Challenges and Opportunities in Understanding Genetics of Fungal Diseases: Towards a Functional Genomics Approach

Mariolina Bruno,a Vasiliki Matzaraki,a,b Frank L. van de Veerdonk,a Vinod Kumar,a,b,c Mihai G. Netea,a,d,e

aDepartment of Internal Medicine, Radboudumc Center for Infectious Diseases (RCI), Radboud University Medical Center, Nijmegen, The Netherlands
bUniversity of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, The Netherlands
cNitte (Deemed to be University), Nitte University Centre for Science Education and Research (NUCSER), Medical Sciences Complex, Deralakatte, Mangalore, India
dDepartment for Genomics & Immunoregulation, Life and Medical Sciences Institute (LIMES), University of Bonn, Bonn, Germany
eDepartment of Genetics, University of Medicine and Pharmacy of Craiova, Craiova, Romania

ABSTRACT
Infectious diseases are a leading cause of morbidity and mortality worldwide, and human pathogens have long been recognized as one of the main sources of evolutionary pressure, resulting in a high variable genetic background in immune-related genes. The study of the genetic contribution to infectious diseases has undergone tremendous advances over the last decades. Here, focusing on genetic predisposition to fungal diseases, we provide an overview of the available approaches for studying human genetic susceptibility to infections, reviewing current methodological and practical limitations. We describe how the classical methods available, such as family-based studies and candidate gene studies, have contributed to the discovery of crucial susceptibility factors for fungal infections. We will also discuss the contribution of novel unbiased approaches to the field, highlighting their success but also their limitations for the fungal immunology field. Finally, we show how a systems genomics approach can overcome those limitations and can lead to efficient prioritization and identification of genes and pathways with a critical role in susceptibility to fungal diseases. This knowledge will help to stratify at-risk patient groups and, subsequently, develop early appropriate prophylactic and treatment strategies.

KEYWORDS
host immune response, fungal infections, polymorphisms, genetic predisposition, immunology, functional genomics

HUMAN GENETIC SUSCEPTIBILITY TO INFECTIOUS DISEASES

Although there has been tremendous progress in medical research and health care, infectious diseases remain a leading cause of morbidity and mortality worldwide (1). Ever-increasing global connectivity together with human demographics and environmental changes has contributed to the emergence of new infectious diseases, such as the recent pandemic with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (2), and the reemergence of existing ones, such as Candida auris infection (3). Human infectious diseases are characterized by an extensive variation in clinical phenotypes among individuals infected by the same agent, indicating that genetic and nongenetic factors determine this variation. Many genetic epidemiological studies in the last half century, ranging from observational studies to more sophisticated twin or segregation studies, pointed to the importance of host heritable factors in susceptibility to infectious diseases. One of the first discovered single-gene traits influencing susceptibility to infection was the sickle hemoglobin variant as a major resistance factor for malaria (4). Stronger evidence came from several early twin studies reporting higher concordance rates in monozygotic than in dizygotic twins for genetic susceptibility to various infectious diseases (5–9). Also, follow-up studies of adopted children in the late 1980s showed they had a markedly increased risk of death from an infectious
disease if one of the biological parents had died prematurely from an infectious cause rather than other causes, such as cancer or cardiovascular diseases (10).

Infectious pathogens, which elicit the host immune response, have long been recognized as the main source of evolutionary pressure (11, 12). Immune-related genes are the most abundant and diverse genes in the human genome (13), suggesting an evolutionary advantage of a varied immunological response to a wide range of infectious pathogens. The study of the genetic contribution to infectious diseases has undergone revolutionary advances over the last decades in line with the development of novel technologies in the field. Traditional linkage studies identified a few important candidate genes (14). With the advent of the genomic era, genome-wide association studies (GWASs) have identified numerous genetic loci in autoimmune diseases (15), but only with limited success in the field of genetics of infectious disease (16). High-throughput technologies and the generation of multi-omics data sets have enabled a powerful multilevel study of the genetics of complex diseases, including infectious disease, to offer a better understanding of the interplay between host, invading pathogen, and environment.

Here, we provide an overview of the available approaches for studying human genetic susceptibility in fungal infections, reviewing current methodological and practical limitations. We will also discuss the use of a systems genomics approach to understanding genetics and molecular pathways underlying the human host defense against fungal infections.

**BURDEN OF FUNGAL DISEASES ON GLOBAL HEALTH**

Human are constantly exposed to fungi: some are colonizing the human host, the so-called commensal fungi, and some are ubiquitous in the environment, the so-called environmental fungi. A fully functional host immune system has effective mechanisms for preventing severe fungal infections, but when the immune system fails, human-pathogenic fungi can cause potentially "opportunistic" life-threatening diseases (17). The burden of fungal diseases on global health is expanding in parallel with an increase in individuals with acquired immune deficiencies or those receiving immune suppressive therapies or myeloablative treatments (18). Human fungal infections cause more than 1.5 million deaths every year (19) and affect more than one billion individuals worldwide (20). The steady increase in the incidence of nosocomial invasive fungal infections has significantly contributed to health-related costs (21). Despite the increasing numbers and the recent outbreak of emerging C. auris infections (3), the impact of fungal diseases on human health still remains underestimated (22, 23). The majority of human fungal infections are caused by *Candida*, *Aspergillus*, and *Cryptococcus* spp. (19). These fungi are ubiquitous, but *Cryptococcus* and *Aspergillus* spp. are also environmental (24), whereas *Candida* spp. are commensal colonizers of mucocutaneous surfaces and the gastrointestinal tract (25).

The diagnosis of fungal infections can be problematic due to clinical challenges in fungal isolation and identification (26, 27). Therapeutic challenges are raised by the fact that no vaccines are yet available, current antifungal therapeutic options remain limited, and, on top of that, multidrug-resistant fungal species are arising (28). As a result, the mortality rate of patients with invasive fungal infection remains unacceptably high, reaching 40% to 50% (29). Risk factors to develop fungal infections have been well described (30–33), and certain high-risk groups of patients can be further classified according to specific risk scores, which include a large panel of clinical and laboratory parameters linked to disease susceptibility or clinical phenotype evolution (34–37). However, not all patients at risk develop fungal disease, and a large variability in clinical evolution has been reported among patients with the same predisposing factors. This observation suggests that human genetic variation plays roles in the susceptibility to fungal infections and in the severity of the outcome. Indeed, several monogenic disorders resulting in severe primary immunodeficiencies, as well as mutations and common polymorphisms in immune genes, have been associated with an increased susceptibility to mucosal and/or invasive fungal infections, which have been
reviewed elsewhere (38). Despite significant advances over the last few years in identifying genetic variations leading to immune imbalances, which lead to increased susceptibility to fungal infections, there are still many challenges to fully understand the genetic architecture of fungal infections. To overcome these challenges, a systems genomics approach has been followed to identify risk loci and molecular pathways underlying host immune defense and disease pathogenesis. By integrating multiple molecular data sets that reveal interindividual variability, it is possible to prioritize and identify genes and pathways with a critical role in susceptibility to fungal diseases. Ultimately, this knowledge will help to stratify at-risk patient groups and, subsequently, develop early appropriate prophylactic and treatment strategies against opportunistic fungal infections.

OVERVIEW OF HOST IMMUNE RESPONSE TO FUNGAL PATHOGENS

Opportunistic fungal infections are characterized by the interaction between the host, the invading fungus, and the environment, which is sustained by a complex and dynamic equilibrium of several interconnected factors. The microbiological and environmental factors taking part in this delicate interaction, such as the role of the commensal microbiome, the dynamic fungal morphological adaptations, and genomic mutations, are well reviewed elsewhere (39). Despite the differences in the pathogenesis of infection between environmental and commensal fungal species, there are several common host immune defense mechanisms. To infect the human host, the fungal pathogen must be able to overcome three levels of host defense; the first is a physical barrier that consists of the skin and mucosa. The second barrier, presented by the innate immune system, is largely dependent on the recognition of evolutionarily conserved fungal cell wall components (pathogen-associated molecular patterns [PAMPs]). These PAMPs are recognized by various pattern recognition receptors (PRRs) circulating, such as pentraxin-3 (PTX3) or mannose binding lectin (MBL), or present on the surface of innate immune cells, such as macrophages, monocytes, NK cells, and neutrophils. In particular, the mannan cell wall component is mainly recognized by the macrophage mannose receptor (MMR), the C-type lectin-like receptor dectin-2, and Toll-like receptor 4 (TLR4). TLR2 binds to phospholipomannan, and the dectin-1 receptor recognizes β-glucan. Coordinated engagement of PRRs and following intracellular signaling pathways mediated by several kinases and adaptor molecules, such as spleen tyrosine kinase (Syk) and caspase recruitment domain-containing protein 9 (CARD9), results in the activation of innate immune effector mechanisms. Those mechanisms include phagocytosis, generation of reactive oxygen species (ROS) by NADPH oxidase and reactive nitrogen species (RNS) by myeloperoxidase (MPO) that promote the killing of the fungus and, finally, to production of pro- and anti-inflammatory cytokines. Proinflammatory cytokines, such as interleukin-1 beta (IL-1β) and tumor necrosis factor alpha (TNF-α), have important roles in the host defense against fungal infections. IL-1β is transcribed as an inactive form (pro-IL-1β) and further processed into its active mature form via the NLRP3 inflammasome, a multiprotein complex, which is also crucial for antifungal host defense (40). TNF-α enhances antifungal activities by promoting phagocytosis and neutrophil recruitment (41, 42). In turn, the release of cytokines, combined with antigen presentation activity of myeloid cells, is crucial for activation of the adaptive T-cell immunity, in particular, Th1 and Th17 subsets (43), representing a third longer-term barrier against fungal infection (44). Gamma interferon (IFN-γ) produced by Th1 lymphocytes had been shown to have a central role in the resistance against systemic fungal infections (43); Th17 responses have been proven to be crucial for human anti-Candida mucosal host defense and granulocyte recruitment, but they can contribute to detrimental immunopathology during fungal infections (45, 46). In an immunocompetent host, the majority of the invading microorganisms are detected and destroyed within minutes or hours by the innate immune defense mechanisms. An overview of host immune responses against fungal infection is presented in Fig. 1. Invasive fungal infections are mainly found in patients with a weakened immune
system, either due to reduced cellular immune effector mechanisms or defects in epithelial barriers.

**APPROACHES TO STUDY THE GENETICS OF FUNGAL INFECTIONS**

Although the above-mentioned factors are important, they do not explain all infections, and only a minority of patients at risk will actually develop disease, suggesting the critical role of Genetics in determining disease susceptibility. Indeed, several approaches, from classical family-based and candidate-gene approaches, to novel ones, such as genome-wide association studies (GWASs) and integrative approaches, have attempted to decipher the genetic factors to mucosal and/or invasive fungal infections. An overview of these approaches is presented in Fig. 2.
Classical approaches. (i) Mendelian susceptibility to fungal infections: a family-based approach. Classical approaches, such as family-based approaches to study genetic factors, have captured rare mutations that confer a mendelian (monogenic) form of predisposition to fungal infections. Much of our understanding about genetic susceptibility to specific fungal pathogens has been achieved through family-based studies of certain rare primary immunodeficiencies presenting as clinical manifestation signs of a mucosal or invasive fungal infection (47, 48). A prototypical example is chronic granulomatous disease (CGD), a rare inherited disorder (frequency, ~1/200,000) caused by mutations in genes encoding four of five protein subunits of the phagocyte NADPH oxidase, namely, the X-linked CYBB gene (gp91phox) and the autosomal recessive CYBA (p22phox), NCF1 (p47phox), and NCF2 (p67phox) genes (49). Patients with CGD fail to produce ROS and suffer from recurrent life-threatening bacterial and fungal infections, especially invasive aspergillosis (IA) (50), accounting for one-third of all deaths in CGD patients (51). Notably, patients with mutations in the NCF4 (p40phox) gene do not develop IA, as they are still able to produce ROS (52).

Another example is myeloperoxidase (MPO) deficiency, which is the most common inherited phagocytic disorder (frequency, ~1/2,000) (53). The vast majority of MPO-deficient patients are asymptomatic; however, a complete enzymatic deficiency predisposes them to invasive candidiasis (54). More recently, CARD9 deficiency has emerged as an important and fungus-specific susceptibility factor for both mucosal and invasive fungal infections (55), without predisposing individuals to other infectious or noninfectious sequelae. More than 15 missense and nonsense mutations in the CARD9 gene (56) result in Th17 deficiency and altered dectin-1 signaling as well as defective neutrophil recruitment to certain anatomical sites, including the central nervous system (CNS) (57). A few inborn monogenic disorders that predispose individuals to invasive fungal infections (IFIs) (but not fungal specific) were previously described: specific mutations in the transcription factor GATA2 cause the so-called “MonoMAC syndrome” characterized by monocytopenia, B-cell and natural killer (NK)-cell lymphopenias, myelodysplasia, and increased susceptibility not only to mycoses but also to papillomaviruses and nontuberculous mycobacteria (NTM) of low virulence potential (58). Genetic mutations in genes involved in the IL-12/IFN-γ signaling pathway, extensively reviewed in reference 29, have
been shown to predispose individuals not only to NTM but also to fungal infections by intracellular fungi (59). Such fungal infections include especially those whose eradication relies on an effective interaction between phagocytes and Th1 lymphocytes (e.g., *Histoplasma capsulatum*, *Paracoccidioides brasiliensis*, and *Cryptococcus neoformans*) (59, 60).

An intact host mucocutaneous barrier depends on functional IL-17 signaling. Chronic mucocutaneous candidiasis (CMC) is another primary immunodeficiency characterized by recurrent or persistent skin, mucosal, or nail infections by *Candida* spp., mainly *C. albicans*. CMC refers to a heterogeneous group of disorders, all caused by impaired Th17 responses and subsequent defective mucosal and skin antifungal host defense mechanisms. CMC can be caused by direct mutation in IL-17F and IL-17RC signaling resulting in mucosal but not systemic candidiasis, such as IL-17F and IL-17RC, which are specific for CMC, as well as IL-17RA and the adaptor ACT1 (TRAF3IP2), also predisposing individuals to bacterial infections (61–64). Other genetic mutations in several genes variously involved in Th17 differentiation can be causal for CMC and are generally associated with other syndromic manifestations. Such examples include the loss-of-function STAT3 mutation which causes hyper-IgE syndrome (65), biallelic mutations of the Th17 differentiation master regulator RORC (66), autosomal dominant STAT1 mutations, which lead to defective Th17 responses by indirectly impairing STAT3 activity (67, 68), and CARD9 mutations (56). Other CMC-associated monogenic diseases include (but are not limited to) autosomal recessive DOCK8 deficiency (69), X-linked severe combined immunodeficiency disorder (SCID), 22q11.2 deletion (athymic DiGeorge syndrome), and diseases from mutations of many other genes, nicely reviewed elsewhere (70). Interestingly, autoimmune polyendocrinopathy candidiasis ectodermal dystrophy (APECED) syndrome, caused by AIRE mutations and characterized by the presence of neutralizing autoantibodies against IL-17F and IL-22, presents CMC as the sole infectious consequence (71). To sum up, primary immunodeficiencies offer unique opportunities for a better understanding of the genetic and immunological components of fungal infections, which will help to develop novel immune-based therapeutic approaches against these infections.

(ii) Nonmonogenic susceptibility to fungal infections: a candidate gene approach.

Another classical approach widely adopted in several genetic studies of complex diseases, including fungal diseases, is the candidate gene approach. The selection of the candidate genes usually relies on in vitro murine or patient’s experimental data by hypothesis-driven biological plausibility. The majority of candidate gene studies include a case-control design. To avoid any form of confounding and population heterogeneity, case and controls need to be accurately matched, and the sample size should be adequate to ensure reproducibility and statistical power (72). The vast majority of candidate gene studies for susceptibility to fungal infections have focused on immune-related genes involved in innate recognition of microbes, acquired immunity, intracellular signaling pathways, or different cytokines. Immune-related genes are a special case in the genome because, depending on the geographic region, the selective pressure on them has been different; that is the reason why most of those genes are highly polymorphic and, subsequently, highly prone to population stratification biases (73).

Several single nucleotide polymorphisms (SNPs) in immune-related genes have been described that increase or decrease the risk of fungal diseases in patients with an acquired immunocompromised status (74–76). Two of the most studied pathological conditions characterized by an immunocompromised status are systemic candidiasis in the intensive care unit (ICU) and invasive aspergillosis (IA) in allogeneic hematopoietic stem cell transplant (HSCT) recipients, and most studies that identified SNPs associated with fungal infections have been conducted with these kinds of patients.

Since other excellent recent reviews already described in more detail SNPs influencing susceptibility to fungal infections (74–76), here and in Table 1 we report only some representative associations which have been described in the last 14 years supported by strong functional evidences. One of the most studied immune gene families that encode receptors on innate immune cells that recognize fungal antigens is that of the Toll-like receptors (TLRs). Three SNPs in the *TLR1* gene were significantly associated
with candidemia susceptibility (77), while SNPs in the TLR4 gene were associated with both IA (78, 79) and systemic candidiasis (80). A stop codon in DECTIN1 (Tyr238X) has been associated with increased risk for IA after HSCT (81) but not for invasive candidiasis after HSCT (82). The same stop codon polymorphism was further associated with mucocutaneous fungal infections (83) and oral and gastrointestinal colonization by Candida species in HSCT patients (82). Two frequent polymorphisms (281A/G and 734A/C) in the PTX3 gene have been associated with increased risk of developing IA in both HSCT donors (84) and solid organ transplant recipients (85). These SNPs have also been functionally validated using in vitro studies with patients’ primary neutrophils, showing impaired Aspergillus phagocytosis and killing (84). SNPs in the NOD2 gene regulate susceptibility to IA after HSCT, and NOD2 deficiency affords resistance to IA (86). In addition, genetic variation for the monocyte/macrophage-targeted chemokine receptor CX3CR1 and the neutrophil-targeted chemokine receptor CXCR1 has been shown to be crucial for fungal infections, particularly those caused by Candida spp.: carrying the allele M280 in the CX3CR1 gene in homozygosity was associated with an increased risk for disseminated candidiasis.

### TABLE 1

<table>
<thead>
<tr>
<th>Gene(s)</th>
<th>Polymorphism(s)</th>
<th>Chromosome location</th>
<th>Reported associations</th>
<th>Functional evidence</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>TLR1</td>
<td>rs4833095, rs5743618, rs5743611</td>
<td>4p14</td>
<td>Increased candidemia susceptibility</td>
<td>Impaired cytokine release by primary monocytes</td>
<td>77</td>
</tr>
<tr>
<td>TLR4</td>
<td>rs4986790, rs4986791</td>
<td>9q33.1</td>
<td>Increased susceptibility to IA</td>
<td>Delayed immune cell reconstitution after HSCT (78)</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Validation study in a separate cohort</td>
<td>79</td>
</tr>
<tr>
<td>CLEC7A</td>
<td>rs16910526</td>
<td>12p13.2</td>
<td>Increased candidemia susceptibility</td>
<td>Increased C. albicans-induced IL-10 in PBMCs</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Increased susceptibility to IA</td>
<td>Diminished A. fumigatus-induced IFN-γ and IL-10 in PBMCs</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Decreased A. fumigatus-induced IL-10 and IL-1β in PBMCs</td>
<td>82</td>
</tr>
<tr>
<td>PTX3</td>
<td>rs2305619, rs3816527</td>
<td>3q25.32</td>
<td>Increased susceptibility to IA after HSCT</td>
<td>Lower phagocytosis efficiency and A. fumigatus killing in neutrophils</td>
<td>84</td>
</tr>
<tr>
<td>NOD2</td>
<td>rs2066842</td>
<td>16q12.1</td>
<td>Reduced susceptibility to IA after HSCT</td>
<td>Lower A. fumigatus-induced cytokine production in PBMCs</td>
<td>86</td>
</tr>
<tr>
<td>CX3CR1</td>
<td>rs3732378</td>
<td>3p22.2</td>
<td>Increased candidemia susceptibility</td>
<td>C. albicans-induced renal failure in Cx3cr1−/− mice</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Diminished AKT and ERK signaling and decreased blood monocyte counts</td>
<td>89</td>
</tr>
<tr>
<td>CXCR1</td>
<td>rs2234671</td>
<td>2q35</td>
<td>Increased candidemia susceptibility</td>
<td>Impaired C. albicans killing and neutrophil degranulation</td>
<td>90</td>
</tr>
<tr>
<td>CLEC1A</td>
<td>rs2306894</td>
<td>12p13.2</td>
<td>Increased susceptibility to IA after HSCT</td>
<td>Lower A. fumigatus-induced IL-1β and IL-8 production in macrophages</td>
<td>14</td>
</tr>
<tr>
<td>IL1B</td>
<td>rs16944</td>
<td>2q14.1</td>
<td>Increased invasive mold infection</td>
<td>Reduced Aspergillus-induced IL-1β, TNF-α, and IL-22 production in PBMCs</td>
<td>91</td>
</tr>
<tr>
<td>IL1RN</td>
<td>rs419598</td>
<td>2q14.1</td>
<td>Increased invasive mold infection</td>
<td>Reduced Aspergillus-induced IL-1β and NF-κB production in PBMCs</td>
<td>91</td>
</tr>
<tr>
<td>IFNG</td>
<td>rs2069705</td>
<td>12q15</td>
<td>Decreased susceptibility to IA after HSCT</td>
<td>Improved Aspergillus killing and higher IFN phytohemagglutinin-induced IFN-γ production in PBMCs</td>
<td>144</td>
</tr>
</tbody>
</table>
but not mucosal or recurrent vulvovaginal candidiasis (RVVC) (87) in two different patient cohorts (88) and leads to impaired human monocyte trafficking and survival (89). The mutant CXCR1-T276 allele was associated with increased susceptibility to disseminated candidiasis and impaired neutrophil degranulation and fungal killing capacity (90). Last, but not least, genetic variation in pro- and anti-inflammatory cytokines has also been shown to be associated with susceptibility to fungal diseases. An important example is represented by IL-1 family genes: polymorphisms or certain haplotypes in IL-1β, IL-1α, and IL-1Ra were associated with an increased risk of developing IA in solid organ recipients (91) and in leukemic patients (92) as well as decreased Aspergillus fumigatus-induced cytokine production (91). Candidate gene studies have historically paved the way for personalized medicine and prophylactic anti-fungal treatment in high-risk patients. However, these studies present limitations, such as, to name a few, population stratification issues, lack of replication among different studies and across populations, poor functional evidence, and noncorrection for multiple testing as well as small sample size, which result in limited statistical power (73, 75).

"Novel" approaches. For decades, studies of genetic susceptibility to infectious diseases have been looking at inherited monogenic defects causing spontaneous infections and have been screening for single polymorphisms in candidate genes. However, such studies were performed in relatively small patient cohorts and were usually based on hypothesis-driven in vitro or previous knowledge in the field.

(i) Moving to unbiased genome-wide approaches to study genetics of fungal infections. The advent of the genomic era with advances, such as the mapping of human genetic variation compiled by the international HapMap project (93) and the 1000 Genomes project (94), and the development of several high-throughput sequencing (HTS) platforms for (deep) sequencing and imputation tools have all contributed to a better understanding of genetics in various human complex diseases in diverse populations. Such advances have also been applied to fungal infections. For example, next-generation sequencing (NGS) and whole-exome sequencing (WES), which sequences all of the protein-coding regions of genes in a genome, have become among the most widely used, unbiased, “hypothesis-generating” novel methods for studying the rare monogenic defects underlying susceptibility to fungal infections. For example, van der Veerdonk et al. identified the STAT1 gene as a cause of chronic mucocutaneous candidiasis by using an NGS approach (68), and this was validated by Liu et al. who identified heterozygous germ line mutations in the STAT1 gene in 47 patients with autosomal dominant chronic mucocutaneous candidiasis by using WES (95). WES in a case of a leukemic patient presenting an unusual invasive mucormycosis has revealed several putative polymorphisms in immune-related genes (e.g., PTX3, TLR6, NOD2, DDX58 (RIG-I), and CCR5) potentially influencing mucormycosis infection (96). Moreover, exome sequencing has been implemented as a discovery tool for genetic diagnosis of primary immunodeficiencies (PIDs) manifested as fungal infections described in a Dutch hospital (97). Collectively, these studies show that WES is a promising and affordable approach for discovering novel disease-causing genes and allelic polymorphisms influencing disease susceptibility targeting a small number of individuals or even single patients. In addition, sequencing of just the exome of patients would allow identification of rare variants. Early studies using exon sequencing to identify rare variants in other infectious diseases, which were focused on the TLR4 gene in meningococcal disease and on five TLR genes in tuberculosis, showed an excess of rare (and some more frequent) coding changes in patients compared to controls (98, 99). Therefore, WES can potentially open up new avenues for discovering rare variants that predispose individuals to fungal infections.

However, the majority of low-frequency and/or rare variants that have been associated with infectious diseases, including systemic Candida infections, are noncoding variants (intronic or intergenic) (100, 101). To explore the role of common noncoding variants, follow-up studies on the genetics of fungal diseases made use of genomic tools, such as genotype imputation, custom genotyping arrays, and whole-genome sequencing, to reveal novel associations between phenotypes and variants. For example, a pilot association study performed a screen of ~120,000 SNPs across 186 genetic...
loci related to immune function among hospitalized patients with candidemia compared to those in healthy and patient-matched controls and revealed significant associations between novel SNPs in the CD58, TAGAP, and LCE4A-C1orf68 genes and candidemia susceptibility (101). Of note, the presence of two or more high-risk SNPs within these loci resulted in an ~20-fold increased risk of developing candidemia, indicating a possible synergistic effect on increasing the infection risk (101). A large GWAS of volunteers contributing DNA from the 23andMe database identified three significant associations between yeast infection and variants downstream of the PRKCH gene, within DSG1, and C1orf177 (102). Another pilot GWAS, which was performed in children with dermatophytosis caused by the fungal species Trichophyton tonsurans, identified SNPs in eight genes involved in leukocyte activation, melanocyte function, and extracellular matrix remodeling that have been significantly associated with increased infection rate (103). All these studies indicated the role of common variants in contributing to variability in the susceptibility to fungal diseases. Despite significant progress over the last few years in identifying susceptibility genes for fungal infections, there is still much genetic information unexplored, and the molecular mechanisms underlying susceptibility are not fully understood due to challenges that are discussed below.

(ii) Limitations of studying the genetics of fungal diseases. GWASs for identifying genetic risk factors in fungal infections have not been as successful as those in other complex diseases, such as autoimmune diseases (104), because of several limitations. One of the major limitations in studying the genetics of fungal infections is the lack of power due to relatively small patient cohorts. Large sample sizes are required in order to obtain sufficient statistical power to detect true disease associations (105). The collection of a patient cohort is also complicated by the possible presence of asymptomatic infections or different ethnicities. Patient cohorts must be ethnically homogeneous and well phenotyped in order to identify phenotype- and population-specific associations. Considering the genetic substructure of human populations, it is crucial to consider that the allele frequency differs substantially among ethnic groups, and in certain cases, for example, in the African ancestry, it is possible to find larger variation and a lower linkage disequilibrium (106). The admixture of ethnic groups (107) as well as subtle differences in the ethnic compositions of cases and controls (108) can lead to false-positive results. While careful matching of demographic factors can reduce the number of false-positive results, statistical methods (nicely reviewed in reference 107) can now be applied to address this issue and mitigate these caveats.

Another limitation is that most GWASs have been focused on identifying only common variants whose minor allele frequency (MAF) is >5% (105, 109, 110), missing low-frequency or rare variants. To identify rare variants, next-generation sequencing (NGS) studies with relatively small cohorts followed by testing of associated variants in larger cohorts might be a promising complementary strategy (110). After validating the SNP in a validation cohort, a “wet-lab” functional validation of the disease-causing effect of this genetic variant is critical and required to confirm the causal relationship between genetics and phenotype.

Another limitation is that a GWAS alone, while providing significant associations between a genetic variant and a disease, cannot explain the biological consequence or pinpoint the causal gene, especially when noncoding genetic variants are discovered (111). A possible approach for exploring the link between a GWAS genetic variant and its effect is to statistically correlate variants with measured biological quantitative data by performing quantitative trait loci (QTL) analysis. For example, a statistical correlation between a genetic variant to gene expression is called expression-QTL (eQTL) analysis (112), to cytokine production is cytokine-QTL (cQTL) analysis (113), to DNA methylation is methylation-QTL (meQTL) analysis (114), among others (115). Of note, eQTL and cQTL analyses have already been implemented for studying interindividual variability in cytokine production in response to fungal pathogens (113, 116). In particular, the GOLM1 gene was associated with C. albicans-induced IL-6 production, and a genetic variant within this locus was also associated with increased susceptibility to candidemia (113).
In addition, it is becoming increasingly clear that the outcome of an infectious disease reflects the dynamic interaction between humans, pathogen genotypes, and the environment (117). The host-fungal interaction exhibits features of a dynamic system that may exert genetic effects known as genotype-by-genotype (GxG) interactions (16). Those GxG interactions led to a slow host-pathogen coevolution (especially in cases of a commensal fungi such as *C. albicans*); this phenomenon might justify the host heterogeneity in the frequency of polymorphisms and haplotypes among populations (118). At the same time, the fungal pathogen can rapidly acquire mutations to adapt to host polymorphisms in a specific population, resulting in considerable genomic variation across fungi from different geographic regions (119–121). In turn, rapid pathogen evolution or host-pathogen coevolution might have caused a fluctuation over time of the disease susceptibility genes across populations, as mathematically modeled by Lambrechts et al. (122), and may have directly played a role in the limited success of GWASs on fungal (or, in general, infectious) disease susceptibility. Last but not least, classical GWASs can detect only the genetic component of the three-way interplay between the host immune system, different pathogen morphotypes, and the environment. In particular, host and pathogen genetic variability interaction with environmental influences are even more challenging to model, and they can be collectively defined as gene-environment (GxE) interactions (123). For example, environmental factors such as pH and/or an imbalanced microbiome influence the susceptibility to develop recurrent vulvovaginal candidiasis (RVVC) (124). Specific interactions between commensal bacteria and fungi could play an important role in the development of invasive candidiasis (125). Therefore, a more integrative and multilevel analysis of host, pathogen, and environmental variation is required to account for all these interactions while studying the pathogenesis of a fungal disease.

(iii) Overcoming limitations: the introduction of functional genomics approaches. Given the complexity of host-pathogen interactions, conventional experimental approaches that study only individual molecular components (either of the host or pathogen) cannot provide a comprehensive picture of these interactions. The development of high-throughput data acquisition technologies and the possibility to integrate multi-omics data sets have laid the foundations for a new discipline: systems biology (126, 127). The increasing use of systems biology is tightly intertwined with that of functional genomics, which represents a novel, more powerful multilevel manner for studying the genetics of complex diseases (128, 129) (Fig. 2). Thus, the integration of high-throughput multi-omics data (transcriptomics, proteomics, metabolomics, lipidomics, etc.) with genetics can be used to prioritize genes for follow-up functional experiments to better understand their role in host immune defense and identify molecular pathways that underlie disease pathogenesis (Table 2). Several studies have applied a functional genomics approach to understand host genetic susceptibility to fungal infections, where genome-wide data (also called “static biomarkers” [129]) were integrated, validated, or complemented with other multi-omics data sets in the context of the disease, where host-pathogen interactions are dynamically changing. Table 3 lists the studies in the last 5 years that identified a genetic variant associated with fungal infections using a systems genomics approach.

A specific role of the type I interferon pathway in anti-*Candida* host defense was supported by integrating transcriptional analysis and functional genomics (130) using *Candida*-stimulated human immune cells. Of note, the importance of this pathway was validated through immunological and genetic studies in both healthy volunteers and patients with systemic candidiasis or suffering from CMC. Moreover, polymorphisms in type I interferon genes modulated *Candida*-induced cytokine production, and they were correlated with susceptibility to systemic candidiasis (130). The first transcriptome-wide association study (TWAS) (131) of the fungal immunology field identified molecular pathways underlying candidemia susceptibility using unbiased transcriptomics data, which were then validated in a patient cohort. Significant associations between *CCL8*, *STAT1*, *PSMB8*, and *SP110* polymorphisms and susceptibility to candidemia were identified by integrating transcriptomics data, and a candidemia GWAS followed up by functional *in vitro* validation in the context
of Candida infection (130). Another study suggested that RIG-I-like receptor (RLR) MDA5 has a critical role in anti-Candida host immune defense by integrating genetic, transcriptomic, and immunological data generated from mouse and human studies (132). The additive value of integrating multiple molecular data sets became even more apparent by two follow-up studies where genes and pathways underlying candidemia susceptibility were prioritized. In the first study, suggestive genetic associations together with transcriptomic data prioritized novel pathways implicated in candidemia susceptibility, including the complement and hemostasis pathways (100). In the second study, integration of GWAS data with variants that affect cytokine levels (cytokine-QTLs) from different Candida-stimulated cell types prioritized lipid and arachidonic acid metabolism as potential mechanisms that affect monocyte-derived cytokines to influence susceptibility to candidemia (133).

Although African populations suffer the most from infectious diseases, they are still underrepresented in studies of disease susceptibility (117). The first genome-wide association study of susceptibility to cryptococcosis in HIV patients was carried out with genotype data from 524 patients of African descent. This study identified six loci upstream of the CSF1 gene (encoding macrophage colony-stimulating factor [M-CSF]) that were significantly associated with the disease susceptibility and validated in a separate cohort. Functional data from RNA sequencing (RNA-seq) of human peripheral blood mononuclear cells (PBMCs) stimulated with C. neoformans and in vitro experiments with HIV patients’ PBMCs confirmed the crucial role of M-CSF for anti-Cryptococcus host defense mechanisms (134).

Given that genetic variants significantly associated with a disease are often regulated in a context- and cell-specific way (135), with the development of single-cell RNA-seq, it has become possible to prioritize genes in a cell-type-specific fashion. For example, by combining bulk and single-cell transcriptome data in response to Candida stimulation with GWAS data on candidemia susceptibility, the LY86 antigen has been prioritized and further validated to exert a protective role against candidemia risk (136). Furthermore, genes and cellular processes that contribute to the pathogenesis of RVVC, including cellular morphogenesis and metabolism, and cellular adhesion were identified through the integration of genomic approaches and immunological studies in two independent cohorts of patients with RVVC and healthy individuals (137). In particular, the role of SIGLEC15, a lectin expressed by various immune cells that binds sialic acid, in Candida recognition and RVVC susceptibility was also validated in the same study with both in vitro and in vivo functional assays (137). Wang et al. in the Hi-HOST Phenome Project (H2P2) identified two SNPs significantly associated with
basic fibroblast growth factor (FGF2) production in response to Mucor circinelloides and C. albicans, posing those allelic variants as potential candidates for the antifungal host immune response (138). However, they did not validate whether and how the presence of these SNPs are associated with an increased risk of fungal infections.

Overall, such an integrative functional approach is valuable in the context of not only fungal infections but also other infectious diseases, for which the limited size of patient cohorts limits the power of the GWAS. The reasons for using such an approach are 3-fold. First of all, this approach makes use of large-population-based cohort studies in the context of the disease that can be excellent models to get a powerful analysis to understand disease pathophysiology. Second, it is very versatile and provides independent layers of evidence intersecting with each other: from the multi-omics untargeted molecular candidate to the experimental or clinical evidence (top down) and vice versa (bottom up) in a multidisciplinary and collaborative way. Last but not least, the ultimate aim of functional genomic studies is to provide “actionable data” with translational potential. Since this can be a pathway-based targeted approach, it is possible to validate and clinically translate those findings to patients. Knowing the underlying pathways of human host defense allows, for example, for the identification of ways to prevent the disease, the development of novel diagnostic tools to be used in patient risk stratification, and the identification of new potential therapeutics. For a robust implementation of such a host-oriented therapy, it is particularly crucial to make sure that the results are validated in a physiologically relevant model, preferably relevant primary models of disease or appropriate patient samples or clinical strains. It has been shown that not always what has been validated in human cell lines (139), in mice (140), or in a laboratory pathogen strain (141, 142) holds true in patient’s cells or fungal clinical isolates.

**FUTURE PERSPECTIVES**

Over the last decades, the study of the genetics of infectious disease susceptibility has been revolutionized, and it has been developed more rapidly thanks to new technologies. This progress was important at multiple levels. First, it has made the research process more effective, comprehensive, and productive, providing valuable new findings on host-pathogen interactions. Such an evolution of the field combined with an interdisciplinary approach

---

<table>
<thead>
<tr>
<th>Gene</th>
<th>Polymorphism(s)</th>
<th>Chromosome location</th>
<th>Reported association</th>
<th>Functional validation/evidence</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOLM1</td>
<td>rs11141235</td>
<td>9q21</td>
<td>Increased candidemia susceptibility</td>
<td>cQTL locus: lower C. albicans-induced IL-6 production</td>
<td>113</td>
</tr>
<tr>
<td>IFIH1</td>
<td>rs1990760, rs3747517</td>
<td>2q24.2</td>
<td>Increased candidemia susceptibility</td>
<td>Reduced C. albicans-induced IL-10 in PMCs</td>
<td>132</td>
</tr>
<tr>
<td>MAP3K8</td>
<td>rs1360119</td>
<td>10p11.23</td>
<td>Increased candidemia susceptibility</td>
<td>Reduced IL-6, IL-8 and IFN-γ in serum of candidemia patients</td>
<td>100</td>
</tr>
<tr>
<td>SPTBNS (eQTL of PLA2G4B)</td>
<td>rs8028958</td>
<td>15q15.1</td>
<td>Increased candidemia susceptibility</td>
<td>Lower C. albicans-induced IL-6 and ROS in PBMCs</td>
<td>133</td>
</tr>
<tr>
<td>LY86</td>
<td>rs9405943</td>
<td>6p25.1</td>
<td>Increased candidemia susceptibility</td>
<td>Lower migration towards MCP-1 of monocytes with knockdown of LY86</td>
<td>136</td>
</tr>
<tr>
<td>SIGLEC15</td>
<td>rs2919643</td>
<td>18q21.1</td>
<td>Increased RVVC susceptibility</td>
<td>Increased C. albicans-induced IL-17A, IL-22 and IFN-γ</td>
<td>137</td>
</tr>
<tr>
<td>MFHAS1</td>
<td>rs139408032</td>
<td>8p23.1</td>
<td>NA</td>
<td>cQTL locus: higher M. circinelloides-induced FGF-2 production</td>
<td>138</td>
</tr>
<tr>
<td>FRMD4A</td>
<td>rs61836093</td>
<td>10p13</td>
<td>NA</td>
<td>cQTL locus: higher C. albicans-induced FGF-2 production</td>
<td>138</td>
</tr>
<tr>
<td>CSF1</td>
<td>rs1999713</td>
<td>1p13</td>
<td>Decreased cryptococcosis susceptibility in HIV patients</td>
<td>Upregulation of CSF1 upon C. neoformans stimulation of human PBMCs; higher phagocytosis and killing of C. neoformans in PBMCs from HIV patients pretreated with M-CSF</td>
<td>134</td>
</tr>
</tbody>
</table>

*NA, not applicable.*
can be a useful tool to identify potential novel therapeutic drug targets. In this respect, a recent study has shown that the proportion of drug mechanisms with direct genetic support increases significantly across the drug development phases, indicating that prioritizing a genetically supported drug target could double the success rate in drug discovery (143). A stratification of patients based on genetic profiling would pinpoint the patients with high risk of disease and who will benefit most from the drug. Unless additional clinical trials provide evidence of a treatment effect based on genetic profiling, we should be aware and cautious of the benefits and harms of new drug targets. In addition, a host-directed therapeutic target may also result in a weaker selection pressure on pathogens, potentially making it more difficult for a pathogen to evolve beyond the control of the host immune response.

Integrating such a plethora of omics data would catalyze the identification of diagnostic markers that might be useful for severity stratification or eligibility for specific treatments. Considering the host variability in immune-related genes, personalized medicine based on an individual’s genetic profile, such as immunotherapy-based interventions or targeted antifungal prophylaxis in genetically susceptible individuals, are leading to increasingly more powerful precision medicine. Nonetheless, risk stratification approaches guiding the clinical decision-making process based on a patient’s individual susceptibility profile are expected to be promising. From a more basic science and biotechnological aspect, new technologies are gaining ground in the study of the genetics of infectious diseases, such as single-cell sequencing at the transcriptome level and whole-genome sequencing for the primary immunodeficiencies at the genomic level.

It is expected that in the coming years, novel technologies that will help dissect the interaction of host genetics and the metagenomic (microbiome and also mycobiome) make-up of an individual will be further integrated, as an increasing number of studies will investigate these complementary genomes of an individual. Some of the available technologies that can potentially be implemented and integrated with the genetic level are the organ-on-chips approach, which would allow us to better dissect the human-fungus-environment interaction in a more dynamic manner, which is more comparable to human physiology. These novel tools in a systems genomic approach framework will also be used to decipher the pathophysiology of emerging fungal infections (e.g., Candida auris). In addition to this, such an approach needs to be employed more extensively in populations of non-European ancestry.

ACKNOWLEDGMENTS

We thank Diletta Rosati for the help with the figure realization.

F.L.V.D.V. was supported by a Vidi grant of the Netherlands Association for Scientific Research, the Europeans Union’s Horizon 2020 research and innovation program under grant agreement number 847507, HDM-FUN, and the La Caixa foundation (identifier [ID] 100010434). M.G.N. was supported by an ERC Advanced Grant (number 833247), a Spinoza Grant of the Netherlands Organization for Scientific Research, and a Competitiveness Operational Program Grant of the Romanian Ministry of European Funds (FUSE).

REFERENCES


August 2021 Volume 89 Issue 8 e00005-21 iai.asm.org 14

Minireview Infection and Immunity


Mariolina Bruno obtained her M.D. from Sapienza University of Rome (Italy) and received a degree in Life Sciences from the Sapienza School from Advances Studies (SSAS). She is currently a final-year Ph.D. candidate in the Department of Internal Medicine of Radboudumc in Nijmegen (The Netherlands). She has been working on a project aimed at defining how *C. auris* is recognized by immune cells. She is currently investigating host susceptibility factors for *Aspergillus* infection in patients with chronic pulmonary aspergillosis (CPA) and chronic granulomatous disease (CGD), with a particular focus on immunometabolism.

Mihai G. Netea was born and studied medicine in Cluj-Napoca (Romania). He completed his Ph.D. at the Radboud University Nijmegen (The Netherlands) on studies investigating the cytokine network in sepsis. After working as a postdoc at the University of Colorado, he returned to Nijmegen where he finished his clinical training as an infectious diseases specialist and where he currently heads the division of Experimental Medicine, Department of Internal Medicine, Nijmegen University Nijmegen Medical Center. His main research interests are pattern recognition of fungal pathogens and the induction of antifungal immunity, primary immunodeficiencies in innate immune system, and the study of the memory traits of innate immunity.