Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Chapter 13

Host Genetics

It Is Not Just the Virus, Stupid

Martin T. Ferris[1], Mark T. Heise[1,2], Ralph S. Baric[2]

[1]Department of Genetics, University of North Carolina, Chapel Hill, NC, USA; [2]Department of Microbiology and Immunology, University of North Carolina, Chapel Hill, NC, USA

Chapter Outline

1. INTRODUCTION

Much of the focus of viral pathogenesis is on viral virulence factors; however, virulence can only be defined in the context of a virus–host combination. Furthermore, there is abundant evidence that highly pathogenic viruses have been selected for specific host genetic variants within the target population. The identification and mechanistic dissection of host genetic polymorphisms therefore has powerful implications for our understanding of viral diseases, including the identification of novel therapeutic targets. In this chapter, we review the classical approaches used to identify host allelic variations that regulate disease susceptibility. We also discuss next-generation molecular technologies and computational approaches that are poised to revolutionize our understanding of the impact of host genetic variation on viral diseases at the population level.

Viruses require efficient pathways for transcription, translation, assembly, and release, while simultaneously avoiding detection and clearance by the host innate immune system. As discussed in Chapter 12 (The Virus–Host Interactome), host proteins that promote the survival or replication of a virus are referred to as host factors. With the advent of high-throughput screening technologies, it has become feasible to identify in a single experiment large numbers of host factors that are required to promote viral replication. In addition, high-throughput screens are used to identify constitutive or inducible cellular factors that exhibit antiviral activity (referred to as restriction factors). These high-throughput screens often use genetically deficient yeast or siRNA knockdown technologies. Although in vitro screens allow for the rapid identification of candidate host or restriction factors, an understanding of the role of these factors in viral pathogenesis requires in vivo studies.

The use of mice carrying genetically defined deletions of specific genes (gene knockouts) has identified a number of host genes that play a critical role in limiting viral disease. Many of these genes are related to immune responses, such as those involved within the complement pathway or in the development of adaptive immune responses. However, many other host genes influence intrinsic biological activities such as transcription, translation, and intracellular transport. Removal of many of these genes has divergent effects that are dependent upon the virus. A given host gene may play a protective, pathologic, or neutral role during infection, and this will vary between viruses. Individual viruses co-opt divergent host mechanisms, and the specific host response can have radically different results depending upon that interaction. Such approaches have revealed the complex interactions between viral pathogens and the host immune system and highlight the need to study viral infections through integrated systems.

The use of classical genetic approaches has also provided significant insight into the role that specific host genes or pathways play in modulating disease severity. Much of this progress has come from studies comparing mouse strains with differing susceptibility to specific viral pathogens and the subsequent mapping of the

Viral Pathogenesis. http://dx.doi.org/10.1016/B978-0-12-800964-2.00013-6

polymorphic genes responsible for strain variation. For example, studies in the 1920s identified mouse strains that differed in their susceptibility to infection by a number of flaviviruses, including West Nile virus and Japanese encephalitis virus, revealing that differential disease responses are driven by a single autosomal-dominant genetic locus. Follow-up studies using positional cloning showed that animals carrying a functional allele of the *OAS1b* gene are highly resistant to flavirvirus replication (Brinton and Perelygin, 2003). Similar studies identified mouse genes associated with resistance to influenza virus (*Mx1*) (Staeheli et al., 1988) and murine cytomegalovirus (*Ly49H*) (Lee et al., 2001).

Such Mendelian traits are also prevalent in humans and reveal an interesting complexity in disease responses. For example, although persons who are homozygous for defective alleles of the chemokine receptor *CCR5* (*CCR5Δ32*) are highly resistant to HIV infection (Huang et al., 1996), these individuals may be at increased risk for severe neurologic infection with West Nile virus (Glass et al., 2006). Similarly, the ABH fucosyltransferase 2 (*FUT2*) allele contributes to resistance against norovirus infection, whereas other alleles of FUT2 encode a necessary receptor for these viruses (Lindesmith et al., 2003). In fact, virus evolution to altered forms of *FUT2*-regulated blood group carbohydrate receptor ligands (e.g., A, B, H) results in new pandemic outbreak strains that target distinct human subpopulations depending on the type of *FUT2* allele (Lindesmith et al., 2008). These findings highlight the importance of studying polymorphisms that contribute to specific viral responses.

2. COMPLEX AND POLYMORPHIC GENETIC INTERACTIONS

Although there are cases where viral disease outcomes are governed largely as a Mendelian trait, the majority of viral disease traits are much more complex. That is, multiple genetic, environmental, and demographic effects interact to determine an individual's disease response. Furthermore, many of the genetic variants contributing to trait differences are due to polymorphisms that have more moderate effects rather than the extreme phenotypes typically seen with gene-knockout approaches (Figure 1). Within a polymorphic population, there can be multiple genetic variants affecting the expression and function of a single gene. In this way, three different alleles of the translation factor *Eif4E* all confer bean plants with some level of resistance to clover yellow vein virus. Importantly, each of these alleles is effective against a different subset of virus strains, showing the complexity of disease responses that can be driven by multiple segregating natural polymorphisms (Hart and Griffiths, 2013).

A statistically complex framework, developed from elegant agricultural experiments over 80 years ago, posited that even if traits are not under strict Mendelian control, those under some form of genetic control must be heritable. That is, individuals who have some resistance to a viral pathogen tend to have offspring that are also somewhat resistant to the same viruses. Assessment of such heritability was initially used to measure the overall impact of host genetic variants on phenotypic traits, then for the selection of favorable traits (more below), and eventually as the basis of genetic mapping approaches.

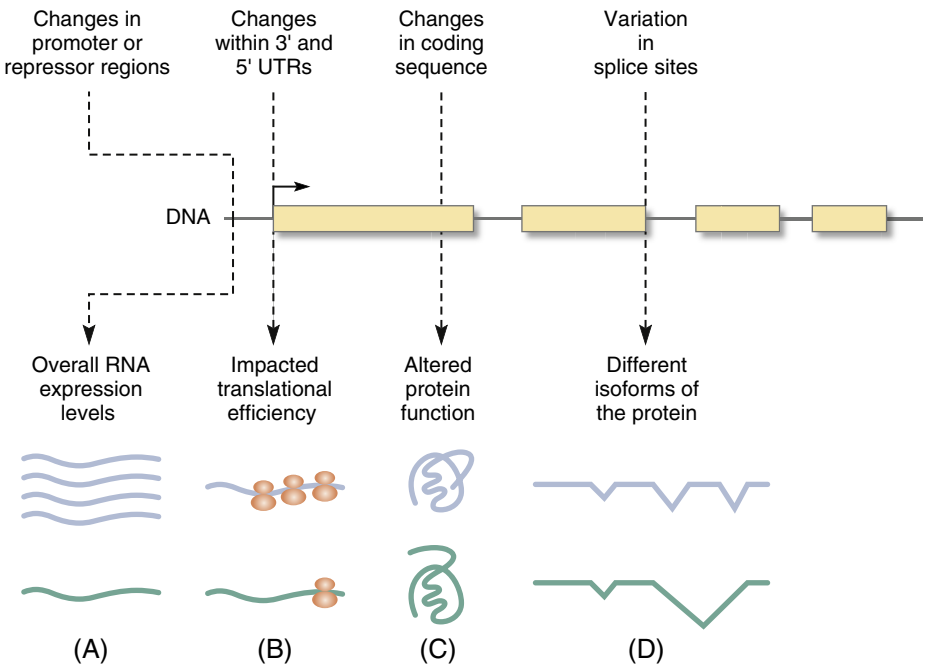

FIGURE 1 Effects of natural genetic variation. In contrast to gene knockouts, natural polymorphisms can impact genes in a variety of ways. These include (A) changes in promoter or repressor regions that impact overall RNA expression levels; (B) changes within 3′ and 5′ UTRs that impact translational efficiency; (C) changes in coding sequence that alter protein function, and (D) variation in splice sites that create different isoforms of the protein. Importantly, multiple changes can segregate within the same gene in a population, driving multiple expression and functional differences in disease responses.

These types of analyses were enthusiastically applied within the agricultural community and are still heavily used today. Controlled agricultural breeding allows for the explicit choice of breeding pairs, such that the next generation of animals will be enriched for a desired phenotypic trait. By iterating this process over generations, eventually the breeders fix large numbers of genetic variants within this population, and all animals have the advantageous traits of interest. Indeed, this is how milk and meat production, as well as crop yields, are regularly maximized. Also, it was quickly realized that these methods could be applied to disease resistance, such as the breeding of resistance to potato viruses. Furthermore, with the development of a variety of molecular markers, the classical need to use phenotypic markers to aid in the selection of resistant breeders has been reduced, allowing for more robust and rapid approaches. However, such approaches had only limited feasibility within human populations (and were used solely in the assessment of heritability) until mapping approaches were developed.

3. BIOMEDICAL ADVANCES IN VIRAL DISEASES

A key insight from genetic studies is that virus–host interactions and outcomes vary depending upon the virus. For example, the complement pathway is a complex pathway involved in the control of pathogens through several recognition and response arms. Depending on the recognition or response arms activated, differential disease outcomes can result (Stoermer and Morrison, 2011). As an example, the mannose binding lectin (MBL) complex, which binds sugar moieties on pathogens to activate the complement cascade, exhibits a direct antiviral activity against flaviviruses in part by blocking viral fusion to cells. In contrast, MBL not only fails to protect against Ross River virus (RRV) infection, but MBL deposition within RRV-infected joint and muscle tissue promotes complement activation and subsequent inflammatory tissue destruction (Gunn et al., 2012). Since the MBL genes are highly polymorphic in humans, and MBL levels correlate with RRV-induced disease severity, genetic variation in human MBL genes may be associated with susceptibility to RRV-induced disease. These types of studies also have important implications for the development of antiviral therapies targeting the complement cascade. Although therapies designed to block complement activation might be beneficial in the case of RRV, this type of intervention might have adverse consequences for individuals infected with West Nile virus or dengue virus.

The development of technologies to create gene-specific knockouts, especially in mice, dramatically increased the ability to determine how specific host genes affect viral pathogenesis. Studies using knockout mice have contributed to the identification and characterization of a wide array of host genes that play either protective or pathologic roles during viral infection (Table 1). As noted above, components of the host complement cascade play a protective role during West Nile virus infection, but exacerbate RRV-induced arthritis. Similarly, a deficiency in the chemokine receptor *CCR2* reduces early stage lung pathology during influenza virus infection, but exacerbates disease following Chikungunya virus infection by altering the inflammatory infiltrate and resultant arthritic disease. Importantly, in addition to providing insights into the role that a specific molecule plays in the pathogenesis of specific viral diseases, these types of studies have revealed important principles about the role of host genes in viral pathogenesis (Sidebar 1).

Although gene-specific knockouts are a powerful tool for investigating the role of specific genes or host pathways in viral pathogenesis, targeted knockout mice are not generally used as a screening platform for identifying novel genes that regulate the response to viral infection. Many of the advances in innate and intracellular responses to viruses have been derived either from high-throughput in vitro screens, or manipulation of the more tractable *Caenorhabditis elegans* or *Drosophila* systems. For example, such studies identified the highly evolutionarily conserved Toll-like receptor (TLR) pathway. However, it is possible to conduct unbiased screens using mice, such as by

TABLE 1 Gene Removal and Discordant Pathologic Responses

Gene	Protective Role	Pathologic Role
MBL (*MBL-1/MBL-2*)	West nile virus (Fuchs et al., 2010)	Ross river virus (Gunn et al., 2012)
C3	West nile virus (Mehlhop et al., 2005), influenza virus (Kopf et al., 2002)	Ross river virus (Morrison et al., 2008)
CCR5	West nile virus (Glass et al., 2006), influenza virus (Dawson et al., 2000)	Vaccinia virus (Rahbar et al., 2009)
Ifn-γ	Mouse hepatitis virus (Pewe and Perlman, 2002)	Herpesvirus (Schijns et al., 1994)
CCR2	Chikungunya virus (Poo et al., 2014)	Influenza A virus (Dawson et al., 2000)

Sidebar 1 Key findings from genetic manipulation studies

1. A large number of immune-related genes influence viral pathogenesis either by providing protection or enhancing pathology (Table 1). In addition, many host genes with no obvious immune-related functions can also impact viral disease. These can include viral receptors, such as the Sindbis virus receptor, NRAMP, but also host genes with effects at other stages of viral infection, such as transcription and translation, as with the translation factor Eif4E and its role in clover yellow vein virus pathogenesis (Hart and Griffiths, 2013).
2. Genes within host response pathways are often redundant. Many host immune pathways, such as the complement pathway (Stoermer and Morrison, 2011) or the TLR pathway (Hoebe, 2009), have multiple-sensing molecules, as well as a conserved set of downstream signaling molecules and response pathways. In this way, a specific virus might interact with multiple sensors within the same pathway. Thus, dissection of such pathways can often require complex experiments and the generation of double (or even triple) knockouts in the same animals
3. Many genes play different roles when interacting with different pathogens. Just as there can be redundant-sensing pathways for some pathogens, the specific molecules that a virus interacts with can modulate disease progression. In this way, abrogation of a specific pathway member can either enhance or ameliorate disease (Table 1). In more extreme examples, such as the role of *Hc* in influenza virus infection, a molecule can be protective for one virus, but have no role in the pathogenesis of a closely related virus. Such results illustrate the difficulty in understanding the general role of immune pathways when describing viral immunity and pathogenesis.

introducing mutations into the mouse genome through the use of *N*-ethyl-*N*-nitrosourea (ENU). ENU is an alkylating agent that induces point mutations in sperm progenitor cells at a predicable frequency. When male mice are treated with ENU and then bred to wild-type female mice (G_1 mice), their offspring carry a subset of the mutations derived from ENU mutagenesis. Through subsequent rounds of breeding, ENU-induced mutations affecting immune phenotypes that drive differential viral disease responses can be identified (Figure 2).

Historically, the identification of causal ENU-induced mutations has required a time- and labor-intensive mapping effort. However, with the advent of high-resolution whole-exome or genome sequencing, the efficiency of identifying causal ENU mutations has been significantly enhanced, thereby increasing the ability of this approach to identify genes impacting specific phenotypes. The power of this system is illustrated by the identification of a large number of innate immune genes that modulate the host response to herpesvirus infection, as well as the identification of several nonobvious host factors that regulate homeostasis and protect against viral disease. For example, ENU mutagenesis initially identified *UNC93B1* as a mediator of herpesvirus resistance in mice that is essential for signaling via multiple TLRs. Subsequent studies also demonstrated that UNC-93B deficiency in humans is associated with enhanced susceptibility to herpesvirus infection. Multiple ENU-induced mutations within the same gene can have strikingly different effects on phenotype, such as variant mutations affecting TLR signaling (Hoebe, 2009).

These types of studies, often termed reductionist, have provided new avenues for combating viral diseases, and have uncovered a variety of complex mechanisms and pathways involved in host immune responses and cellular function. Since the majority of these studies have been conducted within only one or a few genetic backgrounds, more complex genetic systems are needed to fully disentangle host responses to viral pathogens.

4. GENETIC MAPPING TO IDENTIFY VARIANTS AFFECTING COMPLEX TRAITS

Regardless of the number and type of markers used, genetic mapping follows the same principles (Figure 3). Individuals are typed at a series of markers (whose positions relative to each other are known), and these individuals are also phenotyped for traits of interest. Whether these are closely related populations (e.g., a large family pedigree of humans or an F2 cross between two inbred strains of animals) or larger population-level association studies, all such studies rely on recombination within gametogenesis to break apart genome structure. As a result, only markers physically close to the polymorphic loci of interest will remain associated or linked throughout the study population. Therefore, at each marker, the significance of the association between the trait values and the two alleles at that marker is determined. When polymorphisms controlling a trait are unlinked to a marker (e.g., they are on different chromosomes), then there will be completely random association between the two variants at the marker and the trait of interest. However, when tested markers exist on the same chromosome as a polymorphism controlling a trait, there will be an increased association of one of the two marker variants and an increase in the trait of interest. Furthermore, for the markers near the trait-controlling polymorphism, there will be a tighter and tighter association between the marker variant and an increase in the trait of interest. Indeed, with sufficient numbers of individuals, it is possible to identify polymorphisms driving subtle effects.

Genetic markers have undergone development from restriction fragment length polymorphisms or micro/minisatellite markers, to well-annotated single nucleotide

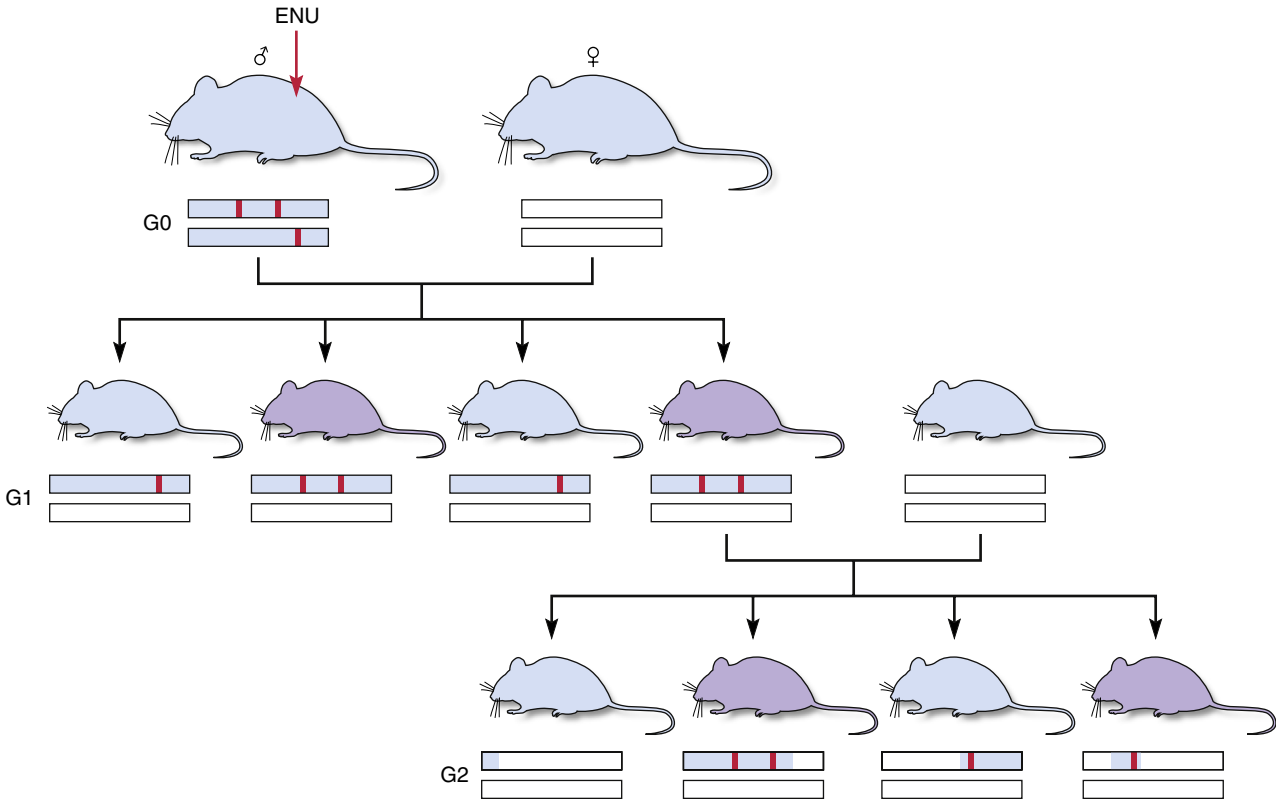

FIGURE 2 *N*-ethyl-*N*-nitrosourea (ENU) mutagenesis and differential disease responses. Following ENU mutagenesis of the testes of a male mouse, a large number of mutations are accumulated within the sperm of the animal (red bars on black chromosomes). This male is then bred to a female of a closely related strain (denoted by white chromosomes). G1 offspring of this mating receive a variety of mutations from the ENU-mutagenized sperm, some of which can confer novel phenotypes, such as susceptibility to a viral pathogen. Here, purple animals have acquired a common mutation showing enhanced pathology following viral infection, as compared to wild-type animals. Depending on the mode of action of the mutation (e.g., dominant in this case), these phenotypic variants can be seen in the G1 generation, or in later generations through backcrossing. Furthermore, by conducting further crosses, such as to a wild-type female mouse, a new mapping population (G2) can be used to identify the underlying causal mutations driving these phenotypes.

polymorphisms (SNPs) or even haplotype reconstructions (a set of tightly clustered heritable polymorphisms). Using current methods, it is possible to more finely identify both the chromosomal locations and the relative linkages between genetic factors. Careful analysis is needed to differentiate multiple genetic variants affecting many traits in close proximity to each other from pleiotropy (a single variant affecting multiple traits). However, analyses of host genetic contributions at specific loci suggest that there are often multiple variants close together that commonly affect a given trait, or set of related traits.

For example, many early mapping studies showed that the major histocompatibility complex (MHC) locus affects a wide variety of immune responses including complement function, B-cell and T-cell responses, and allergic responses. This genome region contains a large number of genes involved in antigen presentation, the complement cascade, innate immune responses, and development all in close physical proximity. Furthermore, as the sequence and annotation of mammalian genomes has progressed, other loci with tightly linked genes of related function have been identified. Such tightly linked genes hint at co-evolutionary pressures, but more practically they create difficulty in identifying specific polymorphic genes influencing phenotypes.

The development of genome-wide association studies (GWAS) is based on the assumption that historical recombination events within a population have broken apart all but the tightest associations. (This is in contrast to linkage analysis, which seeks to maintain long-term linkages between markers and phenotypes.) Therefore, only associations with causative (or intimately proximal polymorphisms) will be identified. The GWAS approach is something of a double-edged sword. In order to identify these very tight associations, many markers are necessary. This necessitates incredibly large cohorts of individuals with which to attain statistical significance. However, when associations do exist within these cohorts, those associations can detect incredibly small effects (those influencing disease phenotypes by <1% of the overall trait variation). Indeed, the GWAS literature shows that large numbers of identified SNPs have small-to-moderate effects on disease phenotypes and traits.

GWAS requirements, such as large numbers of individuals and the often subtle nature of SNPs on phenotypic traits, have limited the ability to investigate the role of human

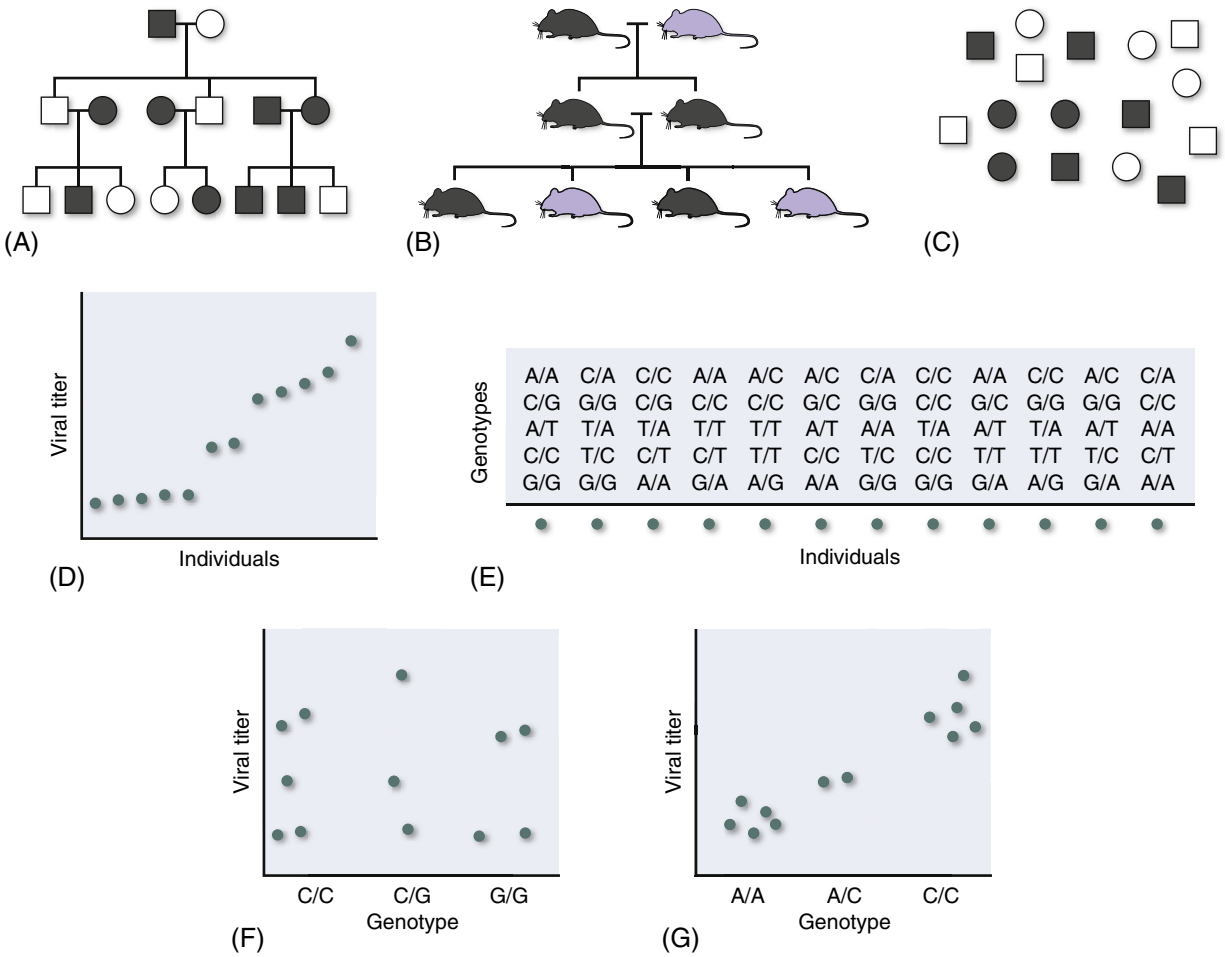

FIGURE 3 Genetic mapping involves the phenotyping and genotyping of a related population. (A) Family pedigrees or (B) crosses of experimental animals are used for linkage analysis. (C) Large cohorts of unrelated individuals are used for genome-wide association studies. Individuals are then (D) phenotyped and (E) genotyped at a number of genomic markers. Genetic mapping assesses the statistical strength of association between phenotypes and genotypes to distinguish (F) nonsignificant relationships from (G) those markers linked to genetic variants causing the phenotypic differences.

polymorphisms within the context of infection. Confounding factors—such as variation in viral doses and exposures, the effect of viral genetic variants on disease, the often narrow windows of symptomatic infection, and the numbers of individuals infected during an outbreak—can cloud subtle host genetic contributions to differential disease outcomes. Despite such limitations, for chronic viral infections, such as those caused by HIV or hepatitis C virus, GWAS studies have proven useful for identifying host genes that are associated with variation in viral control, disease progression, or treatment responses.

Perhaps most illustrative of the power of these approaches for chronic infections are studies of viral load and disease progression following HIV infection (see also Chapter 9, HIV and AIDS). GWAS studies of the response to HIV have identified a number of genes within the MHC that are associated with control of viral load. These include the HLA complex 5 gene (*HCP5*) and *HLA-C*, where a SNP in the 3′ UTR regulates *HLA-C* expression through microRNA interactions. Several GWAS studies have identified genes associated with long-term nonprogression or rapid progression to HIV disease, including *HCP5*, the *CXCR6* chemokine receptors, *PROX1* (a regulator of T cell IFN-γ production), and the SMAD family interacting protein PARD3. These studies provide new insights into the potential role of specific polymorphic genes in HIV pathogenesis, and they illustrate that HIV disease involves the complex interaction of multiple polymorphic genes that contribute to variation in disease progression (van Manen, van 't Wout, and Schuitemaker, 2012). In the case of hepatitis C virus, GWAS studies have identified a polymorphism upstream of *Il28B* that influences spontaneous virus clearance (Thomas et al., 2009). This SNP is also associated with improved responses to certain antiviral treatments.

Given the need to have extremely large cohorts of individuals at similar disease stages in order to study viral

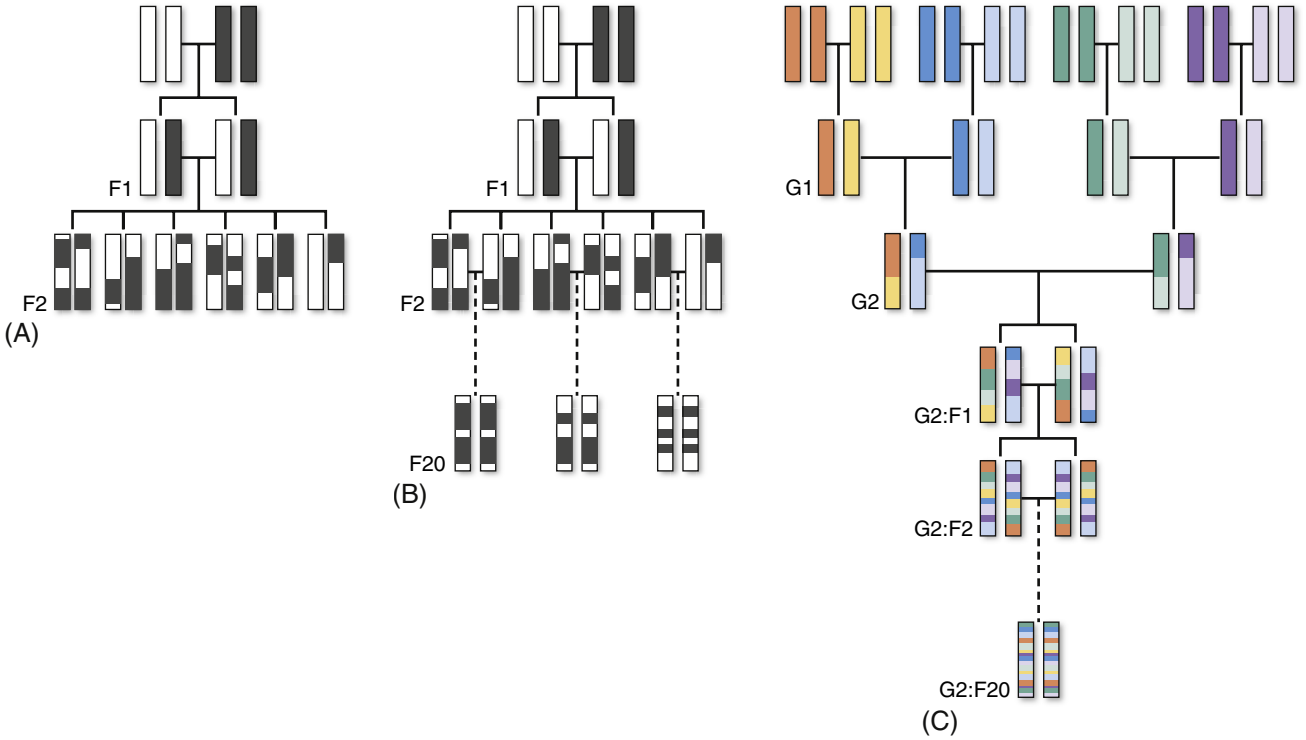

FIGURE 4 Recombinant Inbred models of genetic diversity. (A) Classical mapping populations, such as F2s, involve nonpermanent genotypes of individuals. By crossing F2 animals, followed by inbreeding (B) classical recombinant inbred panels can be generated. These panels are sets of lines all derived from common ancestors, but containing unique haplotype combinations of these founder lines, and allow for the assessment of genetic effects across time points, treatments, and conditions of viral infections. In order to better model the genetic diversity and complexity present within the human population, the Collaborative Cross panel (C), derived from eight classical inbred founder strains, was bred from a complex funnel design to incorporate genetic information from each founder line within each CC line. Following this funnel breeding, a large number of unique lines are generated and can be used for experimental analysis and genetic mapping of complex traits.

diseases on genome-wide scales, more targeted approaches have been used to carefully probe candidate genes, which are thought to be important in specific viral diseases. For example, initial candidate studies of the signaling lectin *L-SIGN/CLEC4M* identified a polymorphism associated with SARS-coronavirus resistance, and these findings appeared robust in an in vitro model (Chan et al., 2006). However, further analysis of patient cohorts failed to replicate this association, a cautionary note in interpreting genetic studies (Zhi et al., 2007).

5. RECOMBINANT INBRED PANELS

Animal models are critical for understanding the interaction between human pathogens and the host immune system (see Chapter 10, Animal Models). Mice offer the most tractable experimental system, in part because of well-developed molecular and genetic methods. Inbred mouse strains can be used to map genome locations contributing to differential disease responses between inbred strains. For certain large-effect genes, it is possible to identify causal genes contributing to disease differences in these populations. Genes such as *Mx1*, *Oas1b*, *CEACAM1a*, and *Cmv-1* (*Ly49H*) have been identified as major players in specific viral diseases. In addition, inbred mouse strains offer the ability to perform reproducible studies designed to elucidate the mechanisms by which a specific gene impacts disease. Researchers have designed recombinant inbred panels to extend studies using animals with different genetic variants.

Classical recombinant inbred (RI) panels (e.g., the C57BL/6J by DBA/2J (BxD) and A/J by C57BL/6J (AxB and BxA)) were derived from pairs of inbred lines. The generation of RI panels is initially identical to the generation of F2 animals for genetic mapping. However, instead of testing these F2 animals for viral disease, pairs of F2 animals are mated with a sibling, and the resultant animals undergo brother–sister mating until a set of new inbred lines are derived from this initial cross (Figure 4). Each new inbred line in this panel has an inbred genome composed of different segments from each of the two founder inbred lines. Furthermore, since each new inbred line is generated from two unique F2 animals, each one of the new RI lines has private recombination events. In this way, a panel of recombinant inbred lines provides a large number of inbred and reproducible mouse lines that are all related to each other. Each set of lines shares a subset of polymorphisms they

received in common from the two founder lines, while at other loci they have different polymorphisms. Therefore, a whole RI panel can be used to conduct genetic mapping, or individual lines within a panel can be used to study specific aspects of viral disease.

A major advantage of RI panels is the ability to compare or contrast treatments across time points or viral infections. For example, use of the BxD panel to compare host responses to highly pathogenic H5N1 influenza virus and a less-pathogenic H1N1 influenza virus identified completely different polymorphic host loci controlling exacerbated disease. A major host genetic locus driving H5N1 susceptibility is a defective *Hc* (hemolytic complement) allele present within the DBA/2J strain. However, this defective *Hc* allele has no effect on disease responses following H1N1 infection.

Such comparative studies again highlight the complexity of infectious disease responses, relying on host genetic differences as well as virus–host genetic interactions. They also show the importance of understanding both the basal (uninfected) status of the immune system as well as how different viruses interact with specific variant host pathways to influence disease outcome. This was shown in a study of the whole-genome transcriptional responses of C57BL/6J and DBA/2J lines before and after H5N1 infection (Boon et al., 2009). Whereas many transcripts are commonly regulated in response to infection between these two lines, over one-third of transcripts have baseline differential expression between the two strains. Furthermore, approximately half of the genes are underneath quantitative trait loci (QTL) driving differential H5N1 responses. Therefore, high-priority candidates are primarily differentiated by basal expression differences versus differences in induction between strains. Such analyses significantly improve the identification of "core" immune or cellular responses that respond to viruses across genetic backgrounds. For instance, there are over 1800 transcripts that are commonly regulated following H5N1 infection of C57BL/6J and DBA/2J mice. The data also identify those transcriptional responses that are specific to individual RI lines (Figure 5).

RI panels and GWAS studies have demonstrated that genetic complexity, driven by multiple genetic variants, is a hallmark of many disease response phenotypes. In addition, the complex gene–gene interactions that exist between polymorphic loci can lead to radically different disease responses. GWAS studies have led to the recognition that complex gene–gene interactions (e.g., epistasis and dominance) contribute much of the genetic control of complex diseases. New methodologies, which require great increases in sample sizes and statistical power, are therefore needed to identify these interactions.

As a result, there is an increased interest in developing next-generation genetic reference populations that capture many of these attributes. Although such panels have been used in work on *Arabidopsis* and *Drosophila*, they have only recently been attempted for a mammalian model to study human biomedical traits. Specifically, the first such mammalian genetic reference population that incorporates more genetic complexity than traditional RI panels is the Collaborative Cross (CC) recombinant inbred panel (Collaborative Cross Consortium, 2012). This panel was derived from eight inbred mouse strains (A/J, C57BL/6J, 129S1/SvImJ, NOD/ShILtJ, NZO/HILtJ, CAST/EiJ, PWK/PhJ, and WSB/EiJ), which together contain high levels of genetic variants (approximately 40 million SNPs and four million small insertions/deletions, similar to levels of common human genetic variants), uniformly distributed across the genome. By including the wild-derived, inbred strains CAST/EiJ, PWK/PhJ, and WSB/EiJ, the CC contains input from the three major *Mus musculus* subspecies: *castaneus*, *musculus*, and *domesticus*; and these three strains contribute the majority of the genetic polymorphisms within the cross. Furthermore, as there are eight equal haplotype contributions at each locus, the minor allele frequencies are at roughly 12.5% (as opposed to under 1% in human populations), improving the ability to detect the role of rare variants and epistatic interactions on disease outcomes.

By breeding animals from these eight inbred strains together in a funnel design, and then inbreeding the resultant offspring together, new inbred lines are created. Due to the recombination events that occur throughout the funnel breeding and inbreeding processes, the genome of each new RI line is a mosaic of the eight founder strains. By repeating this process over and over, while shuffling the positions of the eight founder strains at the top of the funnel (Figure 4), multiple independent CC lines have been created.

A related population, Diversity Outbred mice, was started from incipient CC animals. Instead of defined inbred lines, this population is maintained fully outbred. Although not a classical genetic reference population (in that specific genotypes cannot be maintained and studied across treatments), Diversity Outbred mice can be used to study specific allele frequencies at a population level across different treatments. Furthermore, Diversity Outbred mice provide a useful experimental model to mimic outbred populations, such as humans.

High-resolution genotyping of CC and Diversity Outbred animals make it possible to accurately describe the genomic composition of the each population. These genetic details have two main advantages over other approaches. First, instead of genetic mapping based on individual SNPs, mapping can be conducted on founder haplotypes. In this way, false-positive associations between genome regions and phenotypes are reduced, true associations increase in significance, and it is possible to rapidly narrow QTL regions down to a smaller number of likely candidate genes

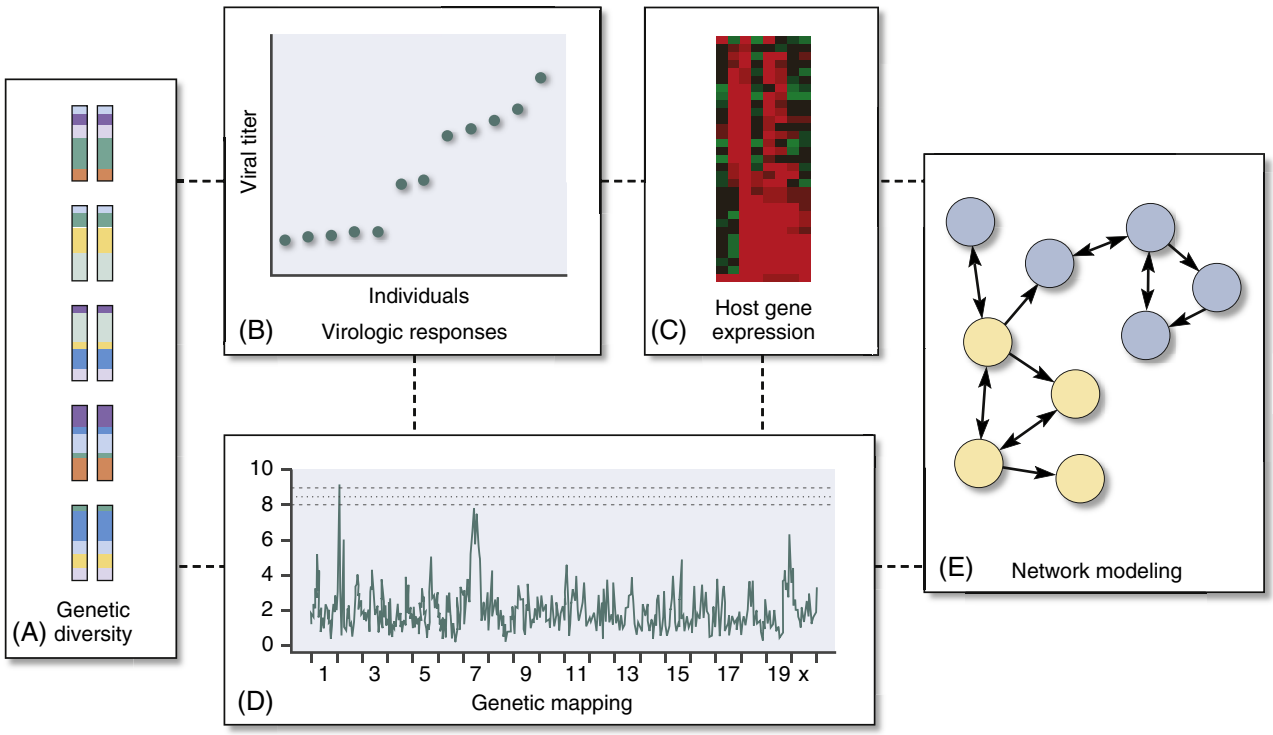

FIGURE 5 Systems genetics incorporates genetic complexity to explain differential disease responses. An advantage of genetically tractable, yet complex experimental systems such as RI panels, is the ability to explicitly integrate host genetic (A), virologic (B), and transcriptional responses (C) in order to identify polymorphic genetic loci (D) that contribute to differential virologic responses, and to develop transcriptional networks (E) that shed mechanistic insight into these polymorphic responses.

and features. In a study of H1N1 influenza virus infection of a progenitor population of CC animals, these approaches identified sets of high-priority candidate genes driving differential host responses, and showed that multiple antiviral *Mx1* alleles segregating within this population can differentiate antiviral and clinically protective disease responses (Ferris et al., 2013). Second, this genomic sequence and composition information is critical for correctly assessing microarray and RNA-seq data, avoiding misaligned sequences from RNA-seq, and correctly assessing microarray probes that are impacted by SNPs.

These genetic reference populations are powerful tools for identifying polymorphisms contributing to differential viral disease outcomes. Indeed, several studies have used these two populations to identify individual polymorphisms or polymorphic pathways that contribute to different immunological or disease responses (Phillippi et al., 2014). An additional use of such genetic reference populations is the detection of epistatic and other complex genetic interactions. The tractability of genetic reference populations can allow for direct assessment of these genetic interactions.

The shuffling of susceptibility alleles across the CC population has recently been used to identify a novel mouse model of Ebola virus hemorrhagic fever (Rasmussen et al., 2014). None of the founder lines of the CC, or other classical mouse strains, have provided an effective model of severe Ebola virus disease. However, several F1 crosses between CC lines develop a severe Ebola virus disease, including hepatic involvement, coagulation defects, and rapid mortality. These results further emphasize the utility of studying genetically variant models to understand disease processes. Furthermore, with the development of gene-editing approaches (discussed below), it is becoming possible to more fully study putative epistasis across a range of genetic backgrounds.

Elsewhere in this book, systems biology approaches to viral pathogenesis are described in detail (see Chapter 11, Systems Virology). These approaches integrate a variety of data types with whole-genome measures of transcription and translation to better understand the complex networks and pathways that are altered during infection, specifically targeting those that are causative for host-mediated protection or disease exacerbation. In addition to the genetic mapping approaches described above, genetic reference populations provide a novel and powerful overlay to systems biology approaches. Integrating variant disease pathologic, virologic, and transcriptional responses to virus infection, across a range of genetically distinct host backgrounds, allows for the direct assessment of the role of genetic variation on host response networks (Figure 5).

6. REPRISE

Throughout this chapter, we have described the approaches that have provided much of our understanding of host genetic determinants of viral disease. These studies have shown the need for sophisticated approaches to identify, manipulate, and characterize genetic variants contributing to complex responses. It is clear that many polymorphisms underlying complex diseases are pleiotropic (a single polymorphism impacts multiple aspects of the disease response), epistatic (some responses occur only when an individual has multiple interacting polymorphisms), and epigenetic (polymorphisms that either influence, or are activated or repressed by, genome modifications such as methylation).

Gene deletion approaches have allowed for the dissection of pleiotropic gene responses (e.g., many gene knockouts ameliorate weight loss and inflammation, but have no effect on viral titers; others have an effect on viral titers, but not on weight loss or inflammatory responses). However, it is only with the advent of the large next-generation genetic reference populations that epistatic interactions can reliably be detected. Such systems allow for the induction, characterization, and manipulation of epigenetic states, clarifying the role and mechanisms of epigenetic responses in complex viral diseases.

Molecular and experimental advances have provided additional approaches. One is the ability to use tissue-specific promoters to generate constitutive or tissue-specific gene knockouts. Inducible systems are often able to rescue embryonic lethal gene knockouts. Another is the ability to directly modify the genome via CRISPR/Cas9 or Talens gene-editing technologies, which enable the introduction of a single mutation, or combination of mutations, into a host genome (Niu et al., 2014; Yang et al., 2013). The influence of specific polymorphisms on disease responses can then be assessed across a variety of genetic backgrounds. By engineering mutations into specific backgrounds, or engineering all epistatic mutations into a naïve background, rare epistatic phenotypes can be more easily validated and their molecular underpinnings investigated.

Although there is concern that mouse models may not accurately recapitulate human disease, some notable cases have validated the power of using mouse models. For example, a polymorphism in human *Ifitm3* was identified through the sequencing of patients hospitalized for influenza infection, and a splice variant encoding a truncated Ifitm3 protein is enriched within these hospitalized individuals (Everitt et al., 2012). The anti-influenza effects of *Ifitm3* were then recapitulated in cell culture and with a knockout mouse model. That said, it remains unclear how often mouse knockouts will successfully translate to human polymorphisms. Furthermore, since epigenetic modifications can also impact the host response to environmental stimuli, it will be important to assess whether these types of effects can be accurately modeled in mice or other experimental systems.

Genetic methods have significantly contributed to our understanding of the influence of host genes on disease outcomes. It is clear that host-pathogen coevolution and selection has altered the population genetic structure of most species, and there is abundant evidence that host genetic variation has a major impact on viral pathogenesis. Much of the progress in this area has come through the use of gene-specific knockouts or forward genetic screens. Recent advances in systems genetics, including QTL analysis in humans, whole-genome sequencing, genetically complex mouse genetic reference populations, and genome-editing techniques, promise to significantly enhance our ability to study the polymorphic genes and pathways that influence disease outcome. These approaches are making a critical contribution to our understanding why different humans are either resistant or susceptible to specific viral infections.

REFERENCES

Boon AC, deBeauchamp J, Hollmann A, Luke J, Kotb M, Rowe S, et al. Host genetic variation affects resistance to infection with a highly pathogenic H5N1 influenza A virus in mice. J Virol October 2009;83(20):10417–26.

Brinton MA, Perelygin AA. Genetic resistance to flaviviruses. Adv Virus Res 2003;60:43–85.

Chan VS, Chan KY, Chen Y, Poon LL, Cheung AN, Zheng B, et al. Homozygous L-SIGN (CLEC4M) plays a protective role in SARS coronavirus infection. Nat Genet January 2006;38(1):38–46.

Collaborative Cross Consortium. The genome architecture of the collaborative cross mouse genetic reference population. Genetics February 2012;190(2):389–401.

Dawson TC, Beck MA, Kuziel WA, Henderson F, Maeda N. Contrasting effects of CCR5 and CCR2 deficiency in the pulmonary inflammatory response to influenza A virus. Am J Pathol June 2000;156(6):1951–9.

Everitt AR, Clare S, Pertel T, John SP, Wash RS, Smith SE, et al. IFITM3 restricts the morbidity and mortality associated with influenza. Nature March 25, 2012;484(7395):519–23.

Ferris MT, Aylor DL, Bottomly D, Whitmore AC, Aicher LD, Bell TA, et al. Modeling host genetic regulation of influenza pathogenesis in the collaborative cross. PLoS Pathog February 2013;9(2):e1003196.

Glass WG, McDermott DH, Lim JK, Lekhong S, Yu SF, Frank WA, et al. CCR5 deficiency increases risk of symptomatic West Nile virus infection. J Exp Med January 23, 2006;203(1):35–40.

Gunn BM, Morrison TE, Whitmore AC, Blevins LK, Hueston L, Fraser RJ, et al. Mannose binding lectin is required for alphavirus-induced arthritis/myositis. PLoS Pathog 2012;8(3):e1002586.

Hart JP, Griffiths PD. A series of eIF4E alleles at the bc-3 locus are associated with recessive resistance to clover yellow vein virus in common bean. Theor Appl Genet November 2013;126(11):2849–63.

Hoebe K. Genetic dissection of toll-like receptor signaling using ENU mutagenesis. Methods Mol Biol (Clifton, N.J.) 2009;517:239–51.

Huang Y, Paxton WA, Wolinsky SM, Neumann AU, Zhang L, He T, et al. The role of a mutant CCR5 allele in HIV-1 transmission and disease progression. Nat Med November 1996;2(11):1240–3.

Kopf M, Abel B, Gallimore A, Carroll M, Bachmann MF. Complement component C3 promotes T-cell priming and lung migration to control acute influenza virus infection. Nat Med April 2002;8(4):373–8.

Lee SH, Girard S, Macina D, Busa M, Zafer A, Belouchi A, et al. Susceptibility to mouse cytomegalovirus is associated with deletion of an activating natural killer cell receptor of the C-type lectin superfamily. Nat Genet May 2001;28(1):42–5.

Lindesmith L, Moe C, Marionneau S, Ruvoen N, Jiang X, Lindblad L, et al. Human susceptibility and resistance to norwalk virus infection. Nat Med May 2003;9(5):548–53.

Lindesmith LC, Donaldson EF, Lobue AD, Cannon JL, Zheng DP, Vinje J, et al. Mechanisms of GII.4 norovirus persistence in human populations. PLoS Med February 2008;5(2):e31.

van Manen D, van 't Wout AB, Schuitemaker H. Genome-wide association studies on HIV susceptibility, pathogenesis and pharmacogenomics. Retrovirology August 24, 2012;9(70):4690–9.

Mehlhop E, Whitby K, Oliphant T, Marri A, Engle M, Diamond MS. Complement activation is required for induction of a protective antibody response against West Nile virus infection. J Virol June 2005;79(12):7466–77.

Morrison TE, Simmons JD, Heise MT. Complement receptor 3 promotes severe ross river virus-induced disease. J Virol November 2008;82(22): 11263–72.

Niu Y, Shen B, Cui Y, et al. Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. Cell 2014;156:836–43.

Pewe L, Perlman S. Cutting edge: CD8 T cell-mediated demyelination is IFN-gamma dependent in mice infected with a neurotropic coronavirus. J Immunol (Baltimore, Md.: 1950) February 15, 2002;168(4):1547–51.

Phillippi J, Xie Y, Miller DR, Bell TA, Zhang Z, Lenarcic AB, et al. Using the emerging collaborative cross to probe the immune system. Genes Immun January 2014;15(1):38–46.

Poo YS, Nakaya H, Gardner J, Larcher T, Schroder WA, Le TT, et al. CCR2 deficiency promotes exacerbated chronic erosive neutrophil-dominated chikungunya virus arthritis. J Virol June 2014;88(12):6862–72.

Rahbar R, Murooka TT, Fish EN. Role for CCR5 in dissemination of vaccinia virus in vivo. J Virol March 2009;83(5):2226–36.

Rasmussen AL, Okumura A, Ferris MT, Green R, Feldmann F, Kelly SM, et al. Host genetic diversity enables ebola hemorrhagic fever pathogenesis and resistance. Science (New York, N.Y.) November 21, 2014;346(6212):987–91.

Schijns VE, Haagmans BL, Rijke EO, Huang S, Aguet M, Horzinek MC. IFN-gamma receptor-deficient mice generate antiviral Th1-characteristic cytokine profiles but altered antibody responses. J Immunol (Baltimore, Md.: 1950) September 1, 1994;153(5):2029–37.

Staeheli P, Grob R, Meier E, Sutcliffe JG, Haller O. Influenza virus-susceptible mice carry mx genes with a large deletion or a nonsense mutation. Mol Cell Biol October 1988;8(10):4518–23.

Stoermer KA, Morrison TE. Complement and viral pathogenesis. Virology March 15, 2011;411(2):362–73.

Thomas DL, Thio CL, Martin MP, Qi Y, Ge D, O'Huigin C, et al. Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. Nature October 8, 2009;461(7265):798–801.

Yang H, Wang H, Shivalila CS, et al. One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. Cell 2013;154:1370–9.

Zhi L, Zhou G, Zhang H, Zhai Y, Yang H, Zhang F, et al. Lack of support for an association between CLEC4M homozygosity and protection against SARS coronavirus infection. Nat Genet June 2007;39(6): 692,4; author reply 694–6.