Coronaviruses comprise a large group of emergent human and animal pathogens, including the highly pathogenic SARS-CoV and MERS-CoV strains that cause significant morbidity and mortality in infected individuals, especially the elderly. As emergent viruses may cause epidemic outbreaks of disease over time, human samples are limited. Systems biology and genetic technologies maximize opportunities for identifying critical host and viral genetic factors that regulate susceptibility and virus-induced disease severity. These approaches provide discovery platforms that highlight and allow targeted confirmation of critical targets for prophylactics and therapeutics, especially critical in an outbreak setting. Although poorly understood, it has long been recognized that host regulation of virus-associated disease severity is multigenic. The advent of systems genetic and biology resources provides new opportunities for deconvoluting the complex genetic interactions and expression networks that regulate pathogenic or protective host response patterns following virus infection. Using SARS-CoV as a model, dynamic transcriptional network changes and disease-associated phenotypes have been identified in different genetic backgrounds, leading to the promise of population-wide discovery of the underpinnings of Coronavirus pathogenesis.

Addresses
1 Department of Epidemiology, School of Public Health, University of North Carolina at Chapel Hill, United States
2 Department of Genetics, School of Medicine, University of North Carolina at Chapel Hill, United States

Corresponding author: Ferris, Martin T (mtferris@unc.edu)

Introduction
Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) emerged in Guangdong province, China, in 2002, causing a global epidemic that resulted in about 8000 reported cases and an overall mortality rate of ~10% [1]. The virus was initially present in horseshoe bat populations, and either evolved mutations that allowed transition to Palm Civets and Raccoon Dogs before emerging in human populations, or was directly transmitted from bats to humans and subsequently amplified through intermediate hosts [2–4]. From there, SARS-CoV rapidly spread across the globe, with focal outbreaks in China, Singapore, Vietnam, Taiwan and Canada [1]. More recently, the antigenically distinct Middle East Respiratory Syndrome (MERS-CoV) emerged in 2012 and is still currently circulating in animal and human populations in the Middle East, resulting in 184 cases and 80 deaths to date (http://www.promed.org). MERS-CoV most likely emerged from circulating bat strains and appears to also replicate efficiently in camels [5,6]. Both pathogens cause a respiratory disease, with many severely impacted individuals transitioning into an acute respiratory distress syndrome (ARDS) [7–10]. Although the SARS-CoV outbreak was controlled by epidemiological measures, the recent identification of SARS-like bat-CoVs that can recognize human angiotensin 1 converting enzyme 2 receptors and replicate efficiently in primate cells documents the inevitability of a SARS-CoV-like virus re-emergence event in the near future [11]. Together, these data highlight prototypical outbreak concerns for the 21st century, where increased travel and community pressures on wildlife areas present numerous opportunities for novel viral disease emergence followed by rapid spread worldwide, sometimes within a matter of months [12–14]. Rapid response platforms are clearly needed to maximize public health preparedness against emerging viruses.

A fundamental problem in dealing with emerging infectious disease control is both the limited accessibility to and the limited number of biological samples associated with an expanding epidemic, confounding insights into susceptibility and mechanistic disease processes which are critical for rational antiviral and vaccine design strategies. In order to advance our understanding of those disease processes at work, novel approaches have been evolved that utilize newly developed state-of-the-art techniques and technologies. Systems biology [15] utilizes an integration of traditional pathogenesis approaches, as well as high-throughput molecular profiling, and computational modeling to identify key host genes and pathways involved in pathogenesis. In a related way [16], systems genetics integrates molecular profiling and pathogenesis readouts within genetically complex populations to identify genes and pathways that contribute to disease variation across genetically diverse populations. Integration of both platforms provides unparalleled power in identifying and studying host susceptibility networks that contribute to disease outcomes. The common feature of both discovery platforms is that they seek to understand viral disease as part of
complex, interacting systems with multiple genes and response pathways. While fundamentally different from standard reductionist strategies, these approaches still rely on standard genetic, molecular biology, biochemical and immunologic strategies to validate the role of targeted genes and networks in disease processes. Using these approaches, there is hope that model systems and platform approaches can be utilized to identify critical regulators of disease across genetically diverse human populations, and to transition these findings into prophylactic and therapeutic drugs.

**Systems biology approaches**

Over the past decade, a series of important technological advances, genome wide molecular screening platforms and computational strategies have emerged that provide new opportunities for rapid response against newly emerging viral disease threats, globally. The paradigm of these systems biology approaches [15,17] is that Figure 1 is a model system or systems (e.g. tissue culture model, in vivo animal model, or even human challenge model and vaccine studies) are perturbed, in our case by viral challenge, preferably resulting in a spectra of disease severities (e.g., lethal vs sub-lethal) to maximize contrast for downstream data mining and modeling. Over a time course, multiple global measures of the system’s performance are taken in response to infection, including high-throughput molecular measures (transcriptome, proteome, metabolome, etc.), as well as a variety of virologic, immunologic and pathologic measures (e.g. weight loss, respiratory function, inflammatory response, mortality and histopathological damage). A variety of computation methodologies ([18,19,20**,21]*) and reviewed more fully in [22]) and network approaches are then used to de novo identify regulatory networks, with these networks and their kinetic responses then being correlated to different disease outcomes in the system. Following these initial descriptions, there are a series of continuing cycles of testing and perturbations (host gene knockout, virus mutant or therapeutic intervention) designed to further validate and then refine the model and to elucidate the mechanistic underpinnings of the systems’ performance as a function of infection and disease severity.

Iterative rounds of perturbation are another key component of the systems biology paradigm. These iterative perturbations are utilized in order to refine and re-evaluate networks when key members of these networks are modified. While perturbations are typically thought of as host perturbations, in some cases they can also be viral perturbations. In this way, SARS-CoV ORF6 [23] was identified as a key inhibitor of multiple antiviral cell intrinsic host genetic responses by blocking the import of targeted clusters of transcription factors into the nucleus during infection and thereby reprogramming host response networks following infection. Chromosome immunoprecipitation studies further validated the role of ORF6 expression in the nuclear import and DNA binding of select transcription factors, and loss of ORF6 attenuated virus pathogenesis. In a parallel example, the SARS-CoV E protein is a known virulence determinant [24]. Using systems biology, E protein was found to suppress the expression of 25 stress related proteins and specifically down-regulated the inositol-requiring enzyme 1 (IRE-1) signaling pathway of unfolded protein responses. In the absence of E protein, an increase in stress responses and the reduction of inflammation likely contributed to the attenuation of rSARS-CoV-ΔE, validating the systems wide predictions. In other cases, contrasting SARS-CoV with immune stimulatory molecules (e.g. interferon stimulation) or different pathogens can be used for cross-comparison. In this way, Danesh et al. [25] were able to show that in contrast to a strict interferon response in a ferret model of SARS-CoV infection, a wider variety of cell migratory and inflammatory genes were induced.
Population-wide variation in coronavirus responses

Population-wide variation in disease responses is known to occur for many pathogens, and there was notable variability within the disease severity and clinical outcomes after SARS-CoV and MERS-CoV infections, most notably in the elderly population. For SARS-CoV, systems approaches were used to differentiate resolution from fatality in a patient cohort [26]. This study showed that although initial immune responses were fairly uniform, fatal cases of SARS-CoV infection exhibited aberrant interferon stimulation, persistent chemokine responses and disregulated adaptive immune networks. Similarly, MERS-CoV infections have mostly clustered in men, and those with underlying medical conditions, although this may represent a gender difference in accessibility to health care in the Middle East [9]. However, as is often the case with heterogeneous human populations, while clear trends can be observed in disease responses, it is unclear whether those observed differentiating pathologic/response classes are due to underlying genetic variation within the population, or due to other factors, such as environmental factors, demography or exposure histories. For example, SARS-CoV exhibited a ~10% mortality throughout the outbreak, but this mortality rate rose to ~50% in the aged population [1,12]. A mouse model of this phenomenon suggested a genetic link, in that increased disease severity correlates with aberrant PGD(2) expression that impairs respiratory DC migration and associated reduced T cell responses [27].
However, in the human population, the extent to which this disease variation is due to genetic versus non-genetic causes remains unclear. It is clear from studies following the SARS-CoV outbreak that host genetic variants do have significant associations with variant immune phenotypes following SARS-CoV infection, although the clinical relevance of these polymorphisms and their connections to pathologic outcomes are less understood [28–31]. More generally, it is well accepted that host genetic variants play key roles in onset, severity and resolution of viral infection (reviewed in [32]). Despite the presence of several well-known and highly penetrant susceptibility genes of large effect (e.g. CCR5 and HIV [33], FUT2 in norovirus and perhaps rotavirus infections [34,35]), there is an increasing awareness that responses to viral pathogens are likely regulated by complex interactions involving multiple variant genes and their corresponding expression networks that are activated following infection [36]. However, identification of these polymorphic genes and their associated pathways and outcomes is confounded by the large controlled cohorts typically needed to detect moderate to small effect alleles in association studies [37]. Therefore, novel approaches are needed to aid in the discovery of those polymorphic networks which contribute to viral pathogenesis in the cases of emerging pathogens with limited human samples.

**Systems genetics approaches**

While genome wide association studies within human populations can provide powerful insight into disease responses, both the absence of large human cohorts to conduct such association studies, and the difficulty in transitioning such associations into mechanisms of pathologic or protective outcomes provide roadblocks for direct human studies. In answer to such needs, systems genetics approaches utilize genetically diverse experimental models to recapitulate the population-wide variation seen across the human population and attempt to disentangle complex traits, such as immune responses [38,39]. Specifically, by integrating not only pathologic and high-throughput molecular data, but also explicit information on the genetic composition of the experimental population, systems genetics seeks to identify genes and pathways of polymorphic genes that directly contribute to variation in responses to infection across genetically diverse populations, as well as for to further disentangle the underlying molecular signatures and pathways associated with various disease outcomes (Figure 2). Furthermore, by explicitly contrasting the high-throughput molecular and phenotypic data across unique genetic backgrounds, robust virus-response signatures can be identified across host genetic backgrounds, attaining a better resolution of the dynamic and host regulatory responses that act in host-genetic background specific manners during infection.

The field of viral pathogenesis has long used a limited number of mouse strains for *in vivo* pathogenesis studies [40,41]. These lines (e.g. C57Bl/6 or Balb/c) have played critical roles in the development of animal models and reagents that are useful for the study of host responses; however, they do not recapitulate the genetic variation present within the outbred human population, which is critical to disease responses. Recently, newly developed mouse resources were explicitly designed for systems genetics analysis as well as better capturing the genetic variation seen within human populations. Specifically the Collaborative Cross (CC) [42] recombinant inbred panel and Diversity Outbred (DO) [43] population are novel mouse resources which combine the utility of experimental mouse models with the genetic variability critical to contrasting experimental models with human responses. The CC and DO are complimentary resources (Figure 3) with levels of natural genetic variation roughly consistent with common variants segregating across the human population (~107 single nucleotide polymorphisms and ~106 small insertion/deletions), and characterized by relatively uniform distributions of variation across the genome. The large number of CC lines, and the continual generation of novel genomes of DO mice give rise to an incredibly large number of combinations of genetic variants across those genomes. These attributes are critical for first, mapping of genetic variants associated with infectious outcomes, second, creating novel genetic background with which to study transcriptional and regulatory networks, third, describing new models of virus diseases and pathologies, and fourth, accurate modeling of the human population’s genetic composition while maintaining experimentally tractable systems [44]. Importantly for systems genetics approaches, the CC and the DO not only facilitate initial discovery, but by allowing for the generation of new crosses and animals with similar allele frequencies but in new combinations, they also allow for the validation of the role of specific polymorphic genes and further mechanistic study (Figure 3).

Systems genetics approaches have been used extensively in studying the responses to influenza [44–46,47**]. Overall, these studies have found that multiple host polymorphisms contribute to differential disease outcomes following influenza infection, that some of these polymorphisms act in virus strain-specific manners, and that different subsets of transcripts associate with specific disease responses following these infections. Furthermore, by integrating these systems genetics approaches throughout multiple timepoints, Nedelko et al. [47**] were able to show that polymorphisms worked at specific points throughout the infection process, pointing to further complexity in the role of genetic regulation underlying differential disease outcomes. Together, these studies highlight the incredible power and precision that systems genetics approaches can provide, especially when
blended with systems biology and computational modeling.

Systems approaches have classically used traditional transcriptome profiling, such as microarray and mRNA seq. However, there is increasing evidence that non-coding RNAs play roles in regulating immune responses [48,49], and can have direct impact on viral infection [50]. Relevant to Coronavirus pathogenesis, two studies of contrasting IAV and SARS-CoV induced long [51] and small [52] non-coding RNAs were recently conducted within a subset of the founder animals of the CC, focusing on founder lines from the three genetically distant subspecies of *Mus musculus*, which have distinct responses to both SARS-CoV and IAV infection. Both of these studies found that there were pervasive changes in the expression levels of these noncoding transcripts during infections. Importantly for systems genetics approaches, they showed that these two pathogens led to differential regulation of these noncoding RNAs and that the levels...
of differential expression for these noncoding RNAs vary depending on host genetic background. This work highlights that unique interactions between specific viral infections and host genetic variation drive differential disease outcomes, and through the use of systems genetics approaches, host responses and the critical pathways causing various pathologic outcomes can be defined.

With a growing appreciation for the overall roles of noncoding RNAs in regulating immune responses and pathogenesis [53], as well as evidence that polymorphisms within noncoding RNAs can directly impact pathologic outcomes during infection, such as clearance of Hepatitis B infection [54], the investigation and detection of noncoding RNAs in future systems genetics...
approaches will provide a rich investigative environment for investigating how host genetic variation shapes immune responses and pathologic outcomes.

**Future prospects**
As illustrated throughout this study, the integration of systems approaches in traditional studies on viral pathogenesis provides immensely powerful tools with which to identify the host factors critical for pathologic or protective outcomes following viral infections in experimental systems. A key challenge for the field is to transition targets generated by systems approaches into therapeutics and prophylactics. Recently this has been seen for both MERS-CoV [55**], and H7N9 avian influenza [56], through cell culture models. In both cases, application of systems approaches and contrasting infections (MERS-CoV and SARS-CoV; H7N9 and H3N2 influenza) were used to identify pathways differentially regulated between related pathogens, and then this information was applied to select and test potential antiviral compounds which were able to inhibit both the target and related virus in the case of Coronaviruses [55**], or just the specific H7N9 target virus but not the related H3N2 virus [56]. Future approaches in these veins, and transitioning such results to *in vivo* systems genetic platforms such as the CC will further improve our capacity to combat conventional and new viral diseases of the future.

A longstanding divide in the scientific community has been bridging the gap between experimental systems and human populations. Indeed, some commonalities exist between murine and human immune responses [57,58], such as the role of IFITM3 in both human and mouse responses to influenza [58]. However, there are other studies highlighting discordance between humans and mice [59]. While systems approaches identify key genes, both their focus on pathways and systemic responses, and the explicit integration of genetic variation will allow for more robust descriptions of how pathogens cause variant disease responses within and across species. These results will increase the likelihood that, while individual genes might not be key regulators of disease across species, there will be commonly identified pathways regulating disease that can be identified in experimental models and transitioned into human systems. In support of this hope, Mitchell [21*] was able to show common transcriptional signatures between human cells and mice following highly pathogenic flu and SARS infections. Similarly, Sims [23] found conserved signals between immortalized Calu3 cells and primary airway epithelial cultures. Furthermore, systems based approaches studying influenza vaccine responses within humans were able to identify the CaMKIV kinase pathway as critical for these responses, and this molecule was validated in murine knockout systems [57]. The further advancement and refinement of such approaches in experimental systems, combined with state-of-the-art experimental approaches such as gene editing [60], as well as molecular profiling and disease data gathered from human cohorts [61], hold keys for transitioning bench-top findings to clinical results. Given the expanding nature of viral emergences, due to increased connectivity and ease of travel, the continuing refinement and further development of systems approaches combined with the advanced methodological approaches being developed should provide novel avenues with which to quickly address the added complexity of host genetic variation in combating emerging pathogens.

**Acknowledgements**
We acknowledge SP and RG for assistance with figures. The authors were supported by the Division of Intramural Research, National Institute of Allergy and Infectious Diseases under award number U19AI100625. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

**References and recommended reading**
Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest


By isolating novel Coronaviruses from wild bats, and showing that these viruses (a) use the same ACE2 receptor as SARS-CoV and (b) can specifically utilize the human ACE2, this paper highlights the need for vigilance and the development of methodologies to quickly respond to novel disease outbreaks.


21. By utilizing de novo network assembly approaches, and a series of escalating doses of SARS-CoV, the authors were able to identify and validate the role of serpine1 and the urokinase pathway as protective in SARS-CoV infection. This study highlighted the ability of systems biology approaches that can be used not only in vitro, but also in dissecting in vivo Coronavirus pathogenesis responses.


23. By explicitly integrating multiple pathogens and multiple pathogen strains in this analysis, the authors were able to identify sets of transcripts and key regulators that acted in virus-specific and pan-virus ways. Furthermore, they were able to show that these approaches could be used to predict responses derived from in vitro systems into ex vivo primary human airway cultures.


31. Wang SF, Chen KH, Chen M, Li WY, Chen YJ, Tsao CH, Yen MY, Huang JC, Chen YM: Human-leukocyte antigen class I Cw 1502 and class II DR 0301 genotypes are associated with resistance to severe acute respiratory syndrome (SARS) infection. Viral Immunol 2011, 24:421-426.


Utilizing the genetically diverse, recombinant inbred BxD panel of mice, the authors were able to show that host responses to influenza A virus were under the control of multiple polymorphisms. Importantly for both systems genetics approaches, and the study of host polymorphisms in human populations, many of these polymorphisms acted at specific times post-infection.


The authors utilize a systems biology approach and contrasting cell culture infections of SARS-CoV and MERS-CoV to identify critical networks controlling these infections. Most importantly, they were able to identify kinase inhibitors which attenuated growth of both viruses, highlighting the ability to transition systems approaches to therapeutics.


