Human Virome Workshop

National Institutes of Health

Office of Strategic Coordination—NIH Common Fund

April 29, 2022

Virtual Meeting

Revised May 31, 2022

This meeting summary was prepared by Rose Li and Associates, Inc., under contract to the National Institutes of Health's (NIH) Office of Strategic Coordination (OSC). The views expressed in this document reflect both individual and collective opinions of the meeting participants and not necessarily those of NIH OSC. Review of earlier versions of this meeting summary by the following individuals is gratefully acknowledged: Bethany Stokes, Elizabeth A. Finch, Nancy Tuvesson.

Purpose

This workshop was designed to assess the feasibility and need for developing a National Institutes of Health (NIH) Common Fund program that aims to characterize the human virome in order to yield greater understanding of the viruses we harbor and their impact on immune function and human health.

Discussants

External Discussants: Kathy Burns (Dana Farber), Michael Busch (Vitalant/UCSF), Rick Bushman (UPenn), Charles Chiu (UCSF), Ronald Collman (UPenn), Erick Delwart (UCSF), Stephen Francis (UCSF), Soren Gantt (CHU), Geoffrey Ginsburg (NIH), Scott Handley (WashU), Lori Holtz (WashU), Kate Jeffrey (Moderna), H. Benjamin Larman (JHU), Efrem Lim (Arizona State), Corinne Maurice (McGill), Sallie Permar (Weill Cornell)

NIH Discussants: Beena Akolkar (NIDDK), Rohan Hazra (NICHD), Hye-Sook Kim (NCCIH), Leia Novak (NIAID), Mugdha Samant (NIH)

Session I: Healthy Human Virome

Moderators: Drs. Leia Novak, NIAID, and H. Benjamin Larman, JHU

This session focused on commensal viruses that are largely considered to have a neutral or mutualistic effect on the human host; this category includes viruses that cause persistent infections without known pathology (e.g., anelloviruses), endogenous retroviruses, and viruses of the microbiome.

Technical Advances and Gaps

Scientific advances related to technology, particularly those in high throughput DNA sequencing, have transformed the research community's ability to understand and evaluate the human virome. However, many reference databases are heavily focused on viromes of pathogens, not commensal or mutualistic viruses. In addition, the virome sequencing research community lacks standardization of reference databases; currently databases use different assay protocols and controls, leading to results that cannot be compared across studies or identify possible sample contaminations (i.e., whether the virus detected is part of the virome or introduced via contamination). Because of the lack of standardized reference sequences and of accurate mapping abilities, the field can only qualify results (i.e., identify whether a virus is present or not) and cannot quantify the results. Other areas that require standardization include virome library preparation, virome sequencing, and control materials.

Discussants considered how to best evaluate how the virome and microbiome interact and impact each other. They noted that computational approaches do exist to predict these interactions, but emphasized that the gold standard for assessing these relationships is via cell culture infection studies. They emphasized that the computational tools needed to characterize viromes do not currently exist. Machine learning methods could be leveraged to compare viromes across studies and help to identify contamination.

Need for Large-Scale Longitudinal Studies

Most virome-related studies are small in size, are cross-sectional, and focus on a single cohort or sample type (e.g., blood, stool, or skin), which can significantly limit the generalizability of findings and the ability to identify major discoveries related to the diversity of the human virome. Longitudinal studies in larger cohorts are needed to expand the current understanding of the human virome. Samples collected

during blood donations could serve as a longitudinal cohort that could be metagenomically assessed over time to determine how the virome composition changes in healthy individuals.

Discussants agreed that a cohort of 100,000s of samples is needed in order to robustly evaluate the virome with sufficient power to understand complex diseases associated with multiple genetic and environmental factors. For more straightforward questions, smaller sample sizes would be appropriate.

Areas Requiring More Study

Discussants identified several cross-cutting gaps in the virome field including (1) an understanding of virome composition and regulation across different geographies, tissues, and the lifespan and in terms of environmental and host perturbations and (2) the relationship between healthy homeostasis of the immune system, metabolism, and host genetic variation. Discussants emphasized that the virome research community has not yet reached consensus on how to define a "healthy" virome, and developing that definition will inform future studies of both healthy and disease-specific viromes. They identified the following topics as requiring more study:

- How do active mobile genetic elements contribute to human disease?
- What is the relationship between mobile genetic elements, the commensal virome, and the immune system?
- How do immune cells develop in response to commensal viruses?
- How is the commensal virome impacted by the exposome (e.g., environment, diet), including other viruses, over time?
- How is the virome changed in immunocompromised individuals?
- How do the levels and distributions of commensal viral populations change across geography, demography, body sites, exposomes, and populations with specific comorbidities?
- Can the virome serve as biomarkers for human therapeutics, and does the virome impact therapeutics and biomarkers?
- How does the virome benefit the human immune system?
- How does the virome of blood transfusions clinically impact the recipient? What about the viromes of organ transplants?

Discussants also noted that the biomedical research community is just beginning to study how human endogenous retroviruses (HERVs) may contribute to disease. The HERV research field must also determine how HERVs interact with the commensal virome, pathogens, and the host immune system, as well as develop a better understanding of HERV and host control pathway evolution.

Session II: Infant Virome

Moderators: Drs. Hye-Sook Kim, NCCIH, Rohan Hazra, NICHD, and Sallie Permar, Cornell University

Knowledge Gaps

Discussants highlighted the importance of understanding how the timing of an infection impacts the infant virome, as well as central tolerance and disease pathologies (e.g., polio) later in life. Like in the previous session, they identified the need for larger, more robust studies because most infant virome studies have evaluated only a few individuals. They emphasized the need for high-grade longitudinal studies of the infant virome in order to understand the normal evolution of the infant virome over time and in relation to age. Understanding timing will also help identify any major critical windows of viral development in infants.

Discussants also emphasized the need to better investigate the interaction between host genetics and common viruses, such as the importance of human leukocyte antigen (HLA) in determining the response to common viruses (e.g., herpesviruses). However, studying gene–environment interactions (including interactions with common viruses) requires large-scale sample sizes, which can be difficult to obtain.

Discussants noted that differences in maternal–infant vertical transmission between viral species are important to understand; although the impact of vertical transmission is known for many eukaryotic pathogens, it is not well understood for the virome more broadly, including bacteriophages. In addition, discussants emphasized that understanding how vertical transmission differs among various viral species is important.

Some studies have found that microbiome populations within the gastrointestinal tract show regional specificity; however, the research community has not yet evaluated how microbiome communities in different anatomical regions change over time. Discussants also noted that very few non-primate animal models serve as appropriate model organisms for infant virome studies because, for example, mice do not produce a fecal pellet until after weaning. Non-human primates (NHPs) may be best suited for infant virome studies, but identifying additional animal models for such studies will benefit the field.

"Healthy" Infant Virome

Similar to the previous session, discussants emphasized the need to understand the definition of a "healthy" infant virome, as well as adult virome, before other more granular studies can be performed. For example, once the definition of "healthy virome" is established, researchers can investigate the impacts of host influences and other factors (e.g., genetics, geography, environmental exposures, adverse childhood experiences) on the virome, as well as interactions among factors (e.g., genetics and environment), may lead to disease over time. Researchers could also understand how to measure those factors in order to develop virome biomarkers and therapeutics that slow or inhibit disease progression.

Breast Milk Virome

Discussants reviewed the virome of breast milk. A recent longitudinal study found that less animal cell viruses were found in babies fed breast milk (compared to those fed formula) by 4 months of age, indicating that breast milk may suppress viral activities. They also noted the importance of studying mothers and infants over time after an infection to understand long-term impacts of that infection on both hosts.

Bacteriophage Induction

Discussants considered how phage induction may play a significant role early in infant virome development; however, the origin of these induction signals is poorly understood. Oxygen levels may play a role as a signal because the gut does not operate anaerobically early in life. Discussants noted that many of the signals related to bacteriophage induction during early years of life are unknown, whereas more is known about adult stages; however, evaluating induction early in life may be challenging. Some findings could be extrapolated from evolutionary microbiology perspectives related to seeding in new environments, such as the infant virome environment.

Future Opportunities

The *All of Us* program will begin to recruit pediatric participants (infants through age 17) in 2023 with the goal to enroll 150,000-200,000 participants and follow them longitudinally; this cohort will serve as an invaluable resource to study the infant virome. *All of Us* will also begin to incorporate geographical data, as well as data related to social determinants of health and environment, into its databases to

accompany the omics data generated. Discussants added that longitudinal antibody profiling in the setting of the *All of Us* would be highly valuable to the field. Discussants also recommended that *All of Us* recruit mother and infant dyads.

Discussants also discussed monoclonal antibodies and vaccination as intervention strategies to delay or prevent the acquisition of virome-related pathologies, but noted that these could have unintended consequences and that many studies (e.g., causal studies) are required before implementation.

Session III: Disease and Virome

Moderators: Drs. Mugdha Samant, NIH, Beena Akolkar, NIDDK, and Frederic Bushman, UPenn

Prioritized Disease States and Sample Types

Longitudinal studies of the human virome are particularly important because of the tremendous degree of inter-individual variability. Discussants identified the following notable disease states that would benefit from an analysis of the virome in large longitudinal cohorts in order to improve prevention or treatment: allergies, diabetes, complex immune disorders, neurological disorders, childhood diseases, common comorbidities of aging, diseases that cause immunocompromised states (e.g., HIV, immunosuppression), chronic diseases or diseases with long-lasting effects (e.g., COVID-19), cancers that affect younger populations (i.e., those below age 30 years), unexplained fever, and respiratory illnesses. Discussants noted that studying diseases with long-lasting effects is particularly important in the context of children who may be infected at a young age and continue to experience effects throughout their lifespan.

Discussants emphasized that the most pertinent sample types to investigate in studies of virome-disease interactions are stool, urine, breast milk, nasal swabs, blood, lung, gut biopsy, oral cavity, vagina, and brain samples. Nasal swab samples can be collected frequently and longitudinally and can capture both epithelial cells from the host and viral RNA and DNA; collecting samples longitudinally can help researchers evaluate whether a detected virus is causal or associated with a given pathology. Breast milk samples could help researchers better understand how the infant virome and microbiome are seeded and colonized early in life. Discussants noted that sample collection procedures must be standardized to ensure that researchers are collecting and processing tissues consistently and to allow for cross-study comparisons. Discussants emphasized that in addition to samples, collecting contextual data about a subject's environment is also critical.

Leveraging Approaches from the Microbiome Field

Successful approaches and findings from microbiome research studies may be applicable to characterization of the virome and to further understand the relationships between the virome and disease. Replicating how the microbiome was defined for the virome may help advance the virome field's understanding of specific disease. The microbiome field achieved significant successes in studies using microbiome-free mouse models, and a virome-free mouse model may offer similar benefits for the virome field.

Areas Requiring More Study

Discussants identified topic areas that require further study and would benefit from further characterization of the virome:

- What is the definition of a healthy state versus a disease state?

- How does the virome correlate with disease states? Is the virome protective or causative for a particular disease?
- If a disease is associated with the virome, is the disease linked to one virus or several viruses?
- How do treatments for a disease impact the virome, particularly drugs that modulate the immune system?
- How does the virome impact the gut–brain axis and, in turn, impact neurological disorders?
- What can be learned about the virome in populations with low incidence of specific diseases, such as inflammatory bowel disease?
- How does the virome and viral proteome impact host immunity and cause disease?
- How does environment (e.g., pollutants, toxins, climate, weather) and demography (e.g., sex, gender, age, race/ethnicity, geography, social determinants of health) impact the virome and lead to disease?
- What are seasonal or temporal changes that occur within the virome and how do those changes relate to disease?
- How do acute perturbations (e.g., diet, infections, medications) impact the host virome?

Gaps and Opportunities

Discussants emphasized the need for standardization of procedures for isolation, purification, and processing of virus-like particles, including what tissues are collected and the sequencing read lengths used for both RNA and DNA. They noted that the field may benefit from developing separate sequencing and processing protocols for phage viruses and eukaryotic viruses. One major challenge is that processing and sequencing virome samples is complicated and highly expensive.

Discussants highlighted that, in addition to positive control samples, studies must also collect negative control samples in order to eliminate the possibility of contamination and evaluate any possible batch effects. Another challenge for virome studies is identifying a matched control group of patients (e.g., household control, age, geographical region). Discussants recommended the development of a global viral atlas that includes samples and sequencing data from subjects across the globe. Such an atlas could be used in pandemic and outbreak preparedness. One discussant suggested that the field conduct an international and large-scale virome study by sampling sewage systems in various cities, performing metagenomics sequencing, and correlating these virome data with clinical data collected in those cities over time.

Discussants also noted that virome sequencing should occur in tandem with genetic sequencing and immunophenotyping to identify how host genetics impact the virome and vice versa. They noted that the virome field could leverage existing single cell RNAseq data, spatial transcriptomics data, and other types of sequencing data to extract virome-related information. Discussants also highlighted opportunities to leverage germ-free and virome-free mouse models to evaluate functionality and facility differences to aid in standardization, and to culture individual viruses to facilitate functionality studies. They identified the need to develop experimental systems (e.g., animal models, cell cultures) and functional tools to evaluate the virome in relation to specific diseases, in particular loss-of-function and gain-of-function models for the virome to evaluate functional changes. In addition, funding opportunities to facilitate development of models and tools are needed.

Discussants also noted that anelloviruses are the most common human viruses, yet little is known about their biology; this topic requires further basic science investigation, particularly to understand how these common viruses achieve viremia without causing disease and how these viruses could potentially be beneficial, not just commensal, to the host. The virome field must develop a robust cell culture

system for anelloviruses, as well as other virus types (e.g., adenoviruses, astroviruses, picornaviruses), because currently one does not exist and such a cell culture platform would be immensely valuable to the virome research community. In addition, mouse models that can be used to evaluate anelloviruses also do not exist and must be developed to advance the field.

Discussants noted that the virome field could convene experts across biomedical disciplines to help a potential NIH Common Fund-supported virome characterization program result in a comprehensive and interdisciplinary platform to be used in many future research studies.

Sample Strategies and Repositories

Discussants noted that the Researching COVID to Enhance Recovery (RECOVER) program collects postmortem samples of individuals with COVID-19, which could be leveraged for studies of the virome. The International HundredK+ Cohorts Consortium (IHCC) and The Environmental Determinants of Diabetes in the Young (TEDDY) study may also serve as repositories of samples for virome studies. As previously stated, the *All of Us* program also collects longitudinal samples, as well as contextual data, and will soon begin to enroll pediatric patients; this cohort could be leveraged for virome studies as well.

Question & Answer Session

Moderator: Frederic Bushman, UPenn

Standardization

Discussants noted that the field currently measures viral exposures in a variety of methods and that uniform measuring may help evaluate data across populations and studies. Many researchers in the virome field have discussed the development of standards for virome analysis pipelines. Discussants also noted that a standardized inducer of gut bacteria does not yet exist.

Leveraging Existing Work

Discussants agreed that significant virome-related work has been completed in other fields, including the microbiome and ecology fields, and that work, including data generation and protocol development, can be leveraged and reapplied to the human virome field. Methods and findings from other fields, particularly ecological and microbial perturbations studies, may be highly informative for virome studies. Discussants noted that within upcoming years, more computational advances will likely lead to better tools and databases for use by the virome research community. They also noted that the efforts of researchers to characterize anelloviruses may be useful to those in industry that aim to develop viral-based therapeutics.

Longitudinal Studies

Discussants reemphasized the importance of collecting longitudinal samples and evaluating them to solve important virome research objectives. They noted that many ongoing COVID-19 studies and clinical trials will likely result in longitudinal samples that could be used particularly to evaluate the impact of long-lasting COVID-19 disease effects on infected adults and children as well as to compare post-COVID-19 samples to those collected before the COVID-19 pandemic.

Functional Characterization

Discussants echoed previous statements related to the importance of conducting functional characterization of various viruses, which requires that each individual virus is cultured and functionally evaluated via additional models (e.g., animal models). In addition, discussants emphasized the need to study the functionality and effects of each individual viral protein. Functional analyses must be

performed in conjunction with sequencing because sequencing studies cannot provide all necessary information.