Trends in Microbiology

CellPress

Review

These Are the Genes You're Looking For: Finding Host Resistance Genes

Jeffrey S. Bourgeois,[1,2] Clare M. Smith,[1,2,3] and Dennis C. Ko [1,2,4,*,@]

Humanity's ongoing struggle with new, re-emerging and endemic infectious diseases serves as a frequent reminder of the need to understand host–pathogen interactions. Recent advances in genomics have dramatically advanced our understanding of how genetics contributes to host resistance or susceptibility to bacterial infection. Here we discuss current trends in defining host–bacterial interactions at the genome-wide level, including screens that harness CRISPR/Cas9 genome editing, natural genetic variation, proteomics, and transcriptomics. We report on the merits, limitations, and findings of these innovative screens and discuss their complementary nature. Finally, we speculate on future innovation as we continue to progress through the postgenomic era and towards deeper mechanistic insight and clinical applications.

Highlights

Understanding host–microbe interactions remains critical as growing antibiotic resistance, new outbreaks, and re-emerging pathogens put lives at risk.

Microbiologists and infectious disease researchers have leveraged and advanced the genomic technological breakthroughs of the last decade.

Advances in clustered regularly interspaced short palindromic repeats (CRISPR/Cas9)-mediated genome editing and reduced sequencing costs have made CRISPR screens the dominant loss-of-function and gain-of-function screening platform.

Advances in sequencing, high-throughput technologies, and resources for model organisms and humans have dramatically improved natural diversity screens. Such resources include consortia with repositories for electronic medical records and human cell lines, and model-organism diversity panels.

Integrating pathogen diversity into host resistance screens helps to further define the genetic landscape of host–pathogen interactions.

An Expanding Toolbox for a Critical Discipline

The importance of microbiological research is constantly reaffirmed as drug resistance rejuvenates dangerous old foes, ancient plagues re-emerge, and novel pathogens arise and drive world-wide pandemics. Fascination with the microscopic world stretches back to the mid-1600s when Hooke and van Leeuwenhoek first used microscopy to observe microbes. Later advances in germ theory by Pasteur, Henle, Koch, and others, solidified the importance of studying microbes, while the advent of molecular biology reshaped how microbes are studied. Now, advances in genomics are fueling significant microbiological breakthroughs. Genetic tools, like the mouse Collaborative Cross, are being applied to microbiology even as microbiology generates new genetic tools, like clustered regularly interspaced short palindromic repeats and CRISPR Associated Protein 9 (CRISPR/Cas9). This interplay is redefining the microbial research toolbox.

In this review we focus on how cutting-edge genetics and genomics screens are being applied to identify genetic determinants of host resistance to bacterial infection. We discuss screens that harness CRISPR/Cas9, natural diversity, proteomics, and transcriptomics to understand host–bacterial interactions. Our goal is to demonstrate the power of these screening approaches and to encourage continued innovation in host–bacterial interactions research.

Harnessing CRISPR/Cas9 to Identify Host Resistance Genes

Genetic screens have fundamentally improved our understanding of biology, but, until recently, these studies were limited in mammalian cell lines. Previous genetic screens in these cells were restricted to siRNA knockdown systems – which can have off-target effects, variable efficacy, batch effects, and limited throughput – or haploid cell screens, which are limited to a small number of cell lines [1]. While these studies were impactful, the sum of their limitations make it technically challenging and financially infeasible to identify genes with small effect sizes that contribute to polygenic and clinically relevant traits.

[1]Department of Molecular Genetics and Microbiology, School of Medicine, Duke University, Durham, NC, USA
[2]University Program in Genetics and Genomics, Duke University, Durham, NC, USA
[3]Duke Human Vaccine Institute, School of Medicine, Duke University Durham, NC, USA
[4]Division of Infectious Diseases, Department of Medicine, School of Medicine, Duke University, Durham, NC, USA

*Correspondence: dennis.ko@duke.edu (D.C. Ko).
@Twitter: @denniskoHiHOST

By overcoming several of these weaknesses, CRISPR knockout screens have recently become the dominant method to identify host genes and pathways involved in the host response to bacterial infection. CRISPR screens come in two forms, traditional arrayed screens and pooled selection experiments, and the vast majority of host–pathogen interaction CRISPR screens have used the latter. Pooled CRISPR selection screens first deliver a pool of single guide RNAs (sgRNAs) into a population of cells expressing Cas9, then select, based on a cellular phenotype, and finally compare each guide's frequency in the selected and unselected samples using sequencing (Figure 1). By screening multiple guides within a pool and targeting multiple sgRNAs to a single gene, this pooled approach reduces noise from off-target effects, batch effects, or failed knockout compared with siRNA screens or arrayed CRISPR screens. Further, because CRISPR/Cas9 technology is rapidly being adapted to a number of primary and immortalized cell lines across a variety of species, this screening platform can be used to detect cell type- and cell line-dependent resistance or susceptibility factors. CRISPR screens can also be performed *in vivo*, as multiple studies have knocked out genes in murine immune progenitor cells, returned them to irradiated mice, and measured the impacts on immune cell development [2,3]. Together, these strengths make CRISPR/Cas9 screens invaluable for understanding host immunity.

While genetic knockout systems are powerful tools they do have limitations. For instance, genetic ablation can have unintended consequences on the genetic landscape that mask phenotypes

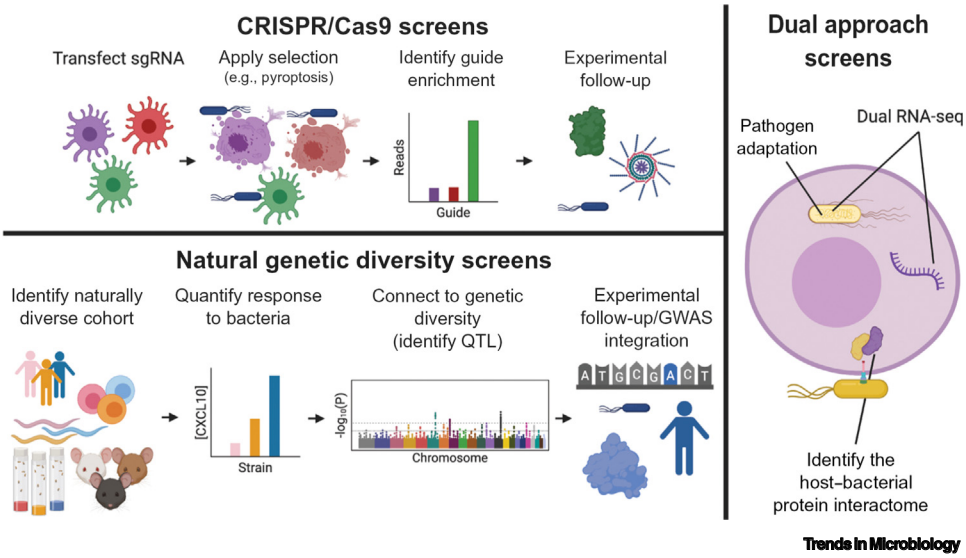

Figure 1. Interrogating the Host: Modern Screens for Identifying Host Resistance Genes. Top. Clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR Associated Protein 9 (Cas9) screens enable high-throughput screening of genes by transfecting a pool of single-guide RNA (sgRNA) into cells and performing a selection for traits associated with host–bacterial interactions. In the provided example, different colors represent cells transfected with different sgRNAs, and the green guide knocked out a gene required for pyroptosis. As such, the guide DNA is enriched following infection with pyroptosis-inducing bacteria and is identified for additional experimental follow-up work. Bottom. Natural Genetic Diversity Screens enable researchers to identify genetic loci that confer resistance to bacterial infection. After identifying a diverse population and phenotyping individual responses to a bacterial pathogen, phenotypic diversity can be statistically associated to genotypic diversity. Following identification of genetic loci associated with phenotypic diversity (quantitative trait locus, QTL), follow-up work examines which polymorphisms and/or genes contribute to host resistance. Further, comparison and integration of individual studies into additional screens can provide a more complete picture of a QTL's role in host biology. Right. Dual approach screens involve simultaneous screening of host and bacterial genomes, transcriptomes, and proteomes to understand infection biology. These approaches essentially employ studying the joint host–bacterial interaction to learn about the pressures that the bacteria overcome in the host and/or the systems they manipulate to promote virulence. These approaches can be critical hypothesis-generating tools. Abbreviation: CXCL10; C-X-C motif chemokine ligand 10: GWAS, genome-wide association study.

[4,5]. Early stop codons caused by CRISPR/Cas9 can upregulate associated genes through compensatory processes paired to nonsense-mediated mRNA decay [6]. Further, CRISPR/Cas9 screens cannot identify essential genes that moonlight in immunity. Notably, some work has attempted to circumvent the latter limitation by enabling temporal control of the Cas9 machinery [7]. Despite these drawbacks, screens using and improving on CRISPR/Cas9 have probed bacterial sensing, phagocytic uptake, bacterial manipulation of the intracellular niche, and toxin susceptibility.

Several groups have harnessed CRISPR screens to examine how host immune cells respond to bacterial pathogen-associated molecular patterns (PAMPs). Parnas *et al.* performed a whole-genome screen (125 793 sgRNAs targeting 21 786 mouse genes) to untangle the response of myeloid cells to lipopolysaccharide (LPS) stimulation [8]. The screen identified known factors in LPS signaling, including *TLR4*, *MYD88*, and *ZFP36*, as well as novel positive and negative regulators of tumor necrosis factor (TNF). Follow-up experiments revealed that oligosaccharyltransferase (OST) complex regulates Toll-like receptor (TLR)4 signaling. Interestingly, an independent CRISPR screen in a murine pro-B cell line also identified that the OST complex regulates TLR5, 7, and 9 signaling [9], suggesting that arginine glycosylation is a conserved regulator of TLR signaling. Notably, these techniques rediscovered genes with established involvement in sensing, raising confidence in the results of these screens.

Once a PAMP is sensed, the immune response must be regulated to avoid excessive inflammation and/or sepsis. Several groups sought to understand this regulation using CRISPR screening, including one that identified the complement peptidase *Cpb1* as a positive regulator of Caspase-11-mediated pyroptosis in response to LPS, *Salmonella*, and *Shigella* in murine macrophages [10]. Their experimental validation revealed that CPB1 enables the C3/C3aR signaling pathway to amplify TLR4 and interferon-α/β receptor (IFNAR) signaling cascades, enhancing Caspase-11 activity and sepsis severity in mice and humans. An additional series of CRISPR screens revealed that autophagy genes modulate TNF signaling to counteract interferon-gamma- (IFN-γ)-mediated cell death and sepsis in murine models [11]. These studies demonstrate the power of CRISPR screens to identify genes involved in harmful immune responses *in vitro,* which could also serve *in vivo* as biomarkers or druggable targets for sepsis.

Beyond sensing and signaling, many groups have used CRISPR/Cas9 screens to examine how host genes facilitate intracellular bacterial lifestyles, particularly in phagocytic host cells. One recent screen identified host genes required for *Salmonella enterica* uptake into PMA-differentiated human THP-1 macrophages [12]. This study complemented a phagocytosis screen that used magnetic sorting to identify genes required for uptake of beads or other substrates [red blood cells (RBCs), myelin, zymosan] conjugated with ferrous nanoparticles by differentiated human U937 macrophages [13]. Both screens identified *NHLRC2* as a driver of phagocytosis, with Yeung *et al.* [12] demonstrating that the protein likely plays a role in both macrophage phagocytosis and differentiation. Finally, a third screen examined the later stages of phagosome development and found that the bicarbonate transporter *SLC4A7* is required by phagocytes to acidify their vacuoles and kill bacterial pathogens [14]. Together, these screens uncovered dozens of genes involved in phagocytosis; however, they also highlight an old screening maxim, 'You get what you screen for'. Each group performed a screen for phagocytosis, but what they identified varied based on cell type, phagocytic substrate, timing, and other experimental parameters. Thus, considering these factors is critical when designing screens and interpreting results.

While the studies above focused on host genes required to resist bacterial infection, several screens also identified host proteins that are hijacked by bacteria to aid in pathogenesis. For

instance, a herculean series of screens identified hundreds of host factors required for *Legionella pneumophila* to enter phagocytic cells, replicate, and eventually kill the host cell [15]. This study provides a comprehensive view of how previously identified host factors and newly identified genes work together to facilitate pathogenesis. Additional work has focused on host factors that enable bacteria-mediated cytotoxicity. For instance, one CRISPR screen that focused on *Vibrio parahaemolyticus*-induced cell death found that host cell surface sulfation allows the bacteria to adhere and kill cells using its type III secretion system 1, while host cell surface fucosylation enables the type III secretion system 2 translocon to enter the host membrane and facilitate cytotoxicity [16]. Similar work with enterohemorrhagic *Escherichia coli* demonstrated that host sphingolipid biosynthesis is critical for both type III secretion- and Shiga toxin-mediated cell death [17]. Finally, a CRISPR screen by Tao *et al.* identified certain family members of the Wnt receptor Frizzled (*FZD1/2/7*) as the predominant colonic receptors for the *Clostridium difficile* toxin TcdB [18]. These studies highlight CRISPR screen identification of disease-risk loci, which are potential targets for host-directed therapeutics to reduce bacteria-induced pathology.

Altogether, this work demonstrates the impact that CRISPR screens have already had on understanding host resistance to bacterial infection. However, this likely represents the tip of the iceberg of what can be learned by harnessing this approach. Researchers across disciplines are constantly working to advance the technology and expand the number of questions that can be asked (Box 1). As these techniques develop, so will our understanding of host–bacterial interactions.

As mentioned above, significant advances have been made performing CRISPR screens *in vivo* [2,3]; however, loss-of-function screens in whole model organisms are still predominantly

Box 1. Moving beyond Single-Gene Knockouts in CRISPR Screens

The many advantages of CRISPR screens have inspired researchers to expand the technology to address additional questions. Two major advances in CRISPR screens have included using Cas9 constructs designed to modulate gene expression rather than ablate it and knocking out multiple genes to understand genetic networks. While these approaches have had limited use in host–bacterial screening, they are worth discussing as they will likely become more prevalent in the immediate future.

Through engineering of Cas9 machinery, researchers have developed systems to increase (CRISPRa) or decrease (CRISPRi) host gene expression without directly disrupting the gene [83]. In particular, CRISPRa technology represents a major innovation as it is one of few techniques that can be used for gain-of-function screening. Further, unlike traditional cDNA overexpression systems, CRISPRa modulates expression at the native locus and thus keeps *cis*-regulatory effects intact. While no bacterial host resistance screens have been performed using CRISPRa, a recent publication used it to identify norovirus resistance genes [84]. Interestingly, while one weakness of CRISPRa and CRISPRi is that different sgRNAs can yield different intensities of overexpression or knockdown, some groups have actually engineered sgRNAs that intentionally differ in efficiencies in order to study the impacts of titrated gene dosage on phenotypes [85].

In addition to advances in modulating expression, some studies seek to examine how perturbing multiple genes in a single cell impacts host phenotypes (Figure I). Approaches using standard CRISPR/Cas9, CRISPRa, and CRISPRi technologies have all attempted to cotransfect sgRNAs into cells, determine which sgRNAs were expressed, and measure the impacts of those coexpressed sgRNAs on the host. There are several different approaches to examine combinatorial effects, primarily differing in how to determine which sgRNAs were transfected into a cell. One technique, called CRISP-Seq, pairs CRISPR screening, fluorescent sorting, and single-cell RNA sequencing [2]. By placing genes for fluorescent markers and unique guide indexes on the sgRNA plasmids, the authors were able to use cell sorting and single-cell sequencing to connect the sgRNA(s) expressed in each cell to the resulting changes in the transcriptome. While CRISP-Seq has great promise, the number of sgRNA combinations tested is limited by single-cell RNA sequencing throughput as well as the number of available fluorescent markers. Further, techniques that depend on guide indexes are subject to spontaneous recombination during lentiviral preparation than can decouple the unique guide indexes and sgRNA [86]. An alternative approach, called direct-capture perturb-seq, leverages single-cell RNA sequencing paired to a modified library preparation to directly identify expressed sgRNAs during sequencing [87]. By directly sequencing guides, this method lacks some of the limitations associated with CRISP-Seq, though it is still bound by single-cell RNA sequencing throughput.

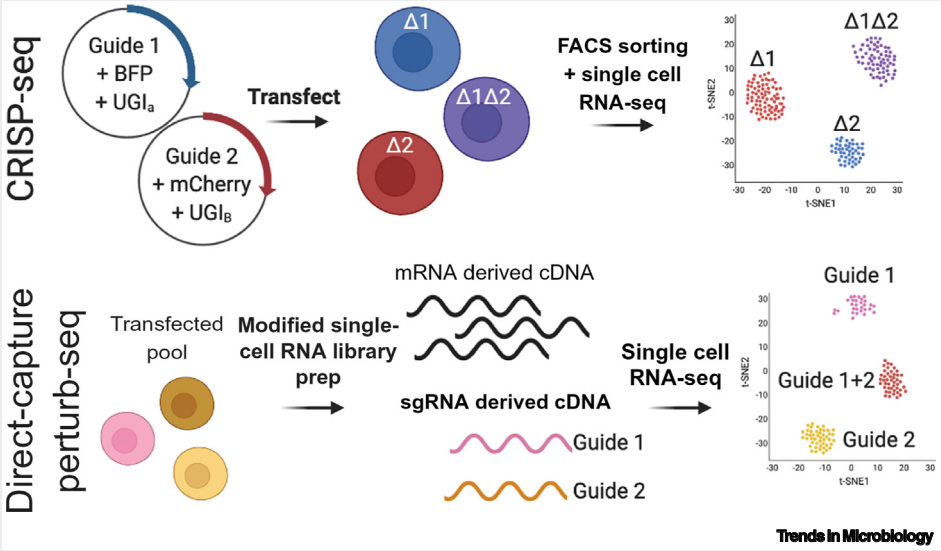

Figure I. Approaches for Multigene Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) Knockout Screens. Two approaches to assay the impacts of multigene deletions include CRISP-seq [2] and Direct-capture perturb-seq [87]. In CRISP-seq (top), single-guide RNA (sgRNA) plasmids include a unique fluorescent marker and unique guide index (UGI). This allows two unique ways of determining which sgRNA(s) were transfected into each cell assayed by single-cell RNA-seq. First, during the library generation, flow cytometry can be used in order to determine which fluorescent markers are expressed in the cell. Second, the UGI can be detected directly during RNA sequencing and mapped back to the sgRNA plasmid. Together, these features enable indirect identification of the sgRNA expressed in the cell as well as the resulting transcriptome changes in each. By contrast, direct-capture perturb-seq (bottom) simply uses a modified single-cell RNA-seq library preparation that enables sgRNA to directly sequence each expressed sgRNA alongside the transcriptome. Abbreviation: FACS, fluorescence-activated cell sorting.

performed using random mutagenesis [e.g., *N*-ethyl-*N*-nitrosourea (ENU) mutagenesis]. These tried-and-true techniques have been thoroughly reviewed elsewhere [19,20], but we want to draw special attention to the role these systems have played in understanding host–bacterial interactions in zebrafish. This is perhaps best illustrated by a pair of landmark papers by Tobin *et al.* who used an ENU mutagenesis screen to identify a critical role for *lta4h* in modulating zebrafish susceptibility to *Mycobacterium marinum* infection [21]. Interestingly, they found that *lta4h* activity must be carefully balanced as excessive activity results in accumulation of the proinflammatory LTB_4 eicosanoid, while reduced activity leads to accumulation of the anti-inflammatory lipoxin A_4 [22]. Tobin *et al.* confirmed this optimal immune balance in humans as human SNPs associated with *LTA4H* gene expression confer heterozygous advantage [21], and the *LTA4H* genotype appears to serve as a predictive biomarker for the efficacy of the anti-inflammatory drug dexamethasone [22]. This marriage of zebrafish and human genetics, leading to clinical translation, demonstrates the power of random mutagenesis screens. However, the future of loss-of-function screening in zebrafish is unclear as impressive work has been done to develop zebrafish CRISPR screens [23–26], though these techniques have not yet been applied to host–pathogen biology.

Leveraging Natural Genetic Diversity to Identify Host Resistance Genes

An additional limitation of genetic knockout systems is that they cannot reveal which genes contribute to 'natural' interindividual differences. This is because the majority of interindividual differences are not large genetic ablations but rather SNPs [27]. Therefore, complimentary techniques that harness natural diversity are necessary to identify naturally occurring host resistance and susceptibility genes (Figure 1).

The technical and statistical frameworks that leverage natural variation to identify host resistance genes differ but are broadly categorized into linkage analysis and association analysis [28]. For many years, linkage analysis has been the predominant method for studying Mendelian traits, where highly penetrant variants of large effects are mapped using genotyped family pedigrees. This kind of analysis is still very powerful, especially in finding rare variant 'inborn errors of immunity' [29], such as mutations in interleukin (IL)-12 and IFN-γ signaling that predispose to mycobacterial and *Salmonella* infection [30]. However, for identifying more common resistance alleles for polygenic traits, genome-wide associations studies (GWAS) are the predominant discovery method. In GWAS, members of a genetically diverse population are assessed for a phenotype (e.g., infection status, disease severity, cytokine response during infection, leukocyte counts, etc.). Then, markers across the genome are tested for association with the trait, asking how well each marker explains the observed phenotypic variation when accounting for genetic relatedness and other covariates (usually through principal components analysis) [31]. These studies depend on linkage disequilibrium maps [32,33] which are used to select tag SNPs (to reduce the number of markers that need to be genotyped) and to understand individual association signals using the patterns of genetic inheritance along chromosomes. These techniques have identified loci associated with genetic resistance, susceptibility, and quantitative host–bacterial interactions [quantitative trait loci (QTL)]. We discuss three applications of natural variation in host–bacterial studies: (i) leveraging natural variation in model organisms through association mapping, (ii) leveraging natural genetic variation in humans using GWAS, and (iii) examining natural genetic variation *in vitro* using cellular GWAS.

Identifying Host Bacterial Resistance Genes by Using Natural Diversity in Mouse Models

Many studies have paired the strengths of invertebrate (Box 2) and vertebrate models of infections with the natural diversity present within these species to study host–pathogen interactions.

Box 2. Natural Diversity Screens Using Animal Models

Caenorhabditis elegans and *Drosophila melanogaster* have frequently been used in natural diversity screens aimed at understanding host–pathogen interactions, as researchers have collected, laboratory-adapted, phenotyped, and sequenced wild isolates of these species. A classic example used natural diversity to identify that a single amino acid substitution in the *C. elegans* neuropeptide Y receptor homolog (NPR-1) drives a differential response to bacterial prey in wild isolates [88]. More recent work has sought to understand how the microbivorous worms 'avoid' pathogen exposure in the wild. One group leveraged a diversity panel to identify two separate single amino acid substitutions in a neuronal ubiquitin E3 ligase (HECW-1) that affect *Pseudomonas aeruginosa* avoidance [89]. Later work has also leveraged diversity to identify QTLs for *P. aeruginosa* [90] and *Serratia marcescens* [91] avoidance, though no causal genes were identified.

Early work harnessing *Drosophila* diversity to study bacterial susceptibility performed candidate diversity screens targeting immune genes on the second [92,93] and X [94] *D. melanogaster* chromosomes. Following these studies, a novel resource containing hundreds of genotyped strains was devised for genome-wide *D. melanogaster* association mapping: the *Drosophila melanogaster* Genetic Reference Panel (DGRP) [95]. One study using the panel found that genetic variation in the *kri*, *S6k*, *mad2* locus, and the *BomBc1* antimicrobial peptide locus associated with resistance to *Enterococcus faecalis* [96]. Another expanded our understanding of fly gut immune defense by pairing the DGRP with *Pseudomonas entomophila* [97]. Researchers found that animals that showed higher survival cleared the infection faster, launched a more controlled reactive oxygen species response, maintained protein synthesis, and preserved gut stem cell activity. Additionally, 27 survival QTLs were mapped, including several SNPs in the *Nrk* and *Gyc76C* genes, which the authors confirmed play key roles in immunity.

By using the standardized DGRP, it becomes easier to compare results across studies. For instance, one impressive screen infected the DGRP with *Metarhizium anisopliae* or *P. aeruginosa* and identified a number of sex-dependent QTLs associated with survival, including in the *kri*, *S6k*, *mad2* locus that associated with the *E. faecalis* survival above [98]. Further, by mining DGRP studies, they found that geotaxis, oxidative stress, blood glucose, starvation, and sleep patterns all correlate with postinfection survival.

Curiously, despite significant amounts of genetic diversity in laboratory zebrafish strains [99], we are unaware of any work that has leveraged this diversity to study infectious disease. This may represent an exciting future direction for the field.

Here we focus on host genetic screens of bacterial systems that use the Collaborative Cross (CC) mouse model, which consists of ~100 recombinant inbred lines derived from eight diverse founder strains (C57BL6/J, 129S1/SvlmJ, A/J, NOD/ShILtJ, NZO/HiLtJ, CAST/EiJ, PWK/PhJ, and WSB/EiJ) [34]. These diverse yet inbred strains allow nearly genetically identical replicates and mechanistic follow-up. While QTLs revealed by the CC lines are quite large (the average recombination distance is ~10 Mb), individual CC strains that are outliers for a trait of interest can be crossed to produce F_1 and F_2 progeny for more sensitive QTL mapping. This section will complement recent reviews on the impact of CC mice on host–pathogen biology [35,36] as we unpack a number of recent host–bacterial papers with exciting implications.

Several studies have identified natural variation in responses and susceptibilities of CC mice to *Klebsiella pneumoniae* [37], *S. enterica* serovar Typhimurium (*S.* Typhimurium) [38], *Pseudomonas aeruginosa* [39], and the periodontitis-associated bacteria *Porphyromonas gingivalis* and *Fusobacterium nucleatum* [40]. Even with relatively small cohorts (35–48 CC lines), the *S.* Typhimurium and *K. pneumoniae* studies identified putative QTLs underlying susceptibility [37,38]. Additional work by one of the authors of this review demonstrated that the CC mice encompass a broad phenotypic spectrum after *Mycobacterium tuberculosis* infection and further showed that host genetic background modified the BCG (bacillus Calmette–Guérin) vaccine's protective efficacy against tuberculosis [41]. Intrinsic host resistance to *M. tuberculosis* and the response to the BCG vaccine were not correlated, demonstrating the genetic complexity of *M. tuberculosis* protection.

As noted above, an advantage of the CC is the ability to follow-up phenotypes with mechanistic experiments. For instance, in the original *M. tuberculosis* study [41], the mouse line CC042 had unusual susceptibility to infection. Further work, using F_2 progeny to map the QTL, revealed that CC042 mice have a 15-base deletion in the *Itgal* gene that resulted in impaired lymphocyte trafficking and *M. tuberculosis* susceptibility [42]. Interestingly, work by a separate group on CC042 identified that the same deletion drives susceptibility to *S.* Typhimurium [43], highlighting a conserved role for the protein in pathogenesis. Unexpectedly, the CC042 *Itgal* deletion is absent in the original CC parental strains and instead arose *de novo* during inbreeding. The high frequency of these *de novo* mutations during the inbreeding process may have ramifications on reproducibility across CC generations [44]. Importantly, these studies demonstrate that experimental follow-up can surmount the large QTLs obtained from CC studies and reveal very specific genetic mechanisms for differences in susceptibility.

There have been two notable attempts to connect CC bacterial pathogenesis studies to human GWAS cohorts. The first studied periodontal models, using *P. gingivalis* and *F. nucleatum* infection [45]. Unfortunately, even with F_2 progeny, the identified QTLs were massive – the two significant QTLs contained 80 genes, while the eight suggestive QTLs contained 1309 genes. Thus, while the authors note some overlap of their QTLs to loci identified in aggressive periodontitis and chronic periodontitis human GWAS cohorts, the implications are unclear. Additionally, work from Lorè *et al.* examined *P. aeruginosa* susceptibility using the CC system and identified a QTL containing 31-protein-coding genes [46]. By examining the associations between their QTL and SNPs in the human syntenic region from a cystic fibrosis GWAS, the authors identified two SNPs in *DPYD*, rs10875080 and rs11582736, which associated with a modest delay in the age of first *P. aeruginosa* infection in the human cohort. Follow-up work has yet to connect *Dpyd* to resistance in the CC lines. Together, these data demonstrate both the promise of the CC mice to identify possible human resistance genes and the challenge of going from QTL intervals to causal variants.

The CC mapping power problem has been highlighted throughout this section and is due in part to the number of strains available. Recent power analyses in the CC indicate that, for studies of polygenic traits explained by many loci of small effect, more CC strains are required than may be currently available [47]. In lieu of more CC strains, another approach may be to leverage more refined phenotypes ('endophenotypes') that explain a larger proportion of the variance. While the broader genetics field has taken an expression (e)QTL approach, the infectious disease field has the opportunity to move away from polygenic traits, such as survival or burden, and instead leverage bacterial expression as a novel refined intermediate phenotype that is likely to be caused by very specific host variants. The advent of diverse but reproducible model populations, such as the CC, in combination with next-generation bacterial technologies, make these dual host–pathogen mapping approaches possible.

Identifying Host Resistance Genes Using Natural Human Diversity

GWAS of human infectious diseases present several unique challenges. First, even with covariates (i.e., age, sex, comorbidities, and unknown factors with principal components analysis), a successful GWAS requires large population sizes, generally in the thousands. Second, obtaining exposed controls is challenging, so many infectious disease GWAS simply use population controls (e.g., blood donor banks [48–50]). This approach has a significantly lower logistical and financial cost compared with obtaining carefully matched controls who were exposed to the pathogen and resisted infection without a protective memory response, but it results in significantly lower power. Third, integrating individual GWAS into the broader scientific literature is difficult as differences in populations, sample sizes, disease definition, and study design can severely impair reproducibility across cohorts. Despite these challenges, GWAS can reshape medicine by informing drug design and providing possible clinical biomarkers. For example, an early GWAS of hepatitis C virus (HCV) infection [51] led to the development of a genetic biomarker that helped direct recombinant IFN treatment [52].

Researchers have boosted sample size by assembling massive databases of genotyped individuals with clinical information from electronic medical records (e.g., UK biobank [53], eMERGE network [54]) and self-reporting (23andMe [55]), and performing simultaneous GWAS across many phenotypes. For example, a study by 23andMe recruited a cohort of >200 000 European-descent participants to fill out a survey about their experiences with common infectious diseases [56]. The group identified 59 genome-wide significant SNPs associated with 17 different infectious diseases. They also identified pleiotropic SNPs that affect multiple diseases. For example, rs1978060, in *TBX1*, is associated with tonsillectomies, childhood ear infections, and myringotomy, which implicates the gene broadly in bacterial susceptibility. Further, they found that SNPs in HLA class I genes tended to correlate with viral disease susceptibility, while bacterial diseases were more likely to have significant SNPs in HLA class II genes, in line with each protein's canonical role in immunity. Notably, they also found associations in autoimmunity-associated genes, highlighting potential tradeoffs between infection protective alleles and inflammation homeostasis alleles.

Even without massive databases, dozens of bacterial susceptibility GWAS papers have been published over the last decade (Table 1). The specific loci identified for each disease have been reviewed elsewhere [57,58], so we will focus on how GWAS results are validated and used to enhance our understanding of host–pathogen biology and/or genomics.

Validation

One of the most critical questions following GWAS is whether the identified SNPs are true hits or a consequence of sampling bias. The gold standard for validating SNPs is to replicate the

Table 1. Human Genome-wide Association Studies for Bacterial Infection

Pathogen(s) and/or Disease(s)	Refs
Mycobacterium tuberculosis and/or tuberculosis	[48–50,60,106–113]
Mycobacterium leprae/leprosy	[114–118]
Staphylococcus aureus, *S. aureus* bacteremia, and/or *S. aureus* infective endocarditis	[119–123]
Streptococcus pneumoniae, *Neisseria meningitidis*, pneumococcal or meningococcal bacteremia/sepsis/meningitis, and/or pneumonia- associated sepsis	[102,124–127]
General bacteremia and/or non-typhoidal *Salmonella* (NTS) bacteremia	[77,128]
Oral bacterial pathogen burden (in periodontal disease patients)	[129–131]
Enteric fever/typhoid fever	[132]
Helicobacter pylori	[133]
Pseudomonas aeruginosa (in cystic fibrosis patients)	[134]
Common infectious disease (23 pathologies tested)	[56]

association in additional, unrelated cohorts. Unlike the original GWAS, which requires genome-wide significance (traditionally $P < 5 \times 10^{-8}$) to overcome the large number of multiple comparisons, targeted replication of the association can have weaker statistical significance as the correction burden is lower. Examples of successful replication are discussed throughout this section.

Importantly, failure to replicate GWAS can occur for a number of reasons. For example, the 23andMe study replicated known associations with shingles and tonsillectomy but failed to replicate several other reported genome-wide associations [56]. This may be due to differences in populations or criteria for patient inclusion, as well as false positives. Population differences are particularly difficult to overcome as differences in allele frequencies and in linkage disequilibrium among populations can result in insufficient power for SNPs that are rare in the replication population or even make replication impossible if the SNP is not present. Further, complex gene × gene and gene × environment interactions may prevent certain alleles from having an impact in certain populations. Alternatively, differences in disease definition across studies can have dramatic impacts on patient inclusion, leading to profound differences in the actual phenotype being measured. Determining whether a replication failure is due to the heterogeneity of infectious disease studies or false positives can be difficult and is most adequately addressed by carefully considering the differences between the specific cohorts in question.

Finally, the 'winner's curse', in which the observed effect sizes of GWAS hits are often larger than their true biological effects, can substantially contribute to lack of replication. Analysis of the results of 100 quantitative trait GWAS papers that attempted replication by Palmer and Pe'er [59] revealed that 31% replicated at Bonferroni-corrected *P*-value thresholds, compared with 75% predicted by the effect sizes in the discovery cohorts. Winner's curse accounted for most of this difference, resulting in a predicted replication rate of 37%. When analysis was restricted to only the 39 papers that attempted replication within the same continental ancestry, and reported numbers of the replication cohort for each locus being tested, the replication rate increased to 43%, consistent with the predicted replication rate considering winner's curse (45%). Thus, as with all science, interpretation of GWAS findings is a continual process, and as replication, functional, and clinical follow-up studies are conducted, the most robust findings will stand the test of time.

Biological Insight

Understanding the mechanistic underpinnings of GWAS findings can take different forms. Some researchers focus on how implicated gene(s) affect infection. Such experiments involve knocking

down or knocking out the gene and assessing how cells or mice change their response to the pathogen. Other studies focus on uncovering (i) which genetic variant out of the many in linkage disequilibrium causes a locus to associate with pathogen resistance, and (ii) how that variant regulates the associated gene(s) to induce resistance (e.g., coding mutation, enhancer mutation, splicing variant, etc.).

We discuss Curtis *et al.* as an example of how GWAS can generate insight into infection biology. In their work, they identified *ASAP1* as a novel human resistance factor against *M. tuberculosis* [50]. After identifying several SNPs in *ASAP1* as putative hits in their original GWAS (~11 000 Russian individuals), the group validated these SNPs in an additional Russian dataset and in a previously published cohort from Ghana and The Gambia. To connect the variation to *ASAP1* function, the group examined how host genotype and *Mycobacterium* infection affect expression in immune cells. They found that the rs4733781 risk allele associates with reduced baseline and postinfection expression of *ASAP1* in dendritic cells. Finally, they demonstrated that knockdown of *ASAP1* in patient-derived dendritic cells reduced dendritic cell matrix digestion and motility. These data suggest that ASAP1 helps to activate the adaptive immune system, as suppression hinders migration of dendritic cells to lymphatic tissues.

By contrast, work by Zheng *et al.* exemplifies how GWAS can provide insights into genetic regulation as well as host–pathogen interactions [60]. In their work, they performed an initial GWAS (~2100 Han Chinese individuals), followed by two smaller replication studies to identify two SNPs associated with tuberculosis. Following up on rs6114027, located in an *TGM6* intron, they revealed that the risk allele associates with reduced *TGM6* expression in patient immune cells. Based on luciferase assays, they demonstrated that the intronic risk allele suppressed expression relative to the protective allele. Further work with patients and mouse models confirmed the importance of *TGM6* in *M. tuberculosis* resistance. Together, these data both identified an allele-specific and causal role for the rs6114027-containing intron in regulating *TGM6* expression, as well as demonstrated the importance of the gene during *M. tuberculosis* infection.

Identifying Host Resistance Genes Using Cellular GWAS

While the previous section demonstrates the wealth of knowledge obtained through traditional GWAS, challenges of heterogeneity still hinder these studies. Further, there are inherent limitations to what phenotypes can be measured in humans, making it difficult to understand how the SNPs contribute to specific molecular and cellular changes that impact pathophysiology. Several groups have utilized complementary cellular GWAS methods to identify SNPs that alter host cell–pathogen interactions following stimulation or infection of primary immune cells [61–66], differentiated induced pluripotent stem cells (iPSCs) [67], or lymphoblastoid cell lines (human B cells immortalized *ex vivo* through Epstein–Barr virus infection) [68–71]. Shared advantages of these systems are (i) consistent dose and timing across cell lines, (ii) controlled bacterial genetics and/or PAMP sources, (iii) same donors across technical and biological replicates, and (iv) a greater number of testable phenotypes. Collectively, these systems have identified a plethora of novel SNPs and genes that contribute to host resistance to infection.

One form of cellular GWAS attempts to uncover the genetic factors that contribute to variable transcriptional responses to pathogenic stimuli. Similar to studies that identify QTL associated with gene expression (eQTL), these studies seek to understand response eQTL (reQTL), or eQTL that emerge in response to stimulation. For instance, work on patient-derived monocytes stimulated with IFN-γ or LPS [61], or patient-derived dendritic cells stimulated with LPS, influenza virus, or IFN-β [62], revealed a number of reQTLs. Similar work found reQTLs and response chromatin accessibility QTLs (response cQTLs) in iPSC-derived macrophages following stimulation

with IFN-γ and/or *S*. Typhimurium [67]. Interestingly, these studies found overlap between their reQTLs and SNPs identified in bacterial, inflammatory, and/or autoimmune disease GWAS. More recent work has identified QTL associated with miRNA regulation during monocyte stimulation [72]. These studies demonstrate that understanding response QTLs provides mechanistic information that complements traditional GWAS.

Other work has examined how variation affects the host immune response. Specifically, the Human Functional Genomics Project has identified genetic variation that affects pathogen sensing and cytokine production by immune cells *ex vivo*. Li *et al.* exposed peripheral blood mononuclear cells from over 100 European donors to LPS, heat-killed *M. tuberculosis*, or heat-killed *Candida albicans* [64]. Following stimulation, the authors measured six cytokines and identified six novel QTLs, including two QTLs involved in *M. tuberculosis*-induced IL-8 and one QTL involved in LPS-induced IL-10. They found high levels of pleiotropy, with all genome-wide hits also associating with additional pathogen-induced cytokine responses. Following integration of their data in the National Human Genome Research Institute (NHGRI) GWAS catalog, their SNPs also associated with infectious disease susceptibility, autoimmunity, and curiously, heart disease. Later work by the same group expanded the concept to measure cytokines from macrophages, peripheral blood mononuclear cells, and whole blood from over 400 individuals in response to 18 different killed pathogens, TLR-ligands, or metabolic stimuli [65]. The group found 18 genome-wide associated SNPs in 17 independent loci with a high degree of replication using their previous study [64] as a replication cohort. One of the top hits from this study was located in the *TLR1*, *TLR6*, and *TLR10* locus, and affected IL-6 production following peripheral blood mononuclear cell (PBMC) stimulation with polyinosinic:polycytidylic acid (poly I:C). This locus is under positive selection, suggesting its importance in host resistance. Interestingly, an earlier study also found a number of SNPs that correlate with IL-6 production at this locus following stimulation of whole blood with Pam_3CSK_4 [66]. This convergence highlights the importance of TLR 1/6/10 signaling in cytokine production and demonstrates the ability of even highly divergent cellular GWAS to identify common loci.

One of our laboratories carries out cellular GWAS screening with a system called Hi-HOST (High-throughput Human *in vitrO* Susceptibility Testing) using live pathogens. The original Hi-HOST screen was performed using 352 lymphoblastoid cell lines (from donors of western European or Nigerian descent) infected with *S.* Typhimurium or *Salmonella* Typhi [68]. We screened for QTLs involved in host cell invasion, pyroptosis, bacterial replication, and intracellular bacterial survival. We found that rs514182, an eQTL for the methionine salvage enzyme *APIP*, associated with pyroptosis during *S.* Typhimurium invasion [69]. Follow-up work revealed that APIP negatively regulates pyroptosis by reducing the abundance of the methionine-derived metabolite methylthioadenosine. The same locus is associated with systemic inflammatory response syndrome, as well as nontyphoidal *Salmonella* (NTS) bacteremia and sepsis [73]. Finally, metabolites in the salvage pathway regulated by APIP, in particular methylthioadenosine and *S*-adenosyl homocysteine, are reliable sepsis biomarkers [73] that also directly regulate *S.* Typhimurium virulence [74,75]. Another hit from the Hi-HOST screen was rs8060947, an eQTL for *VAC14*, which correlated with *S.* Typhi invasion [70]. VAC14 reduces cholesterol levels in the plasma membrane, leading to reduced *S.* Typhi docking and invasion. This SNP also associated with typhoid fever susceptibility, malarial risk [76], and modestly with NTS, *S. pneumoniae*, *E. coli*, and *Acinetobacter* bacteremia [77].

Hi-HOST recently evolved into the Hi-HOST Phenome Project (H2P2), which screened an additional 528 lymphoblastoid cell lines from four populations, with eight different stimuli (seven live bacterial and fungal pathogens and one bacterial toxin), for 79 phenotypes [71]. Several of the

17 genome-wide significant hits demonstrated pleiotropy, including rs953897, located in the transcriptional repressor ZBTB20. rs953897 associated with *Chlamydia trachomatis* burden and 20 additional H2P2 phenotypes at $P < 0.05$. *ZBTB20* knockdown validated three of the strongest associations: *C. trachomatis* burden and induced IL-6 production, and *S.* Typhimurium-induced pyroptosis. Further, integration of the H2P2 data into the eMERGE PheWAS and other clinical datasets revealed that a number of the top H2P2 SNPs associate with human disease, including the *ZBTB20* SNP (rs953897) with HCV susceptibility, and an SNP in *CXCL10* (rs2869462) with inflammatory bowel disease. Future work is exploring the biology behind the H2P2-identified SNPs and further adapting the Hi-HOST screening platform.

While this review has focused on methods that interrogate the host to identify genes that contribute to resistance, there is a major caveat in studying host–pathogen interactions in this way: these approaches generally require treating pathogens as static, homogeneous stimuli, rather than dynamic and diverse components of the host–pathogen interaction. There is substantial value in incorporating pathogen biology into approaches to identify host resistance genes, including through bacterial diversity-informed GWAS, as described in Box 3.

Interrogating the Host–Pathogen Interface through Dual RNA-seq

In this last section we discuss techniques that use the bacteria as a biological sensor to understand the pressures inflicted by the host (Figure 1). This can take many forms, including examining the bacterial–host protein interactome or tracking bacterial transcriptional and genomic adaptation to the host (Box 4), or by simultaneously measuring how the bacteria and host respond to one another during infection. We discuss the latter by examining the impact of dual RNA sequencing (Dual RNA-seq) on understanding host–pathogen interactions. Notably, while this method does not directly implicate host genes as being important for resistance, these screens do identify host processes that appear to place significant selective pressures on the bacteria and therefore represent an important hypothesis-generating tool.

Improvements in RNA-sequencing technology has enabled capturing and sequencing of host and pathogen RNA during *in vitro* and *in vivo* infection. This technology, called dual RNA-seq or dual-seq, directly examines correlations between host and bacterial transcripts, providing preliminary evidence of how the two entities respond to each other. Dual-seq has been reviewed elsewhere [78], so here we briefly discuss how the technology can identify host processes that restrict bacterial virulence. In contrast to traditional dual-seq studies that measure the host and pathogen transcriptional profiles during infection, measuring restriction requires comparing the transcriptional profiles of resistant and susceptible hosts. Work by Thänert *et al.* used dual-seq to understand differential *Staphylococcus aureus* susceptibility in C57BL/6 and A/J mice [79]. Infected tissues from susceptible A/J mice showed evidence of elevated inflammation and hypoxia, correlating with bacterial upregulation of genes involved in anaerobic metabolism and acid resistance. This likely drives the sepsis-related mortality in these mice. By contrast, *S. aureus* in resistant C57BL/6 mice increased expression of *de novo* amino acid synthesis pathways, antimicrobial peptide resistance genes, and stress response genes, implying that nutritional and innate immunity restrict bacterial virulence.

Beyond interanimal variation, other work has probed heterogeneity across cells within a single animal. One study examined *M. tuberculosis* permissibility in alveolar (permissive) and interstitial (restrictive) lung macrophages and revealed a link between host metabolism and bacterial fitness [80]. The transcriptional differences followed traditional M1/M2 polarization schema, with permissive alveolar macrophages displaying high fatty acid metabolism and M2-like gene expression, which triggers fatty acid metabolism, growth, and iron storage in the invading *M. tuberculosis*.

Box 3. Using Bacterial Diversity Informed GWAS to Identify Host Resistance Genes

In our discussion of animal models of natural diversity, single genetically identical strains of pathogen were used to identify loci associated with pathogen outcome. However, in human cohorts, participants can have dramatically different bacterial strains, serovars, or even species pooled into the same experimental group. This can make it difficult to identify associated loci as pathogen diversity can confound the host response. An alternative approach to traditional GWAS incorporates pathogen diversity by genotyping both host and pathogen DNA from participants to stratify results by bacterial lineage (Figure I). This approach has been applied to *M. tuberculosis,* where one study ran GWAS on their entire patient cohort, as well as separate GWAS on patients infected with the Beijing lineage or those infected with a non-Beijing lineage [49]. While no significant associations were noted in the pooled GWAS, the researchers were able to find a statistically significant association with susceptibility to the non-Beijing lineages (rs1418425, intergenic between *CD53* and *LRIF1*, odds ratio 1.62). This association was found in their primary and replication cohorts; however, it is interesting to note that the association was only significant among older individuals.

Beyond this approach of stratifying GWAS by bacterial taxonomy, a second approach involves using hypothesis-free genome-to-genome analyses to identify interactions between host and pathogen genetics by measuring associations between human and pathogen SNPs (Figure I). This information can then be used to stratify GWAS results using interacting genes. The genome-to-genome approach has proved fruitful for viral human GWAS with HIV [100] and HCV [101]. In these viral studies multiple associations were found between HLA genes and viral proteins, as well as an association between rs12979860 (an intronic SNP in *IFNL4*) and the HCV protein NS5A. However, the increased multiple comparison burden from the larger size of bacterial genomes has thus far made this approach recalcitrant for bacterial human GWAS, though it has been attempted in a small pneumococcal disease cohort [102]. Future work with large bacterial GWAS cohorts should consider sequencing host and bacterial DNA, as identifying host–bacterial interacting genes could provide important insight into both host resistance alleles and host–pathogen interactions.

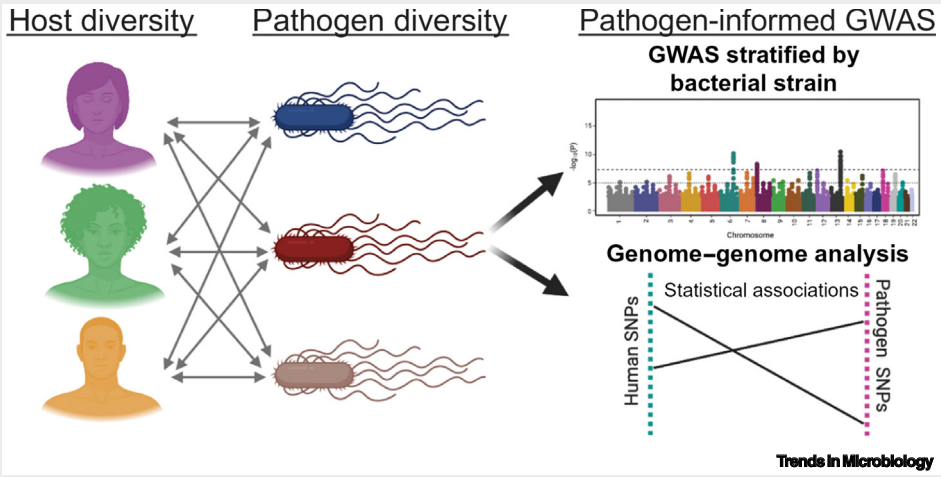

Figure I. Pathogen-Informed Genome-wide Associations Studies (GWAS). Infections occur in the context of both host and pathogen diversity. Simultaneous sequencing of host and pathogen DNA enables pathogen diversity to be captured during GWAS studies. This can take two forms. First, GWAS can be stratified by bacterial strain, so that the heterogeneity of the infected state is significantly reduced and biological differences between individuals are more likely driven by host genetic diversity. Second, hypothesis-free genome-to-genome analysis looks for statistical associations between the bacterial and host genomic diversity. These statistical associations may mark sites of host–pathogen evolutionary conflict and implicate host genes in the host–bacterial interaction.

By contrast, the restrictive interstitial macrophages are more M1-like, sequester iron, reduce potassium and chloride transport, and drive expression of stress-response, iron-salvaging, and dormancy genes in their associated bacteria. Similar results were found in a dual-seq study examining *S.* Typhimurium growth dynamics, as they found that *Salmonella* uses the effector SarA to drive macrophages into a permissive M2-like state and enable rapid intracellular bacterial growth [81]. This suggests that clinical manipulation of host metabolism or macrophage polarization could impact bacterial virulence and disease progression.

Box 4. Identifying Important Host Processes by Studying the Bacteria

Many bacteria manipulate host processes to thrive within an intracellular space. Understanding which host proteins bacteria have evolved to interact with can shed light on which host genes are necessary for virulence. Two techniques can be used to uncover interactions between bacterial and host proteins: coimmunoprecipitation approaches or BioID-based screens. A recent preprint used the former by generating 15 *S*. Typhimurium mutants containing chromosomally tagged effector proteins and infecting HeLa cells (a human epithelial cell line) and RAW264.7 cells (a murine macrophage cell line) [103]. Following infection, secreted effectors were immunoprecipitated, and host-interacting proteins were identified using mass spectrometry. A similar BioID screen ectopically overexpressed five BirA-tagged *S.* Typhimurium effectors in human epithelial cells to identify effector-interacting proteins [104]. Notably, there were considerable differences in the results from these studies, with some of the interactions from the infection experiment failing to replicate in the BioID paper. The lack of replication may be the result of effector overexpression failing to model infection conditions. Thus, the strengths and weaknesses of these systems must be considered when designing protein interactome screens.

The above examples rely on the fact that bacterial proteins have evolved over countless generations to manipulate host proteins. However, genomic approaches can also be used to study adaptation in real time. For example, recent work by Crofts *et al.* used a human infection model of *Campylobacter jejuni* to understand how the pathogen's biology changes during infection [105]. The authors infected volunteers with a genetically diverse population of *C. jejuni* and collected feces from participants in order to (i) compare the *C. jejuni* transcriptome within humans to the transcriptome in other conditions, and (ii) track how selective pressures in humans influenced *C. jejuni* evolution. The authors noted that genes involved in hydrogen peroxide detoxification, iron acquisition, and antimicrobial peptide resistance were upregulated, suggesting that the bacteria were responding to host innate and nutritional immunity. By studying the pathogen's genomic adaptation, the authors found that a number of mutants showed consistent selection patterns across patients. Notably, mutations in seven out of eight flagellin modifier genes in the *C. jejuni* genome were under selection in humans, suggesting an important role for flagellin modifications and/or host flagellar sensing during *C. jejuni* infection. Together, these experiments uncovered clues from the bacterial response to humans that reveal host selective pressures.

Outstanding Questions

Do CRISPR screen results replicate under different conditions? What are the implications of the concordance or discordance?

Do results from Collaborative Cross mice translate to human cohorts, and how can such results be efficiently integrated?

How do loci from Collaborative Cross or human GWAS studies influence host–pathogen interaction outcomes on the molecular and cellular levels?

How can targeted genetic engineering be paired with natural host diversity to identify genetic modifiers of genes?

How do loci that impact host susceptibility to bacterial infection impact other aspects of host biology? Are there consequences for chronic disease in humans?

What impact does microbial genetics have on the results from screens?

How can microbial diversity be screened simultaneously with standard host screens to expand the clinical relevance of the results?

Improving single-cell RNA-sequencing technology will have dramatic impacts on the ability for these studies to uncover cellular heterogeneity in bacterial susceptibility. Previous work has validated this tool for dual-seq [82], though it remains underutilized. Future work could apply single-cell dual-seq to dissect why neighboring cells within a single tissue have drastically different outcomes to bacterial infection.

Concluding Remarks

While this review focused on host–bacterial interactions, these screening approaches have also been adopted and advanced by virologists, mycologists, and parasitologists to study host–pathogen biology. The work presented here demonstrates that screens can dramatically advance the field. However, it is important to note that simply compiling lists of associations does not, by itself, illuminate how genes confer resistance to infection. Rather, following-up screens with additional studies provides key mechanistic insight into host–pathogen biology (see Outstanding Questions).

We discussed how innovative technology has advanced the field; however, new approaches are constantly on the horizon. For example, as discussed in Box 1, advances in CRISPR/Cas9 technology will likely expand the breadth of questions answered by CRISPR screens. Additionally, while many screens above treat the invading pathogen as a biological constant and/or a signaling PAMP, significant insight can be obtained by simultaneously screening host and pathogen diversity. As discussed in Box 3, this concept has already been acknowledged in human GWAS studies; however, future *in vitro* and animal model screens should also consider inclusion of pathogen diversity (whether natural diversity or bacterial knockout pools) to significantly improve modeling of complex host–pathogen interactions. In particular, we think that including bacterial diversity in approaches such as the Collaborative Cross model, Hi-HOST, or dual-seq studies could yield fascinating results. Similarly, expanding the unique phenotypes measured in natural diversity screens will improve our understanding of host susceptibility, particularly enabling pleiotropic loci identification via screen integration. Finally, improvements on methods such as single-cell RNA sequencing, epigenetic modification detection, iPSC differentiation, high-throughput

microscopy, flow cytometry, and computational toolkits will revolutionize the screens listed in this review as they expand the number of quantifiable phenotypes, provide single-cell resolution, and increase throughput and statistical power. Thus, more than 350 years after the birth of the field, it remains an exciting time to be a microbiologist.

Acknowledgments

The authors would like to thank Kyle Gibbs and Michael Kiritsy for their thoughtful comments following a thorough reading of the manuscript, as well as Dr Liuyang Wang and Ben Schott for sharing their thoughts on current trends in screening during the conceptualization of this review. J.S.B. was supported by National Institutes of Health (NIH) 1F31AI143147. C.M.S. was supported by the Duke University School of Medicine Whitehead Scholars Award. D.C.K. was supported by NIH R01AI118903, R21AI146520, R21-AI144586, R21AI133305, and U19AI084044. All figures were created with Biorender.com.

References

1. Perreira, J.M. *et al.* (2016) Functional genomic strategies for elucidating human–virus interactions: will CRISPR knockout RNAi and haploid cells? *Adv. Virus Res.* 94, 1–51
2. Jaitin, D.A. *et al.* (2016) Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-Seq. *Cell* 167, 1883–1896 e15
3. LaFleur, M.W. *et al.* (2019) A CRISPR-Cas9 delivery system for *in vivo* screening of genes in the immune system. *Nat. Commun.* 10, 1668
4. Teng, X. *et al.* (2013) Genome-wide consequences of deleting any single gene. *Mol. Cell* 52, 485–494
5. Rossi, A. *et al.* (2015) Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature* 524, 230–233
6. El-Brolosy, M.A. *et al.* (2019) Genetic compensation triggered by mutant mRNA degradation. *Nature* 568, 193–197
7. Chylinski, K. *et al.* (2019) CRISPR-switch regulates sgRNA activity by Cre recombination for sequential editing of two loci. *Nat. Commun.* 10, 5454
8. Parnas, O. *et al.* (2015) A Genome-wide CRISPR screen in primary immune cells to dissect regulatory networks. *Cell* 162, 675–686
9. Sato, R. *et al.* (2017) Requirement of glycosylation machinery in TLR responses revealed by CRISPR/Cas9 screening. *Int. Immunol.* 29, 347–355
10. Napier, B.A. *et al.* (2016) Complement pathway amplifies caspase-11-dependent cell death and endotoxin-induced sepsis severity. *J. Exp. Med.* 213, 2365–2382
11. Orvedahl, A. *et al.* (2019) Autophagy genes in myeloid cells counteract IFNgamma-induced TNF-mediated cell death and fatal TNF-induced shock. *Proc. Natl. Acad. Sci. U. S. A.* 116, 16497–16506
12. Yeung, A.T.Y. *et al.* (2019) A genome-wide knockout screen in human macrophages identified host factors modulating *Salmonella* infection. *mBio* 10, e02169-19
13. Haney, M.S. *et al.* (2018) Identification of phagocytosis regulators using magnetic genome-wide CRISPR screens. *Nat. Genet.* 50, 1716–1727
14. Sedlyarov, V. *et al.* (2018) The bicarbonate transporter SLC4A7 plays a key role in macrophage phagosome acidification. *Cell Host Microbe* 23, 766–774 e5
15. Jeng, E.E. *et al.* (2019) Systematic identification of host cell regulators of *Legionella pneumophila* pathogenesis using a genome-wide CRISPR screen. *Cell Host Microbe* 26, 551–563 e6
16. Blondel, C.J. *et al.* (2016) CRISPR/Cas9 screens reveal requirements for host cell sulfation and fucosylation in bacterial type III secretion system-mediated cytotoxicity. *Cell Host Microbe* 20, 226–237
17. Pacheco, A.R. *et al.* (2018) CRISPR screen reveals that EHEC's T3SS and shiga toxin rely on shared host factors for infection. *mBio* 9, e01003-18
18. Tao, L. *et al.* (2016) Frizzled proteins are colonic epithelial receptors for *C. difficile* toxin B. *Nature* 538, 350–355
19. Varshney, G.K. and Burgess, S.M. (2014) Mutagenesis and phenotyping resources in zebrafish for studying development and human disease. *Brief Funct. Genomics* 13, 82–94
20. Caignard, G. *et al.* (2014) Mouse ENU mutagenesis to understand immunity to infection: methods, selected examples, and perspectives. *Genes (Basel)* 5, 887–925
21. Tobin, D.M. *et al.* (2010) The lta4h locus modulates susceptibility to mycobacterial infection in zebrafish and humans. *Cell* 140, 717–730
22. Tobin, D.M. *et al.* (2012) Host genotype-specific therapies can optimize the inflammatory response to mycobacterial infections. *Cell* 148, 434–446
23. Varshney, G.K. *et al.* (2015) High-throughput gene targeting and phenotyping in zebrafish using CRISPR/Cas9. *Genome Res.* 25, 1030–1042
24. Shah, A.N. *et al.* (2015) Rapid reverse genetic screening using CRISPR in zebrafish. *Nat. Methods* 12, 535–540
25. Wu, R.S. *et al.* (2018) A rapid method for directed gene knockout for screening in G0 zebrafish. *Dev. Cell* 46, 112–125 e4
26. Trubiroha, A. *et al.* (2018) A rapid CRISPR/Cas-based mutagenesis assay in zebrafish for identification of genes involved in thyroid morphogenesis and function. *Sci. Rep.* 8, 5647
27. 1000 Genomes Project Consortium (2015) A global reference for human genetic variation. *Nature* 526, 68–74
28. Botstein, D. and Risch, N. (2003) Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat. Genet.* 33, 228–237
29. Casanova, J.L. and Abel, L. (2020) Lethal infectious diseases as inborn errors of immunity: toward a synthesis of the germ and genetic theories. *Annu. Rev. Pathol.* Published online April 14, 2020. https://doi.org/10.1146/annurev-pathol-031920-101429
30. Bustamante, J. *et al.* (2014) Mendelian susceptibility to mycobacterial disease: genetic, immunological, and clinical features of inborn errors of IFN-gamma immunity. *Semin. Immunol.* 26, 454–470
31. Bush, W.S. and Moore, J.H. (2012) Chapter 11: Genome-wide association studies. *PLoS Comput. Biol.* 8, e1002822
32. Consortium, G.P (2010) A map of human genome variation from population-scale sequencing. *Nature* 467, 1061–1073
33. International HapMap 3 Consortium (2010) Integrating common and rare genetic variation in diverse human populations. *Nature* 467, 52–58
34. Churchill, G.A. *et al.* (2004) The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat. Genet.* 36, 1133–1137
35. Noll, K.E. *et al.* (2019) The Collaborative Cross: a systems genetics resource for studying host–pathogen interactions. *Cell Host Microbe* 25, 484–498
36. Abu Toamih Atamni, H. *et al.* (2018) The Collaborative Cross mouse model for dissecting genetic susceptibility to infectious diseases. *Mamm. Genome* 29, 471–487
37. Vered, K. *et al.* (2014) Susceptibility to *Klebsiella pneumoniae* infection in collaborative cross mice is a complex trait controlled by at least three loci acting at different time points. *BMC Genom.* 15, 865
38. Zhang, J. *et al.* (2018) Identification of new loci involved in the host susceptibility to *Salmonella* Typhimurium in collaborative cross mice. *BMC Genom.* 19, 303

39. Lore, N.I. *et al.* (2015) Host genetic diversity influences the severity of *Pseudomonas aeruginosa* pneumonia in the Collaborative Cross mice. *BMC Genet.* 16, 106
40. Shusterman, A. *et al.* (2013) Genotype is an important determinant factor of host susceptibility to periodontitis in the Collaborative Cross and inbred mouse populations. *BMC Genet.* 14, 68
41. Smith, C.M. *et al.* (2016) Tuberculosis susceptibility and vaccine protection are independently controlled by host genotype. *mBio* 7, e01516-16
42. Smith, C.M. *et al.* (2019) Functionally overlapping variants control tuberculosis susceptibility in Collaborative Cross mice. *mBio* 10
43. Zhang, J. *et al.* (2019) A loss-of-function mutation in the integrin alpha L (Itgal) gene contributes to susceptibility to *Salmonella enterica* serovar Typhimurium infection in Collaborative Cross strain CC042. *Infect. Immun.* 88, e00656-19
44. Srivastava, A. *et al.* (2017) Genomes of the mouse Collaborative Cross. *Genetics* 206, 537–556
45. Nashef, A. *et al.* (2018) Integration of murine and human studies for mapping periodontitis susceptibility. *J. Dent. Res.* 97, 537–546
46. Lorè, N.I. *et al.* (2020) Collaborative Cross mice yield genetic modifiers for *Pseudomonas aeruginosa* infection in human lung disease. *mBio* 11, e00097-20
47. Keele, G.R. *et al.* (2019) Determinants of QTL mapping power in the realized Collaborative Cross. *G3 (Bethesda)* 9, 1707–1727
48. Grant, A.V. *et al.* (2016) A genome-wide association study of pulmonary tuberculosis in Morocco. *Hum. Genet.* 135, 299–307
49. Omae, Y. *et al.* (2017) Pathogen lineage-based genome-wide association study identified CD53 as susceptible locus in tuberculosis. *J. Hum. Genet.* 62, 1015–1022
50. Curtis, J. *et al.* (2015) Susceptibility to tuberculosis is associated with variants in the ASAP1 gene encoding a regulator of dendritic cell migration. *Nat. Genet.* 47, 523–527
51. Ge, D. *et al.* (2009) Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature* 461, 399–401
52. Liu, S. *et al.* (2012) New protease inhibitors for the treatment of chronic hepatitis C: a cost-effectiveness analysis. *Ann. Intern. Med.* 156, 279–290
53. Bycroft, C. *et al.* (2018) The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562, 203–209
54. McCarty, C.A. *et al.* (2011) The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC Med. Genom.* 4, 13
55. Eriksson, N. *et al.* (2010) Web-based, participant-driven studies yield novel genetic associations for common traits. *PLoS Genet.* 6, e1000993
56. Tian, C. *et al.* (2017) Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. *Nat. Commun.* 8, 599
57. Correa-Macedo, W. *et al.* (2019) The interplay of human and *Mycobacterium tuberculosis* genomic variability. *Front. Genet.* 10, 865
58. Mozzi, A. *et al.* (2018) Genetic susceptibility to infectious diseases: Current status and future perspectives from genome-wide approaches. *Infect. Genet. Evol.* 66, 286–307
59. Palmer, C. and Pe'er, I. (2017) Statistical correction of the winner's curse explains replication variability in quantitative trait genome-wide association studies. *PLoS Genet.* 13, e1006916
60. Zheng, R. *et al.* (2018) Genome-wide association study identifies two risk loci for tuberculosis in Han Chinese. *Nat. Commun.* 9, 4072
61. Fairfax, B.P. *et al.* (2014) Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* 343, 1246949
62. Lee, M.N. *et al.* (2014) Common genetic variants modulate pathogen-sensing responses in human dendritic cells. *Science* 343, 1246980
63. Nedelec, Y. *et al.* (2016) Genetic ancestry and natural selection drive population differences in immune responses to pathogens. *Cell* 167, 657–669 e21
64. Li, Y. *et al.* (2016) Inter-individual variability and genetic influences on cytokine responses to bacteria and fungi. *Nat. Med.* 22, 952–960
65. Li, Y. *et al.* (2016) A functional genomics approach to understand variation in cytokine production in humans. *Cell* 167, 1099–1110 e14
66. Mikacenic, C. *et al.* (2013) Variation in the TLR10/TLR1/TLR6 locus is the major genetic determinant of interindividual difference in TLR1/2-mediated responses. *Genes Immun.* 14, 52–57
67. Alasoo, K. *et al.* (2018) Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. *Nat. Genet.* 50, 424–431
68. Ko, D.C. *et al.* (2009) A genome-wide *in vitro* bacterial-infection screen reveals human variation in the host response associated with inflammatory disease. *Am. J. Hum. Genet.* 85, 214–227
69. Ko, D.C. *et al.* (2012) Functional genetic screen of human diversity reveals that a methionine salvage enzyme regulates inflammatory cell death. *Proc. Natl. Acad. Sci. U. S. A.* 109, E2343–E2352
70. Alvarez, M.I. *et al.* (2017) Human genetic variation in VAC14 regulates *Salmonella* invasion and typhoid fever through modulation of cholesterol. *Proc. Natl. Acad. Sci. U. S. A.* 114, E7746–E7755
71. Wang, L. *et al.* (2018) An atlas of genetic variation linking pathogen-induced cellular traits to human disease. *Cell Host Microbe* 24, 308–323 e6
72. Rotival, M. *et al.* (2020) Population variation in miRNAs and isomiRs and their impact on human immunity to infection. *Genome Biol.* 21, 187
73. Wang, L. *et al.* (2017) Human genetic and metabolite variation reveals that methylthioadenosine is a prognostic biomarker and an inflammatory regulator in sepsis. *Sci. Adv.* 3, e1602096
74. Bourgeois, J.S. *et al.* (2018) Methylthioadenosine suppresses *Salmonella* virulence. *Infect. Immun.* 86, e00429-18
75. Husna, A.U. *et al.* (2019) The multifunctional enzyme S-adenosylhomocysteine/methylthioadenosine nucleosidase is a key metabolic enzyme in the virulence of *Salmonella enterica* var Typhimurium. *Biochem. J.* 476, 3435–3453
76. Malaria Genomic Epidemiology Network (2019) Insights into malaria susceptibility using genome-wide data on 17,000 individuals from Africa, Asia and Oceania. *Nat. Commun.* 10, 5732
77. Gilchrist, J.J. *et al.* (2018) Genetic variation in VAC14 is associated with bacteremia secondary to diverse pathogens in African children. *Proc. Natl. Acad. Sci. U. S. A.* 115, E3601–E3603
78. Westermann, A.J. *et al.* (2017) Resolving host–pathogen interactions by dual RNA-seq. *PLoS Pathog.* 13, e1006033
79. Thänert, R. *et al.* (2017) Host-inherent variability influences the transcriptional response of *Staphylococcus aureus* during *in vivo* infection. *Nat. Commun.* 8, 14268
80. Pisu, D. *et al.* (2020) Dual RNA-Seq of Mtb-infected macrophages *in vivo* reveals ontologically distinct host–pathogen interactions. *Cell Rep.* 30, 335–350 e4
81. Stapels, D.A.C. *et al.* (2018) *Salmonella* persisters undermine host immune defenses during antibiotic treatment. *Science* 362, 1156–1160
82. Avital, G. *et al.* (2017) scDual-Seq: mapping the gene regulatory program of *Salmonella* infection by host and pathogen single-cell RNA-sequencing. *Genome Biol.* 18, 200
83. Kampmann, M. (2018) CRISPRi and CRISPRa screens in mammalian cells for precision biology and medicine. *ACS Chem. Biol.* 13, 406–416
84. Orchard, R.C. *et al.* (2019) Identification of antinorovirus genes in human cells using genome-wide CRISPR activation screening. *J. Virol.* 93, e01324-18
85. Jost, M. *et al.* (2020) Titrating gene expression using libraries of systematically attenuated CRISPR guide RNAs. *Nat. Biotechnol.* 38, 355–364
86. Xie, S. *et al.* (2018) Frequent sgRNA-barcode recombination in single-cell perturbation assays. *PLoS One* 13, e0198635
87. Replogle, J.M. *et al.* (2020) Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nat. Biotechnol.* 38, 954–961

88. de Bono, M. and Bargmann, C.I. (1998) Natural variation in a neuropeptide Y receptor homolog modifies social behavior and food response in *C. elegans*. *Cell* 94, 679–689
89. Chang, H.C. *et al.* (2011) Natural polymorphisms in *C. elegans* HECW-1 E3 ligase affect pathogen avoidance behaviour. *Nature* 480, 525–529
90. Martin, N. *et al.* (2017) Natural genetic variation in the *Caenorhabditis elegans* response to *Pseudomonas aeruginosa*. *G3 (Bethesda)* 7, 1137–1147
91. Glater, E.E. *et al.* (2014) Multigenic natural variation underlies *Caenorhabditis elegans* olfactory preference for the bacterial pathogen *Serratia marcescens*. *G3 (Bethesda)* 4, 265–276
92. Lazzaro, B.P. *et al.* (2004) Genetic basis of natural variation in *D. melanogaster* antibacterial immunity. *Science* 303, 1873–1876
93. Lazzaro, B.P. *et al.* (2006) Genetic variation in *Drosophila melanogaster* resistance to infection: a comparison across bacteria. *Genetics* 174, 1539–1554
94. Hill-Burns, E.M. and Clark, A.G. (2009) X-linked variation in immune response in *Drosophila melanogaster*. *Genetics* 183, 1477–1491
95. Mackay, T.F. *et al.* (2012) The *Drosophila melanogaster* genetic reference panel. *Nature* 482, 173–178
96. Chapman, J.R. *et al.* (2020) The genetic basis of natural variation in *Drosophila melanogaster* immune defense against *Enterococcus faecalis*. *Genes (Basel)* 11, 234
97. Bou Sleiman, M.S. *et al.* (2015) Genetic, molecular and physiological basis of variation in *Drosophila* gut immunocompetence. *Nat. Commun.* 6, 7829
98. Wang, J.B. *et al.* (2017) The genetic basis for variation in resistance to infection in the *Drosophila melanogaster* genetic reference panel. *PLoS Pathog.* 13, e1006260
99. Balik-Meisner, M. *et al.* (2018) Population genetic diversity in zebrafish lines. *Mamm. Genome* 29, 90–100
100. Bartha, I. *et al.* (2013) A genome-to-genome analysis of associations between human genetic variation, HIV-1 sequence diversity, and viral control. *eLife* 2, e01123
101. Ansari, M.A. *et al.* (2017) Genome-to-genome analysis highlights the effect of the human innate and adaptive immune systems on the hepatitis C virus. *Nat. Genet.* 49, 666–673
102. Lees, J.A. *et al.* (2019) Joint sequencing of human and pathogen genomes reveals the genetics of pneumococcal meningitis. *Nat. Commun.* 10, 2176
103. Walch, P. *et al.* (2020) Global mapping of *Salmonella enterica*–host protein–protein interactions during infection. *bioRxiv* Posted May 5, 2020. https://doi.org/10.1101/2020.05.04.075937
104. D'Costa, V.M. *et al.* (2019) BioID screen of *Salmonella* type 3 secreted effectors reveals host factors involved in vacuole positioning and stability during infection. *Nat. Microbiol.* 4, 2511–2522
105. Crofts, A.A. *et al.* (2018) *Campylobacter jejuni* transcriptional and genetic adaptation during human infection. *Nat. Microbiol.* 3, 494–502
106. Thye, T. *et al.* (2010) Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. *Nat. Genet.* 42, 739–741
107. Thye, T. *et al.* (2012) Common variants at 11p13 are associated with susceptibility to tuberculosis. *Nat. Genet.* 44, 257–259
108. Mahasirimongkol, S. *et al.* (2012) Genome-wide association studies of tuberculosis in Asians identify distinct at-risk locus for young tuberculosis. *J. Hum. Genet.* 57, 363–367
109. Sveinbjornsson, G. *et al.* (2016) HLA class II sequence variants influence tuberculosis risk in populations of European ancestry. *Nat. Genet.* 48, 318–322
110. Chimusa, E.R. *et al.* (2014) Genome-wide association study of ancestry-specific TB risk in the South African Coloured population. *Hum. Mol. Genet.* 23, 796–809
111. Png, E. *et al.* (2012) A genome wide association study of pulmonary tuberculosis susceptibility in Indonesians. *BMC Med. Genet.* 13, 5
112. Sobota, R.S. *et al.* (2016) A locus at 5q33.3 confers resistance to tuberculosis in highly susceptible individuals. *Am. J. Hum. Genet.* 98, 514–524
113. Luo, Y. *et al.* (2019) Early progression to active tuberculosis is a highly heritable trait driven by 3q23 in Peruvians. *Nat. Commun.* 10, 3765
114. Zhang, F.R. *et al.* (2009) Genomewide association study of leprosy. *N. Engl. J. Med.* 361, 2609–2618
115. Zhang, F. *et al.* (2011) Identification of two new loci at IL23R and RAB32 that influence susceptibility to leprosy. *Nat. Genet.* 43, 1247–1251
116. Wong, S.H. *et al.* (2010) Leprosy and the adaptation of human toll-like receptor 1. *PLoS Pathog.* 6, e1000979
117. Wang, Z. *et al.* (2016) A large-scale genome-wide association and meta-analysis identified four novel susceptibility loci for leprosy. *Nat. Commun.* 7, 13760
118. Liu, H. *et al.* (2015) Discovery of six new susceptibility loci and analysis of pleiotropic effects in leprosy. *Nat. Genet.* 47, 267–271
119. Moreau, K. *et al.* (2018) Human genetic susceptibility to native valve *Staphylococcus aureus* endocarditis in patients with *S. aureus* bacteremia: genome-wide association study. *Front. Microbiol.* 9, 640
120. Nelson, C.L. *et al.* (2014) A genome-wide association study of variants associated with acquisition of *Staphylococcus aureus* bacteremia in a healthcare setting. *BMC Infect. Dis.* 14, 83
121. Ye, Z. *et al.* (2014) Genome wide association study of SNP-, gene-, and pathway-based approaches to identify genes influencing susceptibility to *Staphylococcus aureus* infections. *Front. Genet.* 5, 125
122. DeLorenze, G.N. *et al.* (2016) Polymorphisms in HLA class II genes are associated with susceptibility to *Staphylococcus aureus* infection in a white population. *J. Infect. Dis.* 213, 816–823
123. Brown, E.L. *et al.* (2015) Genome-wide association study of *Staphylococcus aureus* carriage in a community-based sample of Mexican-Americans in Starr County, Texas. *PLoS One* 10, e0142130
124. Kenyan Bacteraemia Study Group (2016) Polymorphism in a lincRNA associates with a doubled risk of pneumococcal bacteremia in Kenyan children. *Am. J. Hum. Genet.* 98, 1092–1100
125. Rautanen, A. *et al.* (2015) Genome-wide association study of survival from sepsis due to pneumonia: an observational cohort study. *Lancet Respir. Med.* 3, 53–60
126. Davila, S. *et al.* (2010) Genome-wide association study identifies variants in the CFH region associated with host susceptibility to meningococcal disease. *Nat. Genet.* 42, 772–776
127. Martinon-Torres, F. *et al.* (2016) Natural resistance to meningococcal disease related to CFH loci: Meta-analysis of genome-wide association studies. *Sci. Rep.* 6, 35842
128. Gilchrist, J.J. *et al.* (2018) Risk of nontyphoidal *Salmonella* bacteraemia in African children is modified by STAT4. *Nat. Commun.* 9, 1014
129. Divaris, K. *et al.* (2012) Genome-wide association study of periodontal pathogen colonization. *J. Dent. Res.* 91, 21S–28S
130. Rhodin, K. *et al.* (2014) Chronic periodontitis genome-wide association studies: gene-centric and gene set enrichment analyses. *J. Dent. Res.* 93, 882–890
131. Offenbacher, S. *et al.* (2016) Genome-wide association study of biologically informed periodontal complex traits offers novel insights into the genetic basis of periodontal disease. *Hum. Mol. Genet.* 25, 2113–2129
132. Dunstan, S.J. *et al.* (2014) Variation at HLA-DRB1 is associated with resistance to enteric fever. *Nat. Genet.* 46, 1333–1336
133. Mayerle, J. *et al.* (2013) Identification of genetic loci associated with *Helicobacter pylori* serologic status. *JAMA* 309, 1912–1920
134. Li, W. *et al.* (2014) Unraveling the complex genetic model for cystic fibrosis: pleiotropic effects of modifier genes on early cystic fibrosis-related morbidities. *Hum. Genet.* 133, 151–161