine or pathogen, the persistence and distribution of the vaccine in the body, as well as the nature of the innate signals triggered by the vaccine. Vaccines such as YF-17D that induce a primary immune response in individuals who have not previously been exposed to the vaccine or virus, will mount a delayed plasmablast response. Furthermore, because YF-17D consists of a live virus that causes an acute viral infection during which viral loads in the serum can be detected for a week or longer, the systemic and sustained presence of the virus may result in a more prolonged and robust plasmablast response. Finally, the triggering of multiple TLRs and innate receptors by YF-17D[3,39] may result in potent activation of the innate immune system that results in a prolonged antibody response. Indeed, consistent with this notion, our previous work has shown that synthetic nanoparticles containing antigens plus ligands that signal through TLR4 and TLR7 induces synergistic increases in antigen-specific, neutralizing antibodies compared to immunization with nanoparticles containing antigens plus a single TLR ligand[40].

Of note, yellow fever and other flaviviruses have a specific capability to inhibit interferon signaling via multiple mechanisms, including suppression of JAK–STAT signaling[41], which could potentially cause the observed delay in interferon responses following YF-17D vaccination. Interestingly, the vaccinia virus also has several mechanisms for

---

Fig. 5 | Time-adjusted transcriptional predictors of antibody responses. a, AUC bar plot of antibody response prediction performance per dataset for the elastic-net classifier trained on inactivated influenza datasets only. Data are presented as mean values ± 95% confidence intervals (CIs). n = 2000 bootstrap replicates. **b**, Heat map of high-versus-low antibody responder difference across vaccines of modules differentially expressed (FDR < 0.05) between high and low antibody responders to inactivated influenza vaccination. **c**, Kinetics of the predictive power of M156.1 across vaccines. For each vaccine/time point combination, the AUC was computed based on difference in the geometric mean of the fold changes of the genes in the M156.1 between high and low responders (Methods). **d**, Weighted ROC curves for a logistic regression classifier using M156.1 expression either at day 7 in all vaccines (day 7) or at the vaccine-specific peak expression time point. FPR, false positive rate; TPR, true positive rate. **e**, Per-vaccine AUC bar plot for a logistic regression classifier using M156.1 expression either at day 7 in all vaccines (yellow) or at the vaccine-specific peak expression time point (green indicates peak at day 7, and pink indicates peak at other time points). Data are presented as mean values ± 95% CIs. n = 2,000 bootstrap replicates.

inhibition of interferon responses, including preventing activation of IRF-3 and NF-κB and dephosphorylation of STAT1/2 transcription factors[42]. Although early response data was not available, the smallpox vaccine containing vaccinia also induced some degree of delayed interferon response following vaccination (Fig. 4d).

While YF-17D demonstrated delayed induction of interferon signatures, induction of B and T cell signatures at day 1 was much earlier than typically observed with other vaccines. This timing is most likely too early to represent an antigen-specific response but could reflect nonspecific activation or recruitment of naïve cells into the circulation.
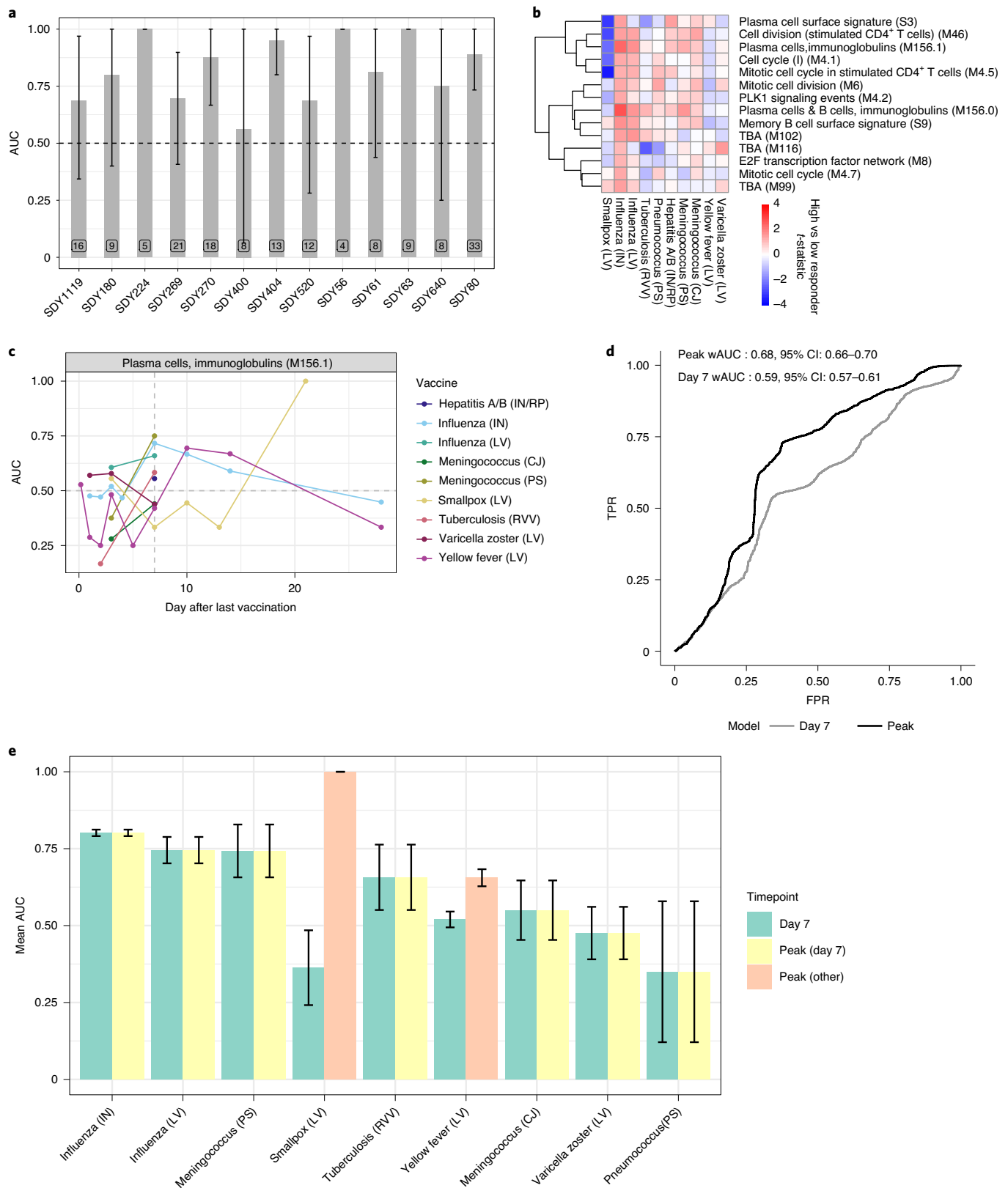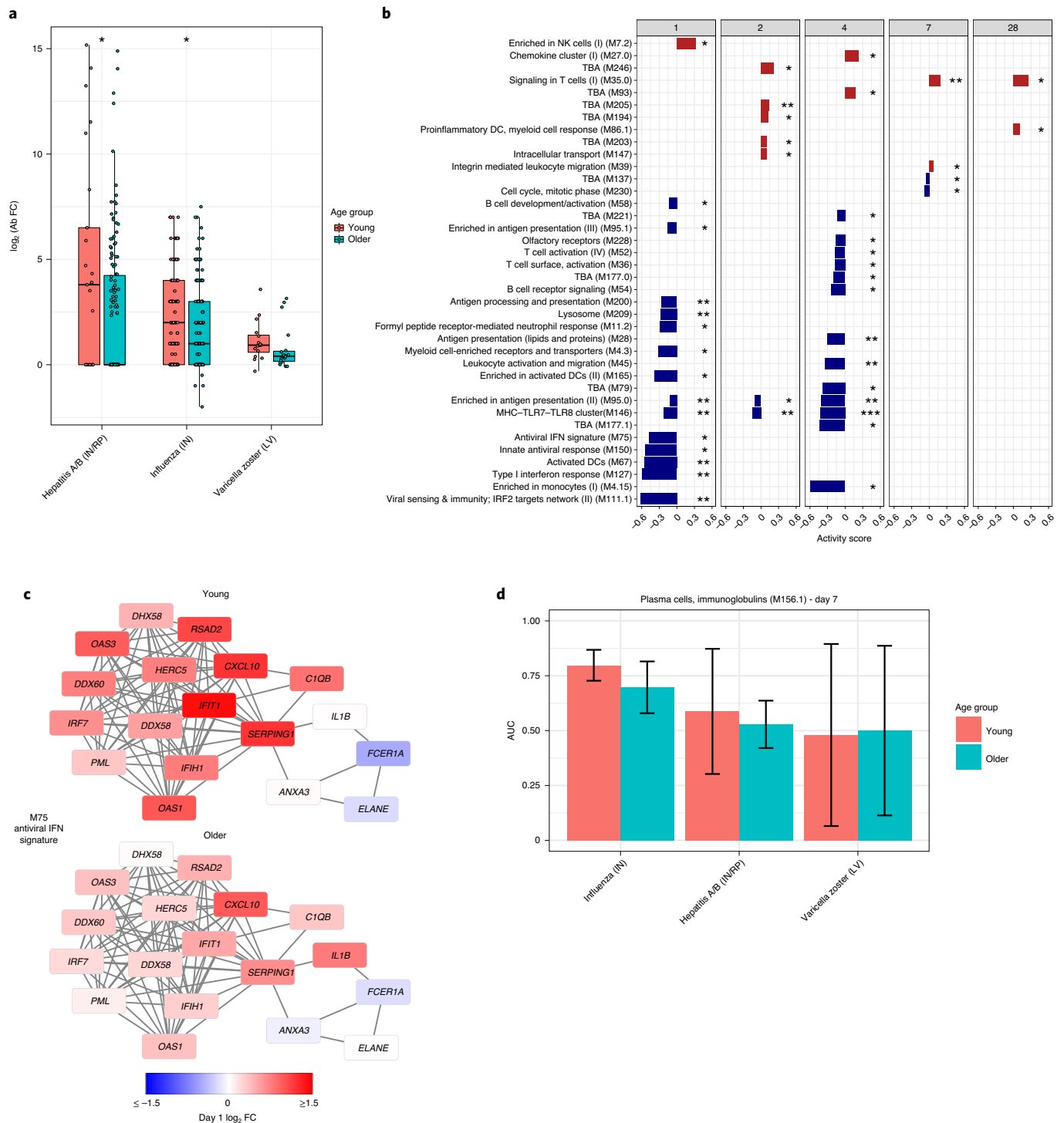
Fig. 6 | Impact of aging on transcriptional responses to vaccination. a, Box plots of day 30 antibody responses to vaccination in young (≤50 years) and older (≥60 years) participants across vaccines. The center line represents the median, box limits represent upper and lower quartiles, and whiskers indicate 1.5 times the interquartile range. Hepatitis A/B (IN/RP)−young, n = 25; hepatitis A/B (IN/RP)−older, n = 135; influenza (IN)−young, n = 123; influenza (IN)−older, n = 175; varicella zoster (LA)−young, n = 16; varicella zoster (LA)−older, n = 19.

b, Modules differentially expressed between young and older participants in response to inactivated influenza vaccination (QuSAGE FDR < 0.05). **c**, Network plot of module M75 on day 1 following inactivated influenza vaccination in young and older participants. Each edge represents a coexpression relationship, as described in Li et al.[10]; colors represent the day 1 log₂ FC. **d**, Bar plot of the day 7 AUC of module M156.1 across vaccines in young and older participants. Data are mean values ± 95% CIs. n = 2,000 bootstrap replicates.

Alternatively, these signatures could be a result of increased relative proportions of adaptive cells in the blood due to extravasation of innate cells into tissues at the site of injection. Further investigation at a cellular level is required to address these hypotheses and elucidate

the mechanisms by which YF-17D exerts such unique early effects on the adaptive immune system.

Finally, our analysis of predictive signatures of antibody responses (Fig. 5) indicates that vaccine response kinetics play an

important role in determining such signatures. Here we have illustrated this principle for a single plasma cell transcriptional module; however, future analyses may enable detection of additional and more accurate signatures. We have previously proposed the concept of a 'vaccine chip' that could measure defined biomarkers and be used to predict protective immune responses across vaccines[25]. We proposed that this chip would be designed to measure expression of a select set of genes or modules, subsets of which would predict a particular type of functional or protective immune response (for example, neutralizing antibody titers, effector CD8$^+$ T cell responses, frequency of polyfunctional T cells, response bias in T_H1 versus T_H2 subsets of helper T cells, and so on)[25]. The results of the present study pave the way for the development of a simple PCR assay (a 'vaccine chip') that can be used to monitor plasmablast signatures in vaccinees. Indeed, PCR-based assays are practical and widely deployable globally, as witnessed during the coronavirus disease 2019 pandemic. Thus, they are likely to be of greater practical value than a fluorescence-activated cell sorting-based assay, involving multicolor fluorescence-activated cell sorting analysis to define plasmablast signatures, as such an assay may be too complex to set up in the field. The concept of time-adjusted signatures based on vaccine-specific response kinetics will be useful in the development of future signatures. In practice, small phase I/II trials could be used to define response kinetics and enable the successful application of a 'vaccine chip' to predict immune responses in subsequent trials.

Due to the considerable costs needed to perform a clinical trial of sufficient size, such vaccine studies are rarely performed with more than one vaccine. Here, we have demonstrated that meta-analysis of vaccine trials can provide valuable insights into the common and unique aspects of immune responses across vaccines. These findings complement those of a companion manuscript by S.F., L.T., M.M., D.C., B.G., et al. describing a set of pre-vaccination transcriptional immune states within the same set of participants that influence responsiveness to vaccination. Combined with the Immune Signatures Data Resource[26], these computational approaches and repositories will enable future research into the mechanisms of vaccine-induced immunity to inform development of improved adjuvants and vaccines.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41590-022-01328-6.

References

1. Hagan, T., Nakaya, H. I., Subramaniam, S. & Pulendran, B. Systems vaccinology: enabling rational vaccine design with systems biological approaches. *Vaccine* **33**, 5294–5301 (2015).
2. Pulendran, B. Systems vaccinology: probing humanity's diverse immune systems with vaccines. *Proc. Natl Acad. Sci. USA* **111**, 12300–12306 (2014).
3. Querec, T. D. et al. Systems biology approach predicts immunogenicity of the yellow fever vaccine in humans. *Nat. Immunol.* **10**, 116–125 (2009).
4. Gaucher, D. et al. Yellow fever vaccine induces integrated multilineage and polyfunctional immune responses. *J. Exp. Med.* **205**, 3119–3131 (2008).
5. Nakaya, H. I. et al. Systems biology of vaccination for seasonal influenza in humans. *Nat. Immunol.* **12**, 786–795 (2011).
6. Bucasas, K. L. et al. Early patterns of gene expression correlate with the humoral immune response to influenza vaccination in humans. *J. Infect. Dis.* **203**, 921–929 (2011).
7. Zak, D. E. et al. 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. USA* **109**, E3503–E3512 (2012).
8. Obermoser, G. et al. Systems scale interactive exploration reveals quantitative and qualitative differences in response to influenza and pneumococcal vaccines. *Immunity* **38**, 831–844 (2013).
9. Banchereau, R. et al. Transcriptional specialization of human dendritic cell subsets in response to microbial vaccines. *Nat. Commun.* **5**, 5283 (2014).
10. Li, S. et al. Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. *Nat Immunol.* **15**, 195–204 (2014).
11. Tsang, J. S. et al. Global analyses of human immune variation reveal baseline predictors of postvaccination responses. *Cell* **157**, 499–513 (2014).
12. Fourati, S. et al. Pre-vaccination inflammation and B cell signalling predict age-related hyporesponse to hepatitis B vaccination. *Nat. Commun.* **7**, 10369 (2016).
13. Li, S. et al. Metabolic phenotypes of response to vaccination in humans. *Cell* **169**, 862–877 (2017).
14. Kazmin, D. et al. Systems analysis of protective immune responses to RTS,S malaria vaccination in humans. *Proc. Natl Acad. Sci. USA* **114**, 2425–2430 (2017).
15. Rechtien, A. et al. Systems vaccinology identifies an early innate immune signature as a correlate of antibody responses to the Ebola vaccine rVSV-ZEBOV. *Cell Rep.* **20**, 2251–2261 (2017).
16. Kotliarov, Y. et al. Broad immune activation underlies shared set point signatures for vaccine responsiveness in healthy individuals and disease activity in patients with lupus. *Nat. Med.* **26**, 618–629 (2020).
17. Arunachalam, P. S. et al. Systems vaccinology of the BNT162b2 mRNA vaccine in humans. *Nature* **596**, 410–416 (2021).
18. Nakaya, H. I. et al. Systems analysis of immunity to influenza vaccination across multiple years and in diverse populations reveals shared molecular signatures. *Immunity* **43**, 1186–1198 (2015).
19. Thakar, J. et al. Aging-dependent alterations in gene expression and a mitochondrial signature of responsiveness to human influenza vaccination. *Aging* **7**, 38–52 (2015).
20. Avey, S. et al. Seasonal variability and shared molecular signatures of inactivated influenza vaccination in young and older adults. *J. Immunol.* **204**, 1661–1673 (2020).
21. Ravindran, R. et al. Vaccine activation of the nutrient sensor GCN2 in dendritic cells enhances antigen presentation. *Science* **343**, 313–317 (2014).
22. Hagan, T. et al. Antibiotics-driven gut microbiome perturbation alters immunity to vaccines in humans. *Cell* **178**, 1313–1328 (2019).
23. Oh, J. Z. et al. TLR5-mediated sensing of gut microbiota is necessary for antibody responses to seasonal influenza vaccination. *Immunity* **41**, 478–492 (2014).
24. Pulendran, B. Learning immunology from the yellow fever vaccine: innate immunity to systems vaccinology. *Nat. Rev. Immunol.* **9**, 741–747 (2009).
25. Pulendran, B., Li, S. & Nakaya, H. I. Systems vaccinology. *Immunity* **33**, 516–529 (2010).
26. Diray-Arce, J. et al. The Immune Signatures Data Resource: a compendium of systems vaccinology datasets. Preprint at *bioRxiv* https://doi.org/10.1101/2021.11.05.465336 (2021).
27. Bhattacharya, S. et al. ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. *Sci. Data.* **5**, 180015 (2018).
28. Scherer A. *Batch Effects and Noise in Microarray Experiments: Sources and Solutions* (John Wiley and Sons, 2009).
29. Gasparini, R. & Panatto, D. Meningococcal glycoconjugate vaccines. *Hum. Vaccin.* **7**, 170–182 (2011).

30. Granoff, D. M., Rathore, M. H., Holmes, S. J., Granoff, P. D. & Lucas, A. H. Effect of immunity to the carrier protein on antibody responses to *Haemophilus influenzae* type b conjugate vaccines. *Vaccine* **11**, S46–S51 (1993).

31. Aran, D., Hu, Z. & Butte, A. J. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **18**, 220 (2017).

32. Plotkin, S. A. Correlates of protection induced by vaccination. *Clin. Vaccine Immunol.* **17**, 1055–1065 (2010).

33. Furman, D. et al. Apoptosis and other immune biomarkers predict influenza vaccine responsiveness. *Mol. Syst. Biol.* **9**, 659 (2013).

34. Sobolev, O. et al. Adjuvanted influenza-H1N1 vaccination reveals lymphoid signatures of age-dependent early responses and of clinical adverse events. *Nat. Immunol.* **17**, 204–213 (2016).

35. Moro-Garcia, M. A., Alonso-Arias, R. & Lopez-Larrea, C. When aging reaches CD4+ T cells: phenotypic and functional changes. *Front Immunol.* **4**, 107 (2013).

36. Frasca, D. & Blomberg, B. B. Effects of aging on B cell function. *Curr. Opin. Immunol.* **21**, 425–430 (2009).

37. Panda, A. et al. Age-associated decrease in TLR function in primary human dendritic cells predicts influenza vaccine response. *J. Immunol.* **184**, 2518–2527 (2010).

38. Kollmann, T. R., Levy, O., Montgomery, R. R. & Goriely, S. Innate immune function by Toll-like receptors: distinct responses in newborns and the elderly. *Immunity* **37**, 771–783 (2012).

39. Querec, T. et al. Yellow fever vaccine YF-17D activates multiple dendritic cell subsets via TLR2, 7, 8 and 9 to stimulate polyvalent immunity. *J. Exp. Med.* **203**, 413–424 (2006).

40. Kasturi, S. P. et al. Programming the magnitude and persistence of antibody responses with innate immunity. *Nature* **470**, 543–547 (2011).

41. Thurmond, S., Wang, B., Song, J. & Hai, R. Suppression of type I interferon signaling by flavivirus NS5. *Viruses* **10**, 712 (2018).

42. Smith, G. L. et al. Vaccinia virus immune evasion: mechanisms, virulence and immunogenicity. *J. Gen. Virol.* **94**, 2367–2392 (2013).

43. Bulteau, R. & Francesconi, M. Real age prediction from the transcriptome with RAPToR. *Nat. Methods* **19**, 969–975 (2022).

[1]Division of Infectious Diseases, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. [2]Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA. [3]Department of Pathology, Yale School of Medicine, New Haven, CT, USA. [4]Center for Biostatistics, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [5]Emory University School of Medicine, Atlanta, GA, USA. [6]Multiscale Systems Biology Section, Laboratory of Immune System Biology, NIAID and Center for Human Immunology (CHI), NIH, Bethesda, MD, USA. [7]NIH-Oxford-Cambridge Scholars Program, Cambridge University, Cambridge, UK. [8]Division of Transplant Surgery, University of California, San Francisco, San Francisco, CA, USA. [9]Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [10]Precision Vaccines Program, Boston Children's Hospital, Boston, MA, USA. [11]Harvard Medical School, Boston, MA, USA. [12]ImmPort Curation Team, NG Health Solutions, Rockville, MD, USA. [13]Institute for Immunity, Transplantation, and Infection, Stanford University School of Medicine, Stanford University, Stanford, CA, USA. [14]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [15]University of Lausanne and Lausanne University Hospital, Lausanne, Switzerland. [16]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [25]These authors jointly supervised this work: Steven H. Kleinstein, Bali Pulendran. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: steven.kleinstein@yale.edu; bpulend@stanford.edu

The Human Immunology Project Consortium (HIPC)

A. Deckhut-Augustine[17], R. Gottardo[9,15,16], E. K. Haddad[18], D. A. Hafler[3], E. Harris[19], D. Farber[20], S. H. Kleinstein[3], O. Levy[10,11,14], J. McElrath[9], R. R. Montgomery[3], B. Peters[21], B. Pulendran[13], A. Rahman[22], E. F. Reed[23], N. Rouphael[5], M. M. Sarwal[8], R. P. Sékaly[5], A. Fernandez-Sesma[22], A. Sette[21], K. Stuart[24], A. Togias[14] and J. S. Tsang[6]

[17]NIAID, NIH, Bethesda, MD, USA. [18] Department of Medicine, Division of Infectious Diseases and HIV Medicine, Drexel University, Philadelphia, PA, USA. [19] Division of Infectious Disease and Vaccinology, School of Public Health, University of California, Berkeley, Berkeley, CA, USA. [20] Department of Microbiology and Immunology, Columbia University Irving Medical Center, New York, NY, USA. [21]La Jolla Institute for Immunology, La Jolla, CA, USA. [22] Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [23] Department of Pathology, David Geffen School of Medicine at University of California, Los Angeles, CA, USA. [24] Center for Global Infectious Disease Research, Seattle Children's Research Institute, Seattle, WA, USA.

Methods

Gene expression preprocessing

An extensive description of the preprocessing of microarray and RNA-sequencing (RNA-seq) datasets included in the Immune Signatures Data Resource can be found in the associated paper[26]. The dataset includes 2,949 samples from published studies and 228 samples not included in previously published studies. All these samples were assembled into a single resource. Briefly, raw probe intensity data for Affymetrix studies were background corrected and summarized using the RMA algorithm[44]. For studies using the Illumina array platform, background-corrected raw probe intensities were used. For RNA-seq studies, count data were transformed using voom[45] to mimic the distribution of microarray expression intensities. Expression data within each study were quantile normalized and log transformed separately for each study.

Batch correction

An extensive description of the across-studies normalization used to correct for batch effects can be found in the Immune Signatures Data Resource paper[26]. Briefly, a linear model was fit using the pre-vaccination normalized gene expression as a dependent variable and platform, study and blood sample type (that is, whole blood or peripheral blood mononuclear cell) as independent variables. The estimated effect of the platform, study and sample type was then subtracted from the entire gene expression (before and after vaccination) to obtain batch-corrected gene expression.

Identification of differentially expressed genes

To determine differentially expressed genes, P values were first computed within each study using two-sided paired student's t-tests. Next, Stouffer's method was used to combine P values across studies via the sumz function in the metap R package[46], with weighting according to the square root of the study sample size. Finally, combined P values were then adjusted for multiple testing using the Benjamini–Hochberg procedure. Similarly, average gene fold changes for each vaccine at each time point were computed by averaging across studies while using weighting equal to the study sample size.

Gene-set enrichment analysis

The enrichment analysis of BTMs was performed in two steps. First, for every study and time point, enrichment was calculated using QuSAGE[47], providing as contrast 'day X–day 0', where X is the current time point, and also a 'pairVector' containing the subject identifiers so that a paired analysis would be performed. Second, to integrate the results from multiple studies of the same vaccine, we performed a meta-analysis for every vaccine plus time point combination, using the 'combinePDFs' function of QuSAGE.

Gene and module sharing analysis

The sharing number of a gene/module is computed as the maximum number of vaccines it is significantly differentially expressed (FDR < 0.05) in, irrespective of time point. For modules, the P values were calculated using QuSAGE[47] ('Gene-set enrichment analysis'). A null distribution for sharing was generated by performing 10,000 permutations of gene/module labels within each vaccine plus time point group.

Antibody titer measurements and identification of high and low responders

Depending on the study, antibody titers were measured by neutralization assays, hemagglutination inhibition assay or IgG levels measured by ELISA[26]. Because some vaccines include multiple strains of viral antigens, the fold change in the antibody response metric was defined as the maximum fold change (MFC) of any strain in the vaccine at day 28 (±7 d) compared to before vaccination. To minimize the difference in antibody response between studies (for example, due to different

vaccines or different techniques used for antibody concentration assessment), the high and low responders were identified for each study separately by selecting the participants with an MFC equal or above the 70th percentile as high responders and participants with an MFC equal or below the 30th percentile as low responders.

Identification of predictive signatures of antibody responses

Four training/testing setups were used for identification of predictive signatures of antibody responses: (1) inactivated influenza datasets only, leave-one-study-out; (2) training on all inactivated influenza datasets, testing on other vaccines; (3) leave-one-vaccine-out (all datasets combined); and (4) tenfold cross-validation (all datasets combined). All models were trained using elastic-net logistic regression using the 'caret' and 'glmnet' R packages. BTM enrichment scores were calculated for each sample using the single-sample gene-set enrichment analysis (ssGSEA) function and used as input features to the models, filtering for modules with a standard deviation > 75% quantile of the standard deviation. Models were fit using either the day 3 fold-change value or the day 7 fold-change value of the ssGSEA score separately. Tuning parameters and performance metrics were estimated using tenfold cross-validation. CIs were estimated using the 'ci.auc' function from the pROC R package.

When developing predictive models for the time point adjustment approach, logistic regression models were trained using ssGSEA score fold change for module M156.1, either at day 7 or at the time point of peak expression in a given vaccine. AUC CIs were estimated using linear mixed-effects models fitted with 100 Monte Carlo resamples. When computing AUCs across multiple vaccines, a weighted AUC was computed using sample size as the weights. For the analysis of temporal change in the predictive capability of M156.1 (Fig. 4c), a weighted mean AUC (based on number of samples in each study) was computed using the calculateROC function of MetaIntegrator R package based on the geometric mean of gene fold changes in the M156.1 module.

Statistics and reproducibility

ArrayQualityMetrics R package was used for QC of all microarray data. Outlier detection was based on the following statistics: (1) mean absolute difference of M values (log-ratios) of each pair of arrays; (2) the Kolmogorov–Smirnov statistic Ka between each array's signal intensity distribution and the distribution of the pooled data; and (3) the Hoeffding's statistic Da on the joint distribution of A (average) and M values for each array. Samples that failed all three QC statistics were removed from further analysis. As this study only involved reanalysis of public datasets, no statistical method was used to predetermine sample size, experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All data used in this study are available from ImmuneSpace (https://www.immunespace.org/is2.url).

Code availability

R code used to generate the figures is available from ImmuneSpace.

References

44. Irizarry, R. A. et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264 (2003).

45. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).

46. Dewey, M. metap: meta-analysis of significance values. R package version 1.8 http://www.dewey.myzen.co.uk/meta/meta.html (2022).
47. Yaari, G., Bolen, C. R., Thakar, J. & Kleinstein, S. H. Quantitative set analysis for gene expression: a method to quantify gene set differential expression including gene–gene correlations. *Nucleic Acids Res.* **41**, e170 (2013).

Author contributions

Conception or design of the work: B.P. and S.H.K. Analysis: T.H., B.G., L.E.T. and A.L. Acquisition: E.H., H.E.R.M., J.D.-A., P.D., the HIPC, O.L., R.G., M.M.S., J.S.T., M.S.-F., S.F., M.P.M., D.G.C. and D.R. Interpretation of the data and manuscript writing: T.H., B.P., S.H.K. and R.-P.S.

Competing interests

O.L. is a named inventor on patents held by Boston Children's Hospital regarding human in vitro systems modeling vaccine action and vaccine adjuvants. B.P. serves on the External Immunology Network of GSK, and on the scientific advisory board of Medicago, CircBio, Sanofi, EdJen and Boehringer-Ingelheim. S.H.K. receives consulting fees from Northrop Grumman and Peraton. T.H. owns stock in GSK and Pfizer. The remaining authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41590-022-01328-6.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41590-022-01328-6.

Correspondence and requests for materials should be addressed to Steven H. Kleinstein or Bali Pulendran.

Peer review information *Nature Immunology* thanks Galit Alter and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Jamie D. K. Wilson, in collaboration with the *Nature Immunology* team.

Reprints and permissions information is available at www.nature.com/reprints.
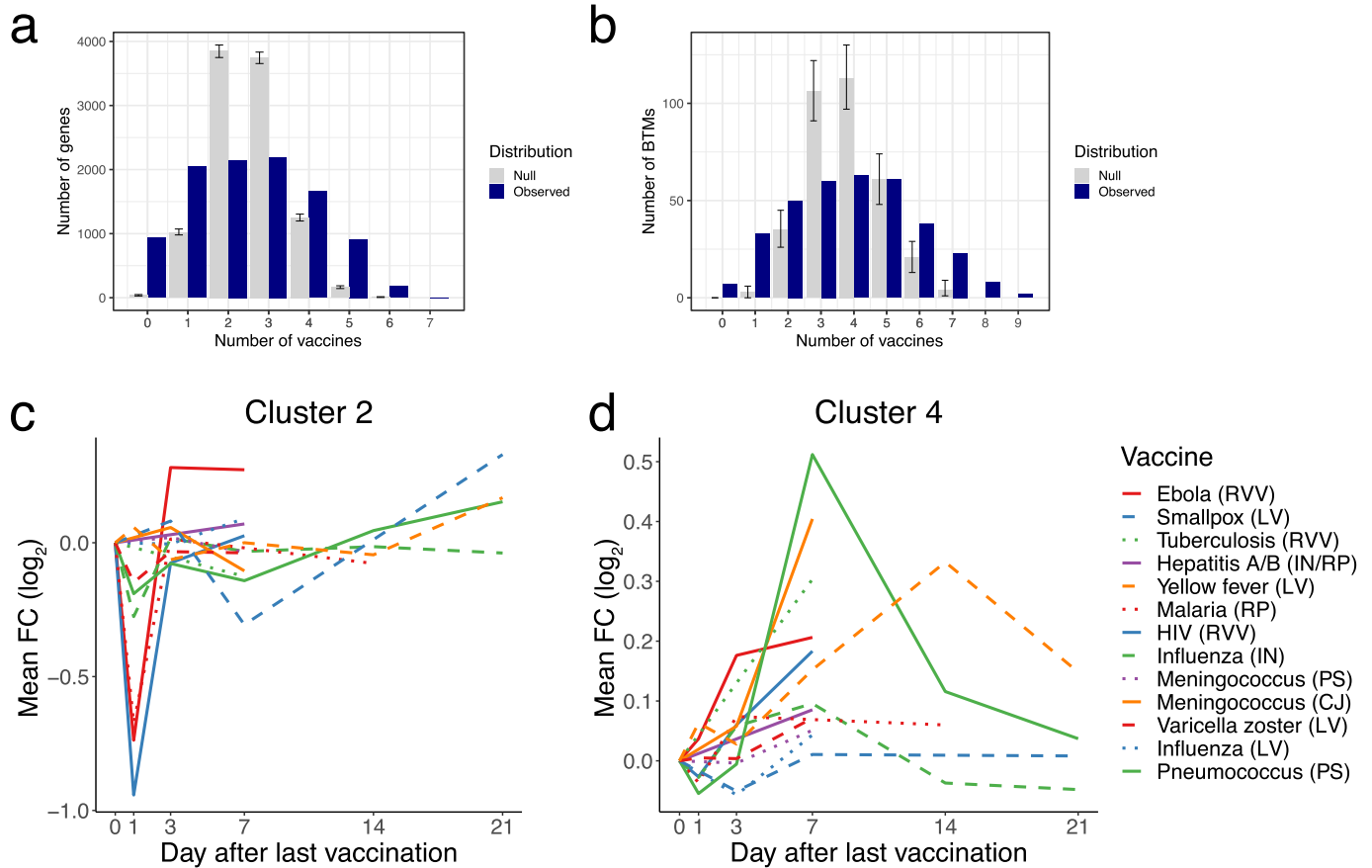
Extended Data Fig. 1 | Overlap in differentially expressed genes/modules and kinetics of common module clusters. a, b) Histograms of overlap in DEGs (A) or differentially expressed modules (B) between vaccines. A gene/module is shared with another vaccine if it is significantly (FDR < 0.05) regulated in the same direction, irrespective of timepoint. Blue bars, number of genesets shared (y-axis) between the same number of vaccines (x-axis). Grey bars represent the null distribution generated by n = 10,000 permutations of gene/module labels within vaccine + timepoint groups. Data are presented as mean values +/− 95% confidence interval. **c**) Kinetics of the mean FC of cluster 2 BTMs across vaccines. **d**) Kinetics of the mean FC of cluster 4 modules across vaccines.
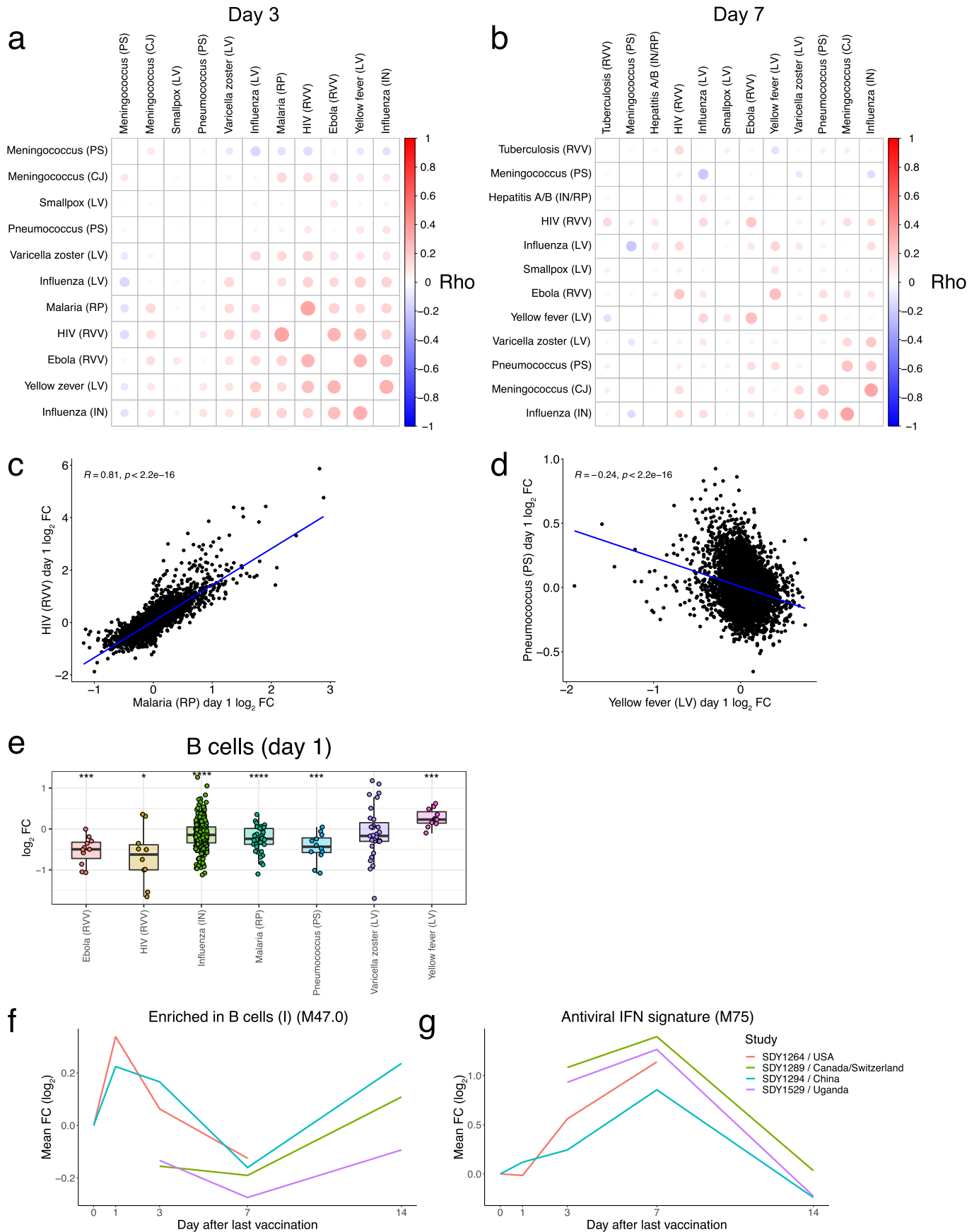
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Gene-level correlations between vaccines and estimated cell frequencies. a) Correlation matrix of pairwise Spearman correlations of Day 3 gene-level fold changes between vaccines. **b**) Correlation matrix of pairwise Spearman correlations of Day 7 gene-level fold changes between vaccines. **c**) Scatterplot of Day 1 gene FCs between HIV and Malaria vaccines. **d**) Scatterplot of Day 1 gene FCs between Yellow Fever and Pneumococcus vaccines. **e**) Boxplot of Day 1 FC in xCell31 estimated B cell frequencies across vaccines. Center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range. Ebola (RVV): n = 11, HIV (RVV): n = 10, Influenza (IN): n = 298, Malaria (RP): n = 42, Pneumococcus (PS): n = 12, Varicella Zoster (LA): n = 31, Yellow Fever (LA): n = 11. **f-g**) Kinetics of the mean FC of modules (F) M47.0 and (G) M75 across YF vaccine studies. In C-D, correlation coefficient and p value determined via Pearson correlation. In E, statistical differences were determined via two-sided paired Student's t-tests within each study and integrating p values across studies within each vaccine using Stouffer's method (see Methods for further details). *p < 0.05, **p < 0.01, ***p < 0.001, **** p < 0.0001.
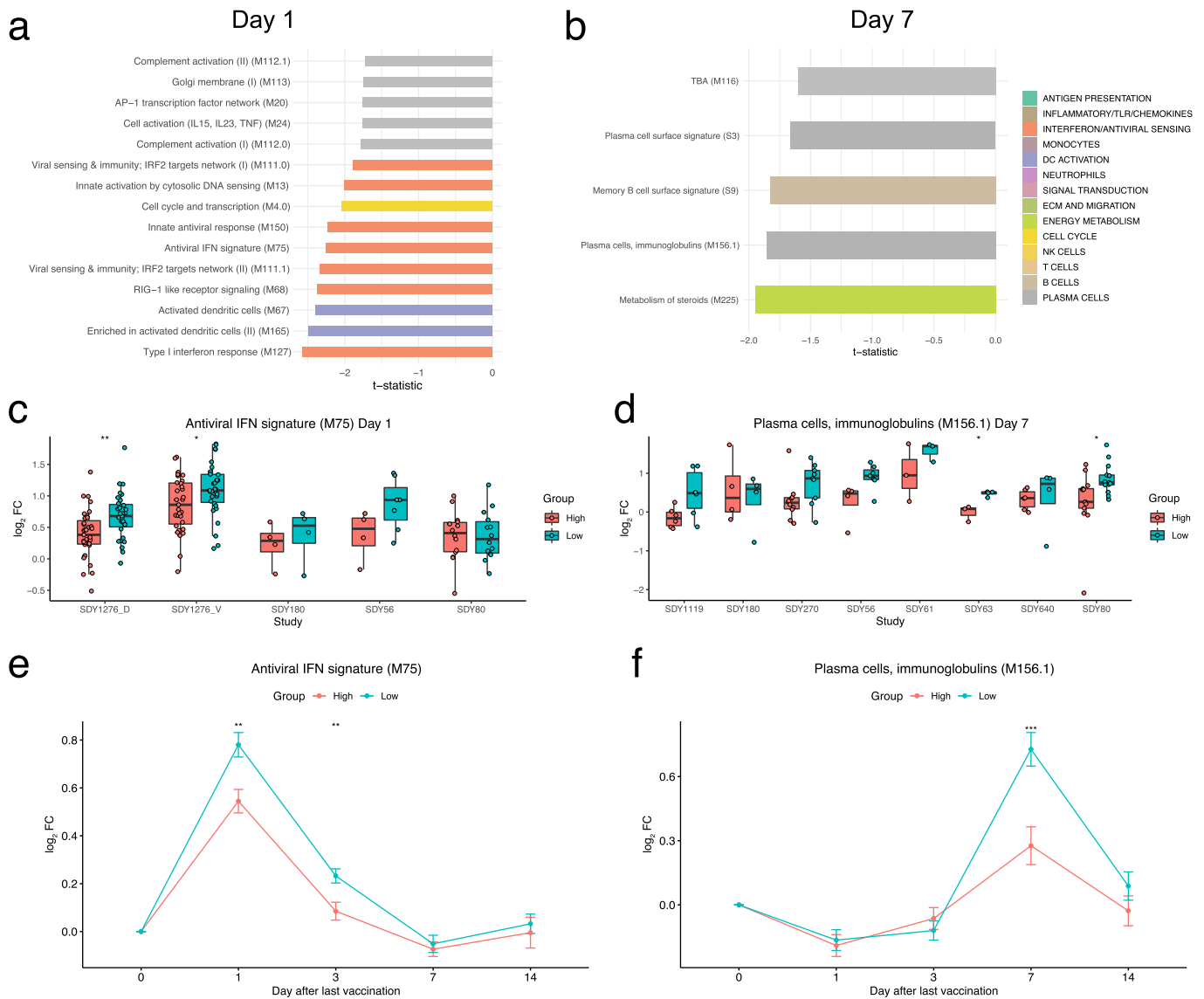
Extended Data Fig. 3 | Impact of pre-existing antibody levels on transcriptional responses to influenza vaccination. a) Differentially expressed modules at Day 1 (FDR < 0.05, t-test between mean fold changes) between participants with high and low baseline antibody titers (negative score indicates increased expression in the low baseline antibody group). Differentially expressed modules at Day 7 between high and low baseline antibody groups. **c, d**) Boxplots of (C) IFN signature module M75 expression at Day 1 and (D) plasma cell module M156.1 expression at Day 7 between high and low baseline antibody groups at Day 1. Center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range. Day 1: SDY1276_D: High − n = 35, Low − n = 31; SDY1276_V: High − n = 31, Low − n = 31; SDY180: High − n = 4, Low − n = 4; SDY56:

High − n = 4, Low − n = 7; SDY80: High − n = 14, Low − n = 14. Day 7: SDY1119: High − n = 6, Low − n = 6; SDY180: High − n = 4, Low − n = 4; SDY270: High − n = 10, Low − n = 9; SDY56: High − n = 4, Low − n = 7; SDY61: High − n = 3, Low − n = 3; SDY63: High − n = 3, Low − n = 4; SDY640: High − n = 6, Low − n = 4; SDY80: High − n = 13, Low − n = 13. **e, f**) Line graphs of (E) M75 and (F) M156.1 expression across time in high and low baseline antibody groups. Error bars represent standard error of the mean. Day 1: High − n = 88, Low − n = 87; Day 3: High − n = 83, Low − n = 88; Day 7: High − n = 49, Low − n = 50; Day 14: High − n = 64, Low − n = 66; In C-F, statistical differences were determined via two-sided unpaired Student's t-tests. *p < 0.05, **p < 0.01, ***p < 0.001.
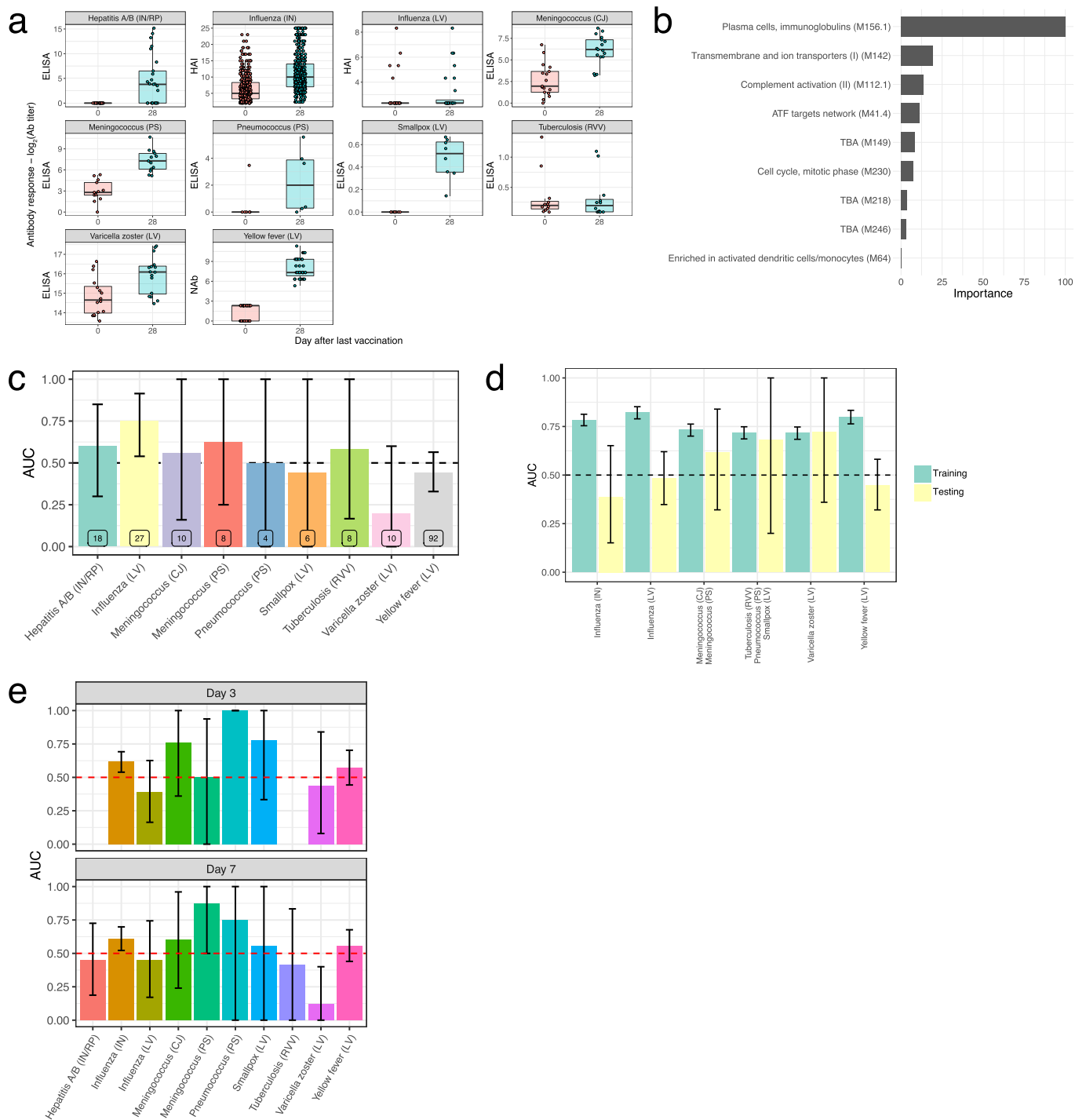
Extended Data Fig. 4 | Antibody response prediction across vaccines. a) Boxplots of Day 30 antibody responses to vaccination across vaccines. Center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range. Hepatitis A/B (IN/RP): n = 25, Influenza (IN): n = 412, Influenza (LA): n = 28, Meningococcus (CJ): n = 17, Meningococcus (PS): n = 13, Pneumococcus (PS): n = 6, Smallpox (LA): n = 8, Tuberculosis (RVV): n = 12, Varicella Zoster (LA): n = 16, Yellow Fever (LA): n = 35. **b)** Barplot of feature importance for the GLM classifier trained on inactivated influenza datasets only. **c)** AUC barplot of antibody response prediction performance across vaccines for the GLM classifier trained on inactivated influenza datasets only. Data are presented as mean values +/- 95% confidence interval. n = 2000 bootstrap replicates. **d)** AUC barplot of antibody response prediction performance of the leave-one-vaccine-out GLM classifier. Data are presented as mean values +/- 95% confidence interval. n = 2000 bootstrap replicates. **e)** AUC barplot of antibody response prediction performance of the 10-fold cross-validation GLM classifier. Data are presented as mean values +/- 95% confidence interval. n = 2000 bootstrap replicates.
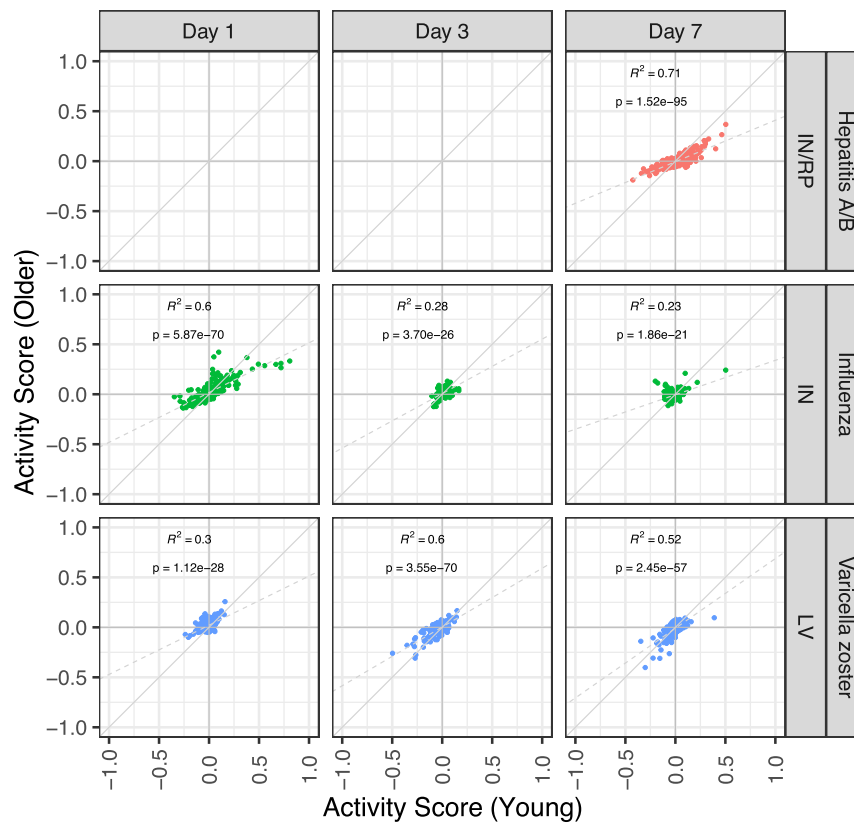
Extended Data Fig. 5 | Comparison of common transcriptional responses between age groups. a) Scatterplots of module activity scores in each vaccine among young (x-axis) and older participants (y-axis) of the most commonly expressed modules (Fig. 2a) on days 1–7. Correlation coefficient and p value determined via Pearson correlation.

nature portfolio

Corresponding author(s): Bali Pulendran

Last updated by author(s): Sep 5, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	All analysis was performed using R version 4.1.2. Packages used include Biobase (v2.54.0), tidyverse (v1.3.2), qusage (v2.28.0), boot (v1.3.28), ggpubr (v0.4.0), ggthemes (v4.2.4), pheatmap (v1.0.12), matrixStats (v0.62.0), reshape2 (v1.4.4), lme4 (v1.1.30), metap (v1.8), circlize (v0.4.15), corrplot (v0.92), ComplexHeatmap (v2.10.0), GSVA (v1.42.0), rlang (v1.0.4), data.table (v1.14.2), limma (v3.50.3), caret (v6.0.93), caretEnsemble (v2.0.1), WeightedROC (v2020.1.31), nlme (v3.1.159), emmeans (v1.8.0), plotROC (v2.3.0), MetaIntegrator (v2.1.3), pROC (v1.18.0), and ArrayQualityMetrics (v3.52.0). Code used to generate all figures presented in the paper can be found at (www.immunespace.org/is2.url).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All data used in this study are available from ImmuneSpace (www.immunespace.org/is2.url).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical test was used to determine the number of samples. Sample sizes were determined by availability of published datasets and established as appropriate to evaluate detection of large vaccine effects by the individual study authors.
Data exclusions	ArrayQualityMetrics R package was used for quality control of all microarray data. Outlier detection was based on the following statistics: i) Mean absolute difference of M-values (log-ratios) of each pair of arrays, ii) the Kolmogorov-Smirnov statistic Ka between each array's signal intensity distribution and the distribution of the pooled data and, iii) the Hoeffding's statistic Da on the joint distribution of A (average) and M values for each array. Using pre-specified criteria within an established public microarray data reuse pipeline, samples that failed all three quality control statistics were removed from further analysis.
Replication	For each of the publicly available datasets, gene expression measurements were performed in a single experiment with multiple biological replicates. All analysis results can be replicated using the publicly available raw data and code (see data and code availability statements).
Randomization	There was no randomization in this study as all analysis was based on publicly available datasets. To account for confounding effects across datasets, we used a linear model to estimate the effect of sequencing/expression platform, study-specific effects (demographic differences and other batch factors), sample type (whole blood/PBMC), age, and sex, as covariates. From here, the study, platform, and sample type effects were eliminated from the entirety of the merged expression matrix.
Blinding	Blinding was not relevant for this study as the goal was to compare transcriptional responses to different vaccines

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging