The missing institution to link infectious and chronic disease

Harshu Musunuri

We often think of infections as temporary– illnesses that sicken us for a short period of time, but which we quickly recover from. Yet a growing body of evidence suggests this understanding is incomplete. The impact of an infection can stick with us, contributing to long-term conditions like dementia, cardiovascular disease, autoimmunity, and cancer. The categories we use clinically, acute versus chronic, are convenient approximations rather than biological realities; in practice, infection outcomes lie on a continuum that ranges from a brief illness resolving in days, to a chronic infection persisting for years, to an acute episode followed by downstream consequences that emerge months or even decades later.

The list of plausible links is by now substantial. Established causal connections include H. pylori and gastric cancer, hepatitis B/C and liver cancer, HPV and cervical cancer, Schistosoma and bladder cancer, Epstein-Barr virus and several lymphomas, and HHV-8 and Kaposi's sarcoma. More recent work has extended the picture into neurological and metabolic disease. EBV is now considered the leading cause of multiple sclerosis, with 32-fold elevated risk following infection, evidence of molecular mimicry between EBNA1 and the brain protein GlialCAM, and animal models that recapitulate disease severity upon EBNA1 exposure. The herpes zoster vaccine has been shown to reduce new dementia diagnoses. P. gingivalis IgA antibody titers predict myocardial infarction independently of classical cardiovascular risk factors. Coxsackievirus B is increasingly implicated in the autoimmune destruction of pancreatic beta cells in type 1 diabetes. Even adenovirus-36 has a reproducible, environmentally driven association with obesity in twin studies. Long COVID has finally made the broader phenomenon legible to the public, although the same kind of post-acute syndrome has long been documented for RSV, Ebola, Zika, and others.

Despite the accumulating signals, the field that should exist to systematically convert these observations into prevention does not. The reason for this is structural rather than scientific.

Establishing that a common pathogen causes a chronic disease, at the standard of evidence required to justify a vaccine program or a public-health recommendation, requires assembling an evidence stack that no single laboratory, grant, or institution is designed to produce. One needs decades-long prospective cohorts with sera banked at multiple time points, standardized multi-pathogen serology and autoantibody panels run identically across those samples, paired tissue archives that can be interrogated for persistent viral sequence and antiviral signatures, HLA and genetic data sufficient for Mendelian randomization, animal models capable of recapitulating disease from a candidate epitope, and natural-experiment epidemiology (such as the accidental Welsh shingles vaccine rollout). Each of these is, on its own, a multi-year program.

Universities cannot run thirty-year programs on five-year grants, and the academic incentive structure rewards the publishable increment rather than the assembled stack. Pharmaceutical companies will not pay to de-risk a target, since they enter only after one already exists. Public health agencies operate downstream of established causality rather than upstream of it. Existing biobanks like UK Biobank and All of Us are extraordinary resources, but they were not built around the infection to chronic disease question, and the assays needed to answer it have not been standardized across them. As a consequence, the evidence required to justify a vaccine against, for instance, P. gingivalis, Ad-36, or coxsackievirus B simply never gets assembled, not because the science is impossible but because no institution exists whose job it is to do so. The infection to chronic disease causal stack is, structurally speaking, an FRO-shaped problem: well-defined, infrastructure-heavy, and public-good in output.

Proposal

I propose a five-year FRO whose mission is to systematically test, and either establish or rule out, causal links between common pathogens and chronic diseases, and to produce public goods that allow others to act on those links. Three concrete deliverables, presented here in order of dependency, would define the work.

Deliverable 1: a modern postulate framework. Before the organization can produce defensible verdicts, the field needs an agreed-upon evidence rubric for what it means to declare an infection causal for a chronic disease. The original Koch's postulates, written for acute bacterial infections, fail in this context, because the pathogen is often cleared by the time disease appears, the mechanism may be autoimmune rather than direct, and exposure is near-universal so simple correlation carries little information. A modernized framework would specify, for each candidate link, the required evidence across several categories: seroepidemiology in prospective cohorts, including exposure-discordant pairs; Mendelian randomization and HLA-stratified risk; autoantibody and molecular-mimicry signatures using techniques such as PhIP-seq; tissue-level evidence of persistent antigen, viral sequence, or antiviral transcriptional signatures; animal recapitulation, ideally including the stricter test of immunizing with a candidate epitope alone; and natural-experiment or interventional epidemiology where vaccines or antimicrobials are available. The deliverable is the framework itself, validated by retrospective application to well-established links such as EBV in MS, HPV in cervical cancer, and H. pylori in gastric cancer, alongside contested cases.

Deliverable 2: a federated infrastructure across existing biobanks. Rather than constructing a new cohort from scratch, the organization would build the standardized assay layer that knits existing cohorts together for this specific question. Concretely, this means a harmonized multiplexed serology panel covering 50 to 100 candidate pathogens, a standardized autoantibody array, and a tissue-level persistence assay protocol, all run identically across partnered biobanks (UK Biobank, All of Us, FinnGen, the Million Veteran Program, and several disease-specific cohorts). The output is a public dataset that links pathogen exposure history, autoantibody profiles, and downstream chronic-disease incidence at a scale no single biobank can produce on its own.

Deliverable 3: end-to-end verdicts on a portfolio of candidate links. Over its lifetime, the organization commits to producing a defensible verdict (causal, not causal, or with a precisely specified set of remaining experiments) on a portfolio of five to ten candidate infection to chronic disease links. Initial candidates might include coxsackievirus B in type 1 diabetes, P. gingivalis in cardiovascular disease, HHV-6 in chronic fatigue syndromes, Ad-36 in obesity, and one or two long-COVID-style post-acute syndromes. The deliverable is not a paper but rather a verdict with its supporting evidence stack, handed off to vaccine developers, regulators, and public-health bodies who can act on it.

Why now

Three things are true today that were not true a decade ago, and their combination is what makes this project tractable for the first time. First, the assay layer has collapsed in cost. PhIP-seq can profile antibody reactivity against hundreds of thousands of peptides for a few hundred dollars per sample, and multiplex serology and spatial transcriptomics have become common techniques. Second, the cohort layer is unprecedented. The simultaneous existence of UK Biobank, All of Us, FinnGen, and the Million Veteran Program puts roughly five million people with linked health records and banked samples within reach of a coordinated effort, and the natural experiments created by COVID and by staggered vaccine rollouts (shingles, HPV, and others) have produced interpretable interventional data essentially for free. Third, the FRO model itself now exists as a working institution-design pattern. These three conditions, together, make the work feasible.

The timing is also right politically. The current framing of chronic disease at the federal level has emphasized environmental drivers such as pesticides and microplastics, and while those factors deserve rigorous study, the

relative neglect of infectious drivers represents a meaningful blind spot. The framing of chronic disease etiology is genuinely of interest to multiple stakeholders, and the institution that produces rigorous causal evidence will play a substantial role in shaping it.

Two-year pilot

While the FRO is a five year commitment, the concept itself could be falsifiable within two years and at modest cost, on the order of ten to fifteen million dollars, through a scoped pilot with three components.

The first component is a convened working group of immunologists, epidemiologists, vaccine developers, and biostatisticians, charged with producing version one of the modernized postulate framework and applying it retrospectively to three well-established links and three contested ones. Success here looks like a framework that domain experts pre-commit to finding defensible, and that identifies the specific missing evidence for each of the contested cases.

The second component is to build and run the harmonized assay layer end-to-end on a single contested link. Coxsackievirus B and type 1 diabetes is a strong starting choice, given the availability of relevant cohorts, well-developed animal models, and a vaccine candidate already in clinical development. The goal is to produce, within two years, either a defensible causal verdict or a precisely specified set of remaining experiments needed to reach one.

The third component is a demonstration of handoff. The pilot succeeds, in the broader systemic sense, only if the assembled evidence stack actually moves the downstream system, whether that means a vaccine developer committing to advance or shelve a program based on the verdict, a regulator engaging seriously with the evidence framework, or a public-health body issuing guidance.

The pilot is informative regardless of outcome. If it produces a defensible verdict and triggers downstream action, the case for the full organization is made. If, on the other hand, the evidence stack remains stubbornly inconclusive even when fully assembled, that itself constitutes an important finding about what the field genuinely lacks and what additional infrastructure or methods would be required.

Complementary reform

A natural extension of the proposal, which the organization could advocate for but does not require in order to function, would be to make it standard practice for vaccine trials and infection natural-history studies to capture autoantibody profiles, persistent-antigen signals, and long-symptom outcomes at baseline and follow-up, using the standardized panels the organization develops. The marginal cost of doing so, once the assay layer exists, is small. The benefit is that every future trial becomes a contribution to the chronic-disease causal map rather than a missed opportunity. As things currently stand, by the time a chronic signal emerges from a cohort, the samples and follow-up needed to investigate it have typically ceased to exist. A modest regulatory or funder-led requirement, layered on top of the standards the organization produces, would change that permanently and at low marginal cost.

The frontier between infectious and chronic disease is one of the most consequential and underexploited surfaces in biomedicine. Prevention is, almost without exception, easier than treating a disease once it has manifested, and a lifetime of poor health is rarely a price worth paying for so-called natural immunity. There is thus a moral imperative to understand the links between seemingly benign infections and the chronic conditions that follow them. Building this institution thus may turn out to be among the most important things we can do for a healthy human lifespan.