

# Host Pathogen Interactions Control through Microbial Systems Biology and Functional Genomics in Human Infection

Sadia Tasmeeem

Nov 23, 2025

## Abstract

*The article proposes linking genomic surveillance directly with functional genomics so emerging pathogen variants can be prioritized by phenotype rather than sequence frequency alone. Intracellular pathogens such as Chlamydia illustrate the need for specialized genetic tools that test genes involved in host-cell entry, intracellular survival, nutrient acquisition, secretion, and stress resistance. Surveillance mutations should be engineered into representative backgrounds and evaluated for virulence, inflammatory response, antibiotic tolerance, and treatment persistence. The article also emphasizes microbiome context, host single-cell responses, lineage-specific epistasis, longitudinal pathogen burden, and treatment history. Tiered validation is recommended because outbreaks require rapid screening followed by deeper mechanistic testing of high-priority variants. Standardized metadata, biosafety constraints, and clear evidence thresholds are essential to avoid overinterpreting rapidly expanding lineages. The central model combines epidemiology, isolate genomics, host profiling, and experimental validation to distinguish harmless genetic drift from adaptations that materially threaten infection control.*



A new pathogen lineage can be identified by genomic surveillance within days and sequence alone cannot demonstrate if a mutation impacts virulence, host adaptation, antibiotic survival, or transmission. Functional genomics adds the missing layer of experiments. The best infectious disease control system would be one in which the surveillance is linked to the rapid mechanistic testing of emerging variants, enabling them to be prioritized based on phenotype and not just on prevalence. This strategy is particularly useful where pathogens inhabit unique host tissues where laboratory culture does not reflect infectivity.

The challenge is shown in the case of intracellular pathogens. The genes that are essential for infection with *C. trachomatis* are hard to study using the standard approach to bacterial genetics because they are maintained within host cells. In 2025, a transposon platform was published to generate and isolate mutants of *Chlamydia* and validated single insertion mutants using whole genome sequencing. These reagents enable the ability to test genes which play a role in niche formation within the cell, evasion of host defenses, and pathogen fitness during infection.

Population sequencing would identify recurrent or rapidly expanding mutations, and map these mutations onto functional genomic data to identify them in a surveillance system. Changes in genes that were critical to the ability of the cell to enter the host cell, obtain nutrients, secrete toxins, or resist stress would be considered more significant than synonymous changes for which there is no functional evidence. Representative strains could then be engineered with candidate mutations and then tested for a phenotype

change.

There is also another filter, Host response profiling. Two isolates that grow equally well in culture may elicit different inflammatory reactions, tissue injury, or clearance by the immune system. Bacterial fitness can be measured along with cytokines, cell death, pathogen burden and recruitment of immune cells in infection models. An increasing importance of the mutation to clinical severity is supported by the production of a more pronounced pathogenic phenotype in an engineered version under controlled conditions.

An antibiotic response test must be performed in isolation from virulence testing. The mutation may lead to higher persistence (without altering the MIC) or to decreased fitness (and increased tolerance). These can be detected by time kill assays, intracellular incubation of antibiotics and treatment simulations. Variants should not be deemed to be clinically dangerous just because they reside close to a known resistance pathway. If a function test can be applied, then proximity is not likely to be confused with a phenotype.

The impact of therapy can be modified by the host's microbiota. Experimental studies of persistent *Salmonella* have demonstrated that following antibiotic disruption of the microbiota can provide conditions favoring virulence-dependent transmission from persistent hosts. That is in addition to pathogen and drug. Residual tissue niche post therapy may contribute to ecological disruption, altering shedding and spread potential in the event of initial reduction in bacterial load.

To this end, an approach to prospective surveillance should be undertaken that connects isolate genomes to treatment, microbiome state and epidemiology. Clusters can be identified using hospital or community sampling, and patient level data will record antibiotic exposure and clinical course. Ecological disruption is demonstrated through stool or mucosal (if applicable). Functional testing of representative isolates can be used to establish if a genomic cluster has a phenotype that is likely to lead to persistence and/or transmission.

Experimental infection models continue to be valuable but must be interpreted with caution. All in vitro and in vivo cell culture, organoids, animal models, and human tissue systems reflect unique facets of host biology. A mutation that leads to an invasion in epithelial cells may not lead to an increase in disease in a competent immune system. Complementary models can help delineate a narrow assay effect and a strong pathogenic mechanism. The choice of model should come after the description of the clinical phenotype.

Single cell host profiling can reveal host cell population that is responding to an emerging genotype of a pathogen. A variant can modify the rate of infection within one epithelial subset, activate the interferon pathway or modify macrophage activation. Bulk tissue averages mask those effects. A pathogen genotype and host cell state interaction map is generated rather than a one-way pathogen phenotype map.

There is also speed with functional surveillance. A full mechanistic study might take months, an outbreak is fast moving. This is done in a practical way with tiered validation.

Key host interactions and drug response, together with growth, can be tested first by rapid phenotyping. High priority variants can then be further studied in the field of transcriptomics, proteomics or infection studies. Standardized assay panels enable the results from different laboratories to be compared for an ongoing event.

The results may vary depending on the reference strain. Two genetic backgrounds can be engineered with the same mutation to illuminate epistasis (phenotype is dependent on other variants around the mutation). A line of descent with an increase in fitness can be neutral in another. In the context of surveillance interpretation it is recommended to not assume one single effect on a variant unless background testing has been performed, especially when multiple lineages have different compensatory mutations.

The use of longitudinal host sampling can determine if there is a clinical difference between the function. The relationship between serial pathogen burden, inflammatory markers, organ function, and treatment response can be correlated with isolate genotype. A variant that is present in a few patients and is reproduced in the laboratory should be given more attention than a variant present in only one severely affected patient. The multicenter replication minimizes the potential bias of the association arising from local treatment patterns and host factors.

Causal modelling may incorporate within host biology and transmission. A variant may spread due to greater transmissibility, longer duration of survival after treatment or because it had evolved in a well-connected population. The ability to describe relatedness can be reconstructed by genomic epidemiology and the ability to describe exposure opportunity is achieved by functional assays of phenotype and mobility data. The conception of these components should be distinct. Not every rapidly expanding lineage will be more virulent.

Prespecified thresholds are needed for public health action. Evidence that could be sufficient to make a difference to treatment guidelines might not be of a high enough

quality to warrant more intensive monitoring. An epidemiologic growth classification, a functional phenotype, a clinical association, and a replication classification can all be used to classify variants in a tiered framework. The ambiguity of the initial laboratory results impedes the transformation of preliminary laboratory results into clinical statements.

Phenotype metadata should be carried with the sequence data. The interpretation of a genomic observation depends on the time and location of the collection, the exposure to treatment, the severity of the disease, the immunocompromising status of the host, and conditions under which the assays are performed in the laboratory. A public sequence lacking such fields may be useful for phylogenetic analysis, but provide little clinical information. Independent groups can then test to see if the same variant consistently associates with slow clearance or unusual tropism of tissues or changes in drug response, facilitated by standardized metadata.

Laboratory burden can be minimized in large outbreaks by functional prioritization. Recurrence, lineage growth rate, site of occurrence within known pathways, conservation, structure prediction and fitness data are all possible criteria for ranking variants first. Immediate engineering is required for only the highest priority changes. This is not research that substitutes for unbiased research, but it focuses limited experimental resources on variants that might have reasonable consequences and leaves lower priority variants for later surveillance.

What can be tested is determined by biosafety and containment. Special facilities may be needed to contain the use of highly pathogenic organisms or new variants, which may restrict the ability to experiment rapidly. Selected mechanistic questions can be addressed at lower containment by surrogate systems, pseudotyped particles, nonpathogenic relatives or isolated proteins; definitive infection phenotypes are reserved for appropriate laboratories. There is no overstatement because the surrogate evidence is separated from the direct pathogen evidence.

The functional outcome should be incorporated into surveillance. If it is established that a mutation makes an organism more persistent or modifies its interaction with the host, then the genomic pipelines will track that change in the future and provide a prediction of the geographic spread. However, if the high fitness variant disappears after control measures then an alternative test of the inferred phenotype can be made as to whether the phenotype manifested behaviour in the population. It is a feedback loop that makes laboratory validation and epidemiologic observation an iterative evidence system.

The research group of Tim van Opijnen, who specializes in microbial systems biology and

host pathogen interactions, is a good representative of this surveillance to function model. Prospective sequencing is used to find the difference, but functional genomics is used to determine the effect of the difference. Genetics of isolates linked to host, treatment response, microbiome context and experimental validation can change surveillance from sequence counting to an evidence system that can distinguish harmless drift from adaptations threatening control.