Section I. Administrative **PROPOSAL: VOLUME I**

1. BAA number	**DARPA – BAA# HR001121S0037**
2. Lead organization	**Icahn School of Medicine at Mount Sinai (ISMMS)**
3. Type of organization	**OTHER EDUCATIONAL**
4. Proposer's reference number	**HR001121S0037-AIM-PA-005**
5. Other team members and type of business for each	**The Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc. (HJF), *OTHER NONPROFIT*; Naval Medical Research Unit TWO (NAMRU-2), *GOVERNMENT ENTITY*; University of Pittsburgh School of Medicine (Pitt), *OTHER EDUCATIONAL;* Princeton University, *OTHER EDUCATIONAL;* University of California, Los Angeles (UCLA), *OTHER EDUCATIONAL;* University of Texas Medical Branch/Galveston National Laboratory (UTMB), *OTHER EDUCATIONAL;* Yale University, *OTHER EDUCATIONAL***
6. Proposal title	**Biomolecular Early Systems Tool for Vaccines (Best-vax)**
7. Technical POC	**Dr. Sealfon, Stuart** Icahn School of Medicine at Mount Sinai One Gustave L. Levy Place, Box 1137 New York, NY 10029-6574 Telephone: 347-866-6317 Fax: 212-289-4107 E-mail: stuart.sealfon@mssm.edu
8. Administrative POC	**Ms. Hadijah Vactor** Director, Grants and Contracts Icahn School of Medicine at Mount Sinai One Gustave L. Levy Place, Box 1075 New York, NY 10029-6574 Telephone: 212-824-8300 E-mail: grants@mssm.edu
9. Award instrument requested	**Cooperative agreement**
10. Places of performance	New York, NY; Bethesda, MD; Singapore; Pittsburgh, PA; Princeton, NJ; Los Angeles, CA; Galveston, TX; New Haven, CT
11. Period of performance	**April 1, 2022 - March 31, 2027**
12. Total funds requested	**Total Funds Requested: $34,774,051** Prime $15,924,265, Subcontractors $18,849,786 **Phase 1 Funds Requested: $19,969,585** Prime $8,503,562, Subcontractors $11,466,023 **Phase 2 Funds Requested: $14,804,467** Prime $7,420,705, Subcontractors $7,383,762
13. Proposal validity period	**180 days**

14. Date proposal was prepared	**October 26, 2021**

B. Official Transmittal Letter.

Section II. Detailed Proposal Information

A. Executive Summary

What is the proposed work attempting to accomplish? We propose to develop a mouse-model rapid-readout R&E tool to predict vaccine strength and long term immune memory (IM) in humans. Using single cell analyses in mouse and human samples in response to vaccination perturbations having varying long term efficacy, we will identify conserved early cellular and molecular processes that initiate the pathway to durable IM. We will study whether the mechanistic nodes identified allow prediction of human individual variation in vaccine responses, and develop a R&E mouse-based tool to predict long term immunity and to accelerate vaccine development and optimization.

How is it done today and what are the limitations? Pre-clinical vaccine candidate testing includes progressive animal model evaluation of antibody activity, cell mediated immunity and pathogen protection. Immune effectors alone do not predict durable IM, which presently takes years of clinical study to assess. The early correlatives of and mechanisms leading to vaccine durability are unknown.

What is innovative in your approach and how does it compare to the current SOA? We use beyond SOA quantitative SC and uniquely accurate bioinformatic and modeling methods to identify the molecular and cellular roadmap to immune memory and to develop robust predictive models.

What are the key technical challenges in your approach and how do you plan to overcome these?

Low generalizability and reproducibility: An important AIM risk is that the R&E tool performance is not reproducible and therefore does not meet benchmarks when transitioned to the IV&V lab or to pharma. We address this critical area by leveraging our culture of rigorous quantitative systems immunology data production, by incorporating replication into the informatics discovery process via external and big data, by experimental validation of new nodes discovered, and by studying a diverse matrix of perturbations with animal studies in two separate institutions.

*p>>N***:** High dimensional molecular data makes causative biological signals difficult to distinguish from noise. We have developed a suite of new methods that help mitigate multiple hypothesis testing and regularize learning models all stages of the analysis pipeline.

Mechanistic prediction of vaccine durability: Systems biology models often do not provide mechanistic insight. Our team has pioneered analysis approaches that generates interpretable models.

Mouse model blood samples: Mouse serial samples for tracing are small. We will adapt bar-coded multiplexed CITE-seq and other sc assays for small volumes, use sensitive analysis methods for rare cell types, and use sacrificed animal samples for larger non-clone tracing-dependent approaches.

Who or what will be affected and what will be the impact if the work is successful? We will accelerate vaccine and protocol development for control of rapidly emerging infectious diseases and maintaining force readiness. We will improve selection of vaccines that induce durable immunity to reduce the need for boosting of warfighters and the public.

How much will it cost and how long will it take? Phase 1: 24 months $XX, phase2: 36 months $YY. Total: 60 months, $ZZ.

B. Goals and Impact

Since the founding of the US military, when General Washington ordered his troops vaccinated for smallpox, vaccination has been a critical element of force readiness. The SARS-CoV-2 pandemic has underscored the critical military need to further accelerate the development and deployment of new vaccines. Since May 2020, together with the US Navy and NMRC, we have led the COVID-19 Health

Action Response for Marines investigation (CHARM), a prospective, longitudinal study of more than 3000 recruits beginning basic training from May-September 2020 at Marine Corp Recruit Depot Parris Island (MCRDPI). At arrival, 9% of recruits were seropositive for SARS-CoV-2 {Letizia, 2021 #135}. However, despite public health measures implemented by the Navy, during only 6 weeks that we monitored each recruit at MCRDPI, a total of 48% became infected with SARS-CoV-2 {Letizia, 2021 #32;Letizia, 2020 #33}. To date, approximately 6 months after basic training, a total of 88% of the 551 long term participants assayed are SARS-CoV-2 seropositive. The COVD-19 pandemic affected training and military preparedness, for example by disabling the Theodore Roosevelt aircraft carrier {Kasper, 2020 #137} and disrupting MCRDPI training during a March-April COVID-19 outbreak preceding CHARM. However the low rate of symptomatic and severe disease caused by SARS-CoV-2 infection in healthy young adults mitigated its warfighter impact. CHARM showed that SARS-CoV-2 spread in a close-quartered military setting at about twice the rate observed in civilian settings. Without a revolutionary improvement in vaccine development time, another pandemic virus that targets younger adults, similar to the 1918 influenza strain, would devastate military readiness. Although the rapid development of SARS-CoV-2 vaccines has been a technical triumph, the current SOA is insufficient to protect the civilian population, to maintain industrial capacity, and to maintain force readiness.

A vaccine deployment bottleneck we propose to address is the absence of methods for preclinical prediction durability of different vaccine preparations and administration protocols. Currently, vaccine durability is only known after long term use in humans. Differences in durability of the SARS-CoV-2 vaccines are only now being determined through long term population follow up. For example, the mRNA-1273 and BNT162b2 mRNA vaccines had indistinguishably high efficacy in their initial clinical trials. However, recent studies show that mRNA-1273 provides better long term protection {Puranik, 2021 #138} (see Fig. 1). Importantly, the reason for these clinically significant differences in long term efficacy between these two vaccines is not known--while the two mRNA vaccines encode identical virus domain sequences, they differ in the timing between the prime and boost dose, human sequence optimization, the dosage of mRNA administered, and the composition of the nano-particle packaging. These factors alone represent a large optimization surface for two nearly identical vaccines that is impractical to explore with long term human studies. Introducing more divergent vaccine preparations represents a further insurmountable obstacle for rational optimization of vaccine durability that our proposed R&E tool will overcome.

Predicting vaccine durability from early immune response biomarkers is a fundamental goal of vaccinology and has been studied for decades (for a recent review see {Van Tilbeurgh, 2021 #133}). While systems immunology has been applied to this question, for example in identifying blood transcriptome signatures predictive of YF-17D vaccine immunogenicity {Querec, 2009 #132}, previous studies have limited cellular resolution. In principle, the advent of single cell multi-omics biology, which has not been harnessed to address this question, provides the basis for a fundamental alteration in the understanding and prediction of long term vaccine immunity. However single cell datasets, by virtue of their massive size and limitations in sample number, introduce unique challenges to avoid overfitting and to provide interpretable mechanistic results. In the Best-vax project, we integrate SOA and beyond SOA single cell systems immunology, single cell biology, bioinformatic approaches and learning methods that have the potential to transform the understanding and predictability of vaccine durability. The Best-vax team is a group of closely collaborating leaders in vaccine development, systems immunology, single cell biology and bioinformatics who together have the unique capabilities to create a multi-omics single cell trajectory from the early days post vaccination to the establishment of durable immunity. This unparalleled reconstruction

and validation of the early mechanisms leading to durable immunity will support predictive modeling to generate a mouse-based R&E tool for vaccine development. If successful, the R&E tool we develop will not only accelerate the pre-clinical stages of vaccine development, but will allow exploration of the durability optimization surface to transform vaccine efficacy. In developing this program, we have worked with the leads of vaccine development at Sanofi and at Moderna. Both companies are providing key assistance and are excited about the prospect of utilizing the tool we propose to develop.

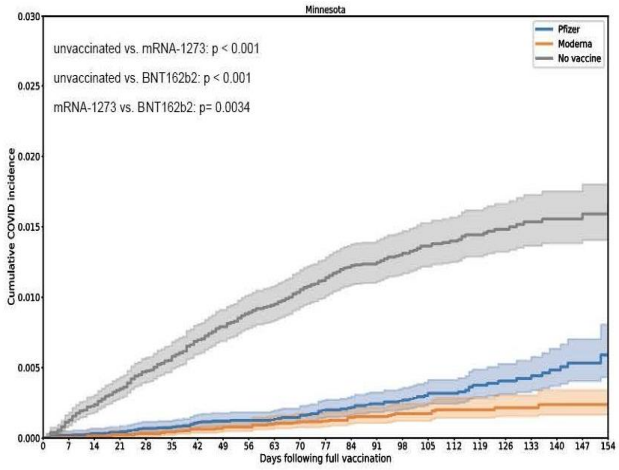


Fig. 1. Diverging efficacy of mRNA-1273 and BNT162b2 mRNA SARS-CoV-2 vaccines over time. (from {Puranik, 2021 #138}).

C. Technical Plan:

a. Approach:

Introduction

Durable antigen-specific IM to infection and vaccination is encoded by long-lived IM cells (plasma cells (PC), memory B cells (MBC), central memory T cells, and stem cell-like memory T cells), that show a continuum of phenotypes and reside largely in bone marrow and spleen {Jameson, 2018 #107}{Palm, 2019 #82}{Youngblood, 2017 #83}{Mamani-Matsuda, 2008 #84}{Pollard, 2021 #85}{Okhrimenko, 2014 #93;Slamanig, 2021 #55}{Wang, 2014 #94}. Following vaccination, early maturation of immune cells occurs in regional lymph nodes (LN), with later development and migration of long-lived IM cells to bone marrow (BM) and spleen. Although the size and number of LN germinal centers (GC), rate of developmental transitions and transcriptional signatures correlate with serological responses to vaccination {Kasturi, 2011 #90;Mesin, 2016 #88;Victora, 2012 #89}{Ma, 2009 #91}, reliable predictors of durable immunity are elusive. The developmental processes leading to the long-lived IM cells can be presumed to initiate during the early days after vaccination, likely influenced by local site reactions and innate immune responses {Kasturi, 2011 #90}{Van Tilbeurgh, 2021 #133}. **We hypothesize early stage immune precursor (IP) cells leading to long-lived IM cells show distinct molecular signatures within**

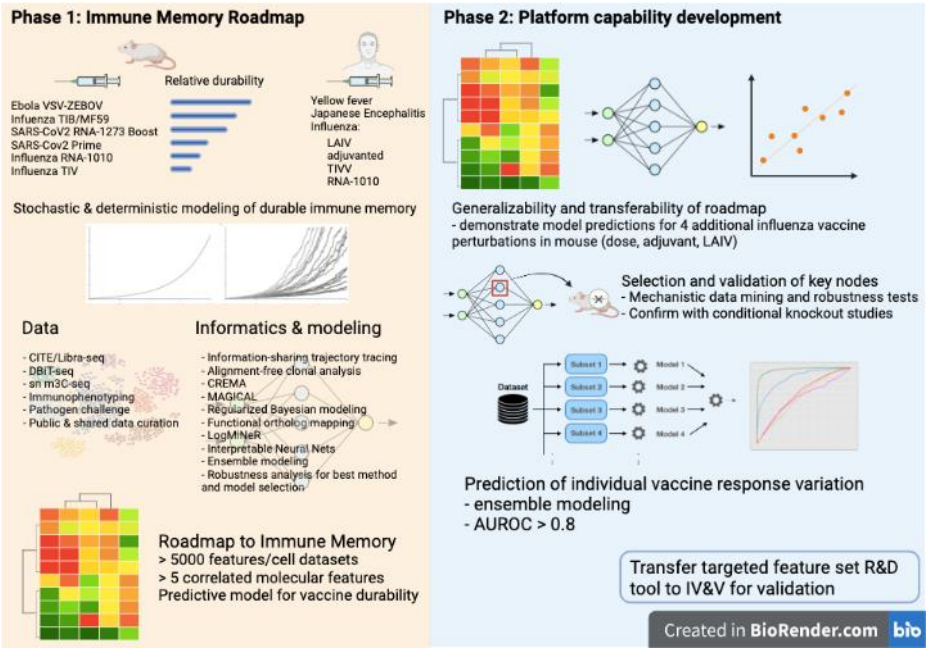
days after vaccination. Accordingly, we propose studies to identify these early state cells and their signatures in conjunction with characterization of other cell compartments in site of injection, regional LN and blood in response to vaccine perturbations of varying durability in order to provide the basis for early prediction of vaccine durability. We will test this hypothesis in mouse vaccination models, with further validation in human samples.

Fig. 1. Program Overview

Research Plan:

Program success rests on three pillars: Obtaining samples from appropriate perturbations, generating high quality data, especially single cell analyses, and selecting and validating the optimum informatic and modeling approaches to address the many fundamental obstacles to identifying mechanistic features and developing a robust research tool. An overview of the proposed research design is presented in Fig. 1.

Establish cell and molecular profiling methods



During the first six months, we will establish the single cell (SC) and immune phenotyping protocols, implementing methods for tissue harvesting and assay or preservation at the two mouse experimental sites, ISMMS AND UTMB. A key determinant for selecting the experimental approach for the SC assays will be the feasibility of preserving the samples after collection, by cryopreservation or other means, or if samples need to be processed for SC CITE-seq/RNA-seq directly after harvest. Sample preservation will allow more flexibility in sample batching and allow for cost saving via pooling of cells for certain steps such as CITEseq antibody staining. However, cryopreservation could affect the cells captured, as some types of cells may not be well preserved and therefore be lost from the analysis. Alternative preservation protocols will be tested, including methanol fixation and commercial buffers (vivoPHIX, RNAssist). A first step will be to determine the impact of preservation on the cell representation and gene expression profile in the different tissue samples. A second goal will be to determine the number of mice it is possible to process in one day; if cell preservation is not possible, maximizing batch size will be preferable to save costs when capturing cells directly after harvest. The pilot experiments will also determine the ability to multiplex cells from different mice, to allow the samples to be mixed in the same surface protein staining and single cell capture batch, to aid in mitigating batch effects. The tested strategies to 'barcode' samples and allow bioinformatic demultiplexing of the samples will consist of (in order of preference) 1) utilizing commercially available oligo-tagged 'hashtag'

antibody reagents; If not enough unique hashtags are available, we will consider in-house antibody labelling or contracted reagent development to increase the number of available hashtags; 2) combinatorial labelling of multiple hashtags per sample; and 3) Oligo-tagged lipid-based multiplexing.

Establish a biological model of good immune memory

We will study the C57BL/6J mouse strain, which provides access to genetic modifications needed for later testing of mechanisms discovered. To cover a wide range of human immune durability and to provide broader relevance of the R&E tool, for the 6 month phase 1 model objectives, ISMMS and UTMB will compare IM generated by a total of six mouse-compatible vaccine perturbations{Nakaya, 2016 #78;Turner, 2021 #50}{Garbutt, 2004 #111;Jones, 2007 #110}{Kavian, 2020 #112} for three pathogens (Ebola, Influenza, SARS-CoV-2) with perturbations including live virus (VSV-ZEBOV), heat-inactivated protein (TIV), adjuvant (TIV-MF59), RNA technology (RNA-1010, RNA-1273) and prime/prime boost administration (Figure 1). Assay of immunological responses from vaccination until 4 months and response to pathogen challenge at 4 months will allow comparison of immune durability across the six vaccine perturbations. The data will be analyzed by modeling, as described below, to demonstrate the required >2.5 fold range of vaccine durability. To assess vaccine durability over a longer time frame, we will also follow a group of vaccinated animals until 8 months post vaccination, assess their IM responses, response to challenge and harvest tissues for potential molecular analysis.

The mouse experiments are shown schematically in Fig. 2. For each of the six vaccine perturbations, one group of 36 mice (18 male, 18 female) and two groups of 10 mice (5 male and 5 female) will be vaccinated and blood will be longitudinally sampled for serology (ELISA and neutralization assays), and B cell and T cell responses using ELISPOTs. PBMCs from each timepoint will be pooled from 3 mice to collect enough cell numbers (~300,000 PBMCs) for ELISPOT assays. 24 mice will be sacrificed at 120 days post vaccination for collection of tissues for long term immune memory characterization (flow cytometry) and biobanking of additional tissues for single cell studies. 10 mice will be challenged with the corresponding pathogen at 120 days post vaccination and monitored for signs of disease and death to evaluate protection. 12 mice will be sacrificed at 240 days post vaccination to generate additional data on vaccine durability and the signatures of long lived IM cells. 10 mice will be challenged at 240 days post vaccination and monitored for signs of disease to evaluate protection. This timeframe precedes immune sensence {High, 2012 #14}.

For generation of data in tissues early after vaccination (1, 3, 7, 14 days) approximately 150 mice per vaccine will be immunized with: VSV-ZEBOV, TIV-MF59 and TIV (experiments starting at contract month 7, Figure X(B)), and with RNA-1010, RNA-1273 prime, RNA-1273 prime-boost (experiments starting at contract month 13, Figure X(B)). Mice (6 at 3 days, 24 at all the other time points per vaccine) will be sacrificed at the specified timepoints and tissues will be processed, assayed and/or biobanked for later use for immunophenotyping and single-cell analysis (Table 1).

Defining good immune memory

By the end of year 1, molecular signatures reflecting good IM will be determined from single cell molecular analyses of lymph node (LN), bone marrow (BM) and spleen cell groups obtained from mice 120 and 240 days after immunization. Analyses include CITE-seq with LIBRA-seq (Linking B cell receptor to antigen specificity through sequencing) for specific reacting cell identification, single nucleus (sn) multiome m3C-seq (methylation and chromatin conformation capture within each nucleus), sn multiome RNA/ATAC-seq and sc resolution spatial transcriptomics DbIT-seq with proteomic registration. LN samples from post vaccine day 1,

3, 7 and 14 will be analyzed to identify markers of long-lived IM cell types that are also detected in IM precursors, indicating detection of the early stages of long lived IM cell formation.

Assemble the roadmap of routes to immune memory

During program year 2, the single cell datasets planned for the first six mouse perturbations will be completed as well as initial single cell datasets from human vaccination studies described below and curation of outside datasets for analysis. These datasets will allow evaluation and selection of the optimum informatics and modeling strategies. The result of this stage of the research will be a set of novel nodes predicting durable IM formation and a model that predicts the level of IM from early responses. The analysis approaches we will utilize are described in later sections.

Phase 2: Demonstration of generalizability and transferability of immune mechanisms from Phase 1 roadmap

Beginning in year 3, to test whether the early responses in mice identified in the initial roadmap reproducibly reflect human long-term immunity, we will study the early immune precursor mechanisms in response to five additional vaccine perturbations in mouse for which human durability data exist. We select influenza vaccine formulations for which there is longitudinal existing data in humans that show different levels and mechanisms of immunogenicity {Kavian, 2020 #112}{van der Most, 2017 #155}{Ambrose, 2010 #156}. By mirroring pre-existing immunity, prime-boost regimens of influenza vaccines in mouse have been shown to recapitulate the relative durability of different vaccines in human {Kavian, 2020 #112}. This study compared inactivated influenza vaccine with those enhanced using high dose inactivated, adjuvanted, or recombinant HA vaccines, and showed different durability across these platforms in human and mouse {Kavian, 2020 #112}, with a more durable response in the case of the adjuvanted vaccine Fluad. The adjuvant AS03 was previously shown to increase persistence of the immune response induced by inactivated influenza vaccines in humans {van der Most, 2017 #155}. LAIV are known to improve durability of protection in children, as compared with inactivated vaccines such as TIV {Ambrose, 2010 #156}. The mechanisms of protection with LAIV are expected to be different than the inactivated influenza vaccines, since the LAIV induce lower levels of neutralizing antibodies and higher levels of T cell responses, adding an additional layer of heterogeneity in vaccine responses which make it a good perturbation for examining generalizability of the roadmap model.

Accordingly, we will administer the following vaccines in the mouse model as at a three-week prime-boost interval: i) TIV, ii) TIV high dose, iii) Recombinant HA (Flublok), iv) Live attenuated influenza vaccines (LAIV, Flumist), v) TIV-AS03 and evaluate the nodes for early immune durability identified and the ability for the model to predict response to pathogen challenge at 4 months post vaccination. The timepoints, tissues and assays for these studies will be selected based on earlier experiments indicating which provide the greatest value for early prediction of durable immunity.

Study of human samples and individual variability in responses:

We will further study generalizability by refining the model to predict vaccine durability and individual variation from analysis of human samples, with data on blood generated by the Best-vax team and analyses of post influenza vaccine human lymph node single cell datasets obtained as part of a different research program by Ali Ellebedy's laboratory that will be shared with the Best-vax team for analysis. Initial human datasets will be obtained the first two years of the program for incorporation into the initial roadmap development and will be completed during Phase 2. These studies will determine the accuracy of models based on nodes in human sample analysis that are functionally homologous to those predictive of durable immune memory in mouse

models for predicting differences in durability between vaccines and among individual human subjects. We include in this evaluation human samples from different vaccines for influenza that have been modeled in mouse as well as two durable vaccines not studied in mouse (Yellow Fever-vax and Japanese Encephalitis Vaccine) to gain insight into generalizability of these key nodes to vaccines not used in the initial discovery process.

Licensed Yellow Fever vaccine: Sanofi Pasteur has agreed to provide cryopreserved samples for Best-vax assays via the Henry Jackson Foundation from a clinical trial evaluating a new Vero-yellow fever vaccine in which the current licensed vaccine (YF-vax) was administered to 18 subjects and PBMC were cryopreserved at day 0,7, 30 and 180 after administration (NCT04142086).

Moderna RNA-1010 influenza vaccine: The Moderna Influenza RNA-1010 vaccine is currently undergoing a phase 1-2 clinical trial from which no biosampling for research is available (NCT04956575). We have worked with Andrea Carfi, the VP & Head of Research, Infectious Disease at Moderna, on a recent single cell study of RNA-1273 coauthored by Moderna and the ISMMS, UTMB and Princeton Best-vax team members {Meyer, 2021 #1}. Moderna is currently planning a new clinical study of RNA-1010 which will allow biosampling for research. Dr. Carfi is supportive of the Best-vax program and will work with us to obtain samples for analysis from this new study.

Japanese Encephalitis and Influenza Vaccine samples from persons deploying in the USINDOPACOM theater: To compare the early mechanisms for other vaccines having differing durability in a key target population for DOD, US Naval Medical Research Unit TWO (NAMRU-2), under the direction of CDR Andrew Letizia, will collect longitudinal NAMRU-2 samples from military personnel receiving the durable Japanese Encephalitis vaccine (JEV) and those receiving the live attenuated and heat inactivated influenza vaccines. NAMRU-2 will spearhead the enrollment and initial human subject's investigation through sample processing of participants who are in or deploying to the USINDOPACOM theater and receiving selected vaccinations. The NAMRU-2 team is composed of personnel based in Singapore and at field activities in sites across SE Asia with primary oversight/management of the USINDPACOM effort held with the NAMRU-2 Singapore office. Enrollment activities will be comprised of on-boarding of regional partners, IRB/HRPO protocol draft and submission, identification of US military and/or local national cohorts for participant enrollments, implementation of enrollment sites (training and logistics), coordination of initial sample processing and shipments to laboratory facilities, and subject matter expertise during data analysis. The primary focus in Phase 1 will be the enrollment of 2 distinct cohorts receiving vaccinations to Japanese Encephalitis Virus (JEV), influenza and possibly SARS-CoV-2: (1) US active duty service members (ADSMs) prior to deployment or already stationed in theater, and/or (2) partner-nation civilians in SE Asia. Blood samples will be obtained and cryopreserved at days 0,3,7,14, 30 and 180 after vaccination.

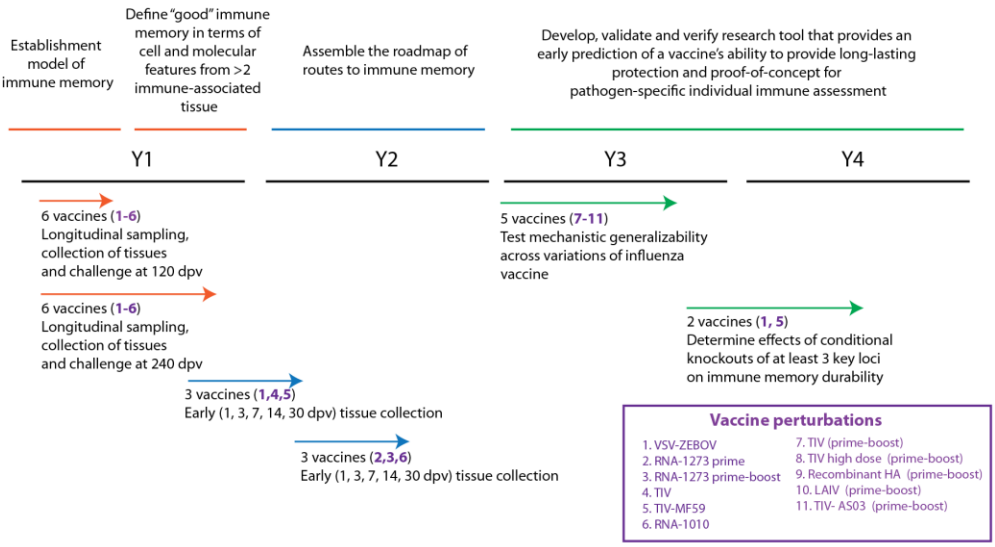
Experimental validation of key MP nodes

To determine contributions of ≥3 conserved pathways to immune memory formation and test correlates of protection in the mouse model, we will use conditional knockout mice for selected nodes identified in the phase I roadmap. VSV−ZEBOV and TIV−MF59 vaccine perturbations will be used in these experiments. The conserved mouse and human early IM nodes will be refined by functional integration of mouse and human data. We expect to be able to procure KO or conditional KO mouse lines to experimentally test a number of novel nodes identified in this roadmap to IM. These studies will be led by Adolfo Garcia-

Sastre who has performed similar studies of the innate immune components contributing to vaccine immunity using B cell transfer from key innate immunity loci knockout to B cell deficient mice (μMT) {Kasturi, 2011 #90}. Cell type specific conditional knockouts or transfer studies will be used with the same vaccination schedule, harvest strategy and immunological assays for TIV+MF59 prime-boost and VSV-ZEBOV to study the effect of knock out of selected nodes. While the final endpoint of these studies will be the effect on response to pathogen challenge at 4 months, we will also gain insight into the relationship and interdependence of the various markers and pathways identified reflecting the development of durable IM.

Figure 2. Timeline of the mouse experiments and vaccine perturbations. dpv: days post vaccination.

Table 1. Overview of mouse model assays planned for roadmap development during 0-24 MAC. The single cell studies are designed dynamically with deeper samping with fewer vaccines using CITE-seq used to guide time point, tissues and cell compartments to concentrate on for subsequent CITE-seq studies and to select the highest value samples for other single cell modalities. In response to the results from initial studies, timepoints and assay allocation may be altered. Orange shading: Tasks 1.1.2 and 1.3 (1.3.5): Establish model and identify features in >2 immune tissues that correlated with good memory. Blue shading: Task 1.3.2: Early responses (1,7,14) and roadmap (30,60,90). X: all vaccine perturbations. x: selected perturbations. x: available for study if indicated by ongoing analyses.



Immunological and Molecular Data Modalities and Methods

IM Assays for vaccine durability:

Quantification of antibody responses: *ELISA.* We will determine serum levels of IgG and IgM specific to the following antigens according to the vaccine of study by ELISA: influenza virus hemagglutinins (HA), SARS-CoV-2 Spike protein (S), Ebola virus glycoprotein (GP). End-point titers will be determined and used for quantification of vaccine efficiency and durability. Influenza serological responses will include avidity and stem antibody assays. *Luminex assays*. We will establish mouse and human IgG multiplex Luminex assays to quantify the IgG protein levels against the different antigens induced by the various vaccines simultaneously, which will allow for more comparable readouts than independent measurements; *Neutralization assays*: the functionality of the antibodies elicited by vaccination will be measured by neutralization assays in serum samples. Serial dilutions of serum from vaccinated animals or from humans will be used to establish the half maximal dilution or inhibitory dilution 50 (ID50). We will use reporter viruses for neutralization assays in mice serum from mice vaccinated with VSV-ZEBOV and SARS-CoV-2 vaccines. In the case of influenza virus vaccines, we will use the virus

			Days post vaccination											
Assay	**Tissue**	**Tissue product**	**1**	**3**	**7**	**14**	**30**	**60**	**90**	**120**	**150**	**180**	**210**	**240**
ELISA	Blood	Serum	X		X	X	X	X	X	X	X	X	X	X
Neutralization	Blood	Serum	X		X	X	X	X	X	X	X	X	X	X
B cell ELISPOT	Blood	Mononuclear cells					X	X	X	X		X		X
T cell ELISPOT	Blood	Mononuclear cells					X	X	X	X		X		X
B cell flow cytometry	Blood	Mononuclear cells			X	X	X			X				x
B cell flow cytometry	LN	Mononuclear cells			X	X	X			X				x
B cell flow cytometry	BM	Mononuclear cells			X	X	X			X				x
B cell flow cytometry	Spleen	Mononuclear cells			X	X	X			X				x
T cell flow cytometry	Blood	Mononuclear cells				X				X				x
T cell flow cytometry	LN	Mononuclear cells				X				X				x
T cell flow cytometry	Spleen	Mononuclear cells				X				X				x
Cytokines	Blood	serum	X		X									
CITEseq	Injection Site	Mononuclear cells	x											
CITEseq	Injection Site	DCs	x											x
CITEseq	Blood	non-naive B cells	x	X	X	x	x	x	x	x				x
CITEseq	Blood	non-naive T cells	x	x	x	x	x	x	x	x				x
CITEseq	Blood	Mononuclear cells	x	x	x	x	x	x	x	x				x
CITEseq	LN	non-naive B cells	x	X	X	x	x	x	x	x				x
CITEseq	LN	non-naive T cells	x	x	x	x	x	x	x	x				x
CITEseq	LN	Mononuclear cells	x	x	x	x	x	x	x	x				x
CITEseq	BM	non-naive B cells								x				x
CITEseq	BM	non-naive T cells								x				x
CITEseq	BM	Mononuclear cells								x				x
CITEseq	Spleen	non-naive B cells								x				x
CITEseq	Spleen	non-naive T cells								x				x
CITEseq	Spleen	Mononuclear cells								x				x
DBiT-seq	LN	intact tissue cells	x		x	x	x	x	x	x				x
sn m3C	Blood	non-naive B cells	x		x	x	x	x	x	x				x
sn m3C	LN	non-naive B cells	x		x	x	x	x	x	x				x
sn m3C	BM	non-naive B cells								x				x
sn m3C	Spleen	non-naive B cells								x				x
sn ATAC/RNA	Blood	non-naive B cells	x		x	x	x	x	x	x				x
sn ATAC/RNA	LN	non-naive B cells	x		x	x	x	x	x	x				x
sn ATAC/RNA	BM	non-naive B cells								x				x
sn ATAC/RNA	Spleen	non-naive B cells								x				x

strains H1N1, H3N2 and B included in the vaccine used in the experiments, and the cytopathic effect will be evaluated to calculate the serum ID50.

B cell responses and germinal center formation and persistence: Evaluation of B cell responses in blood cells from serially sampled mice will be evaluated by measuring the proportion of specific antibody secreting cells by B cell ELISPOTs. B cell phenotyping in blood (mice - terminal time points- and human), and mouse lymph nodes, spleen and bone marrow as well as evaluation of germinal center formation and persistence in lymph nodes and spleen, will be done by spectral flow cytometry Cytek Aurora {Kil, 2019 #52;Turner, 2021 #50;Turner, 2020 #46}. B cell types of interest will be identified by the following phenotypes in mice: Plasmablasts (blood): $CD3^-$ $B220^{lo}$ $CD138^+$ $CD38^+$ $CD27^{++}BCL6^+$ IgD^-; Germinal Center B cells (lymph nodes and spleen): $CD3^-$ $B220^+$ IgD^- $GL7^+CD38^{lo}$; Memory B cells: $CD3^-$ $B220^+$ IgD^- $CD38^{lo}$ $CD27^+$ $CD138^+$; Class switched activated B cells: $CD3^-$ $B220^+$ $CD138^-$ $GL7^-$ $CD38^+$ $CD27^+IgM^-IgD^-$ IgG^+/IgA^+; Long-lived plasma cells (BM only): $CD3^-$ $B220^{lo}$ $CD138^+$ $CD38^+$ $CD27^{++}BCL6^+$ IgD^-. In human, we will identify and characterize the following B cell populations in blood: Plasmablasts: $CD3^-CD19^{lo}$, $CD20^{lo/-}$, $CD27^{hi}$, $CD38^{hi}$, $CD21^+$, $CD138^+$; B memory cells: $CD3^-$, $CD19^+$, $CD20^+$, $CD21^+$, $CD5^-$, $CD27^+$, $CD38^-$; Class switched activated B cells: $CD3^-$, $CD19^+$, $CD20^+$, $CD21^+$, $CD5^-$, $CD2^+$, $CD38^-$ IgM^-IgD^- IgG^+/IgA^+.

T cell responses: Evaluation of T cell responses in blood cells from serially sampled mice and from humans will be evaluated by IFNgamma ELISPOTs. T cell responses will be analyzed in cells from blood (mice - terminal time points - and human), and in mouse draining lymph nodes, spleen and bone marrow by stimulation with pools of peptides corresponding to the relevant antigen of each one of the vaccines, followed by an intracellular cytokine T-cell assay using multidimensional flow cytometry{Gauduin, 2006 #59;Thomas, 1982 #60}. In mice, we will focus on the following populations of interest: CD4 T cells: we will quantify IFNγ, TNFα and expression of the activation markers $OX\text{-}40^+$ $CD25^+$ upon stimulation. CD8 T cells: we will quantify production of IFNγ, TNFα and granzyme B. T follicular helper (Tfh) cells will be identified by the phenotype $CD3^+$ $CD4^+$ $B220^-$ $CXCR5^{++}$ $BCL\text{-}6^+$; expression of activation markers $OX\text{-}40^{++}$ $CD25^+$ will also be quantified in Tfh. In human blood samples: CD4 T cells: we will quantify the expression of IFNγ, TNFα, and activation-induced (AIM) markers $OX\text{-}40^{++}$ $CD25^+$. CD8 T cells: expression of IFNγ, TNFα and granzyme B, and AIMs $CD69^+CD137^+$. Circulating Tfh cells will be identified by the phenotype $CD3^+$ $CD4^+$ $CD19^-$ $CXCR5^{++}$ $PD\text{-}1^+$ $ICOS^+$ $BCL6^+$, and their activation measured by expression of $OX\text{-}40^{++}$ $CD25^+$ (AIM) or $CD154^+$. For determining a 2.5 fold-range among vaccines, durable amplitude will be assessed by the serological titers and neutralizing activity at 4 months and the percentage weight loss or survival rate to challenge. Similar assessment

will be done with activation and cytokine production by CD4 and CD8 T cells. Stability of the response will be compared by the titer decay rate from 2 to 4 months.

Pathogen Challenge: We will challenge mice following Ebola, influenza virus and SARS-CoV-2 vaccines. For influenza, we will use the strain IVR-180 (HA/NA from A/Singapore/2015){Jangra, 2021 #153}, homologous to the H1N1 influenza virus vaccine since 2009; for SARS-CoV-2, variant B.1.351, as this strain has the mouse-adapting N501Y substitution we have shown leads to mouse infectivity{Rathnasinghe, 2021 #116}; for Ebola the mouse adapted virus{Garbutt, 2004 #111;Jones, 2007 #110} done in the Galveston National Lab BSL-4 facility.

SC Systems-level measurements of immune cell responses to vaccine perturbations: Placing cells along a roadmap of IM formation requires comprehensive characterization of individual cells at the level of surface markers, BCR sequence, TCR sequence, transcriptomics, methylation state, chromatin state, chromatin conformation and spatial organization. Our combination of immune memory assays and beyond state of the art sc methods give unprecedented insight into the immune cell types and state to allow reliable reconstruction of the mechanisms leading from MP to IM cells. We use four complementary sc technologies.

Multimodal Cite-seq (~200 surface proteins, transcriptome, V(D)J BCR/TCR in same single cells), with time-subject multiplexed for human cohorts, optimized by Best-vax member John Tsang's lab will be applied to both rare cell type enrichment and unenriched cells to improve map resolution. This method has defined immune transition mechanisms underlying severe COVID-19 (Cell 2021 {Liu, 2021 #18}, see Fig. 4) and pre-vaccination predictors of responses for multiple vaccines{Kotliarov, 2020 #108}. CITE-seq studies will be structured following these steps:

Multi-tissue, single cell dissection of early vaccine responses: linking tissues and blood. For each of the vaccines, groups of six mice will be used per time-point (days 1, 3, 7, and 14 - see Table 1) to dissect multi-tissue early responses, including injection site tissue, draining lymph node (LN), spleen, and peripheral blood. For later time-point groups (days 3, 7, and 14), select blood samples from the previous earlier time-point from the same individual mice (days 1, days 3, and days 7, respectively) will be assessed to evaluate linkage between earlier blood makers to subsequent tissue responses. Multi-modal single cell CITE-seq (simultaneously measuring surface proteins – panels will include ~200 proteins covering major immune cell types and states, transcriptome, BCR/TCR V(d)J) will be applied to dissociated single cell suspensions. Cells from multiple samples will be multiplexed via hashing (and potentially lipid based) barcoding. In addition to sampling from the entire fraction of suspended cells, barcoded cells will be pooled and FACS enriched for more rare subpopulations, including non-naïve B/T cells, dendritic cells (DCs), and potentially CD45- non-immune cells. When appropriate (e.g., days 7 and 14), antigen specific B cells will be labeled via oligo-labeled protein probes (i.e., "LIBRA-seq").

*Tracing the memory responses in tissues and blood***.** Similar to above, but here mouse groups will be pursued for months 1, 4, and 8, and BM, spleen, and blood will be the primary sites for assessing memory responses and durability, including long-lived memory B and plasma cells (see sampling table). The peripheral blood sampling for each of the groups will include earlier serial draws from the same mice as well, including days 7 and 14 for all three groups. LNs will also be assayed in the month 4 group to assess long-lived GCs, which might play a role in sustaining and diversifying long-term humoral responses. To assess whether B cell clones that emerged early detectable in blood can be traced to long-lived memory and plasma cells at later time-points, we will integrate bulk BCR heavy chain sequencing (IgH sequencing) and single-cell V(D)J data. IgH sequencing will be performed on serial blood draws (earlier sampling and final time-point – see above) and tissues (final

time-point). The goal is to link early immune cell phenotypes detectable in blood to those that expanded and eventually become memory and long-lived plasma cells.

*Testing predictors and approaches in humans***:** comparative analysis of yellow fever (YF) and seasonal influenza vaccines. Human blood CITE-seq data will be generated by Shelly Krebs (HJF/WAIR), who has established isolation of and CITE-seq from YF-specific B cells, from blood from 18 YF vaccine recipients (0, 7, 20 and 180 days after vaccination). The Tsang lab will assay a small set of samples from this study to allow correlation of datasets, and will perform the full analyses on the influenza and JEV samples.

Dbit-seq+: As the location of cells in GC contribute to their MP potential{Suan, 2017 #106}, spatial integration of sc multiomics is essential for detection of early mechanisms, Best-vax member Rong Fan's *DBiT-seq+* method, an earlier version of which was described last year in Cell {Liu, 2020 #5} is the **only** spatial transcriptomics (total >22,000 genes)/proteomics platform that provides the essential advance of single cell resolution having up to 5 um resolution and, more recently integrated with CODEX for cell registration (see Fig. 5) The resolution and integration with the CITE-seq datasets are essential for spatial characterization of the sc molecular and cellular processes and interactions underlying GC and early IM development.

Same Cell (multiome) SC ATAC+RNAseq multiome: Building on earlier ultra-high quality single nucleus omics studies (Nature Comm 2021{Ruf-Zamojski, 2021 #10}), the Sealfon lab has established robust same cell single nucleus ATAC/RNAseq[32] that resolves epigenetic regulatory mechanisms underlying cell type and cell response to perturbation. The protocols developed have been applied to a wide variety of tissues in human and mouse, including extensive study of human PBMC in the DARPA-ECHO and DARPA-MBA programs.

Same Cell (multiome) SC chromatin interaction profiling (sn m3C-seq): Best-vax member Chongyuan Luo has developed a method that simultaneously profiles chromatin conformation and DNA methylation in the same nuclei (sn m3C-seq){Luo, 2018 #20}{Lee, 2019 #21}. This technique is critical for linking regulatory changes observed in ATAC-seq and RNA-seq which dramatically increases the signal to noise for detection of epigenetic mechanisms underlying MP cell development.

Curation and integration of public data and data generated from other laboratories:

An important component contributing to the success of the ISMMS DARPA ECHO program was the systematic curation and annotation of public and external data that augmented program data generation to improve the robustness of signatures we developed. Formal curation of outside data is anticipated to be even more valuable for the Best-vax program as 1) external human study measures of vaccine durability are critical for calibrating the R&E tool and 2) mouse/human cross comparison can be extended to datasets from relevant human vaccines and tissues not directly studied in the Best-vax project and 3) validating where possible against external data will augment the robustness of predictive features and the R&E tool developed. This data aggregation will involve:

I. Curation of datasets on vaccines being studied in mouse or human in Best-vax that provide longitudinal data for model fitting to optimize estimation of vaccine durability in humans.

II. Curation of public animal and human molecular vaccination datasets. While this curation will emphasize the vaccines being studied in Best-vax, a broad compendium of results will allow testing for evidence of novel early mechanisms from diverse vaccine perturbations. As part of our DARPA ECHO program, we have already curated a compendium of 22,000 transcriptional profiles from over 150 publicly available human gene expression datasets capturing a wide range of infectious and relevant non-infectious conditions. We will also leverage the "Immune Signatures Data Resource," which integrates longitudinal immune profiling data from human responses to 13 different vaccines {Fourati, 2021 #149, and the "HIPC Dashboard," which includes more than 600 published molecular and cellular signatures associated with human vaccination responses {Smith, 2021 #150}. These

resources were developed by the NIH Human Immunology Project Consortium (HIPC), in projects led by Best-vax investigator Steven Kleinstein.

III. Datasets generated by other AIM performers: If supported by DARPA, we propose to incorporate data from other Assessing Immune Memory performers to improve feature robustness and tool development and robustness. Gustavo Palacios, who is leading a separate ISMMS AIM application with a focus complementary to Best-vax, has agreed in principle to share datasets for augmenting the feature selection and R&E tool validation and robustness.

IV. Datasets provided under MTAs: We will reach out to other investigators to obtain datasets not yet publicly available that are valuable for Best-vax modeling. For example, Ali Ellebedy has performed and expects to continue studies including comprehensive longitudinal single-cell profiling (scRNA-seq+BCR) of human samples from a small number of subjects after influenza vaccination, including LN and BM, tissues we are not proposing to sample in humans. Best-vax investigator Steven Kleinstein, who is spearheading our curation efforts, already collaborates on these studies {Turner, 2020 #46}. Ali has agreed to assist the goals of Best-vax by sharing human influenza vaccine data he has and will be generating. These valuable datasets will allow us to compare the mouse model to human data in these tissues. We can, for example, identify cells in the blood that are associated with predictive signatures derived in mouse and test whether they are clonal precursors of long-lived bone marrow plasma cells, or clonal descendants of cells found in the germinal centers. We will pursue other datasets that allow us to extend or confirm the results we obtain.

Bioinformatics and Modeling

Quantifying vaccine durability

We will establish metrics for the quantification of immune duration and for distinguishing prolonged immunity from the amplitude of the early response. To compare vaccine duration, a metric that captures the effects of amplitude and the transition to durable memory will be used. In order to systematically assess these factors, we will develop a model to fit the data that captures the transition to immune memory and evaluate amplitude increase at time from last vaccine injection, amplitude relative to the 1 month post injection response and the corresponding rate of amplitude decay. In later stages of the program, we will address heterogeneity of subject immune response by including stochasticity in our model design and take advantage of this feature to predict the likelihood of successful individual and population immunity for each vaccine over time.

We propose to utilize a stochastic mathematical model to evaluate the development and stability of B and T cell immunity and overall resistance to pathogen challenge. For the serological and cell mediated immunity models, we will build from existing models (e.g. {Le, 2014 #70}{Hay, 2019 #72} using a stochastic continuous-time markov chain (CTMC) framework including parameter estimation{Fatehi, 2018 #73}. We will apply parameter estimation techniques to learn model parameters such that model output best fits the experimental data

Informatics approaches to identify early mechanistic pathways leading to durable IM.

Overview: For identifying early determinants of durable memory, we will utilize complementary informatics approaches developed by Best-vax researchers, many of which use prior information to reduce data dimensionality and derive explainable models. Importantly, by incorporating external data or prior knowledge into the analysis framework, these methods also improve the relevance and reproducibility of the models and nodes identified across different experiments. We will pursue both contrasts between conditions (early and late, good and bad vaccine responses) as well as reconstructing the sc multi-omics trajectory to resolve the early

molecular transition markers that indicate commitment to durable memory cells. Mapping between human and mouse networks will identify common mechanisms. The methods, outputs and predictive models will be compared using a comprehensive program of robustness analysis to identify generalizable features and to develop a robust and reproducible predictive tool. A description of specific methods we will utilize follows:

Reconstruction of immune roadmap cell trajectories: To prevent the loss of important information that occurs with SOA linear dimensionality reduction and SC trajectory reconstruction methods of SC, while preserving interpretability and generalizability not available in current non-linear methods, the Troyanskaya lab developed a beyond SOA non-linear framework (GraphDR and StructDR) that approximates linear approaches to preserve both information content (especially for identifying rare cell types) and interpretability (Nat. Methods in press, {Zhou, 2020 #29}). By allowing information-sharing across cells, this method leverages high-dimensional information to improve the quality of cell state representation. Rigorous benchmarking experiments demonstrate this framework outperforms 39 other trajectory methods reported (Fig. 3) and demonstrate how valuable it should be for constructing the roadmap to durable immunity.

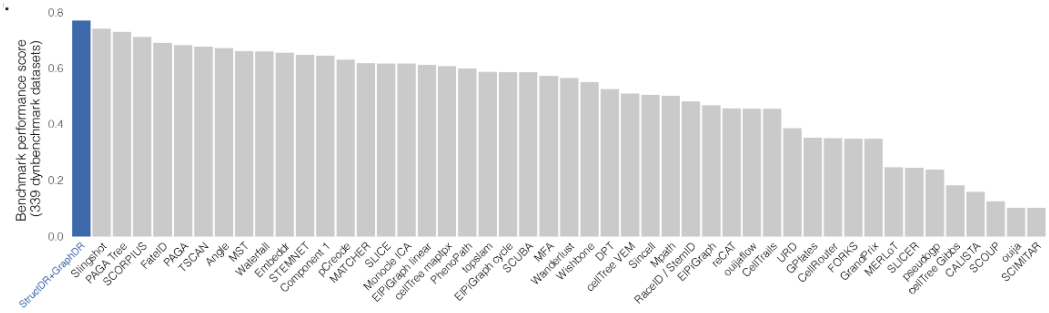


Fig. 3. Benchmarking of cell trajectory analysis methods comparing performance of 40 methods on 339 reference datasets (Nature Methods in press, {Zhou, 2020 #29}. The performance scores are computed based on Saelens et al. 2019 . {Saelens, 2019 #160}.

Incorporating spatial information from the spatial transcriptomics/proteomics DBIT-seq data will allow us to leverage recent methods developed by the Tsang laboratory for integrating multimodal SC data over space and time (see our recent studies {Liu, 2021 #18}{Kotliarov, 2020 #108}). We will incorporate our recent advances in clonal relationship identification{Nouri, 2020 #144}{Nouri, 2018 #145}{Zhou, 2019 #146}, including alignment-free methods, with increased sensitivity and phylogenetic somatic mutation analysis recently developed by the Kleinstein laboratory {Lindenbaum, 2021 #98}. The single cell repertoire will allow tracing BCR and TCR clones across multiple tissues obtained from each individual animal, paralleling the work we have done in CITE-seq data tracing of persistent B cell clones after myasthenia therapy{Jiang, 2020 #100} and migration of B cells after influenza vaccination{Turner, 2020 #147}. Multi-level, time-resolved Bayesian models will be used to integrate time, tissue, and cell type information to assess vaccine-induced changes (e.g. see recent Tsang lab publication {Liu, 2021 #18}), particularly in IM relevant cell types.

Contrasts of early responses: An orthogonal approach we will evaluate to identify new mechanistic features is to contrast the early multi-omics data in good and poor vaccine durability samples in both mouse and human samples. We have developed powerful approaches using massive external data that reduce dimensionality (PLIER{Mao, 2019 #7}), and a new powerful method, MAGICAL, that down samples multi-omics features by coincident changes across omics. We will determine differential features across these contrasts after

downsampling, as well as using deep learning on the larger feature space, as we have done for disease locus prediction{Yao, 2018 #48;Zhou, 2019 #102}. *Multi-omics integration (MAGICAL-Multiome Accessibility Gene Integration Calling After Looping and FENRIR, chromatin activity and gene expression*). To improve multi-omics integration, the Troyanskaya lab developed a Bayesian integrative framework, FENRIR (functional enhancer network for relationship inference of regulatory regions). FENRIR integrates TF binding signals at enhancer regions, chromatin interactions between enhancers and genes, and functional relatedness of genes for diverse tissues and cell types, using data input from ChIP-seq, ChIA-PET, and GIANT gene functional networks (built upon a large compendium of heterogeneous omics data, {Greene, 2015 #67}) and accurately links genes and enhancers. MAGICAL extends FENRIR to provide high accuracy multi-omics differential expression analysis by linking differential sn RNAs with differential accessibility sn chromatin accessibility enhancer sites via ChIA-PET, HiC or C3 datasets (See Fig. 5).

Inference of regulatory mechanisms: To further leverage the complementary single cell data, we will perform linear modeling to correlate distinguishing early molecular events (such as RNA or protein marker expression differences identified from trajectories or contrasts) with epigenetic changes in epigenetic regulatory sites (chromatin, Me, 3C). We will utilize CREMA, recently developed by the Troyanskaya laboratory (see [54]) to the other epigenetic modalities. CREMA links regulatory regions by TF binding site location, regions of interest by co-accessibility analysis and uses linear modeling to elucidate the mechanisms underlying gene expression variation across cells. *CREMA (Control of Regulation Extracted from Multiomic Assays)* leverages multiome datasets to identify mechanisms of gene expression regulation involving both chromatin accessibility and transcription factors. CREMA first selects potential regulatory regions comprising both the proximal promoter region as well as distal regions of interest by co-accessibility analysis across cells. It then identifies transcription factor binding sites in these regions. Finally, relying on the simultaneous measurements of chromatin accessibility and gene expression in individual cells, CREMA uses a linear model to identify key transcription factors and chromatin regions most significantly predictive of gene expression variation across cells. This framework deepens our understanding of early SC regulatory mechanisms early in the vaccine response by expanding to distal regulatory regions and establishing direct linkages between chromatin accessibility, transcription factor presence and gene expression variation.

Interpretable machine learning prediction methods: The final R&E tool requires predictive modeling. We make effective use of all state of the art methods, including elastic net, random forest and GLM as well as deep learning (e.g. {Zhou, 2019 #102;Cofer, 2021 #47}). We have developed new methods that incorporate biological prior knowledge for regularization that we expect to be particularly suitable for the generation of robust and generalizable models (LogMiNeR{Avey, 2017 #117}, xnnet).

xnnet: Interpretable neural net classifier incorporating prior biological information: To overcome p>>N limitations and lack of interpretability in many ML classifiers, for the DARPA ECHO program we developed xnnet, a beyond SOA interpretable neural network based classification method for gene expression data (Fig. 4) that incorporates prior knowledge into hidden nodes to generate a high performing yet interpretable neural network and provide insight into the mechanisms driving the decision process for assignment of samples to classes. We have recently extended xnnet to support integration of any omics data types, including the methylation, ATAC-seq, surface markers, and other datastreams we will be generating with their appropriate prior biological knowledge, to allow prediction from integrated omics datasets.

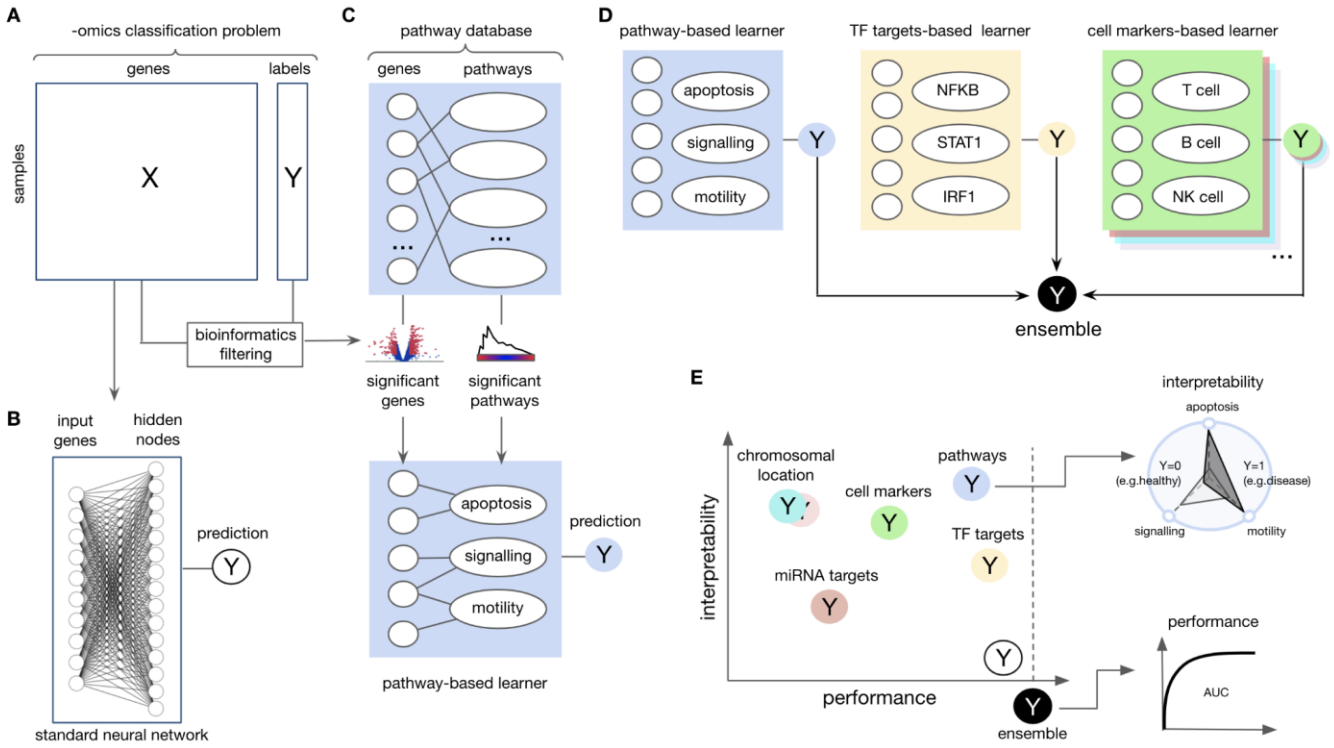


Figure 4 xnnet interpretable ML method. A,B. In contrast to a standard dense and uninterpretable neural network (B), xnnet integrates domain knowledge as gene annotation libraries to build (C) one or more base learners consisting of sparse, easily interpretable neural networks. D. The input nodes are genes; the hidden nodes are gene sets; and edges between genes and gene sets are present only if supported by prior information, which vastly reduces the network complexity. E. Predictions are based on a "super learner", an ensemble model that aggregates predictions from neural networks derived from all the annotation libraries which gives superior performance.

Criteria to identify optimum methods for predicting IM formation

We have proposed a variety of predictive, data driven methods to identify and model mechanistic early pathways that predict durable immunity formation. Our ultimate goal is to have a single well-defined model able to generate accurate predictions and, at the same time, explain the biological mechanisms underlying the predictions. Conventionally, the quality of ML models is evaluated on performance metrics, such as prediction accuracy on validation data. These metrics will play an important role in our methods selection. However, accurate ML models can still lack biological explainability. If the criteria underlying their predictions are obscure, they are not sufficient in order to explain mechanisms of immune memory formation.

Additionally, previous research has shown that models developed in specific contexts frequently do not generalize well to entirely independent data sources. To ensure robustness to the data source, we will evaluate the models in additional curated public and shared datasets. Thus, our strategy for the identification of optimum methods will be based on three dimensions: performance in program data, decision explainability and robustness in independent data sources.

Model evaluation based on performance. To evaluate the performance of informatics approaches, we will use appropriate state of the art metrics and evaluations. For binary classification problems (e.g. good vs poor IM), the most established metrics include accuracy, precision, recall, and ROC AUC. We will assess these metrics for each approach through a rigorous cross-validation and hold-out setup (where a subset of the training data is used as an independent test hold-out set, in addition to the standard cross-validation training). Since -omics data are known to yield noisy measurements, it is also important to assess the robustness of predictions to variable noise levels. We will evaluate the robustness of predictions of each model by perturbing input datasets

through selecting random subsets of data and/or overlaying low levels of noise appropriate for the data modality.

Model evaluation based on explainability. In addition to model performance, we will evaluate the explainability of the different approaches. Broadly speaking, the level of explainability can be seen as the biological and mechanistic insight gained from the model. In contrast to performance, the level of explainability cannot be easily quantified through established metrics. To overcome this problem, we will adopt a combined data- and expert-driven evaluation. Standard models return scores related to the importance of the input features (e.g. highly predictive genes) for model decisions. For example, a linear model returns coefficients for each feature measuring its contribution to the prediction. Thus, feature importance can be used to identify potential key predictors of IM formation. However, focusing on individual predictors may fail to capture the coordinated pathway-level response to vaccination. As such, inferring pathways and processes in an ML framework rather than individual nodes involved in IM formation, is a more robust option. A subset of the proposed methods developed by the Best-vax team belong to a class of principled, directly explainable machine learning models that leverage integration of prior information and domain knowledge. For example, xnnet provides pathway enrichment scores that explain the degree to which a particular pathway guides the decision making in the model (Fig. 4). As a further example, PLIER {Mao, 2019 #7}returns a small set of important predictive latent components, which are associated with specific molecular processes.

The standard and interpretable models considered in our evaluation will produce sets of features and pathways that may explain the formation of IM. To evaluate which explanatory factors provide the most mechanistic insight and actionable predictions, we will consider additional criteria: (1) consistency with known mechanisms, (2) expert curation. The curation will be performed by a domain expert in a blinded manner. The expert will assess the methods' explanatory factors along with a randomly selected set of factors as we previously described {Ju, 2013 #161}{Roussarie, 2020 #162}.

Model evaluation based on robustness in external data.

While most curated public data is at low resolution (the bulk level), it nonetheless provides an opportunity to validate findings from the Best-vax higher resolution (single-cell) data. To evaluate model robustness in external data, we have established a framework that was successfully applied in multiple DARPA ECHO projects. For example, in a DARPA ECHO project to identify a gene signature discriminating Staph infection MRSA vs. MSSA, we obtained two potential signatures based on either single cell multi-omics MAGICAL analysis or with a conventional differential analysis. Using external data, we found that the MAGICAL output signature was the most robust for MRSA/MSSA discrimination. This showed that leveraging public or external bulk datasets is invaluable in selecting analysis approaches that detect more robust biological signals. (Fig. 5). Another example of our systematic approaches to evaluate robustness is shown in Fig. 3. (Nature Methods, in press, where we utilize 339 public datasets to compare performance of 40 methods for trajectory reconstruction. We will systematically assess the robustness to external vaccine-relevant data (see *Data Curation* above) of the results obtained with the Best-vax proposed approaches.

Our criteria for method selection will result in the optimal set of methods with their corresponding inferences for early pathways or features that predict durable immunity formation. Collectively, the outcome of these analyses will lead to 1) a refined set of convergent, conserved early IM pathway nodes to be tested by mechanistic knock-out experiments, 2) robust and generalizable predictive models for durable IM for the R&E tool and 3) prediction of durability from human data.

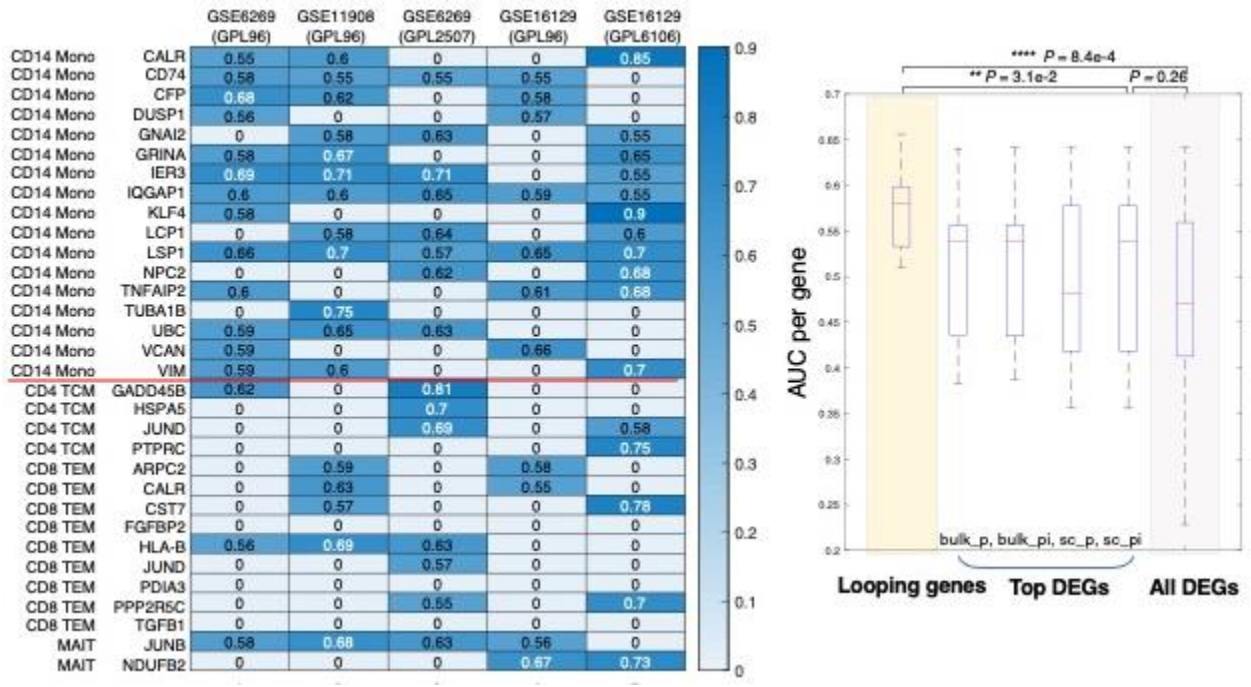


Fig 5. Evaluation of MAGICAL performance based on robustness in external data. MAGICAL infers differential accessibility sites and link differentially expressed genes from single cell multiome data by using an ML method to link genes with regulatory regions via reference Hi-C, CHIA-PET or 3C data. The 36 genes (shown) and linked accessibility sites (not shown) in monocytes and T cell subtypes were identified as differential between MRSA and MRSA PBMC samples assayed by paired scRNA-seq and snATAC-seq (n=8 adult subjects each group). The results were validated by using 5 public bulk transcriptome blood MRSA and MSSA pediatric subject datasets. The median AUC/gene determined as a naive ranking of expression level of each gene in each dataset is shown (left). Despite MAGICAL using single cell celltype assigned data from adults, the genes identified showed good external validation in pediatric bulk datasets, with better validation observed in linked genes identified in monocytes than in T cell subsets. We then compared the AUC per gene obtained on all the combined external datasets obtained by MAGICAL to that obtained by a variety of SOA approaches to identify differentially expressed genes. This validation shows that MAGICAL significantly outperformed the other methods. These results show the effectiveness of an important approach we will use to compare methods using external data, which, as demonstrated, can be invaluable for showing robustness and generalizability, even when the external data is not identical in type to the data being analyzed in the Best-vax program.

Mapping between mouse and human by functional orthology

Building the mechanistic roadmap to select the early mediators and predicting the level of persistent IM is subject to the sparse distribution in a high dimensional feature space (p>>N) that can cause model overfitting. The objective of the orthogonal methods described above is to converge on the most robust biologically relevant mechanisms and readouts to implement in the R&E tool in mouse and to test in human prediction. The correlation of human and mouse mechanisms further powers mechanism identification and predictive feature selection. Because molecular homologues may have different functions in different species, our methods for cross species analysis do not strictly require molecular identity across species, but instead use data driven functional integration in order to compare readouts across species that are linked by mechanism and function (functional analogs).

To assess functional node conservation across mouse and human, we will use a network-based method for delineating functionally analogous orthologs based on functional genomics data. This method was developed by the Troyanskaya lab {Chikina, 2011 #165} and has demonstrated accuracy in transfer of functional knowledge across organisms {Park, 2013 #166}. Functional networks connect different proteins together based on a probabilistic measure of how likely the gene products are to participate in the same

biological process or pathway {Troyanskaya, 2003 #167}{Myers, 2005 #168}{Huttenhower, 2009 #169}{Lee, 2018 #170}. To identify functionally analogous proteins, we use Bayesian integration to combine diverse high-throughput datasets to create a comprehensive network of such functional relationships. These networks can be used to infer the functions of genes based on probabilistic "guilt by association": if a node is connected to many nodes with a certain function it too is likely to be involved in that function. The cross-organism comparisons method relies on directly comparing these functional gene neighborhoods, which we and others have shown can be used to predict accurately gene product function in a single organism {Troyanskaya, 2003 #167}{Myers, 2005 #168}{Huttenhower, 2009 #169}{Lee, 2018 #170}{Guan, 2008 #171}. Intuitively, if two genes from different organisms perform the same function we expect them to be connected to similar types of genes in their species-specific functional network.

Final R&E tool refinement: In developing the best performing model, we plan to estimate the relative value of all predictive data streams and will create an ensemble-based learner. Each component model in the ensemble learner will be weighted by its estimated precision on holdout datasets, and ensemble-based predictor will be evaluated compared to each individual predictor (on separate holdout).

In order to provide a widely accessible final R&E tool for vaccine discovery, to the extent possible we will migrate the key prediction features onto targeted, commonly used platforms, such as flow cytometry, multiplex qPCR, FISH, etc.). After establishment of a final MOP for data generation on one or more targeted platforms, we will retrain and validate the final model, generating the mouse-based R&E tool deliverable to be tested by the IV&V team in year 5.

Data management and sharing processes

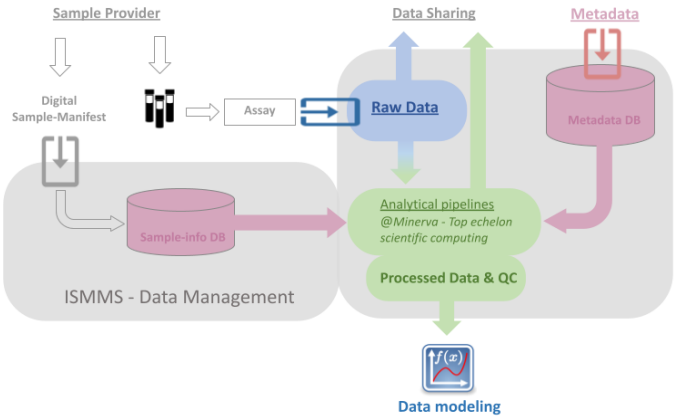
Best-vax will employ integrated sample management, intake and QC of data and cleaned metadata, central sharing among the analysis teams, and formal interim data freeze processes, which are critical for program success. The Sealfon group has developed a stress-tested integrated pipeline (100K sample CHARM study; high throughput NIH consortium studies) encompassing tissue and biological sample management, assay sample preparation, and associated data and metadata management with secure onsite and offsite specimen storage. The sample and data management systems have been designed to ensure a robust and error-free pipeline for handling samples as they are received from varied sample providers and move through our laboratory to eventually become data to be shared with research partners and public repositories. Samples management relies upon the use of a digital sample manifest, which is ingested in an automated fashion into our samples information database (db) upon receipt. Tissue samples and products are logged via a Freezerworks

registry and are physically stored until use in a secure -80C onsite storage or commercial freezer facility with

Figure 6. Data management architecture.

generator backup and 24 hour staffing (Brooks Life Science Systems, see Resources). When samples are assayed and sequencing data are received, a fully automated primary processing and quality assessment pipeline is run. The pipeline assesses each data type on a comprehensive set of associated metrics, producing a qc report that forms the basis for affirming the dataset's quality and transferring it for detailed analysis and modeling. Analysis scripts are systematically documented with tools such as R Notebooks.

An important component of the system is an automated submission module that allows sharing of data with our research and government partners. The overall pipeline for sample and data management, including communication between all systems internally is shown in Fig. 6. The infrastructure described has been carefully developed to meet the needs of our current programs. It was designed to be flexible and



reconfigurable; we will modify it specifically to address the needs of the Best-vax program.

Animal Care Approval: All animal studies will be approved by the Institutional Animal Care and Use Committee (IACUC) before commencing.

Plan for IRB approval:

Human samples obtained from Sanofi (via HJF/WRAIR) and Moderna will be anonymized before transfer to the Best-vax assay teams. These meet the criteria for not human research designation. Consents from these studies will be obtained and confirmation not human research designation will be obtained from the ISMMS IRB and from HRPO before beginning study of these samples. For the vaccinations studies performed by NAMRU-2, they will submit an Institutional Review Board(IRB) human subjects research protocol for review by the US Naval Medical Research Center IRB(IRB00001008) to meet Human Research Protection Office (HRPO) standards (NAMRU-2 FWA# FWA00012695). The protocol will include a description of the research plan, study population, risks and benefits of study participation, recruitment and consent processes, data collection/analysis procedures and an approved informed consent that is compliant with 32 CFR 219.116. All personnel involved with human subjects research will submit and maintain documentation of required trainings. Protocols that enroll foreign national subjects will be first coordinated in-country for local IRB approval before USN IRB and HRPO review and approval.

Fundamental research:
The proposers believe that the scope of this proposal constitutes fundamental research.

b. Rationale
The rationale for many of the study design components are explained in the where they are described in the approach section of the technical plan. Here we address several additional areas of rationale and include data showing assay feasibility.

Justification of preclinical biological model: Preclinical testing of vaccines starts in mouse and progresses through multiple animal models to non-human primates (NHP) {Belser, 2010 #61;Bouvier, 2010 #25;Cohen, 2020 #22;Itoh, 2009 #26;Manicassamy, 2010 #27;Roy Wong, 2021 #23}{Garbutt, 2004 #95}{Suder, 2018 #96}{Roubidoux, 2021 #97}. The Best-vax team studies infection and vaccination in range of mammalian models, including mouse, ferret, hamster, and NHP, e.g.{Bukreyev, 2004 #36;Meyer, 2021 #37;Meyer, 2021 #1}). Developing a vaccine duration tool that can be used for early-stage vaccine, protocol and adjuvant screening will have the greatest impact in improving the rapid identification and development of more durable vaccination if based on a mouse model. Furthermore, only mouse provides tissue resolution and sufficient group numbers to power a study of the large roadmap to IM parameter space. We choose the C57BL/6J mouse strain, which offers access to many genetic modifications if needed to test mechanisms, as the best model for this program. While the inbred C57BL/6J strain does not recapitulate the genetic diversity of humans, this reduced variance increases statistical power for detecting novel early mechanisms. Furthermore, stochastic sources of individual response variation, such as VDJ recombination and somatic mutation, are at least partly reflected in this model{Avnir, 2016 #113}.

We rely on projection onto human data to validate the relevance of early features of the roadmap and mouse-based tool to human. In earlier work, the PI led a team of laboratories in addressing conceptually similar hallucinogen mechanism challenge by using a transcriptome signature in mouse cortex as a bioassay for hallucinogens, which previously could only be measured by human response. Single cell studies reconstructed the neurons, signaling pathways and led to the landmark discovery of the previously unknown dual-receptor target of hallucinogens and of atypical antipsychotics (Nature 2008 {Gonzalez-Maeso, 2008 #11}). The utility of the proposed mouse R&E screening tool is to guide selection among vaccines and protocols to accelerate discovery of vaccines with more durable immunity. Translatability of mouse model responses and human durability is supported in a mouse and human study comparing influenza vaccines. The effect of pre-existing immunity in human was modeled by using a prime-boost regimen in mouse and the relative durability of the vaccines in mouse recapitulated what was found in humans in clinical trials {Kavian, 2020 #112}. The utility of the proposed mouse R&E screening tool is to guide selection among vaccines and protocols to accelerate discovery of vaccines with more durable immunity. **The value of the Bestvax R&E tool on accelerated and improved preclinical vaccine optimization does not require the C57BL/6J mouse to recapitulate the same immunity seen in humans--it only requires that early vaccination mechanistic features measured in C57BL/6J predicts the amplitude and level of durable IM in humans.**

Justification of experimental vaccine perturbations:

Mouse Vaccine Perturbations: The initial six mouse-compatible vaccine perturbations were chosen to cover a range of pathogens and durability (see Fig. 1). While multiple pathogens introduce complexity in the

analysis, they are feasible as our studies are divided among two sites, UTMB for the BSL3/4 pathogen vaccines and ISMMS for the influenza vaccines. For validating the early mechanisms and tools we focus on additional influenza vaccine perturbations to provide a link between the mouse and human studies and because of the high value of improving influenza vaccination durability for a universal vaccine. The human vaccination samples have been selected to provide a range of durability, to test vaccines for durable pathogens not studied in mouse (Yellow fever, JEV) and to provide evidence (and allow improvement) of generalizability of the mechanisms identified and of the R&E tool developed. Vaccination with the live-attenuated Yellow Fever Virus (YFV) strain YF-17D is the benchmark for durable immunity and is included as the gold-standard human durable vaccine. Japanese Encephalitis vaccine has also been demonstrated to provide durable immunity. The influenza vaccines we will study include direct overlap with several of the vaccines modeled in mouse and cover a range of durability. Studying vaccination durability prediction through NAMRU-2 provides data and model testing of high relevance for DOD.

Rationale for Timepoints, numbers and tissues selected for study: Some markers for memory precursors (MP) and IM cell development have been identified, including CCR6 expression in B cell MP in mouse and human {Suan, 2017 #106},ephrin-B1 {Laidlaw, 2017 #140} and Hhex/Tle3 {Laidlaw, 2020 #141}. Epigenetic changes are key elements in both B and T cell differentiation during IM formation {Zhang, 2019 #123}{Akondy, 2017 #87}{Youngblood, 2017 #83}. Despite this progress, the key markers representing the mechanisms leading to high durability memory formation in response to vaccination that would support prediction from early responses are unknown. Furthermore, there is increasing evidence that innate immune cells and unconventional T cell subsets may be useful markers of early processes leading to effective vaccine IM {Van Tilbeurgh, 2021 #133}. A high-resolution sc map of the molecular development from MP in LN to IM cells in BM and spleen, including other cell types and combined with vaccine perturbations of varying durability is needed to identify early predictive mechanisms and markers

The multi-tissue mouse studies are designed to meet interim program milestones, provide adequate temporal sampling of the development of IM and avoid immune senescence {High, 2012 #14}. To model durable and less durable immunity and to establish the roadmap, we will evaluate animals for early immunity signatures at day 1, 3, 7, 14. This timeframe will discriminate between the processes involved in early immunity and durable IM, which will be evaluated at later timepoints. The numbers of samples obtained for each assay is based on our mouse and small animal studies providing insight into variance (e.g. {Meyer, 2021 #1}, see Fig. 7 (Moderna JCI figure). Immunophenotyping and single molecule studies will use n=6 samples for analysis.

Blood is a key tissue for analysis in mouse and human to facilitate translation of signatures and mechanisms across species. We will study the local site and regional LNs at day 1 and profile LNs, BM and spleen until 4 months. Development of GC in regional LN are a key sign of durable immunity to vaccination {Turner, 2021 #50;Turner, 2020 #51;Kil, 2019 #52}{Suan, 2017 #106}. Bone marrow and spleen assays will monitor migration of key cell types identified through other immunological tissues. Long-lived plasma cells, responsible for maintaining circulating antibody levels long term, reside in BM, and spleen has also been implicated in IM formation{Mamani-Matsuda, 2008 #84}{Crotty, 2003 #53;Manz, 1997 #54}{Slamanig, 2021 #55}. In all immune tissues, we will focus on enriched B and T cell analysis, with some broader analysis included to detect important cell types not in B and T compartments.

A previous transcriptome early prediction study of individual variation to YF-vax, which shows only a 10-fold difference in long term immunity levels, was adequately powered with 15 subjects {Querec, 2009 #132}. In order to provide adequate resolution and reproducibility, we will study the 18 YF-vax subjects available and will target 30 subjects for JEV and licensed influenza vaccines obtained by NAMRU-2. NAMRU-2 samples will

be obtained at days 0, 1, 3, 7 and 10 days for deep molecular profiling, and at 1 month, 6 months, and ongoing 6 month intervals for assessing durability. For YF-vax we are using samples already archived from 18 individuals collected at days 0, 7, 30 and 6 months which should be adequate for studying early responses and for evaluating early marker correlation with individual variation that occurs . The Moderna RNA-1010 trial is currently still being planned and the sampling times are not yet determined.

Single cell assay feasibility studies:
All single cell platforms are well established in team laboratories. Shown in Fig. 8-10 are examples of experiments performed using CITE-seq, DBiT-seq and sn mC3-seq.

Feasibility data for computational methods:

We provide feasibility data for MAGICAL in section (a) (see Fig. 5) and for other methods in our publications. Here we demonstrate the application of the xnnet platform described previously (see Fig. 4) to develop an interpretable neural net model for predicting kidney transplant rejection. The learned network (Fig 11a) included many immunological terms associated with the hidden nodes and showed high performance, with a ROC AUC of 0.95 on the held-out samples. Overall, the network clearly separated the two classes (Fig 11b). A frequent limitation of standard classifiers is the inability to explain why a new observation is assigned to a specific class. To address this problem, we mapped each observation from the original input state (the measured genes) to the activation state of the hidden nodes. Each individual sample can be represented as a vector of activation levels corresponding to the different hidden nodes, and visualized as a hidden state radar plot (Fig 11c). This analysis reveals what functions and processes are characteristic of the two groups of samples and how they drive the classification decision.

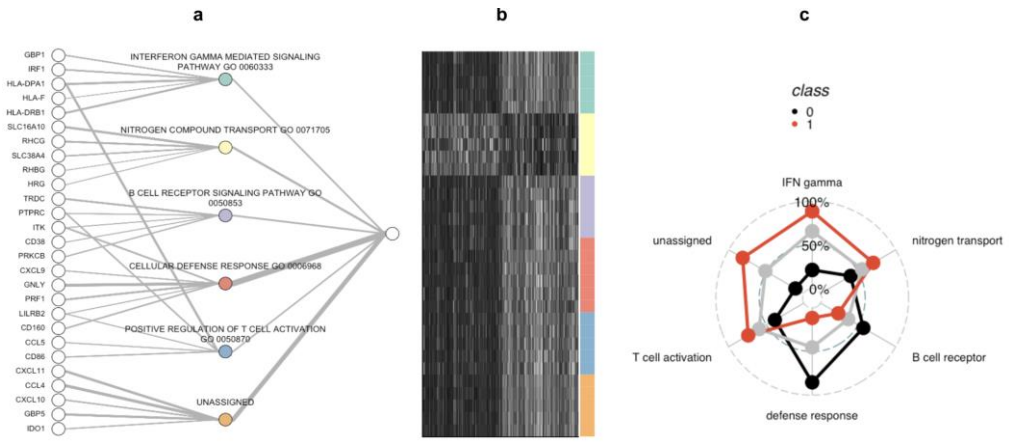


Fig. 11. Interpretable classification by xnnet. To predict kidney transplant recipient rejection, the learned network hidden nodes included various aspects of an immunological response, including Interferon Gamma signalling pathway, B cell receptor, cellular defense response, and regulation of T cell activation (A) and clearly separated the two classes, as evident in the gene expression of selected features (B). Representing each sample as a vector of activation levels corresponding to the hidden nodes in the network, the radar plot summarizes the activation states for groups of samples in the two classes (C).

c. Schedule

To manage the work efficiently, we divide the project into four segments: (1) Sample Acquisition, (2) Data Acquisition, (3) Bioinformatics and Modeling (4) Validation. The segments are further broken down into working areas: *Sample Acquisition*: mouse, human, sample management; *Data acquisition:* immune monitoring, CITE/LIBRA-seq, DBiT-seq Assays, sn m3C-seq, sn RNA/ATAC-seq, public data curation, privately shared data curation; *Bioinformatics and Modeling:* data management, data integration, statistics, BCR/TCR clone analysis, multiomics trajectory reconstruction, human/mouse functional homology mapping, feature selection, ML, mechanism interference; *Validation:* robustness analyses, conditional knockouts, IV&V replication.

MAC 0-6: In the first six months, Sample Acquisition, Data Acquisition, and Bioinformatics and Modeling teams set the foundation of their respective work breakdown structures. ISMMS and UTMB begin and complete six parallel vaccination studies through 120 dpv in C57BL/6J mice for acute (6 MAC Interim Milestone, Tasks 1.1.2.1-1.1.2.6) which establish MOPs for tissue processing. Data production teams optimize protocols and establish MOPs for data generation in mouse tissue from the Sample Acquisition teams (6 MAC Interim Milestone, Tasks 1.1.1, 1.1.3). The ISMMS data management team complete data QC, storage and data transmission pipelines to service Best-Vax and IV&V team needs (6 MAC Interim Milestone, Task 1.2). Lastly, bioinformatics leads at ISMMS model data generated up to 120 dpv from the 6 vaccine perturbations to define IM durability and determine the range of durability across the perturbations. (6 MAC Interim Milestone, Task 1.1.2.7)

MAC 7-12: Data Acquisition teams complete early timepoint vaccine studies (8 MAC Interim Milestone, Task 1.3.2) followed by immune and sc analyses one month later (9 and 10 MAC Interim Milestones, Tasks (1.3.1, 1.3.3, 1.3.4). As the long term mouse study continues, at 12 MAC data production teams complete 8 month serial blood immune profiling and challenge studies on the 6 vaccine perturbations initiated in the first 6 months (12 MAC Milestone, Task 1.3.5, 1.3.6). Sample Acquisition teams continue working to obtain human study samples.

MAC 13-18: In Phase 1 Q3, Data Acquisition teams complete long term follow up C57BL/6J vaccination experiments (13 MAC Interim Milestone, Task 1.7.1) and begin the transition to human data generation as Sample Acquisition teams obtain initial samples for the human vaccination studies (18 MAC Interim Milestone, Tasks 1.5.2, 1.5.4, 1.5.6).

MAC 19-24: In the final quarter of Phase 1, Data Acquisition teams begin and complete data generation for the initial human study samples received across the four parallel studies (24 MAC Milestones, Tasks 1.5.2, 1.5.4, 1.5.6, 1.5.8, 1.6). In addition, they complete additional sc molecular analyses performed on studies through 240 dpv informed by the Bioinformatics and Modeling team, who have begun work on developing roadmap to immune memory (21 MAC Interim Milestone, Task 1.7.6). Modeling teams will implement trajectory analysis, differential analysis and ML analysis to determine functionally homologous early markers of immune memory in mouse and human data (21 MAC Interim Milestone, Tasks 1.7.7, 1.7.8), and integrate relevant knowledge from the completed outside data compendium (24 MAC Milestone, Task 1.4) to deliver a predictive model from early features of level of durable memory after mouse vaccination (24 MAC Milestone, Task 1.7.10).

MAC 25-36: In Phase 2 Year 1, Sample Acquisition teams phase out after obtaining complete sample sets for the human studies (32 MAC Interim Milestone, Tasks 1.9.1, 1.9.2, 1.9.3, 1.9.4). Data Acuisition completse immune and sc data generation for human study samples and transitions to validation studies (34 and 36 MAC, Tasks 1.10, 1.13). Having analyzed all datasets generated in the program thus far, Modeling teams will select the most informative biomarkers of immune memory, specifically the early durable immune pathway mechanistic nodes that contribute to B cell maturation to be tested by knockout studies (36 MAC Milestone, Task 1.12)

MAC 37-48: In Year 4, Modeling and Data Production teams will experimentally validate the key early prediction nodes identified at 36 MAC (46 MAC Milestone, Task 1.14) while Bioinformatics teams complete investigation of early predictors of human/mouse individual variation in vaccine durability. (48 MAC Interim Milestone, Task 1.16). With key IM features and early prediction nodes identified, Modeling teams will develop

a predictive model based on targeted analyses and assemble an R&E tool for predicting the duration of protection following vaccination. The tool will understand the contributions of at least 3 conserved pathways to immune memory formation, provide an early prediction of a vaccine's ability to provide long-lasting protection and proof-of-concept for pathogen-specific individual immune assessment. (48 MAC Milestones, Task 1.17)

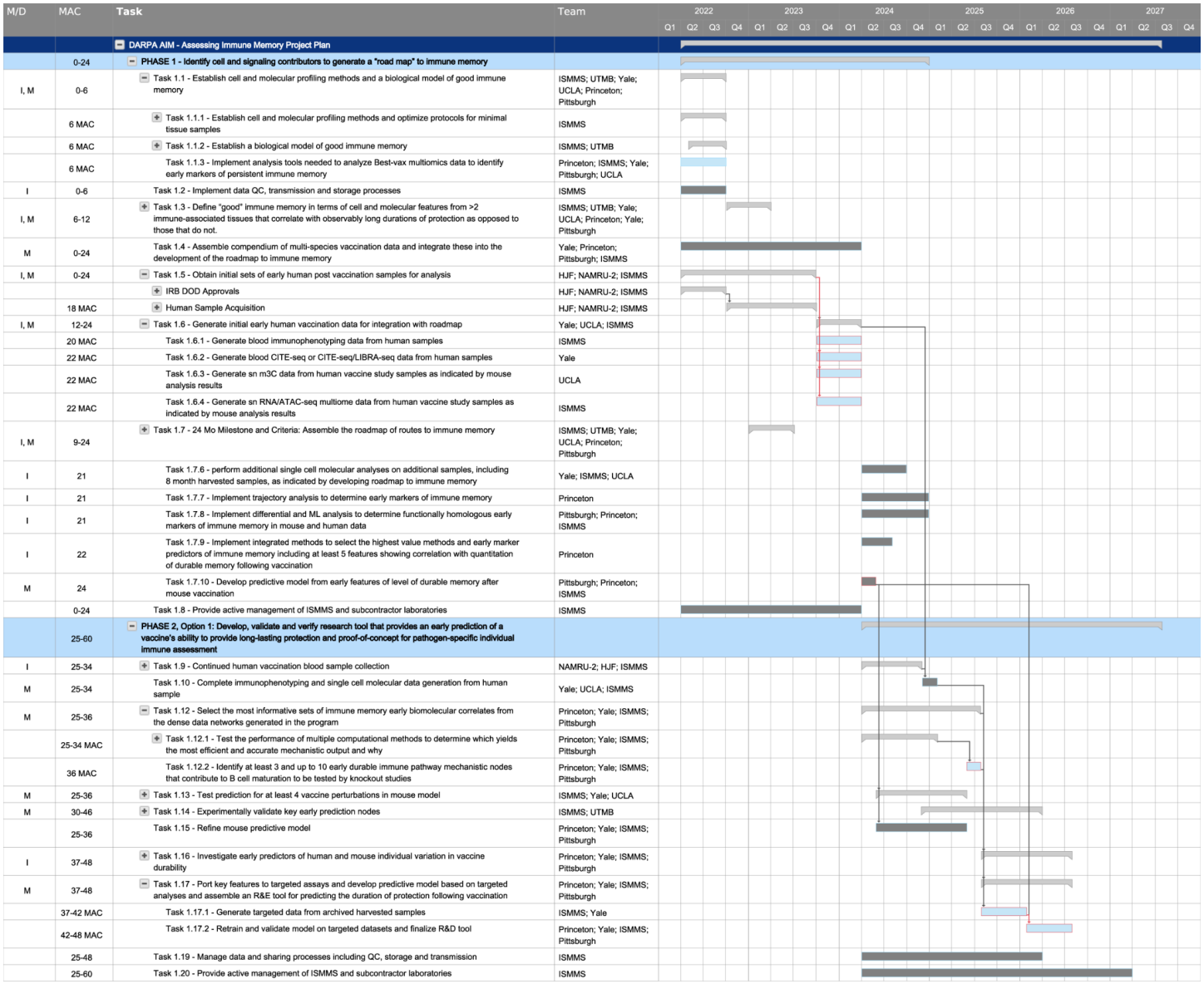
49-60 MAC: After transfer of the R&E tool, the Sample Aquisition, Data Acquisition and Modeling teams will support the IV&V with validation of prediction of individual protection, ensuring AUROC ≥80% of level of protection prior to pathogen challenge is confirmed.

Milestone Dates

- **6 MAC Milestone:** Establishment of cell and molecular profiling methods and protocols and validation of methods and procedures from tissue harvesting to high quality data generation. (Tasks 1.1.1, 1.1.3)
- **12 MAC Milestone:** Complete 8 month serial blood immune profiling and challenge studies on the 6 vaccine perturbations initiated in the first 6 months in VSV-ZEBOV, RNA-1273 Prime-boost and RNA-1273 boost vaccinated mice (Task 1.3.5, 1.3.6)
- **24 MAC Milestone:** Initial samples from all human vaccination studies obtained (Tasks 1.5.2, 1.5.4, 1.5.6, 1.5.8)
- **24 MAC Milestone:** Compendium of multi-species vaccination data developed and integrated into the development of the roadmap to immune memory (Task 1.4)
- **24 MAC Milestone:** Generation of initial early human vaccination data completed for integration with roadmap (Task 1.6)
- **24 MAC Milestone:** Develop predictive model from early features of level of durable memory after mouse vaccination (Task 1.7.10)
- **34 MAC Milestone:** Generation of immune monitoring and single cell molecular data from human samples complete. (Task 1.10)
- **36 MAC Milestone:** Test prediction for 5 additional influenza vaccine perturbations in mouse model (Task 1.13)
- **36 MAC Milestone** Select the most informative sets of immune memory early biomolecular correlates from the dense data networks generated in the program as described in Task 1.12.1 and 1.12.2. (Task 1.12)
- **46 MAC Milestone:** Key early prediction nodes experimentally validated (Task 1.14)
- **48 MAC Milestone:** Port key features to targeted assays and develop predictive model based on targeted analyses and assemble an R&E tool for predicting the duration of protection following vaccination (Task 1.17)
- **48 MAC Milestone:** Determine contributions of ≥3 conserved pathways to immune memory formation
- **48 MAC Milestone:** Research tool that provides an early prediction of a vaccine's ability to provide long-lasting protection and proof-of-concept for pathogen-specific individual immune assessment.
- **60 MAC Milestone:** Validate R&E tool prediction of individual protection at AUROC ≥80% of level of protection prior to pathogen challenge.

Critical Path: The Critical Path begins with defining "good" immune memory (12 MAC Deliverable 1.3) determined from the good IM model (6 MAC Deliverable 1.1.2) generated in six parallel C57BL/6J vaccination studies performed at ISMMS and UTMB (6 MAC Milestone 1.1.1.2-1.1.1.6). It then continues with the development of the Roadmap to Immune Memory (24 MAC Deliverable 1.7) followed by validation of IM Roadmap transferability to humans (36 MAC Deliverable 1.11). The down selection of the best key features of IM (36 MAC Milestone 1.12) utilizing the analyses of the human vaccine studies data (34 MAC Milestone 1.10) obtained from NAMRU-2, HJF and ISMMS (32 MAC Milestone 1.9) are then followed by the investigation of early predictors (48 MAC Interim Milestone 1.16) and development of the R&E tool with key

features (48 MAC Deliverable 1.17). Finally, transfer of the R&E tool (60 MAC Deliverable 1.18) with final acceptance completes the project.



M/D	MAC	Task	Team
		DARPA AIM - Assessing Immune Memory Project Plan	
	0-24	PHASE 1 - Identify cell and signaling contributors to generate a "road map" to immune memory	
I, M	0-6	Task 1.1 - Establish cell and molecular profiling methods and a biological model of good immune memory	ISMMS; UTMB; Yale; UCLA; Princeton; Pittsburgh
	6 MAC	Task 1.1.1 - Establish cell and molecular profiling methods and optimize protocols for minimal tissue samples	ISMMS
	6 MAC	Task 1.1.2 - Establish a biological model of good immune memory	ISMMS; UTMB
	6 MAC	Task 1.1.3 - Implement analysis tools needed to analyze Best-vax multiomics data to identify early markers of persistent immune memory	Princeton; ISMMS; Yale; Pittsburgh; UCLA
I	0-6	Task 1.2 - Implement data QC, transmission and storage processes	ISMMS
I, M	6-12	Task 1.3 - Define "good" immune memory in terms of cell and molecular features from >2 immune-associated tissues that correlate with observably long durations of protection as opposed to those that do not.	ISMMS; UTMB; Yale; UCLA; Princeton; Yale; Pittsburgh
M	0-24	Task 1.4 - Assemble compendium of multi-species vaccination data and integrate these into the development of the roadmap to immune memory	Yale; Princeton; Pittsburgh; ISMMS
I, M	0-24	Task 1.5 - Obtain initial sets of early human post vaccination samples for analysis	HJF; NAMRU-2; ISMMS
		IRB DOD Approvals	HJF; NAMRU-2; ISMMS
	18 MAC	Human Sample Acquisition	HJF; NAMRU-2; ISMMS
I, M	12-24	Task 1.6 - Generate initial early human vaccination data for integration with roadmap	Yale; UCLA; ISMMS
	20 MAC	Task 1.6.1 - Generate blood immunophenotyping data from human samples	ISMMS
	22 MAC	Task 1.6.2 - Generate blood CITE-seq or CITE-seq/LIBRA-seq data from human samples	Yale
	22 MAC	Task 1.6.3 - Generate sn m3C data from human vaccine study samples as indicated by mouse analysis results	UCLA
	22 MAC	Task 1.6.4 - Generate sn RNA/ATAC-seq multiome data from human vaccine study samples as indicated by mouse analysis results	ISMMS
I, M	9-24	Task 1.7 - 24 Mo Milestone and Criteria: Assemble the roadmap of routes to immune memory	ISMMS; UTMB; Yale; UCLA; Princeton; Pittsburgh
I	21	Task 1.7.6 - perform additional single cell molecular analyses on additional samples, including 8 month harvested samples, as indicated by developing roadmap to immune memory	Yale; ISMMS; UCLA
I	21	Task 1.7.7 - Implement trajectory analysis to determine early markers of immune memory	Princeton
I	21	Task 1.7.8 - Implement differential and ML analysis to determine functionally homologous early markers of immune memory in mouse and human data	Pittsburgh; Princeton; ISMMS
I	22	Task 1.7.9 - Implement integrated methods to select the highest value methods and early marker predictors of immune memory including at least 5 features showing correlation with quantitation of durable memory following vaccination	Princeton
M	24	Task 1.7.10 - Develop predictive model from early features of level of durable memory after mouse vaccination	Pittsburgh; Princeton; ISMMS
	0-24	Task 1.8 - Provide active management of ISMMS and subcontractor laboratories	ISMMS
	25-60	PHASE 2, Option 1: Develop, validate and verify research tool that provides an early prediction of a vaccine's ability to provide long-lasting protection and proof-of-concept for pathogen-specific individual immune assessment	
I	25-34	Task 1.9 - Continued human vaccination blood sample collection	NAMRU-2; HJF; ISMMS
M	25-34	Task 1.10 - Complete immunophenotyping and single cell molecular data generation from human sample	Yale; UCLA; ISMMS
M	25-36	Task 1.12 - Select the most informative sets of immune memory early biomolecular correlates from the dense data networks generated in the program	Princeton; Yale; ISMMS; Pittsburgh
	25-34 MAC	Task 1.12.1 - Test the performance of multiple computational methods to determine which yields the most efficient and accurate mechanistic output and why	Princeton; Yale; ISMMS; Pittsburgh
	36 MAC	Task 1.12.2 - Identify at least 3 and up to 10 early durable immune pathway mechanistic nodes that contribute to B cell maturation to be tested by knockout studies	Princeton; Yale; ISMMS; Pittsburgh
M	25-36	Task 1.13 - Test prediction for at least 4 vaccine perturbations in mouse model	ISMMS; Yale; UCLA
M	30-46	Task 1.14 - Experimentally validate key early prediction nodes	ISMMS; UTMB
	25-36	Task 1.15 - Refine mouse predictive model	Princeton; Yale; ISMMS; Pittsburgh
I	37-48	Task 1.16 - Investigate early predictors of human and mouse individual variation in vaccine durability	Princeton; Yale; ISMMS; Pittsburgh
M	37-48	Task 1.17 - Port key features to targeted assays and develop predictive model based on targeted analyses and assemble an R&E tool for predicting the duration of protection following vaccination	Princeton; Yale; ISMMS; Pittsburgh
	37-42 MAC	Task 1.17.1 - Generate targeted data from archived harvested samples	ISMMS; Yale
	42-48 MAC	Task 1.17.2 - Retrain and validate model on targeted datasets and finalize R&D tool	Princeton; Yale; ISMMS; Pittsburgh
	25-48	Task 1.19 - Manage data and sharing processes including QC, storage and transmission	ISMMS
	25-60	Task 1.20 - Provide active management of ISMMS and subcontractor laboratories	ISMMS

Best-Vax program timeline

d. Challenges and Risks

Overfitting and robustness: Important challenges to the assessing immune memory program are the difficulties in identifying reliable signatures from a small cell population that are, robust and generalizable from mouse to human and across vaccine perturbations. Overinterpretation and overfitting is recurrent issue in biological research. Large scale replication studies of high profile cancer papers, for example, have found that as many as 90% are not reproducible despite working with the original authors{, 2017 #158}{Wen, 2018 #159}. This problem results in part from the mismatch between sample size and measurements, which leads to overfitting and unreproducible findings. To overcome this key obstacle, we have rigorous experiments, data generation and QC methods. Importantly, the informatics

approaches we will apply to this problem are designed to use outside information in the discovery process to improve reproducibility and generalizability of the results. The Best-vax computational immunology and bioinformatics researchers are world-leaders in developing methods that improve the robustness of data analysis and machine learning models by incorporation of massive outside data and prior information {Chen, 2021 #8;Cofer, 2021 #47;Greene, 2015 #67;Lee, 2018 #9;Yao, 2018 #48;Zhou, 2019 #102;Zhou, 2020 #29}{Mao, 2019 #7}. These approaches address the p>>N (large number of measurements in relatively few experimental samples) ML problem by using external data for regularization. Analyzing results against public and shared data from many laboratories is an important strategy to avoid problems with reproducibility {Sweeney, 2017 #157}. As we demonstrate in the results shown on MAGICAL in section (a) public data can be valuable for assessing accuracy of results even when the data acquired differs from the data analyzed (e.g. using bulk data from a different cohort to evaluate the accuracy and reproducibility of predictions based on single cell multiomics data analysis). Our team has also developed more accurate and sensitive methods for B cell clonal analysis {Jiang, 2020 #100;Nouri, 2018 #145;Nouri, 2020 #144;Turner, 2020 #46;Zhou, 2019 #146}. The same emphasis on reproducibility and generalizability will be applied to a robustness-based selection among different analysis methods considered to generate robust identification of early pathway mechanisms that reflect and a model that predicts durable immunity

Replication across laboratories: Results of animal immunological studies can vary in different labs, for example due to differences in microbiome {de Jong, 2020 #131}. In order to develop a robust R&E tool that can be readily transferred to different laboratories, an important aspect of our project design is that the animal studies are being performed at two institutions.

Generalizability of results for different vaccine modalities, pathogens and individuals. Protection conferred by vaccination results from complex interactions between innate and adaptive immunity. There is large individual variation in vaccine effectiveness, and the requirements for effective vaccination differs for acute and chronic infections. We study only vaccines for acute infections which have similar demands for an effective early immune response to infection. Furthermore, we hypothesize that there will be core molecular signals of early IM cells that are relevant to different vaccine technologies and to different acute pathogens. To improve the generalizability of the R&E tool, while we focus on influenza vaccines, we are also including vaccines for other diseases as well as a range of vaccine technologies. We also hypothesize that the same early mechanisms predicting the differences in durable memory among vaccines will also reflect inter-individual differences in durable IM from the same vaccine. .

Mapping mouse to human mechanisms: Sequence homologs may serve different functions in mouse and human, whereas different molecules may serve the same function. This makes identification of key analogous nodes in mouse and human challenging. We address this by identifying functional homologs not strictly requiring sequence homology using innovative cross species integration and prediction methods (e.g. Nat. Biotech 2018, Genome Res. 2021 {Cofer, 2021 #47;Yao, 2018 #48}).

Small blood samples in mouse model: Blood is a key link between human and mouse studies. Mouse blood samples are small--about 100 ul if drawn weekly or 180 ul drawn every two weeks. This provides a challenge to doing repeated sampling for single cell studies. We address this by relying on bar-coding CITE-seq for repeated sampling studies, which can be adapted to these small samples, and by obtaining larger blood samples from sacrificed animals for many of the single cell analyses which will not require tracing from the same animal. The developmental mechanisms identified in B cell early in the trajectory towards durable immunity are likely to be recognizable in our experiments without tracing clones, for example via LIBRA-seq labeling or by identifying

changes in comparisons of early samples from vaccines of different durability. Therefore we expect blood analysis from sacrificed mice will provide informative in identifying pathways to immune memory. We are designing the studies to identify early pathway signals in blood whether or not clone tracing or antigen recognition is required.

Determining durability at 4 months in mouse model: It may be difficult to assess long term stability following animals only until 120 dpv, which is a timepoint required to meet the 6 month milestone. To provide further confirmation of the assessment of durability at 120 dpv we will continue to follow the initial range of vaccine perturbations in mouse until 240 dpv. In addition we will perform limited study of the single cell molecular changes indicative of durable responses at this later timepoint.

e. Personnel responsible for each major task

Task	Lead	Support
PHASE 1		
1.1 Methods & IM Mouse Model (0-6MAC)	**Sealfon.**(.05)	**Smith**(.25)Ge(.1)Analyst(.4)sample manag(.15)
Immune Monitoring	**Ramos-Lopez**(.1)	scientist(.5)postdoc(.5)lab manag(.1)tech(.5)
single cell analyses	**Tsang**(.05)	scientist(.35)tech(1.07)**Fan**(.03)postdoc(.5)**Nair**(.15),tech(1.5) **Luo**(.03)posdoc(..5)tech(.08) **Kleinstein** (.02)res assoc(.4)
mouse hi path vaccines	**Bukreyev**(.1)	tech(.75)scientist(1.5)
mouse: IAV vaccines	**Garcia-Sastre**(.04)	**Schotsaert**(.05),postdoc(.8)tech(.4)
1.2 Data Management (0-24MAC)	**Nudelman**(.4)	DB programmer(40%)
1.3 "Good" IM (6-12MAC)	**Sealfon**(.05)	**Smith**(.25)Ge(.1)Analyst(.75)sample manag(.10)
Immune Monitoring	**Ramos-Lopez**(.1)	scientist(.5)postdoc(.5)lab manag(.15)tech(.5)
single cell analyses	**Tsang**(.1)	scientist(.35)postdoc(.75)tech(1.08)**Fan**(.03)postdoc(.5) **Nair**(.15),tech(1.5) **Luo**(.03)posdoc(..5)tech(.08) **Kleinstein** (.02)tech(.45)
mouse hi path vaccines	**Bukreyev**(.1)	tech(.75)scientist(1.5)
mouse: IAV vaccines	**Garcia-Sastre**(.04)	**Schotsaert**(.05),postdoc(.75)tech(.30)
analysis	**Troyanskaya**(.1)	posdoc(2)programmer(.1)**Zaslavsky**(.1)posdoc(.5)
1.4 Outside data compendium (0-24MAC)	**Kleinstein**(.02)	grad student(1)
1.5 Obtain initial human samples (0-24MAC)	**Letizia**(.4*)	**Christy**(.3*) **George**(.10)samp man(.1)**Kreb**(.02)**Modjarrad**(.03)Res support(1.2)
1.6 Initial human data (13-24 MAC)		
Immune Monitoring	**Ramos-Lopez**(.1)	scientist(.5)postdoc(.5)lab manag(.15)tech(.5)samp manag(0.03)
single cell analyses	**Tsang**(.09)	scientist(.25)tech(1.07)**Nair**(.15),tech(.5)**Luo**(.03)posdoc(.2)
1.7 Roadmap construction (9-24MAC)	**Sealfon**(.1)	
mouse hi path vaccines	**Bukreyev**(.1)	tech(.75)scientist(1.5)
mouse: IAV vaccines	**Garcia-Sastre**(.02)	**Schotsaert**(.05),postdoc(.7)tech(.3)
single cell analyses	**Tsang**(.09)	scientist(.27)tech(2.15)**Fan**(.06)postdoc(1)**Nair**(.15),tech(1.5) **Luo**(.03)posdoc(..5)tech(.08)**Kleinstein**(.02)tech(.85)
informatics and modeling	**Zaslavsky**(.2)	**Smith**(.4)Ge(.2)Postdoc(.8)Analyst(.9)**Troyanskaya**(.1)posdoc(2)programmer(.1)
1.8 Active management(0-24MAC)	**Sealfon** (15%)	**Ramos-lopez**(50%), **Zaslavsky**(20%),ops manag(50%), Admin(40%)
PHASE 2 OPTION		
1.9 Human samples(25-32MAC)	**Letizia**(.4*)	**Christy**(.3*) **George**(.10)sample manag(.1) **Kreb**(.02)**Modjarrad** (.03)Res support(1.2)
1.10 Human Data (25-38MAC)		
Immune Monitoring	**Ramos-Lopez**(.1)	scientist(.5)postdoc(.5)lab manag(.15)tech(.5)samp manag(0.05)
single cell analyses	**Tsang**(.05)	scientist(.75)tech(1)**Fan**(.03)postdoc(.5)**Nair**(.15),tech(1.5) **Luo**(.03)posdoc(..5)tech(.08) **Kleinstein** (.02)
1.11 Mouse-human mapping (25-36MAC)	**Troyankaya**(.1)	postdoc(1)Zaslavsky(.1)posdoc(.5)
1.12 Identify nodes (25-36MAC)	**Troyanskaya**(.1)	posdoc(1)Zaslavsky(.1)posdoc(.5)
1.13 Test 5 pertubations	**Sealfon**(.1)	
mouse studies.	**Garcia-Sastre**(.05)	**Schotsaert**(.07),postdoc(1.2),tech(.7)
Immune Monitoring	**Ramos-Lopez**(.1)	scientist(.5)postdoc(.5)lab manag(.15)tech(.5)samp manag(0.05)
single cell analyses	**Tsang**(.05)	scientist(.75)tech(1)**Fan**(.03)postdoc(.5)**Nair**(.15),tech(1.5)**Luo**(.03)posdoc(..5)tech(.(
informatics and modeling	**Troyanskaya**(.1)	postdoc(2)**Zaslavsky**(.2)Smith(.4)Ge(.2)Postdoc(8)Analyst(.9)
1.14 knockout studies		
IAV perturbation	**Garcia-Sastre**(.05)	**Schotsaert**(.07),postdoc(.8),tech(.4)
Ebola perturbation	**Bukreyev**(.1)	tech(.75)scientist(1.5)
Immune Monitoring	**Ramos-Lopez**(.1)	scientist(.5)postdoc(.5)lab manag(.15)tech(.5)samp manag(0.05)
1.15 Model refinement	**Troyanskaya**(.1)	**Sealfon**(.1)**Zaslavsky**(.2)Smith(.4)Ge(.2)Postdoc(8)Analyst(.9)
1.16 Individual variation	**Troyanskaya**(.1)	**Zaslavsky**(.05)Smith(.35)Ge(.15)analyst(.35)postdoc(.8)) **Tsang**(.1)scientist(.3)posdoc(1)tec(.05)**Kleinstein**(.03)tech(1.45)
1.17 Downsampled R&E tool		
Downsampled assays	**Nair**(.15)	scientist(.5)postdoc(.5)lab manag(.1)tech(.5)**Fan**(.03)postdoc(.5) **Ramos-Lopez**(.1)**Tsang**(.05)scientist(.2)postdoc(1)Tech(.5)
Final model	**Zaslavsky**(.2)	**Troyanskaya**(.2)Smith(.4)Ge(.2)Postdoc(8)Analyst(.9)**Kleinstein**(.02)tech(.4)
1.18 Transfer R&E to IV&V	**Zaslavsky** (.1)	**Garcia-Sastre**(.05)**Schotsaert**(.05)**Sealfon**(.05)**Ramos-lopez**(.1)scientist(.25)Nudelm **Tsang**(.05)scientist(.3)**Kleinstein**(.05)tech(.33)**Fan**(.06)posdoc(1)**Troyankaya**(0.5
1.19 Data management	**Nudelman**(.3)	DB programmer(.3)
1.20 Active management	**Sealfon** (.15)	**Ramos-Lopez**(50%), **Zaslavsky**(20%),ops manager(50%), Admin(40%)

*active military personnel

leads and subproject leads are **bold**

D. Management Plan:

Program PI Dr. Sealfon (25%): Dr. Sealfon, director of the Center for Advanced Research on Diagnostic Assays, is the PI of the Best-vax project. He has the primary responsibility to plan, execute, and direct the research in order to reach DARPA AIM objectives. Assisted by his extraordinary management team and effective tracking tools, he maintains granular knowledge of all aspects of program execution and coordination and will make certain proactive interventions are initiated when need to stay on schedule. He is both a strong leader and a team player, who welcomes working closely with the AIM PM and SETA leads to meet and exceed program objectives. Dr. Sealfon has decades of success leading integrated, multidisciplinary research teams. He is the current PI of the DARPA ECHO program (PM: Dr. Eric Van Gieson), CO-I of the DARPA

MBA program (PM: Dr. Eric Van Gieson), and leads the NIH U24 Common Fund Epigenomic Mapping Site for the Molecular Transducers of Physical Activity Consortium (MoTRPAC, NIH program coordinator). Dr. Sealfon's 360° complex research project vision/execution is recognized by government program officers. Dr. Eric Van Gieson, DARPA PM overseeing the ISMMS ECHO program comments that "As prime contractor for a DARPA ECHO program, Dr. Sealfon and his Icahn team have closely managed a program with twelve subcontractors, including tracking, formally reviewing and, if needed, replacing subcontractors in order to continue to excel at meeting program objectives." Program Officer Dr. Tim Gondre-Lewis who oversaw the NIH U19 Modeling Immunity for Biodefense Consortium contracts from 2004-2018 wrote in his most recent Contractor Performance Assessment Reports (CPAR) "Given what I know today about the contractor's ability to execute what they promised in their proposal, I definitely would award to them today."

Project Manager Dr. Ramos (75%) has the ideal scientific and program management background to lead, control, coordinate and evaluate the Best-Vax research teams and to serve as primary contact with the DARPA PM and SETA leads. As an expert at the intersection of innate immune responses to virus infection, adaptive immune responses, and sc technologies, she will lead the integrated management of the subcontract teams and choreograph the synchronized studies occurring across institutions, ensuring coordinated efforts. Dr. Ramos has collaborated with almost all the Best-Vax subcontractors. She has a strong background in large-scale multi-center project coordination, serving as scientific project manager of the NIH funded Human Immunology Project Consortium (HIPC) led by Dr. Fernandez-Sesma during 2014-2018, and she presently a key leader of several complex, time-sensitive projects for the DARPA-NMRC COVID-19 Health Action Response for Marines (CHARM) program.

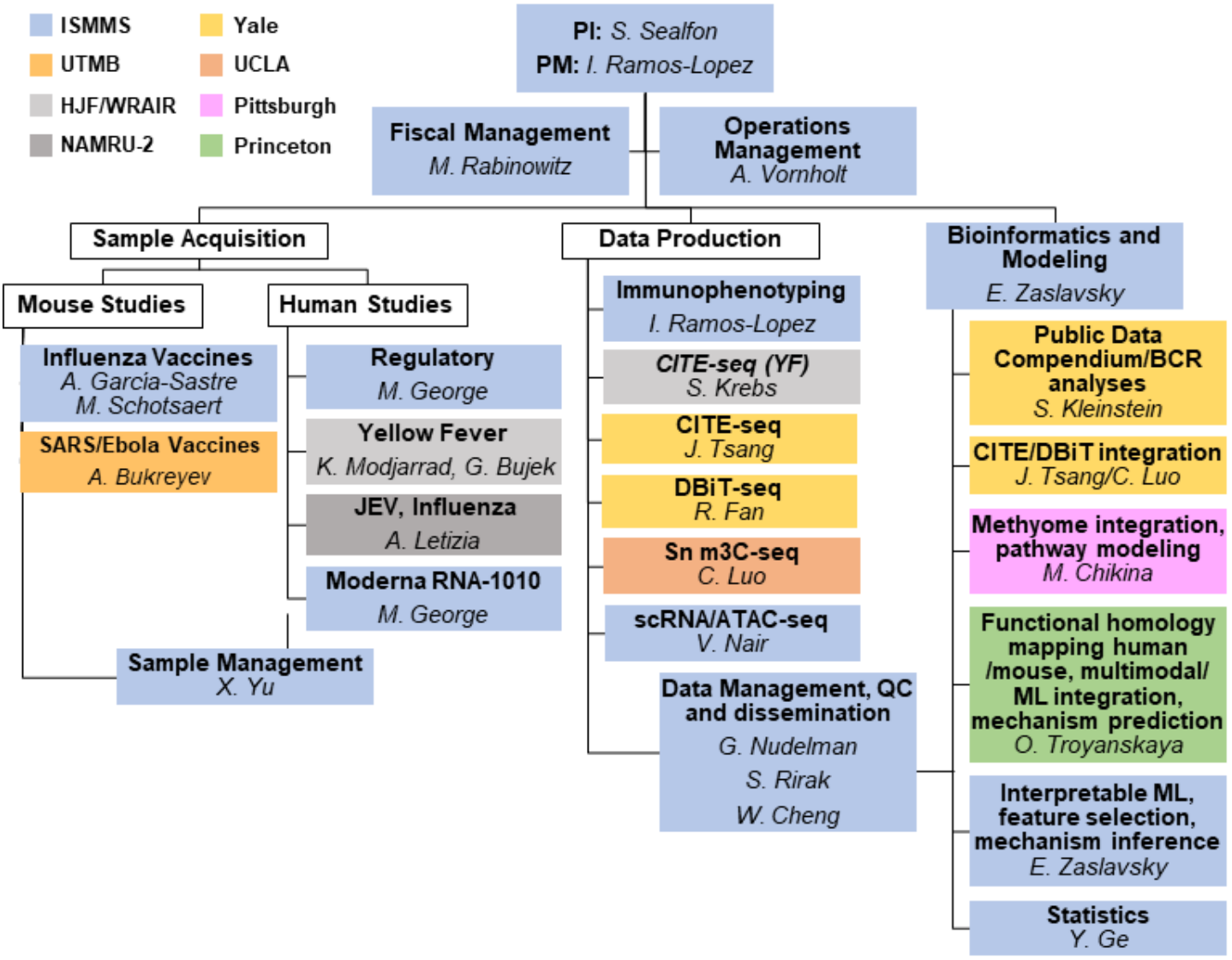
Operations Manager Alexandria Vornholt (50%) will support the Best-Vax team administratively to establish detailed tracking and task monitoring as well as ensure our quick response to DARPA programmatic needs. Mrs. Vornholt has over 10 years of experience in project and operations management and tools. She currently manages the interdisciplinary Genomics, Epigenetics, Transcriptomic working group for the MoTRPAC consortium and she manages DARPA and NIH programs across the Sealfon Lab. With experience producing DARPA-grade reports and managing cross-functional teams in conjunction with breathtaking organizational skills, efficiency and attention to detail, she will be an extraordinary asset to the Best-Vax management team. She will coordinate ISMMS and Best-Vax subcontractor action items and project tracking into a master Smartsheet dynamic management system to closely monitor the technical, programmatic, financial and contractual progress and obligations. She will assist Drs. Sealfon and Ramos in early detection of execution and coordination issues and rapid implementation of corrective actions to achieve program milestones and deliverable on schedule and budget..

Management Structure and Approach:

Our organizational chart is shown. The Best-vax team includes an ideal combination of world-renowned labs in vaccine development, systems and computational immunology, bioinformatics, single cell biology and transitioning to achieve the ambitious goals of AIM. Importantly, our capacity to work together seamlessly and efficiently on complex, high-pressure contract research is already fully established. The Sealfon laboratory has collaborated extensively and closely for decades on dozens of projects and publications with nearly all Best-Vax teams (including Garcia-Sastre, Bukreyev, Troyankaya, Chikina, Kleinstein, Tsang, Letizia), which allows us to leverage unparalleled expertise with highly agile and natural coordination of efforts. We stand out for our ability to actively manage large scale collaborative research and are unique ih our culture of total integration of efforts and personnel across laboratories. In successfully executing the DARPA-Navy CHARM program, we were key to standing up in a month and managing a 3000 subject, 20,000 biosample/month

collection, data production and analysis effort including recruiting, training and deploying field teams, solving supply train issues, massively scaling throughput in real time and managing a large team of subcontractors, collaborators and vendors, all during the peak of the first massive COVID-19 surge in NYC. Within 12 months, we published seven manuscripts at journals with an average impact factor of 26.5 and contributed to COVID-19 control measures in the US military and world-wide. Actively managing Best-Vax to meet program milestones is a comparative cakewalk.

This unique collaborative and performance-oriented culture is augmented by our active management structure and close tracking. Our management tools are SOA, including webex, SLACK, Smartsheet project tracking, as well F2F, text and emails. We are tool agnostic and adapt to what works best for the task. Harnessing these tools and best practices, the PI and PM will host weekly technical and management teleconferences as well as weekly subgroup meetings for each major TA1 component (sample acquisition, mouse challenge studies, assay processing, data analysis) to maintain coordination and agility. These will be supplemented by biweekly meetings with key personnel, in-person project meetings as needed, site visits and AIM program meetings as described.



E. Capabilities:

The Best-vax team, who are DARPA-tested and have collaborated for decades, comprises renowned vaccine developers, systems immunologists, inventors of the most advanced, best-in-class single cell methods, and of mission-critical new informatics and machine learning methods.

Stuart Sealfon (Program PI, Director ISMMS Center for Advanced Research on Diagnostic Assays, h index=66,cites=16454) leads a DARPA ECHO project and plays a key role in a DARPA MBA project. He led an NIH-funded Modeling Immunity for Biodefense consortium from 2001-2016 and is a PI in the NIH Common Fund MoTrPAC physical activity molecular mapping consortium. His research led to two FDA approved drugs Ariahn and Orilissa. A dynamic research leader, has led or co-authored important advances in systems and single cell biology in diverse areas including immunology, virology and vaccinology. With Drs. Bukreyev, Carfi and

Troyanskaya, he recently reported the single cell analysis of SARS-CoV-2 and miRNA 1273 in hamsters (JCI 2021 {Meyer, 2021 #1}) and with Dr. Letizia characterized the immunity and phylogenetics of SARS-CoV-2 in Marine recruits (NEJM 2020, Lancet Resp Med, 2021 {Letizia, 2021 #32;Letizia, 2020 #33}). **Adolfo Garcia-Sastre** (Director ISMMS Emerging Pathogens Institute, Member NAS, h=143;cites=78770) is one of the world's leading vaccine researchers and virologists. Notable recent work includes advances on SARS-CoV-2 therapeutics (e.g. Science 2021 {White, 2021 #30}). Michael Schotsaert (Assistant Prof Micrbiology ISMMS, h=24,cites=2796) is an expert in adjuvants and vaccine technology (e.g.Angew Chem Int 2021{Jangra, 2021 #153})**Alex Bukreyev** (Prof. UTMB, Galveston lab BSL4, H=52,citations=8437) a leading high pathogen vaccine researcher has made seminal contributions to vaccinology, including a SARS vaccine (Lancet 2004 {Bukreyev, 2004 #36}) Ebola vaccine (comparing 5 vaccines in NHP, Sci Trans Med 2021 {Meyer, 2021 #37}) and SARS-CoV-2 (JCI 2021{Meyer, 2021 #1}). **CMDR Andrew Letizia** (Director NMRC-Asia,h=10,cites=303), a military infectious disease researcher led the landmark CHARM study of SARS-CoV-2 infection in Marine recruits (NEJM 2020, Lancet Resp Med, 2021 {Letizia, 2021 #32;Letizia, 2020 #33}). **Kayvon Modjarrad (**founding Director EID branch WRAIR, h=34,cites=4584) an Operation Warp Speed member, co-leads WRAIR's COVID-19 response efforts (Cell 2020 {Joyce, 2020 #40}) and leads/co-leads recent and ongoing vero-cell YF vaccine trials. **Shelly Krebs** (Chief B cell Biology EID WRAIR,h=24;cites=1519) is a rising star in assessing B cell responses to vaccination and antibody effectiveness (Nat. Med 2021, Cell Host Microbe 2021{Dussupt, 2020 #152;Townsley, 2021 #151}). **John Tsang** (Chief, Multiscale Systems Biology Section, NIAID and Co-director, Center for Human Immunology, NIH, h=28,cites=5500) is an innovative leader in predictive systems immunology. Relevant recent accomplishments include single cell systems biology temporal reconstruction to identify mechanisms leading to severe COVID-19 (Cell 2021 {Liu, 2021 #18}) and predictive modeling of individual variation in the antibody response to vaccination (Nat. Med 2020 {Kotliarov, 2020 #108}. **Rong Fan** (Prof Yale Biomed Eng., fellow AIMBE, h=55,cites=15109) is a leading developer of SC and spatial multi-omics profiling technologies, including the only SC resolution spatial transcriptomics/proteomics and the first spatial epigenetic platforms to be used in Best-vax (Cell 2020, Biorxiv 2021 {Liu, 2020 #5})({Deng, 2021 #41}). **Chongyuan Luo** (Assistant Professor UCLA Genetics, h=20,citations=3040 since 2016) is a pioneer in SC epigenetic methods. He developed the single cell methylation-chromatin conformational capture method (Nat. Comm 2018, {Luo, 2018 #20} Nature Meth 2019 {Lee, 2019 #21}). **Venu Nair** (Assoc. Prof. ISMMS Center for Advanced Research on Diagnostic Assays, h=24,cites=1644) leads the MoTrPAC data production laboratory at ISMMS and is an innovator in sc epigenetics {Nair, 2021 #163;Ruf-Zamojski, 2021 #10;Zhang, 2021 #81}. **Yongchao Ge** (Assoc. Prof. ISMMS, h=33,cites=17608) is an outstanding statistician and omics bioinformaticist. **Olga Troyankaya** (Professor Computer Science Princeton, Flatiron Institute. 2020 ACM fellow, h=62,citations=24108) is a leader in developing comprehensive informatics frameworks to extract signal from high dimensional data and a pioneer in the use of Bayesian network analysis to interpret experimental data by leveraging public data. Relevant recent work includes advances in single cell trajectory analysis (Nature Meth in press {Zhou, 2020 #29}, genetic landscape integration (Nature Gen, 2021 {Park, 2021 #31})and multiomics data integration (Cell Systems 2021 {Chen, 2021 #8}). **Steven Kleinstein** (Prof. Yale Pathology and Computer Science, h=49,cites=9512) is a leading computation immunologist who has developed important new analysis tools. He leads an identification of immune and vaccine signatures project for the NIH Human Immune Project Consortium and has made recent major contributions to the characterization of vaccine signatures and B cell repertoires (Science 2021, Nature 2020, J. Immunol 2020 {Brioschi, 2021 #45;Turner, 2020 #46}{Avey, 2020 #66}). **Maria Chikina** (Assist. Prof. Pitt Computational and Systems Biology, h=22,cites=2069) is a rising star in developing computational

techniques that overcome biases and artifacts inherent in large scale omics datasets. Her methods have led to important contributions to mechanistic understanding of germinal center formation (Nature Immunol. 2020, Immunity 2019 {Weisel, 2020 #43;Trivedi, 2019 #44}) and she developed PLIER, a transformative method for dimensionality reduction (Nat. Meth. 2019 {Mao, 2019 #7}).

Specialized facilities: All Prime and subcontractor laboratories have substantial wetlab and/or drylab footprints equipped with SOA instrumentation and have, access to outstanding core department and institutional facilities and world-class high performance computing resources, as described in Book 2. Unique program resources include Dr. Bukreyev's access to the Galveston National Lab BSL3 and BLS4/ABSL-3/ABSL-4 facilities that allow the SARS-CoV-2 and Ebola challenge studies and the NAMRU-2 DoD research laboratory in Singapore which is uniquely positioned to manage the enrollments of a diverse range of US ADSM and local partner-nation participants who are receiving the targeted vaccines

F. Statement of Work (SOW)

Phase 1 (Base). (0-24 MAC) Generate a cell and molecular "road map" to immune memory

Task 1.1 (0-6 MAC; **ISMMS**, UTMB, Yale, UCLA, Princeton, Pittsburgh). Establish cell and molecular profiling methods and a biological model of good immune memory. ***6 MAC Deliverable:*** *Show >2.5-fold variation in immune memory durability among vaccine perturbations in C57BL/6J mice*

- **Task 1.1.1** (0-6 MAC) Optimize immune and single cell methods for mouse tissue studies
  - **Task 1.1.1.1-.2** (0-1 MAC; **ISMMS, UTMB**) Obtain approval for animal research
  - **Task 1.1.1.3** (0-6 MAC; **Yale**, ISMMS, UTMB) Establish CITE-seq/LIBRA-seq
  - **Task 1.1.1.4** (0-6 MAC; **Yale**, ISMMS, UTMB) Pilot assay for tissue harvesting, cell purification and CITE-seq with biosamples from both ISMMS and UTMB
  - **Task 1.1.1.5** (0-6 MAC; **Yale**, ISMMS, UTMB) Pilot assay for tissue harvesting, cell purification and DBiT-seq with lymph node biosamples from ISMMS and UTMB
  - **Task 1.1.1.6** (0-6 MAC; **Yale**, ISMMS, UTMB) Establish DBiT-seq/LIBRA-seq
  - **Task 1.1.1.7** (0-6 MAC; **UCLA**, ISMMS, UTMB) Pilot assay for tissue harvesting, cell purification and sn m3C-seq with biosamples from both ISMMS and UTMB.
  - **Task 1.1.1.8** (0-6 MAC; **ISMMS**, UTMB) Pilot assay for tissue harvesting, cell purification and snRNA/ATAC multiome with biosamples originating from both ISMMS and UTMB.
  - **Task 1.1.1.9** (0-4 MAC; **ISMMS**, **UTMB**) Validate immune monitoring assay panels
- **Task 1.1.2** (0-6 MAC; **ISMMS**, UTMB) Establish a biological model of good immune memory
  - **Task 1.1.2.1** (0-6 MAC, **UTMB**) VSV-ZEBOV vaccination experiments in C57BL/6J mice
    - **Task 1.1.2.1.1** (0-6 MAC, **UTMB**) Mouse-adapted challenge studies 120 dpv
    - **Task 1.1.2.1.2** (0-6 MAC, **UTMB**) Quantitate serological and nAb to 120 dpv
    - **Task 1.1.2.1.3** (0-6 MAC, **UTMB**) Quantitate immune memory in lymph nodes, spleen and bone marrow at 120 dpv
    - **Task 1.1.2.1.4** (0-6 MAC; **UTMB**) Harvest and process blood, lymph node, spleen and bone marrow samples from 120 dpv
  - **Task 1.1.2.2** (0-6 MAC, **UTMB**) RNA-1273 (SARS-CoV-2) prime boost experiments
    - **Task 1.1.2.2.1-.4** same as 1.1.2.1.1-.4
  - **Task 1.1.2.3** (0-6 MAC, **UTMB**) RNA-1273 prime experiments

Task 1.1.2.3.1-.4 same as 1.1.2.1.1-.4

Task 1.1.2.4 (0-6 MAC, **ISMMS**) Perform TIV+M59 (influenza) experiments

Task 1.1.2.4.1 (0-6 MAC, **ISMMS**) Mouse-adapted challenge studies 120 dpv

Task 1.1.2.4.2 (0-6 MAC, **ISMMS**) Quantitate serological and nAb to 120 dpv

Task 1.1.2.4.3 (0-6 MAC, **ISMMS**) Quantitate immune memory in lymph nodes, spleen and bone marrow at 120 dpv

Task 1.1.2.4.4 (0-6 MAC; **ISMMS**) Harvest and process blood, lymph node, spleen and bone marrow samples from 120 dpv

Task 1.1.2.5 (0-6 MAC, **ISMMS**) Perform standard dose TIV (Influenza) experiments

Task 1.1.2.5.1-.4 same as 1.1.2.4.1-.4

Task 1.1.2.6 (0-6 MAC; **ISMMS**) Perform RNA-1010 (Influenza) experiments

Task 1.1.2.6.1-.4 same as 1.1.2.4.1-.4

Task 1.1.2.7 (0-6 MAC; **ISMMS**) Develop metric for defining IM durability and determine the range of durability of the 6 vaccine perturbations up to 120 dpv

Task 1.1.3 (0-6 MAC, **Princeton**, **ISMMS**, Yale, Pittsburgh, UCLA) Implement analysis tools For multiomics data to identify early markers of durable IM

Task 1.2 (0-24 MAC**; ISMMS**) Implement and maintain data management, QC and transmission processes

Task 1.3 (6-12 MAC; **ISMMS**, UTMB, Yale, UCLA, Princeton, Yale, Pittsburgh). Define "good" immune memory in terms of cell and molecular features from >2 immune-associated tissues. ***12 MAC Deliverable:*** *Cell and molecular definition of good immune memory with correlation with duration protection, extending from early to late timepoints after vaccination comprising at least 3 known and 2 novel cell types or markers*

Task 1.3.1 (9 MAC; **Yale**, UCLA, ISMMS) LN and BM SC analyses of LN, BM and spleen in C57BL/6J mice at early timepoints and 120 dpv

Task 1.3.1.1 (9 MAC; **Yale**) CITE-seq of LN, BM, Spleen non-naive B, non-naive T, mononuclear cells

Task 1.3.1.2 (9 MAC, **UCLA**) sn M3C-seq non-naive B BM and spleen

Task 1.3.1.3 (9 MAC **ISMMS**) sn ATAC/RNA multiome non naive B BM and spleen

Task 1.3.2 (8 MAC; **ISMMS, UTMB**) C57BL/6J vaccination study early timepoint molecular studies for tissue harvesting

Task 1.3.2.1 (8 MAC; **UTMB**) RNA-1273 Prime, Prime-boost and VSV-ZEBOV with tissue harvesting and processing at 1, 3, 7, 14 and 30 days

Task 1.3.2.2 (8 MAC; **ISMMS**) TIV, RNA-1010 and TIV-MF59 Prime-boost with tissue harvesting and processing at 1, 3, 7, 14 and 30 days

Task 1.3.3 (9 MAC; **ISMMS**, UTMB) Generate immune monitoring data from mouse early time point samples of vaccination site, LN, blood and blood cytokine analysis

Task 1.3.4 (12 MAC, **Yale**, UCLA, ISMMS) Generate sc data from mouse early time point samples of vaccination site, LN, blood and blood cytokine analysis

Task 1.3.4.1 (10 MAC; **Yale**) CITE-seq or CiTE-seq/LIBRA from vaccination site and from LN and blood in mononuclear cells and cell subsets

Task 1.3.4.2 (10 MAC, **Yale**) DBiT-seq or DBIT-seq/LIBRA from LN

Task 1.3.4.3 (10 MAC, **UCLA**) sn m3C non-naive B cell from LN and blood

Task 1.3.4.4 (10 MAC, **ISMMS**) sn ATAC/RNA multiome non-naive B cell from blood

Task 1.3.4.5 (12 MAC, **Princeton**, **ISMMS**, Pittsburgh, Yale) Analyze multiomics data

to define "good" IM (cell and molecular features correlating with durable protection)

Task 1.3.4.5.1 (12 MAC, **Princeton**, **ISMMS**, Pittsburgh, Yale) compare different analysis platforms to identify which provides the most reliable markers of IM

Task 1.3.4.5.2 (12 MAC, **Princeton**, ISMMS, Pittsburgh, Yale) Identify at least 3 known and 2 novel cell type or molecular markers or pathways that contribute to IM

Task 1.3.4.5.3 (12 MAC, **Princeton**, ISMMS, Pittsburgh, Yale) Demonstrate early and late detection of the IM markers from 1.3.4..5.2.

Task 1.3.4.5.4 (12 MAC, **Princeton**, ISMMS, Pittsburgh, Yale) Test markers by analyses of curated vaccination data (public and DTA shared data

Task 1.3.5 (6-12 MAC; **UTMB**) Complete 8 month serial blood immune profiling and challenge studies on the SARS-CoV-2 and Ebola vaccine perturbations initiated in the first 6 months

Task 1.3.5.1 (6-12 MAC; **UTMB**) Pathogen challenge at 240 dpv

Task 1.3.5.2 (6-12 MAC; **UTMB**) Serological and nAb responses

Task 1.3.5.3 (6-12 MAC; **UTMB**) Quantitate IM in LN, BM, spleen

Task 1.3.5.4 (6-12 MAC; **UTMB**) Harvest and analyze or store tissues and cells

Task 1.3.6 (6-12 MAC; **ISMMS**) Complete 8 month serial blood immune profiling and challenge studies on the 3 influenza vaccine perturbations initiated in the first 6 months

Task 1.3.6.1 (6-12 MAC; **ISMMS**) Pathogen challenge at 240 dpv

Task 1.3.6.2 (6-12 MAC; **ISMMS**) Serological and nAb responses

Task 1.3.6.3 (6-12 MAC; **ISMMS**) Quantitate IM in LN, BM, spleen

Task 1.3.6.4 (6-12 MAC; **ISMMS**) Harvest and analyze or store tissues and cells

Task 1.4 (0-24 MAC; **Yale**, Princeton, Pittsburgh, ISMMS) Assemble compendium of multi-species vaccination data and integrate these into the development of the roadmap to immune memory

Task 1.5 (0-24 MAC; **HJF, NAMRU-2**, **ISMMS**) obtain initial sets of early human post vaccination samples for analysis

Task 1.5.1 (6 MAC; **NAMRU-2**) Obtain IRB and DOD HRPO approvals

Task 1.5.2 (18 MAC; **NAMRU-2**) Deliver initial samples for analysis from human Japanese encephalitis vaccine recipients

Task 1.5.3 (18 MAC; **NAMRU-2**) Deliver initial samples for analysis from human influenza vaccine recipients

Task 1.5.5 (6 MAC; **HJF/WRAIR**) Obtain IRB and DOD HRPO non-human research designation for analysis of archived blood yellow fever vaccination clinical trial samples

Task 1.5.6 (18 MAC, **HJF/WRAIR**) Deliver initial samples for analysis

Task 1.5.7 (12 MAC, **ISMMS**) Obtain IRB and HRPO approval for collection of human blood samples from Moderna influenza RNA-1010 clinical studies

Task 1.5.8 (24 MAC, **ISMMS**) Obtain samples for analysis from RNA-1010 vaccination study

Task 1.6 (13-24 MAC; **Yale,** UCLA, ISMMS) Generate initial early human vaccination data

Task 1.6.1 (20 MAC; **ISMMS**) Generate blood immunophenotyping data from human samples

Task 1.6.2 (22 MAC; **Yale**) Generate blood CITE-seq or CITE-seq/LIBRA-seq data

Task 1.6.3 (22 MAC, **UCLA**) Generate sn m3C data

Task 1.6.4 (22 MAC, **ISMMS**) Generate sn RNA/ATAC-seq multiome data

Task 1.7 (9-24 MAC; **ISMMS**, UTMB, Yale, UCLA, Princeton, Pittsburgh). 24 Mo Milestone and Criteria: Assemble the roadmap of routes to immune memory. ***24 MAC Deliverable****: Roadmap of routes to immune*

memory including quantification of more than 1000 features and proteins per cell from immune cell subpopulations and identification of at least 5 features that correlate with immune memory: 1. List of early features that mark cells for long term MBC or long lived plasma cells 2. Draft of predictive model for vaccine durability using mouse early data 3. List of key nodes selected to test in conditional KO in Phase 2.

Task 1.7.1 (9-13 MAC, **UTMB**) 4 month time-course study with harvesting and processing for molecular analyses after VSV-ZEBOV vaccination (continuation of study from Task 1.3.2.1)

Task 1.7.2 (9-13 MAC, **ISMMS**) 4 month time-course study with harvesting and processing for molecular analyses after TIV+MF59 and TIV vaccination (continuation of study from Task 1.3.2.2 and 1.3.2.3)

Task 1.7.3 (13-16 MAC; **UTMB**) early time point (1, 3, 7, 14 days) and tissue processing with RNA-1273 Prime and Prime boost vaccinations

Task 1.7.4 (13-16 MAC; **ISMMS**) early timepoint (1, 3, 7, 14 days) and tissue processing with RNA-1010 vaccination

Task 1.7.5 (16-20 MAC; **Yale**, ISMMS, UCLA). ***20 MAC Deliverable:*** Demonstration of complete roadmap mouse model molecular data generation on early samples and full timecourse studies.

Task 1.7.6 (18-21 MAC; **Yale**, ISMMS, UCLA) perform additional single cell molecular analyses on additional samples, including 8 month harvested samples, as indicated by prior results

Task 1.7.7 (13-21 MAC; **Princeton**) Implement trajectory analysis for early markers of IM

Task 1.7.8 (13-21 MAC **Pittsburgh**, Princeton, ISMMS) Implement differential and ML analysis to determine functionally homologous early markers of immune memory in mouse and human data

Task 1.7.9 (18-22 MAC; **Princeton, ISMMS**) Implement integrated methods to select the highest value methods and early marker predictors of immune memory including at least 5 features showing correlation with quantitation of durable memory following vaccination

Task 1.7.10 (18-24 MAC; **Princeton**, **ISMMS**, Pittsburgh) Develop predictive model from early features of level of durable memory after mouse vaccination

Task 1.8 (0-24 MAC; **ISMMS**) Provide active management of ISMMS and subcontractor laboratories

Phase 2 (Option 1) (25-60 MAC) Develop, validate and verify research tool that provides an early prediction of a vaccine's ability to provide long-lasting protection and proof-of-concept for pathogen-specific individual immune assessment

Task 1.9 (25-32 MAC; **NAMRU-2, HJF/WRAIR, ISMMS**) Continued human vaccination blood sample collection

Task 1.9.1 (25-32 MAC; **NAMRU-2**) Complete blood sample collection for JEV vaccine

Task 1.9.2 (25-32 MAC; **NAMRU-2**) Complete blood sample collection for influenza vaccines

Task 1.9.3 (25-32 MAC; **HJF**) Complete sample delivery from yellow fever vaccination samples

Task 1.9.4 (25-32 MAC; **ISMMS**) Complete blood sample collection for analysis from from Moderna influenza RNA-1010 vaccination study

Task 1.10 (25-38 MAC; **Yale**, UCLA, ISMMS) complete human immunophenotyping and sc data generation

Task 1.11 (25-36 MAC; **Princeton,** Yale, ISMMS, Pittsburgh) ***36 MAC Deliverables****: 1) Demonstration of transferability of immune mechanisms from Phase 1 roadmap to humans by establishing biomolecular*

*correlates of mouse and immune mechanisms that are representative of durability. 2**)** Identification of at least 3 pathways from unbiased molecular profiles with functional homologue conservation in humans that lead to B cell maturation.*

Task 1.12 (25-36 MAC; **Princeton**, ISMMS, Yale, Pittsburgh) Select the most informative sets of immune memory early biomolecular correlates from the dense data networks generated in the program

Task 1.12.1 (25-34 MAC; **Princeton**, Yale, Pittsburgh, ISMMS) Benchmark computational methods to determine which yields the most accurate mechanistic output and why

Task 1.12.1.1 (25-34 MAC; **Princeton**, Yale, Pittsburgh,ISMMS) Build networks for relevant immune cell types that provide a functional map of mouse and human and then quantify (probabilistically) the extent of functional conservation of each of these nodes (genes/proteins).

Task 1.12.1.2 (25-34 MAC; **Princeton**, Yale, Pittsburgh, ISMMS) Determine estimated precision to rank as an ensemble, with final validation by predictive accuracy of long term immunity

Task 1.12.1.3 (25-34 MAC; **Princeton**, ISMMS) Create multiomics single cell trajectory

Task 1.12.1.4 (25-34 MAC; **Princeton**, **ISMMS**, Yale, Pittsburgh) Identify determinants of immune memory by integrated multiomic contrasts and learning approaches

Task 1.12.1.5 (25-34 MAC; **Princeton**, **ISMMS,**Pittsburgh, Yale) Identify functional coherence of multiple informative readouts within each species (e.g. PLIER, XNNET) and functional homology across species early durable IM mechanism discovery

Task 1.12.2 (36 MAC; **Princeton**, ISMMS, Pittsburgh, Yale) Identify durable immune pathway mechanistic nodes that contribute to B cell maturation to be tested by knockout studies

Task 1.13 (25-44 MAC; **ISMMS**, Yale, UCLA) Test prediction for 5 vaccine perturbations in mouse model. ***48 MAC Deliverable:*** Demonstration of validation of the predictive model.

Task 1.13.1 (25-32 MAC; **ISMMS**) Perform 5 influenza vaccine/adjuvant perturbations and harvest early tissues and timepoint that has been determined to be predictive of immune memory to vaccination in preceding studies and analyses

Task 1.13.2 (28-44 MAC; **Yale**, UCLA, ISMMS) Generate immunophenotyping and sc data

Task 1.13.3 (32-36 MAC; **Princeton**, ISMMS, Pittsburgh, Yale) Predict immune durability

Task 1.14 (30-46 MAC, **ISMMS**, **UTMB**) Experimentally validate key early prediction nodes

Task 1.14.1 (30-46 MAC; **UTMB**) Determine effects of conditional knockouts of at least 2 key loci on immune memory durability with a good vaccine (VSV-ZEBOV)

Task 1.14.2 (30-46 MAC; **ISMMS**) Determine effects of conditional knockouts of at least 5 key loci on immune memory durability with a good vaccine (TIV+MF59 prime-boost)

Task 1.15 (25-36 MAC; **Princeton**, ISMMS, Yale, Pittsburgh) : *Refine mouse predictive model. **36 MAC Deliverable***

Task 1.16 (37-48 MAC; **Princeton**, Yale, Pittsburgh, ISMMS) Investigate early predictors of human and mouse individual variation in vaccine durability

Task 1.16.1 (37-44 MAC; **Princeton**, Yale, Pittsburgh, ISMMS) Identify early predictors of individual response variation in the mouse model and in human samples

Task 1.16.2 (42-48 MAC; **Princeton**, Yale, Pittsburgh, ISMMS) Provide proof of principle for pathogen-specific prediction of individual immune response variation to vaccination by comprehensive analysis of mouse and human data

Task 1.17 (37-48 MAC; **ISMMS**, Princeton, Yale, Pittsburgh) Port key features to targeted assays and develop predictive model based on targeted analyses and assemble an R&E tool for predicting the duration of protection following vaccination. ***48 MAC Deliverable****: Demonstration of R&E tool's ability to accurately predict vaccine duration of protection by predicting duration of protection from early (day 0–10) features of response to vaccine in a preclinical challenge model and demonstrate accuracy of correlates to corresponding samples from a human population (≥70%).*

Task 1.17.1 (37-42 MAC; **ISMMS**, Yale) Generate targeted data

Task 1.17.2 (42-48 MAC; **Princeton**, ISMMS, Yale, Pittsburgh) Retrain and validate model on targeted datasets and finalize R&E tool

Task 1.18 (49-60 MAC; **ISMMS**, Princeton, Yale) ***60 MAC Deliverable****: Transfer R&E tool for testing to IV&V team and work with IV&V team on independent validation and verification of durability and individual variation in durability*

Task 1.19 (25-48 MAC**; ISMMS**) Manage data and sharing processes including QC, storage and transmission

Task 1.20 (25-60 MAC; **ISMMS**) Provide active management of ISMMS and subcontractor laboratories

Milestones and Deliverables

Reporting Deliverables

- ISMMS will prepare and submit a Best Vax technical status report slide deck delivered by email to the AIM team every six weeks. Reports will be submitted on or before the 1st or 15th of each month following the six week schedule. The first technical report will be delivered May 13, 2022.
- ISMMS will prepare and submit a financial status report delivered by email to the designated AIM team each month on an agreed upon date (ex: the 15th of each month). The financial report will include monthly expenses and spending projections.
- ISMMS will prepare and submit a Best Vax quarterly technical status report and summary slide deck to include quarterly progress from all subcontractors. Reports will be delivered by email to the AIM PM on or before July 1, October 1, January 2, and April 1 of each program year.
- Upon individual completion of Phase 1 and Phase 2, ISMMS will prepare and submit a Final Report that summarizes the tasks accomplished, goals achieved, data analyses and future goals. The Final Report will include slides from all subcontractors and will be submitted on or before March 31, 2024 and 2027.

Phase 1 Milestones and deliverables

- **6 MAC Interim Milestone** Vaccination experiments VSV-ZEBOV, RNA-1273 (SARS-CoV-2) prime boost, RNA-1273 prime, TIV+M59 prime-boost (influenza), standard dose TIV (Influenza), and RNA-1010 (Influenza) in C57BL/6J mice (Tasks 1.1.2.1-1.1.2.6)
- **6 MAC Interim Milestone**: Data QC, transmission and storage processes implemented. (Task 1.2)
- **6 MAC Interim Milestone**: Develop metric for defining durability of immune memory and determine the range of durability of the 6 vaccine perturbations (Task 1.1.2.7)
- **6 MAC Milestone:** Establishment of sc methods and protocols for Best-vax (Tasks 1.1.1, 1.1.3)

6 MAC Deliverable: Demonstration of a C57BL/6J model of good immune memory with quantitative immune responses at 4 months after vaccination showing a range of at least 2.5-fold between highly durable and poorly durable vaccine perturbations. (Task 1.1)

- **8 MAC Interim Milestone:** C57BL/6J vaccination study for tissue harvesting for early time-point tissue from two durable (VSV-ZEBOV, TIV-MF59) and one transient vaccine perturbation (TIV) with tissue harvesting and processing at 1, 7 and 14 days complete. (Task 1.3.2)
- **9 MAC Interim Milestone:** Single cell molecular analyses from Lymph nodes (LN), spleen and bone marrow at 4 months after vaccination in C57BL/6J mice (Task 1.3.1)
- **9 MAC Interim Milestone:** Mouse model immune monitoring data from the early time point samples generated. (Task 1.3.3)
- **10 MAC Interim Milestone:** Single cell molecular analyses from the vaccine early time point samples generated. (Task 1.3.4)
- **12 MAC Milestone:** Complete 8 month serial blood immune profiling and challenge studies on the 6 vaccine perturbations initiated in the first 6 months (Task 1.3.5, 1.3.6)

12 MAC Deliverable: Slide deck demonstrating cell and molecular definition of good immune memory from >2 immune-associated tissues showing discriminative correlation with long duration protection with demonstration that these profiles extend from early to late timepoints after vaccination. This defining profile will include characterization of at least 3 known and 2 novel cell type or cell state markers that contribute to immune memory.

- **13 MAC Interim Milestone** Vaccination experiments VSV-ZEBOV, RNA-1273 (SARS-CoV-2) prime boost, RNA-1273 prime, TIV+M59 vaccine (influenza), standard dose TIV (Influenza), and RNA-1010 (Influenza) in C57BL/6J animals for longer term follow up completed. (Task 1.7.1)
- **18 MAC Interim Milestone:** Initial samples from human Japanese encephalitis vaccine recipients, human influenza vaccine recipients, and human yellow fever vaccine recipients obtained. (Tasks 1.5.2, 1.5.4, 1.5.6)
- **20 MAC Interim Milestone:** Generation of blood immunophenotyping data from initial human samples complete. (Task 1.6.1)

20 MAC Deliverable: Slidedeck on complete molecular data generation on early samples and full timecourse studies (Task 1.7.5)

- **21 MAC Interim Milestone:** Additional single cell molecular analyses performed on additional samples, including 8 month harvested samples, as indicated by developing roadmap to immune memory (Task 1.7.6)
- **21 MAC Interim Milestone:** Implement trajectory analysis, differential analysis and ML analysis to determine functionally homologous early markers of immune memory in mouse and human data (Tasks 1.7.7, 1.7.8)
- **22 MAC Interim Milestone:** Implement integrated methods to select the highest value methods and early marker predictors of immune memory including at least 5 features showing correlation with quantitation of durable memory following vaccination. (Task 1.7.9)
- **22 MAC Interim Milestone:** Generation of blood CITE-seq or CITE-seq/LIBRA-seq, data from initial human vaccine study samples. (Task 1.6.2)
- **22 MAC Interim Milestone:** Generation of sn m3C and sn RNA/ATAC-seq multiome data from initial human vaccine study samples as indicated by mouse analysis results complete. (Tasks 1.63, 1.6.4)
- **24 MAC Milestone:** Initial samples from all human vaccination studies obtained (Tasks 1.5.2, 1.5.4, 1.5.6, 1.5.8)
- **24 MAC Milestone:** Compendium of multi-species vaccination data developed and integrated into the development of the roadmap to immune memory (Task 1.4)
- **24 MAC Milestone:** Generation of initial early human vaccination data completed for integration with roadmap (Task 1.6)

- **24 MAC Milestone:** Develop predictive model from early features of level of durable memory after mouse vaccination (Task 1.7.10)

24 MAC Deliverable: Slidedeck and white paper of roadmap of routes to immune memory including quantification of more than 1000 features and proteins per cell from immune cell subpopulations and identification of at least 5 features that correlate with immune memory: 1. List of early features that mark cells for long term MBC or Long lived plasma cells 2. First draft of a predictive model for using mouse early data and predicting vaccine durability 3. List of the selection of key nodes to test in conditional KO in Phase 2. (Task 1.7)

Phase 2 Milestones and Deliverables

- **32 MAC Interim Milestone:** Sample delivery from NAMRU-2, HJF/WRAIR and ISMMS/Moderna RNA-1010 trial (Tasks 1.9.1, 1.9.2, 1.9.3, 1.9.4)
- **38 MAC Milestone:** Generation of immune monitoring and sc data from human samples. (Task 1.10)
- **36 MAC Milestone** Select the most informative sets of immune memory early biomolecular correlates from the dense data networks generated in the program as described in Task 1.12.1 and 1.12.2. (Task 1.12)

36 MAC Deliverable: Demonstration of transferability of immune mechanisms from Phase 1 roadmap to humans through establishing biomolecular correlates of mouse and immune mechanisms that are representative of durability (Task 1.11). Demonstration to include at least 4 vaccine perturbations to validate predictive model (Task 1.13)

36 MAC Deliverable: Identification of at least 3 pathways from unbiased molecular profiles with functional homologue conservation in humans that lead to B cell maturation (Task 1.11)

36 MAC Deliverable: Refined mouse predictive model (Task 1.15)

- **44 MAC Milestone:** Demonstrate prediction for 5 vaccine perturbations in mouse model (Task 1.13)
- **46 MAC Milestone:** Key early prediction nodes experimentally validated (Task 1.14)
- **48 MAC Interim Milestone:** Investigation of early predictors of human and mouse individual variation in vaccine durability completed. (Task 1.16)
- **48 MAC Milestone:** Port key features to targeted assays and develop predictive model based on targeted analyses and assemble an R&E tool for predicting the duration of protection following vaccination (Task 1.17)
- **48 MAC Milestone:** Determine contributions of ≥3 conserved pathways to immune memory formation
- **48 MAC Milestone:** Research tool that provides an early prediction of a vaccine's ability to provide long-lasting protection and proof-of-concept for pathogen-specific individual immune assessment.

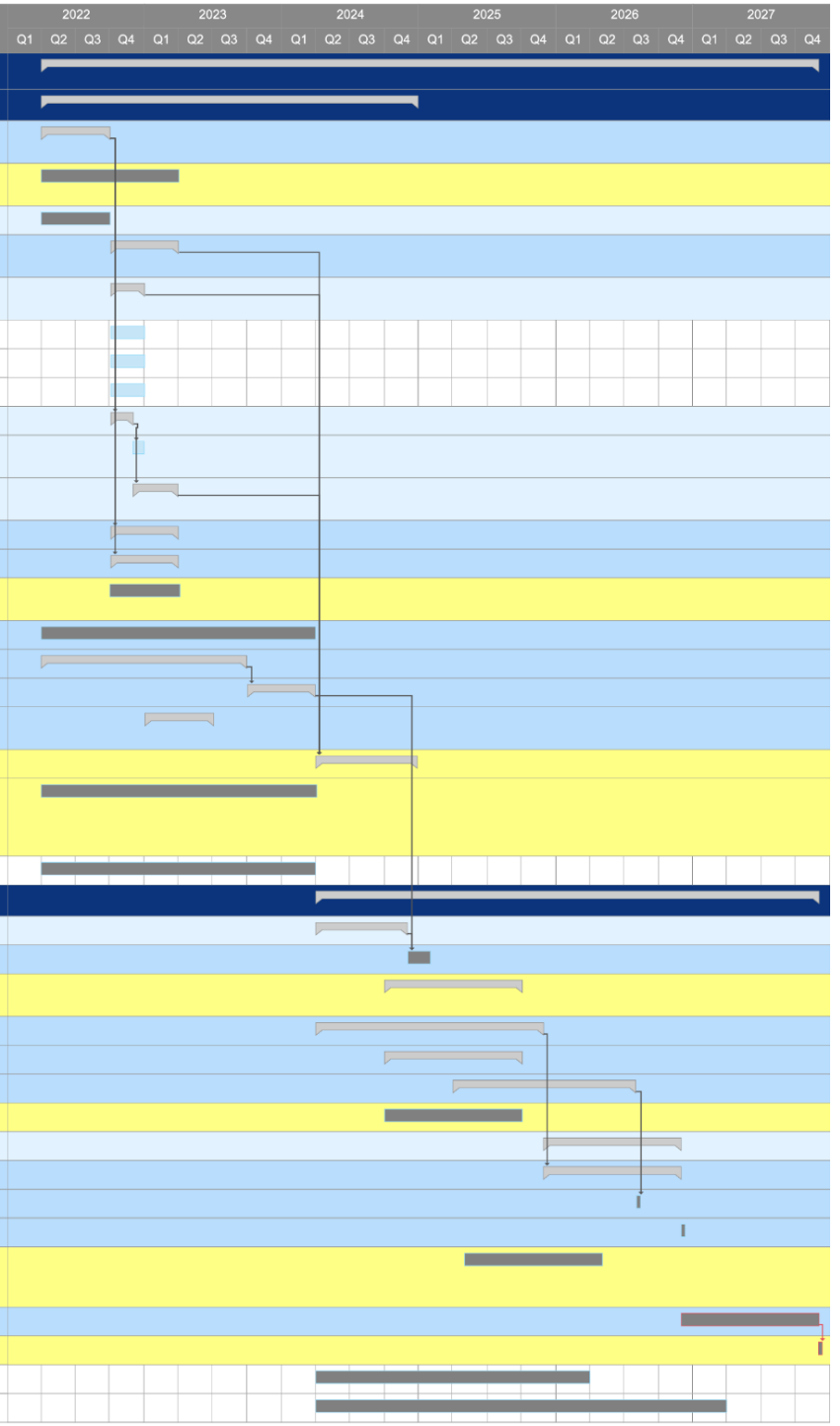
48 MAC Deliverable: Demonstration of research tool's ability to accurately predict vaccine duration of protection by predicting duration of protection from early (day 0–10) features of response to vaccine in a preclinical challenge model and demonstrate accuracy of correlates to corresponding samples from a human population (≥70%). (Task 1.17)

60 MAC Milestone: Validate R&E tool prediction of individual protection at AUROC ≥80% of level of protection prior to pathogen challenge. **60 MAC Deliverable:** Transfer R&E tool for testing to IV&V team and work. Work with IV&V team to perform independent validation

experiments with at least four different vaccine perturbations to predict accurate rank of relative durability. (Task 1.18).

G. Schedule and Milestones

MAC	Task	Team
	DARPA AIM - Assessing Immune Memory Schedule	
0-24	**PHASE 1**	
0-6	Task 1.1 - Establish Methods & IM Mouse Model	ISMMS; UTMB; Yale; UCLA; Princeton; Pitt
6	Show >2.5-fold variation in immune memory durability among vaccine perturbations in C57BL/6J mice	ISMMS; UTMB; Yale; UCLA; Princeton; Pitt
0-6	Task 1.2 - Implement Data Management Processes	ISMMS
6-12	Task 1.3 - Define "good" immune memory	ISMMS; UTMB; Yale; UCLA; Princeton; Yale; Pitt
6-9	Task 1.3.1 - LN and BM SC analyses of LN, BM and spleen in C57BL/6J mice at early timepoints (TPs) and 120 dpv	Yale; UCLA; ISMMS
6-9	Task 1.3.1.1 - CITE-seq of LN, BM, Spleen non-naive B, non-naive T, mononuclear cells	Yale
6-9	Task 1.3.1.2 - sn M3C-seq non-naive B BM and spleen	UCLA
6-9	Task 1.3.1.3 - sn ATAC/RNA multiome non naive B BM and spleen	ISMMS
6-8	Task 1.3.2 - C57BL/6J vaccination study early TP molecular studies for tissue harvesting	ISMMS
6-9	Task 1.3.3 - Generate immune monitoring data from mouse early TP samples of vaccination site, LN, blood and blood cytokine analysis	ISMMS; UTMB
6-10	Task 1.3.4 - Generate sc data from mouse early TP samples of vaccination site, LN, blood and blood cytokine analysis	Yale; UTMB; ISMMS
6-12	Task 1.3.5 - Ebola/SARS-CoV-2 vaccines 8 month serial IM and challenge studies	UTMB
6-12	Task 1.3.6 - Influenza vaccines 8 month serial IM and challenge studies	ISMMS
12	Cell and molecular definition of good immune memory with correlation with duration protection, extending from early to late timepoints after vaccination comprising at least 3 known and 2 novel cell types or markers	ISMMS, UTMB, Yale, UCLA, Princeton, Yale, Pitt
0-24	Task 1.4 - Assemble outside data compendium	Yale; Princeton; Pitt; ISMMS
0-24	Task 1.5 - Obtain initial human samples	HJF; NAMRU-2; ISMMS
12-24	Task 1.6 - Generate initial human data	Yale; UCLA; ISMMS
9-24	Task 1.7 - Construct the roadmap to immune memory	ISMMS; UTMB; Yale; UCLA; Princeton; Pitt
20	Complete roadmap mouse model molecular data generation on early samples and full timecourse studies.	Yale; ISMMS; UCLA
24	Roadmap of routes to immune memory including quantification of 1000+ features and proteins/cell from immune cell subpopulations and identification of 5+ features that correlate with immune memory: 1. List of early features that mark cells for long term MBC or long lived plasma cells 2. Draft of predictive model for vaccine durability using mouse early data 3. List of key nodes selected to test in conditional KO in Phase 2.	ISMMS, UTMB, Yale, UCLA, Princeton, Pitt
0-24	Task 1.8 - Active Management	ISMMS
25-60	**PHASE 2 OPTION 1**	
25-34	Task 1.9 - Continued human sample collection	NAMRU-2; HJF; ISMMS
25-34	Task 1.10 - Complete human sample data generation	Yale; UCLA; ISMMS
36	Task 1.11 - 1) Demonstration of mouse to human transferability 2) Identification of 3+ pathways that lead to B cell maturation.	Princeton; Yale; ISMMS; Pitt
25-36	Task 1.12 - Select the most informative sets of nodes	Princeton; Yale; ISMMS; Pitt
25-36	Task 1.13 - Test prediction for 4 perturbations in mouse model	ISMMS; Yale; UCLA
30-46	Task 1.14 - Knockout Studies	ISMMS; UTMB
36	Task 1.15 - Refined mouse predictive model	Princeton; Yale; ISMMS; Pitt
37-48	Task 1.16 - Investigate individual variation in vaccine durability	Princeton; Yale; ISMMS; Pitt
37-48	Task 1.17 - Downsampled R&E Tool	Princeton; Yale; ISMMS; Pitt
48	Contributions of ≥3 conserved pathways to IM formation	Princeton; Yale; ISMMS; Pitt
48	Research tool that provides early prediction of durability; proof-of-concept individual immune assessment.	Princeton; Yale; ISMMS; Pitt
48	Demonstration of R&E tool's ability to accurately predict vaccine duration of protection by predicting duration of protection from early (day 0-10) features of response to vaccine in a preclinical challenge model and demonstrate accuracy of correlates to corresponding samples from a human population (≥70%)	Princeton; Yale; ISMMS; Pitt
49-60	Task 1.18 - Validate R&D tool for testing to IV&V team	Princeton; Yale; ISMMS; Pitt
60	Transfer final R&E Tool IV&V Validation	Princeton; Yale; ISMMS; Pitt
25-48	Task 1.19 - Data Management	ISMMS
25-60	Task 1.20 - Active Management	ISMMS



H. Transition Plan

Efficient commercialization of IP developed by ISMMS programs is an important institutional mandate. Intellectual property matters, technology transfer and business development for the Best-vax DARPA program will be handled by Mount Sinai Innovation Partners (MSIP). MSIP has a large, experienced legal and business development team. In 2020 we filed 378 patents, many leading to commercial licenses and have also commercialized IP via start ups, for example Sema4, which recently went public and has a ~$2B valuation . The Best-vax subcontracting agreements assure that MSIP will take the lead in transition of products resulting from inventors or co-investors at any program subcontractor. ISMMS is well suited to both perform and coordinate with other institutions translational clinical research necessary as part of a technology commercialization effort. The major IP anticipated involves the R&E tool to accelerate vaccine development. Dr. Sealfon has apprised MSIP that via his contacts with vaccine development leadership at Sanofi Pasteur and Moderna, Inc there is interest in gaining access to future program IP. The MSIP office will assist investigators in identifying additional partners who can assist with advanced development for both medial use and National Security purposes. ISMMS MSIP will lead any outside business discussions as required to facilitate the development and commercialization of new technologies arising from the Best-vax research program.

I. Summary Slides (Does not count towards page limit; two (2) slides maximum):

PowerPoint slide(s) summarizing the proposed effort's vision, goals, impact, scientific/technical approach, and milestone schedule. **Download and use the template provided in Attachment 1 posted with the subject BAA.** Submit the PowerPoint file in addition to Volume I and II of your proposal.