TITLE: A vision for investigating the microbiology of health and disease

AUTHOR: W. Ian Lipkin, MD, Center for Infection & Immunity, Columbia University, New York, NY 10032, USA

Corresponding Author Contact:

W. Ian Lipkin, MD
722 West 168th St. 17th Floor
New York, NY 10032
wil2001@columbia.edu
Phone: 212-342-9033
Fax: 212-342-9044

RUNNING TITLE: Microbial discovery

WORD COUNT: 3,513

AUTHOR AFFILIATION: Columbia University, New York, NY (W.I. Lipkin)

ABSTRACT: The fields of microbial surveillance, discovery and pathogenesis are evolving rapidly with introduction of cultivation-independent molecular diagnostics and highly multiplexed serology, as well as the development of animal models and prospective birth cohorts that can provide insights into host and microbial determinants of health and disease. Here I provide past, present and future perspectives on these fields in the context of my professional and personal relationship with Hilary Koprowski.

Word count: 69

KEYWORDS: microbial surveillance and discovery, pathogenesis, microbiota, diagnostics, birth cohorts

An evening with Hilary was a dance that required physical and mental stamina. He would rapidly and seamlessly shift his focus between such different challenges as musical composition, making pesto, the provenance of a painting or the production of vaccines in transgenic plants. Such musings frequently ran late into the night, punctuated only by the need to fill glasses or attend to the dog at his feet. Hilary had an uncanny talent for extracting the essence of a concept, a task or an object whether it was a dry martini, a deconstructed happy birthday tune or a scientific hypothesis. In another incarnation, Hilary might have written Satie's Gymnopédies, (although he favored complex tone poems) or created Matisse's gouaches découpés; however, events and opportunities led him into the current of science. Many owe their health to his trajectory; nonetheless unlike great works of art, comparable achievements in research can have a short half-life in the public eye. Hilary's son Christopher wrote me and other friends shortly after Hilary's death with profound disappointment that the New York Times had not carried an obituary. I wrote to key people at the Times who cover science and medicine. They agreed that this was an oversight, and referred me to the obit editor. She was persuaded and we ultimately secured the public memorial Hilary wanted and deserved.

Many of my colleagues point to reading Paul De Kreif's book, "Microbe Hunters", during childhood as pivotal in their choice of career. I was more captivated as a boy by tales of archaeologists finding lost cities and explorers charting new routes across the globe. Thus, my road to biomedical research and infectious diseases was circuitous. I wanted to be an anthropologist and only moved into medicine because this seemed to be a good way to have something to offer other cultures in exchange for mining their myths and rituals. Toward the

end of an internal medicine residency at the University of Washington, I interviewed with the Centers for Disease Control Epidemiology Intelligence Service program but elected instead to pursue additional training in neurology at UCSF when informed that I would not be able to focus on work on infectious diseases in the developing world. The emergence of AIDS in San Francisco and the pandemic that followed brought me back on track and highlighted the challenge that has become my life's work: improving global capacity for pathogen discovery and surveillance, and understanding the role of infection and immunity in chronic diseases. I joined Michael Oldstone's group at Scripps for training in virology and pathogenesis. Shortly thereafter he sent me to retrieve Hilary from the Los Angeles airport. I'd known his son Christopher at the University of Chicago and Chris' wife, Mary, from Pittsburgh (she signed out her service to me when I started internship). However, this was no preparation for meeting the patriarch who talked nonstop for 90 minutes about his latest passion, retroviruses and unexplained neurologic diseases. A few months later, Floyd Bloom drew my attention to a paper by Hilary and Rudy Rott suggesting that bipolar disorder was due to infection with an as yet uncharacterized agent [1]. With encouragement from Floyd, Michael and Hilary I decided to tackle this problem using subtractive cDNA cloning -- a project that took three years but ultimately demonstrated the power of genetic methods for diagnostics and discovery and provided a roadmap for proving and disproving links between microbes and disease using molecular methods [2, 3]. Throughout his career, Hilary invested the majority of his effort in prophylaxis and treatment. He nonetheless loved the process of pathogen discovery and the implications of discovery for medicine and public health. The common thread in the presentations at the symposium we are representing here is that each of us is committed to the integration of basic and clinical research to advance the practice of

78 medicine. My paper will focus on current and future perspectives on strategies for identifying
79 microbial targets for that research.
80
81 **Mechanisms of Pathogenesis and Proof of Causation**
82 The democratization of molecular methods for microbial discovery has enabled insights into
83 the pathogenesis of human and animal diseases. However, failure to appreciate the potential
84 pitfalls in the use of sensitive platforms and the need to rigorously test the significance of
85 experimental results has led to erroneous links that have undermined clinical medicine and
86 public health. Prominent examples of such misadventures include MMR vaccine and autism,
87 [4, 5] XMRV and prostate cancer [6, 7] or chronic fatigue syndrome [8, 9]. To pursue
88 meaningful work in pathogen discovery, it is important to have criteria for determining the
89 importance of a finding. Most biologists are familiar with Koch's Postulates for proving a
90 causal relationship between a microbe and a disease--namely, that a microbe must be found
91 in all hosts with disease but not in healthy hosts, and that the microbe must be isolated and
92 shown to cause disease when introduced into a healthy host. The immediate challenge in
93 implementing these postulates is that some agents may not be amenable to culture or that
94 there may be no suitable animal model for infection studies. These postulates were updated
95 to reflect the introduction of molecular methods by allowing for detection of nucleic acid
96 sequences representing a microbe rather than a pure culture that could be used for infection
97 studies. Nonetheless, a fundamental flaw in the postulates remained—namely, that other
98 factors may modify host responses and confound a one-to-one relationship between the
99 presence of an agent and disease. The Hill criteria examine the strength of association
100 between agents and diseases by considering biological plausibility as well as consistency,

specificity, and, where applicable, experimental evidence such as that developed in animal models described by Koch [10]. We have adopted a pragmatic approach that integrates Koch's Postulates and Hill's Criteria in grading causality as possible, probable or confirmed [11]. A **possible** relationship is one where there is evidence of exposure to a microorganism through the isolation and growth in culture from clinical samples, the detection of genetic sequences or proteins, visualization by microscopy, or demonstration of an adaptive immune response to a microorganism. A **probable** relationship is one where there is precedent for a similar disease caused by a similar agent in either the same or a similar host (particularly where there is a cluster of individuals with similar disease), the concentration of the microorganism (or nucleic acid, protein or antibody) is high at the site of pathology, or an antibody response clearly indicates recent exposure. A **confirmed** relationship requires fulfillment of Koch's postulates or demonstration that disease can be abrogated or mitigated through the use of specific drugs, antibodies or vaccines.

In the most straightforward examples, microbes cause disease through cell damage at the site of infection as a result of replication, or by triggering innate or adaptive immune responses. In such examples, the agent (or some nucleic acid or protein component) is likely to be present at the time disease is manifest, facilitating detection and implication. In other situations, more complex mechanisms may pertain. Microbes can induce neoplasia through interference with cell cycle controls. Here too, footprints of the agent may remain in the transformed cells; however, effects may be indirect and reflect inflammation or local metabolic disturbances. Immune responses to the microbe that break tolerance to self or the elaboration of toxins may have profound remote effects that can be challenging to link to the

causative agent. Finally, microbiome research is revealing intricate interactive networks wherein bacteria (and presumably fungi and viruses) have regional and systemic effects on host immunity, metabolism and development. Thus, we can anticipate new postulates and criteria that move beyond implicating microorganisms in disease to assessing the relevance of potential microbial symbionts in health, infectious diseases and responses to drugs and vaccines.

Microbial diagnostics, surveillance and discovery

Molecular assays are rapidly replacing culture in diagnostic microbiology. Culture remains important in drug sensitivity testing and is essential to establishment of in vitro and in vivo systems for pathobiology and vaccine development; nonetheless, molecular methods are rapid, sensitive and may succeed where an infectious agent has fastidious growth requirements. The common molecular assay platform, the polymerase chain reaction, has been adapted to field as well as laboratory use, and is employed for applications as diverse as blood screening for transmissible agents to monitoring the efficacy of antiviral drugs. The presence of a genetic target is revealed through the release of a marker (typically fluorescent) or an amplification product that is detected by a light sensor or mass spectrometer. The majority of PCR assays test for individual infectious agents; however, in recognition of the fact that history and clinical presentation are rarely pathognomonic of infections with individual agents, there is increasing emphasis on multiplex assays that address a panel of agents associated with syndromes. Indeed, multiplex panels are described that detect up to 30 different infectious agents implicated in respiratory diseases, meningoencephalitis and enteric diseases. DNA microarrays have a still higher capacity for

multiplexing. In their earliest versions, hundreds of oligonucleotide probes were spotted onto glass slides; more recently millions of probes have been synthesized in situ onto glass slides or silicon wafers. Although this platform has the potential to represent and detect the known microbial world, processing is cumbersome, requires a large investment in equipment and is two orders of magnitude less sensitive than real time PCR. Thus, while microarrays have enabled characterization of samples wherein microbial target concentrations were high, they have not yet become mainstream tools in clinical microbiology. There are efforts underway to build more sensitive, user-friendly syndrome-focused microarrays that will ultimately lead to ambulatory, bedside and field applications; however, the most exciting innovations in molecular diagnostics at present are in the field of DNA sequencing.

Largely through investments in technology to support human genome projects, the cost for sequencing has decreased from $5,000 per megabase in 2001 with classical Sanger dideoxy methods to $0.5 per megabase in 2012 with the Illumina and Ion Torrent platforms. The reduction in cost as well as an increase in speed has resulted in the discovery of thousands of new microbes; applications in microbial evolution, forensics and surveillance; and new opportunities to explore host-microbe interactions as determinants of disease. The challenge now is less in obtaining sequence data than in finding ways to process those data. Indeed, many recent requests for collaboration focus on analyzing data from already completed sequencing projects. The next phase in the democratization of sequencing will likely involve improvements in access to cloud computing and high-throughput sequencing software.

In contrast to molecular diagnostics where advances in technology have dramatically improved sensitivity, specificity and breadth over the past 20 years, serology is largely unchanged. This lag is important given that molecular methods are inadequate in instances where microbial nucleic acids are not present in an accessible sample or where disease may have been triggered by an agent that is no longer present. The risk of developing disease and its severity after exposure to an infectious agent is modulated by an individual's previous exposures to similar agents or vaccines. Such exposures may confer complete or partial protection or result in increased risk for more severe disease due to antibody-mediated enhancement. Thus, knowledge of an individual's immunological history may influence decisions concerning his/her treatment, vaccination or deployment as a first responder in areas where there is an increased probability of encountering high threat pathogens such as those implicated in hemorrhagic fevers.

Serological assays provide evidence that a host has encountered and responded to a microbe through the activation of specific B-cells. An activated B-cell can respond to femtomolar antigen concentrations and generate up to 109 specific antibody molecules in a week; hence, antibody assays have potential for extraordinary sensitivity. Antibodies are also remarkably stable. To date most serology has been performed using singleplex enzyme-linked immunosorbent assays (ELISA). More recently, Luminex-based systems have been employed that can address up to 100 antigenic targets simultaneously (i.e., 100 individual pathogens, 100 individual antigenic targets for one pathogen, or some variation thereof) [12]. Additionally, arrays are established that comprise spotted recombinant proteins expressed in vitro in *E. coli*, *S. cerevesiae*, baculoviruses, or cell-free, coupled transcription-translation

systems. However, assay development is complex and time consuming; hence, rapid response to an emerging pathogen is impractical. Another method, Luciferase immunoprecipitation systems (LIPS) has been established wherein protein fragments fused to an enzyme reporter are expressed in eukaryotic cells, mixed with serum and immunoprecipitated using protein A/G magnetic beads [13]. A major advantage of this approach is that no secondary antibodies are required; thus, LIPS is ideal for use in pathogen discovery in species where host-specific immunological reagents would otherwise be limiting. Although establishment of a new LIPS assay typically requires only 10-15 days from the time that genomic sequence data becomes available, it does not achieve the degree of multiplexing required to complement what has become standard in direct nucleic acid detection systems.

One strategy to fill this gap is to exploit a programmable peptide chip technology using a similar approach to that employed in oligonucleotide arrays. At present up to 2 million independent polypeptides can by synthesized on a single microarray—a density sufficient to address up to three different proteins of all known vertebrate viruses with overlapping peptides. Such an array could be used not only to survey for exposure to known human pathogens but also for those that emerge through zoonotic transmission.

The successful application of RNA and protein profiling in cancer diagnostics led to speculation that similar methods might be implemented in infectious diseases. Indeed, RNA profiling of peripheral blood has enabled distinction of acute viral versus bacterial respiratory tract infection [14]. Although this strategy may yet prove useful in clinical medicine, it has not

been widely employed and in my experience is insufficient to replace specific microbial diagnostic assays. What it can do, however, is provide insights into mechanisms of disease, prognosis and routes for intervention. Accordingly, the National Institute of Allergy and Infectious Diseases recently initiated the Systems Biology Program for Infectious Disease Research [15].

Social media and situational awareness

The growth of social media has had a profound impact on infectious disease surveillance. Since 1994, ProMED-mail (Program for Monitoring Emerging Infectious Diseases) (http://www.promedmail.org), has delivered continuous updates on new or active outbreaks of infectious diseases to a roster of subscribers that currently comprises more than 60,000 individuals in more than 185 countries. Information is submitted by physicians, veterinarians and public health practitioners worldwide and curated with commentary by an expert panel before posting to subscribers via a listserv. HealthMap (http://www.healthmap.org/en/), a more comprehensive program, assembles reports from news media, ProMED-mail and official documents as well as geo-referenced crowd sourced data into a map that displays real-time updates of disease emergence with hyperlinks to more detailed information. Both ProMED-mail and HealthMap have demonstrated their capacity to identify outbreaks in advance of general media. Future improvements will likely include systems that aggregate medical service utilization data, prescription and over-the-counter drug purchases, school and occupational absenteeism data to provide continuous measures of signal consistent with infectious disease emergence. Zoonoses, infections that originate in wildlife or domestic

238 animals, account for more than 70% of emerging infectious diseases; thus, a substantive
239 surveillance system for humans must also include nonhuman animal surveillance.
240
241 **Microbiota**
242 Our bacterial microbiota comprise 10^{11} cells, outnumbering our own cells by a factor of 10 to
243 1. Less is known about our mycobiome or virome; nonetheless, virus elements comprise
244 approximately 8% of the human genome, and have enabled such important evolutionary
245 advances as placentation. The retroviral element, syncitin, is implicated not only in fusion of
246 fetal and maternal uterine cells to form the placenta but also in inhibiting immune responses
247 that could result in rejection of the fetus. Evidence is accumulating that the microbes that
248 inhabit our oral, gastrointestinal and genitourinary tracts, skin and mucosal surfaces
249 contribute to normal physiology and development and as well as disease. Indeed, our
250 microbiota represent an epigenetic inheritance that influences our digestion, metabolism and
251 propensity to allergy, autoimmunity and neoplasia. The earliest studies of microbiota were
252 phenomenological and described associations between bacterial phyla and clinical
253 syndromes such as obesity [16]. More recently, the use of gnotobiotic animals (for example,
254 mice born in aseptic conditions and subsequently colonized with known flora) has allowed
255 investigators to test specific hypotheses regarding the role of specific microbes in disease [17,
256 18]. Such preclinical studies have inspired fecal transplant interventions in a wide range of
257 disorders ranging from autism to cancer [19]. At the time of writing, the only Food and Drug
258 Administration approved indication for fecal transplantation is in individuals with *Clostridium*
259 *difficile* not responsive to antibiotics [20]. However, this will likely change as more is learned
260 about our microflora. In work as yet unpublished, we have recently found specific vaginal

microbiota associated with stillbirth and risk of HIV, and enteric bacteria associated with specific forms of colon cancer. In addition, some bacteria express products with antimicrobial activity that may have value in medicine [21]. I can envision an era wherein infectious disease specialists prescribe microbial replacement to prevent or treat disease; however, as of yet there are no rigorous methods for defining optimal microflora let alone banking optimal donor microbiota.

Microbes and neuropsychiatric disease

My work in pathogen discovery began in response to Hilary's paper linking Borna disease virus to bipolar disorder. Although this work ultimately severed that link, the investment during the course of that project in developing genetic methods for identification of novel viruses enabled the discovery of more than 600 novel viruses in our laboratory alone and helped launch the application of high-throughput sequencing applications that have transformed basic and clinical microbiology. What lagged, however, was the original effort to investigate the role of microbes in mental illness.

A wide range of infectious agents has been implicated in major depression, bipolar disorder, autism and schizophrenia including bacteria, viruses and parasites. However, no one organism has been found to be specific for any one disorder and in most instances no link is made to any infectious agent. Indeed, evidence from animal models indicates that a key determinant of damage in autism and schizophrenia, two neurodevelopmental disorders for which risk is thought to be established during fetal life [22, 23], may be the innate immune response. Gestational exposure of dams to the viral and bacterial mimics poly (IC) and

lipopolysaccharide (LPS) are sufficient to induce disease in offspring in rodent and primate models of autism and schizophrenia. In the case of poly (IC) the use of nonsteroidal anti-inflammatory drugs aborts damage to neuronal stem cells as well as abnormal behaviors.

Although we cannot replicate such experiments in humans, the advent of large prospective birth cohorts provides an opportunity to test their validity. One example of such a cohort is the Autism Birth Cohort (ABC) [24]. The ABC was initiated to analyze gene-environment-timing interactions in an unselected nationwide birth cohort in Norway wherein cases of disease are prospectively ascertained through population screening. Samples collected serially through pregnancy and childhood include parental blood, maternal urine, cord blood, milk teeth and rectal swabs. Children are continuously screened through questionnaires, referral, and a national registry. Cases are compared with a control group from the same cohort in a 'nested case–control' design. Genetic, proteomic, immunologic, metagenomic and microbiological tools are then used to exploit unique biological samples. Pilot studies with questionnaire data indicate an increased risk of autism in offspring of mothers who report febrile episodes during pregnancy and in offspring who have elevated levels of pro-inflammatory cytokines in cord blood (Hornig and Lipkin, unpublished data). The challenge now is to determine the agent or agents that triggered fever and inflammation and, if applicable, the timing of the exposure and genetic determinants of the host response that culminated in neurodevelopmental damage. Given that the agent(s) may not be present at the time samples were collected (during pregnancy or on the day of birth) it will be essential to pursue not only molecular diagnostics but also serology to look for immunological foot prints of exposure that correlate with disease. In my view, cohorts like the ABC represent the future of pathogen discovery and

pathobiology—a future that Hilary would have enjoyed and exploited in developing interventions to reduce the burden of morbidity and mortality as well as health care costs.

Conclusion

On June 18, 1980, Hilary recounted his version of the development of the polio vaccine for a series established by Fred Rapp entitled Frontiers of Virology. He began and closed his essay with references to Marcel Proust's "Search for Lost Time" (also known as "Remembrances of Things Past"). I will use Hilary's own words in closing this essay dedicated to his memory. "No matter what our fate has been, not sorrow, nor a joy, a triumph, a tragedy, need be irretrievably lost, but may be preserved as a creative process in either art or science, in it freshness and dew." Although he is not with us to participate in discoveries yet to come, each of the essays in this monograph is evidence that Hilary has not been lost. He lives on in the art, music and science that we share.

FIGURE LEGENDS

Figure 1. Hilary Koprowski and the author in November 2002 on the occasion of the author's 50th birthday, for which Koprowski composed a deconstructed rendition of "Happy Birthday" (copyright Warner/Chappell).

REFERENCES

1. Rott R, Herzog S, Fleischer B, et al. Detection of serum antibodies to Borna disease virus in patients with psychiatric disorders. Science **1985**; 228:755-6.

2. Lipkin WI, Travis GH, Carbone KM, Wilson MC. Isolation and characterization of Borna disease agent cDNA clones. Proc Natl Acad Sci U S A **1990**; 87:4184-8.

3. Hornig M, Briese T, Licinio J, et al. Absence of evidence for bornavirus infection in schizophrenia, bipolar disorder and major depressive disorder. Mol Psychiatry **2012**; 17:486-93.

4. Wakefield AJ, Murch SH, Anthony A, et al. Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. Lancet **1998**; 351:637-41.

5. Hornig M, Briese T, Buie T, et al. Lack of association between measles virus vaccine and autism with enteropathy: a case-control study. PLoS One **2008**; 3:e3140.

6. Urisman A, Molinaro RJ, Fischer N, et al. Identification of a novel Gammaretrovirus in prostate tumors of patients homozygous for R462Q RNASEL variant. PLoS Pathog **2006**; 2:e25.

7. Lee D, Das Gupta J, Gaughan C, et al. In-depth investigation of archival and prospectively collected samples reveals no evidence for XMRV infection in prostate cancer. PLoS One **2012**; 7:e44954.

8. Lombardi VC, Ruscetti FW, Das Gupta J, et al. Detection of an infectious retrovirus, XMRV, in blood cells of patients with chronic fatigue syndrome. Science **2009**; 326:585-9.

9. Alter HJ, Mikovits JA, Switzer WM, et al. A multicenter blinded analysis indicates no association between chronic fatigue syndrome/myalgic encephalomyelitis and either xenotropic murine leukemia virus-related virus or polytropic murine leukemia virus. MBio **2012**; 3.

10. Hill AB. The Environment and Disease: Association or Causation? Proc R Soc Med **1965**; 58:295-300.

11. Lipkin WI. The changing face of pathogen discovery and surveillance. Nat Rev Microbiol **2013**; 11:133-41.

12. Vignali DA. Multiplexed particle-based flow cytometric assays. J Immunol Methods **2000**; 243:243-55.

13. Burbelo PD, Ching KH, Bush ER, Han BL, Iadarola MJ. Antibody-profiling technologies for studying humoral responses to infectious agents. Expert Rev Vaccines **2010**; 9:567-78.

14. Zaas AK, Chen M, Varkey J, et al. Gene expression signatures diagnose influenza and other symptomatic respiratory viral infections in humans. Cell Host Microbe **2009**; 6:207-17.

15. Aderem A, Adkins JN, Ansong C, et al. A systems biology approach to infectious disease research: innovating the pathogen-host research paradigm. MBio **2011**; 2:e00325-10.

16. Ley RE, Turnbaugh PJ, Klein S, Gordon JI. Microbial ecology: human gut microbes associated with obesity. Nature **2006**; 444:1022-3.

17. Ridaura VK, Faith JJ, Rey FE, et al. Gut microbiota from twins discordant for obesity modulate metabolism in mice. Science **2013**; 341:1241214.

18. Faith JJ, Ahern PP, Ridaura VK, Cheng J, Gordon JI. Identifying gut microbe-host phenotype relationships using combinatorial communities in gnotobiotic mice. Sci Transl Med **2014**; 6:220ra11.

19. Aroniadis OC, Brandt LJ. Fecal microbiota transplantation: past, present and future. Curr Opin Gastroenterol **2013**; 29:79-84.

20. US Department of Health and Human Services Food and Drug Administration Center for Biologics Evaluation and Research. "Guidance for Industry: Enforcement Policy Regarding Investigational New Drug Requirements for Use of Fecal Microbiota for Transplantation to Treat Clostridium difficile Infection Not Responsive to Standard Therapies". Available at:

http://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/Vaccines/ucm361379.htm. Accessed 29 Sept 2014.

21. Donia MS, Cimermancic P, Schulze CJ, et al. A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. Cell **2014**; 158:1402-14.

22. Rapoport JL, Giedd JN, Gogtay N. Neurodevelopmental model of schizophrenia: update 2012. Mol Psychiatry **2012**; 17:1228-38.

23. Patterson PH. Maternal infection and immune involvement in autism. Trends Mol Med **2011**; 17:389-94.

24. Stoltenberg C, Schjolberg S, Bresnahan M, et al. The Autism Birth Cohort: a paradigm for gene-environment-timing research. Mol Psychiatry **2010**; 15:676-80.

FOOTNOTE PAGE

Conflict of Interest Statement:

The author has no commercial or other association that might post a conflict of interest.

Funding Statement:

There is no funding source associated with this publication.

Corresponding Author Contact:

W. Ian Lipkin, MD
722 West 168th St. 17th Floor
New York, NY 10032
wil2001@columbia.edu
Phone: 212-342-9033
Fax: 212-342-9044