Identifying Host Genetic Risk Factors in the Context of Public Health Surveillance for Invasive Pneumococcal Disease

Jairam R. Lingappa1,*, Logan Dumitrescu2,3, Shanta M. Zimmer4, Ruth Lynfield5, Janet M. McNicholl6, Nancy E. Messonnier6, Cynthia G. Whitney6, Dana C. Crawford2,3

1 Departments of Global Health, Medicine and Pediatrics, University of Washington, Seattle, Washington, United States of America, 2 Center for Human Genetics Research, Vanderbilt University, Nashville, Tennessee, United States of America, 3 Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, Tennessee, United States of America, 4 Emory University, Atlanta, Georgia, United States of America, 5 Minnesota Department of Health, St. Paul, Minnesota, United States of America, 6 Centers for Disease Control and Prevention, Atlanta, Georgia, United States of America

Abstract

Host genetic factors that modify risk of pneumococcal disease may help target future public health interventions to individuals at highest risk of disease. We linked data from population-based surveillance for invasive pneumococcal disease (IPD) with state-based newborn dried blood spot repositories to identify biological samples from individuals who developed invasive pneumococcal disease. Genomic DNA was extracted from 366 case and 732 anonymous control samples. TagSNPs were selected in 34 candidate genes thought to be associated with host response to invasive pneumococcal disease, and a total of 326 variants were successfully genotyped. Among 543 European Americans (EA) (182 cases and 361 controls), and 166 African Americans (AA) (53 cases and 113 controls), common variants in surfactant protein D (SFTPD) are consistently underrepresented in IPD. SFTPD variants with the strongest association for IPD are intronic rs17786286 (allelic OR 0.45, 95% confidence interval (CI) [0.25, 0.82], with p = 0.007) in EA and 5′ flanking rs12219080 (allelic OR 0.32, 95% CI [0.13, 0.78], with p = 0.009) in AA. Variants in CD46 and IL1R1 are also associated with IPD in both EA and AA, but with effects in different directions; FAS, IL1B, IL4, IL10, IL12B, SFTPA1, SFTPB, and PTAFR variants are associated (p ≤ 0.05) with IPD in EA or AA. We conclude that variants in SFTPD may protect against IPD in EA and AA and genetic variation in other host response pathways may also contribute to risk of IPD. While our associations are not corrected for multiple comparisons and therefore must be replicated in additional cohorts, this pilot study underscores the feasibility of integrating public health surveillance with existing, prospectively collected, newborn dried blood spot repositories to identify host genetic factors associated with infectious diseases.

Citation: Lingappa JR, Dumitrescu L, Zimmer SM, Lynfield R, McNicholl JM, et al. (2011) Identifying Host Genetic Risk Factors in the Context of Public Health Surveillance for Invasive Pneumococcal Disease. PLoS ONE 6(8): e23413. doi:10.1371/journal.pone.0023413

Editor: Pieter H. Reitsma, Leiden University Medical Center, The Netherlands

Received January 20, 2011; Accepted July 16, 2011; Published August 15, 2011

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Funding: Genotyping services were provided by the Johns Hopkins University under U.S. Federal Government contract number N01-HV-48195 from the National Heart, Lung, and Blood Institute. Intramural support was also provided through the Centers for Disease Control and Prevention (CDC) Genetics Working Group. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: lingappa@u.washington.edu
† These authors contributed equally to this work.
Variants in SFTPD have the strongest and most consistent associations across EA and AA. SFTPD intronic variants rs17886286 and rs1990374 are underrepresented among EA IPD cases compared to controls (allelic OR 0.45 and 0.60, with p = 0.007 and 0.023, respectively). These variants are in moderate linkage disequilibrium (LD) with each other in EA (SeattleSNPs r² = 0.732) and thus do not represent independent associations. Also, these associations do not follow an additive genetic model given the observation that the genetic effect estimates (odds ratios) for heterozygotes and homozygotes are similar for each of these variants, individually (Table S2). Among AA, two different SFTPD variants (5' flanking rs2219080 and intronic rs17878441) are underrepresented in IPD cases compared to controls (allelic OR 0.32, and 0.36, with p = 0.009 and 0.020, respectively). These SNPs are not in high LD in AA (SeattleSNPs r² = 0.352). The association between SFTPD rs17878441 and IPD is specific to AA given that this SNP is very rare among the EA cohort presented here (1 heterozygote case and 1 heterozygote control among 508 individuals tested, overall MAF = 0.002) and monomorphic in reference populations such as CEU HapMap and SeattleSNPs EA. SFTPD rs1990374 and rs2219080 have consistent directions of effect in EA and AA, although neither is associated in both groups at p≤0.05 (Table 3 and Figure 1).

To explore the joint effects of associated variants in SFTPD, we calculated a weighted Genetic Risk Score (GRS) in EA and AA for significant associations with SFTPD variants (rs1990374 and rs17886286 in EA, and rs2219080 and rs17878441 in AA). Together, the two EA SFTPD variants were associated with IPD at an OR = 0.47 (95%CI: 0.26, 0.84) and p-value = 0.011. Additionally, the two AA SFTPD variants were associated with IPD at an

identified that may be critical in host response to IPD.[5] Public health surveillance for IPD, such as through ABCs, captures data for population-based case cohorts, but without prospective collection of host genetic samples. Most states separately implement population-based dried bloodspot collection from newborns (nDBS) for screening for specific inherited traits. Recently, we reported preliminary results of a pilot study to integrate ABCs invasive bacterial disease data with a state-based nDBS repository to identify genetic samples from cases and controls.[6] Here we use this approach to evaluate candidate host genetic variants as risk factors for IPD.

Results
tagSNPs associated with IPD

We performed single SNP allelic tests of association stratified by race/ethnicity in 543 European-Americans (EA), including 182 cases of IPD and 361 controls (Table 1) for 212 SNPs (Tables 2 and S1); and in 166 African-Americans (AA), including 53 cases and 113 controls for 287 SNPs. Because of the differences in sample size and its impact on power, we evaluated results based on both p-value and consistent direction of effect between the two ancestral groups. Overall, comparing IPD cases to controls, 17 tagSNPs in nine genes (CD46, SFTPA1, SFTPD, IL1B, ILIR1, ILA, IL10, IL12B, FAS) in EA and 11 tagSNPs in six genes (CD46, SFTPB, SFTPD, IL1B, ILIR1, and PTAFR) in AA have common variants that are associated at a liberal threshold of p≤0.05 (Tables 3 and S2, and Figure 1). Of these, three candidate genes, SFTPD, CD46 and ILIR1, have tagSNPs associated with IPD in both EA and AA; these associations are further described below.

Table 1. Study Population Characteristics.

| Characteristic | Surveys cohort | Genotype cohort | Analysis cohort |
|----------------|---------------|----------------|----------------|
|                | Cases (N = 366) | Controls (N = 732) | Cases (N = 330) | Controls (N = 637) | Cases (N = 272) | Controls (N = 546) |
| Demographics   |               |                |               |                    |                |               |
| Median Age in years (range) | 1.11 (0-3.2) | 1.11 (0-3.2) | 1.12 (0-3.2) | 1.11 (0-3.2) | 1.11 (0-3.2) | 1.09 (0-3.2) |
| Female¹ # (%) | 188 (51.4) | – | 171 (51.8%) | – | 144 (52.9%) | – |
| Clinical classification # (%) |               |                |               |                    |                |               |
| IPD            | 366 (100.0%) | – | 330 (100.0%) | – | 272 (100.0%) | – |
| Bacteremia only | 245 (67.0%) | – | 223 (67.6%) | – | 182 (66.9%) | – |
| Pneumonia²     | 53 (14.2%) | – | 43 (13.0%) | – | 37 (13.6%) | – |
| Meningitis²    | 29 (8.2%) | – | 28 (8.5%) | – | 23 (8.5%) | – |
| Other³         | 39 (10.6%) | – | 36 (11.0%) | – | 30 (11.0%) | – |
| Clinical Outcome – Death | 5 (1.3%) | – | 2 (0.6%) | – | 2 (0.7%) | – |
| Race⁴         |                |                |               |                    |                |               |
| European American | 265 (72.2%) | 534 (72.3%) | 240 (72.7%) | 452 (71.0%) | 182 (66.9%) | 361 (66.1%) |
| African American | 62 (16.9%) | 124 (16.8%) | 53 (16.1%) | 113 (17.7%) | 53 (19.5%) | 113 (20.7%) |
| American Indian | 9 (2.45%) | 18 (2.4%) | 9 (2.7%) | 17 (2.7%) | 9 (3.3%) | 17 (3.1%) |
| Asian          | 19 (5.18%) | 38 (5.1%) | 16 (4.8%) | 31 (4.9%) | 16 (5.9%) | 31 (5.7%) |
| Other          | 12 (3.27%) | 24 (3.2%) | 12 (3.6%) | 24 (3.8%) | 12 (4.4%) | 24 (4.4%) |
Table 2. Candidate Gene List.

| Candidate Gene (GenelD) | # Selected | # Genotyped | References |
|-------------------------|------------|-------------|------------|
| **Innate Immune response and signaling pathway genes** | | | |
| CD46 (4179) | 25 | 23 | [6,34,35] |
| CRP (1401) | 5 | 4 | [36-38] |
| FCRGR2A (2212) | 6 | 3 | [39,40] |
| FCRGR3A (2214) | 9 | 1 | [39,40] |
| IKBK4 (51135) | 16 | 15 | [41] |
| L196 (MD-2) (23643) | 24 | 24 | [42] |
| MBL2 (4153) | 10 | 10 | [43,44] |
| MyD88 (4615) | 3 | 2 | [45] |
| SFTPA1 (653509) | 5 | 4 | [46] |
| SFTPB (6439) | 9 | 7 | [47] |
| SFTPD (6441) | 23 | 20 | [46] |
| TLR2 (7097) | 11 | 10 | [42,48] |
| TLR9 (54106) | 7 | 3 | [49] |
| **Pro- and anti-inflammatory mediators** | | | |
| IFNG (3458) | 7 | 7 | [50] |
| IL1A (3552) | 13 | 12 | [51,52] |
| IL1B (3553) | 19 | 13 | [51] |
| IL1R1 (3554) | 35 | 33 | [53] |
| IL4 (3565) | 18 | 15 | [54] |
| IL6 (3569) | 11 | 8 | [55] |
| IL8 (3576) | 3 | 3 | [56,57] |
| IL10 (3586) | 7 | 7 | [58-60] |
| IL12A (3592) | 13 | 11 | [61,62] |
| IL12B (3593) | 9 | 9 | [61,62] |
| IL18 (3606) | 10 | 10 | [63-65] |
| LTA (4049) | 7 | 7 | [66] |
| TNFα (7124) | 5 | 4 | [58,67,68] |
| **Coagulation pathway genes** | | | |
| PROC (5624) | 14 | 5 | [69,70] |
| PTAFR (5724) | 8 | 6 | [71] |
| SERPINE1 (PAI1) (5054) | 9 | 9 | [72,73] |
| **Activation/apoptosis and immune regulatory genes** | | | |
| CD40 (958) | 11 | 10 | [74] |
| CD40LG (959) | 14 | 14 | [74] |
| CTLA4 (1493) | 4 | 3 | [75] |
| FAS (355) | 7 | 7 | [76] |
| FASL (356) | 7 | 7 | [76,77] |
| **All 34 candidate genes** | 384 | 326 | |
Discussion

In our evaluation of a population-based cohort for host genetic variation associated with IPD in children, we identified 27 tagSNPs in 11 genes (CD46, SFTPA1, SFTPB, SFTPD, IL1B, IL1R1, IL4, IL10, IL12B, FAS and PTFAR) associated in EA or AA at a liberal significance threshold of \( p \leq 0.05 \). In particular, in EA and AA, variants in the surfactant protein D (SP-D) encoded by SFTPD (gene ID 6441) are consistently underrepresented in IPD and pneumococcal bacteremia cases compared to controls, suggesting that variants in this gene or those in linkage disequilibrium may confer protection from IPD. This is the first study linking SFTPD gene variation specifically with clinical IPD.

SP-D is a member of the collectin subgroup in the C-type lectin superfamily including surfactant protein A (SP-A) and mannose binding protein. SP-D and SP-A are found primarily in the respiratory tract and other mucosal surfaces and recent data suggests that they impact respiratory infections on multiple levels. Surfactant collectins broadly bind carbohydrates and lipids on the surface of bacteria and viruses, with specific binding of SP-D to S. pneumoniae.
SP-D deficient (sftpd<sup>-/-</sup> knockout) mice are associated with persistent pneumococcal colonization, decreased clearance of bacterial pathogens, and early onset and increased levels of <i>S. pneumoniae</i> bacteremia in colonized mice.[7] Overall, collectins exhibit both pro- and anti-inflammatory effects[8]: SP-D stimulates phagocytosis and scavenging of apoptotic cells with pro-inflammatory consequences.[9] Yet, SP-D and SP-A bind SIRPα[10], TLR2, and TLR4,[11] and CD14[12] through their globular carbohydrate recognition domain (CRD) to down-regulate inflammatory cytokines;[11,13] and sftpd<sup>-/-</sup> knockout mice exhibit high levels of pulmonary inflammation.[14] These findings have led to speculation that collectins have dual roles: if the collectin collaginase tail is bound in the absence of a pathogen stimulus, an anti-inflammatory response results possibly mitigating damage from incidental environmental stimuli.[15] However, when pathogen signals are present, pulmonary collectins may provide pro-inflammatory stimuli for pathogen phagocytosis and NF-κB-mediated cytokine release.[10] Further study is needed to confirm our tagSNP associations and further dissect how protection from IPD by <i>SFTPD</i> variants reflects regulatory functions of SP-D.

Our analysis also identified variants in other innate immune and coagulation pathway genes (e.g., <i>CRP</i> and <i>PTEFAR</i> and inflammatory mediators (e.g., <i>IL1R1</i>, <i>IL1B</i>, <i>IL12B</i> and <i>IL10</i>) that may be associated with IPD. Since, as an exploratory study, we did not correct for multiple comparisons, definitive interpretation of these

---

**Figure 1. Summary of tagSNP associations for European-Americans and African Americans.** Genomic and association characteristics for tagSNPs with p ≤ 0.05 for IPD among either EA or AA are shown. Threshold lines for p = 0.05 (-log<sub>10</sub> (0.05) = 1.3) and OR = 1 are indicated in the relevant forest plots. doi:10.1371/journal.pone.0023413.g001

---

<i>pneumoniae</i> reported.[7] SP-D deficient (sftpd<sup>-/-</sup> knockout) mice are associated with persistent pneumococcal colonization, decreased clearance of bacterial pathogens, and early onset and increased levels of <i>S. pneumoniae</i> bacteremia in colonized mice.[7] Overall, collectins exhibit both pro- and anti-inflammatory effects[8]: SP-D stimulates phagocytosis and scavenging of apoptotic cells with pro-inflammatory consequences.[9] Yet, SP-D and SP-A bind SIRPα[10], TLR2, and TLR4,[11] and CD14[12] through their globular carbohydrate recognition domain (CRD) to down-regulate inflammatory cytokines;[11,13] and sftpd<sup>-/-</sup> knockout mice exhibit high levels of pulmonary inflammation.[14] These findings have led to speculation that collectins have dual roles: if the collectin collaginase tail is bound in the absence of a pathogen stimulus, an anti-inflammatory response results possibly mitigating damage from incidental environmental stimuli.[15] However, when pathogen signals are present, pulmonary collectins may provide pro-inflammatory stimuli for pathogen phagocytosis and NF-κB-mediated cytokine release.[10] Further study is needed to confirm our tagSNP associations and further dissect how protection from IPD by <i>SFTPD</i> variants reflects regulatory functions of SP-D.

Our analysis also identified variants in other innate immune and coagulation pathway genes (e.g., <i>CRP</i> and <i>PTEFAR</i> and inflammatory mediators (e.g., <i>IL1R1</i>, <i>IL1B</i>, <i>IL12B</i> and <i>IL10</i>) that may be associated with IPD. Since, as an exploratory study, we did not correct for multiple comparisons, definitive interpretation of these
findings will require confirmation in larger cohort studies. Nevertheless, our findings support multiple pathways being involved in host response to IPD. Recent studies also suggest that additional genes in the toll-like receptor-signaling pathway (e.g., NFKB, IKB and MAL) may influence response to IPD. [16–18] Furthermore, the collectin MBL2 had variants overrepresented in pneumococcal bacteremia and meningitis, but not for overall IPD. This suggests the possibility of syndrome-specific host genetic associations, but our study was underpowered to definitively evaluate this.

For this analysis, we took an indirect association approach by selecting and genotyping SNPs that are either causative SNPs or in LD with the causative SNP. The latter situation most likely applies to the majority of SNPs found associated with IPD in this study, as 18 of the 27 (67%) associated SNPs are located in introns. Furthermore, of the four SFTPD variants associated with IPD in either EA and/or AA (including rs17886286 and rs1998374), all are intronic. Notably, in SeattleSNPs EA, rs17886286 and rs1998374 are in complete (r² = 1.00) or high (r² = 0.732) LD with rs3088308, a coding non-synonymous SNP, while in SeattleSNPs AA, these SNPs have little to no LD with rs3088308 (r² = 0.002 for rs17886286 and r² = 0.028 for rs1998374) and are not associated with IPD in AA. The two SFTPD variants that are associated with IPD in AA are in moderate LD with a different coding non-synonymous SNP, rs4469829, which is monomorphic in EA (r² = 0.613 for rs17878441 and r² = 0.620 for rs12219080). Furthermore, SFTPD variant rs721917, a non-synonymous SNP known to reduce serum levels of SP-D in EA, is in LD with rs1998374 in AA (SeattleSNPs r² = 0.645) but not EA (SeattleSNPs r² = 0.096). Thus, differences in LD patterns between ancestral populations may help to explain the disparate signals observed in EA compared to AA.

Our primary goal was to assess the feasibility of cross-linking surveillance data with an nDBS repository to perform tagSNP genomic studies, and toward this end we were highly successful: 82% of surveillance cases were linked to an nDBS, and 88% of samples successfully genotyped. Several key issues associated with this experience deserve emphasis. First, the completeness of IPD case surveillance in ABCs through use of active surveillance methods and routine audits of laboratory records combined with the overall low incidence of IPD in the general population support our assumption that controls were at low risk of having had IPD outside the surveillance time-period. Second, efficient linking of surveillance cases to nDBS samples was critical to minimize bias, but this linkage depends on consent requirements for nDBS use, which differ by state and continue to evolve.[19] Third, nearly a quarter of individuals identified as European-American through surveillance were found to have genetic characteristics indicating >10% African ancestry. Given differences in allele frequencies and LD between EA and AA populations, misclassification of ancestry can result in confounding. Although self-reported ancestry can be accurate in some settings,[20] our study underscores that more complete genetic evaluation for ancestry may be important to control for population stratification in some settings. Finally, the tagSNP approach we used is based on the assumption that most disease-causing variation will be captured through LD with common tagSNPs. Recent concern [21,22] that low frequency host variation contributes substantial disease causation may underscore the need for large-scale gene sequencing and not a tagSNP analysis. Our earlier findings demonstrating that gene sequencing can be performed with high-fidelity using DNA from these nDBS samples [6] suggests that application of large-scale sequencing for initial variation discovery could be useful in future replication studies.

Our findings are not corrected for multiple comparisons and therefore should be viewed as preliminary with definitive proof requiring replication in additional cohorts. However, given our study results and the wealth of existing public health surveillance data and existing large repositories of nDBS, replication studies powered to detect associations with IPD and invasive disease caused by other vaccine preventable encapsulated bacteria (e.g., N. meningitidis and H. influenzae) should be feasible and could help further define host genetic risk factors for IPD and other infectious diseases and permit economic and attributable risk analyses to determine the usefulness of such risk factors for implementation of public health prevention interventions.

Materials and Methods

Ethics Statement

The protocol for this study was reviewed and approved by the institutional review boards at the Minnesota State Department of Health and the Centers for Disease Control and Prevention (CDC). Written informed consent was solicited from individuals providing nDBS or their guardians. The ethics approvals allowed for use of anonymous samples in cases where the individuals or their guardians provided written informed consent or where they could not be contacted; nDBS samples from individuals who declined consent were removed from this analysis. After determination of consent status, the database was made anonymous by delinking from surveillance identifiers, at which point CDC closed the project, allowing anonymous genotyping to proceed using the coded nDBS and study database. The University of Washington subsequently provided a certificate of exemption for this study.

Subject Ascertainment

As previously described [6], IPD cases in the state of Minnesota were identified through Active Bacterial Core surveillance (ABCs), coordinated through the Centers for Disease Control and Prevention’s Emerging Infections Program. Cases of IPD were defined as individuals born between January 1, 1997 and December 31, 2000 who had isolation of S. pneumoniae from a normally sterile site during the same time period. nDBS collected from cases during the course of normal newborn screening were identified by cross-linking ABCs identifiers with the Minnesota newborn screening program. For each case nDBS identified, two anonymous control nDBS were selected based on surveillance and newborn screening data by matching case and control race/ethnicity, date of birth, and, when possible, hospital of birth. Parents or guardians of cases were contacted by mail for written consent. Surveillance data and nDBS were included for all cases with parental consent and those who did not respond after two mailings. ABCs data and case and control nDBS were stripped of linkage to personal identifiers.

Study cohort

Overall, 445 individuals were identified as meeting the IPD case definition. Of these cases, stored nDBS were identified for 366 (82%) with a median age of 1.1 years (range 0–3.2 years) and nearly equal numbers of males and females (Table 1). Anonymous nDBS were matched to cases by date of birth and identified race. Of these cases, stored nDBS were matched to cases by date of birth and identified ancestry. Among these 1098 cases and controls (Surveillance Cohort), amplified DNA of adequate quality for genotyping was obtained from 967 (88%) representing 330 cases and 637 controls (Genotyped Cohort). Population substructure analysis indicated that 149 individuals (58 cases and 91 controls) in
the Genotyped Cohort that had been specified in the surveillance data as having European American (EA) ancestry had genotypes consistent with ≥10% African ancestry. These cases were omitted from the final Analysis Cohort leaving 543 EA (182 cases and 361 controls), 166 AA (53 cases and 113 controls), and 109 from other ancestral backgrounds (Table 1). The 709 individuals (239 cases and 474 controls) of EA and AA race/ethnicity are referred to as the Analysis Cohort.

Candidate Gene and tagSNP Selection

We selected 34 candidate genes (Tables 2 and S1) based on publications prior to 2007 reporting an association of the gene or protein product with IPD, pneumonia or sepsis, or laboratory evidence for a role in host response to S. pneumoniae or other encapsulated bacteria (e.g., N. meningitidis). TagSNPs were selected for African-Americans and European-Americans using software LDSelect [23] (r²>0.64; minor allele frequency >5%) and data from SeattleSNPs [24], the Environmental Genome Project [25], Perlegen [26], and the International HapMap Project [27]. Amplified DNA samples were genotyped for 384 tagSNPs with a custom Illumina GoldenGate assay [29] by the Center for Inherited Disease Research (CIDR) as part of the National Heart Lung and Blood Institute (NHLBI) Re-sequencing and Genotyping (R5nG) service.

DNA Extraction and SNP Genotyping

De-identified genomic DNA extracted from 1,022 dried blood spots was whole genome amplified using a multiple displacement method (Molecular Staging, Inc. (MSI), New Haven, CT USA) [29] from a single 3-mm punch from ½ inch dried blood spots. A total of 326 (85%) of 384 selected tagSNPs were successfully genotyped (Tables 2 and S1). Of 326 successfully genotyped, tagSNPs fell into 5 broad biological categories: 13 innate immune response genes with 126 SNPs (39%), 13 pro- and anti-inflammatory mediators with 139 SNPs (43%), three genes in coagulation pathways genes with 20 SNPs (6%), and five genes in activation/apoptosis pathways with 41 SNPs (12%) (Tables 2 and S1). Genes with 50% or more SNPs genotyped but not successfully genotyped included: FCGR3A (8 of 9 SNPs failed genotyping), PROC (9 of 14 SNPs), TLR9 (4 of 7 SNPs) and FCGR2A (3 of 6 SNPs). For the IPD analysis, a total of 212 (65%) and 287 (88%) genotyped SNPs had >95% genotyping efficiency, had a minor allele frequency (MAF) >0.01, and met Harding-Weinberg Equilibrium (HWE) criteria of p>0.0001 among EA and AA, respectively. The majority of SNPs genotyped for CD40LG (10 of 14) were out of HWE (p<0.0001) and subsequently omitted from further analyses. Genotyping associated with the bacteremia case definition included: 120 cases and 361 controls were included for EA with 213 SNPs passing QC, and 39 cases and 113 controls among AA with 281 SNPs passing QC. We also report genotyping data from EA for the 25 cases of pneumonia and 17 cases of meningitis compared to 361 controls, with 211 SNPs passing QC for each of those analyses.

Study Population Demographics and Ancestry

Population stratification was evaluated using STRUCTURE (version 2.2, http://pritch.bsd.uchicago.edu/structure.html) prior to performing SNP quality control or testing for association. [30,31] Discrepancies were noted between genotype inferred ancestry and surveillance database assigned ancestry using 274 SNPs with a minor allele frequency >1%. Samples labeled as European-American that were inferred to have >10% African-American ancestry were omitted from further analysis.

Statistical Analysis

Only SNPs that passed quality control (minor allele frequency >1%, Hardy-Weinberg Equilibrium (HWE) p>0.0001 in controls, and genotyping efficiency >95%) were considered for analysis. All analyses were stratified by race/ethnicity, and tests of association were performed using PLINK [32] or STATA (version 10). Due to the small number of non-European, non-African samples, we did not analyze data from those samples. Exclusion of individual case or control samples due to genotyping quality control criteria and genetic ancestry disrupted the original case-control matching. In order to maximize the sample size for this exploratory analysis, tagSNP associations were assessed through an unmatched comparison of IPD cases to controls. Two-tailed χ² tests were used with p<0.05 considered significant without correction for multiple comparisons. Both allelic and genotypic tests of associations were performed for all SNPs. LD estimates were obtained through SeattleSNPs Genome Variation Server at [http://gvs.gs.washington.edu/GVS/](http://gvs.gs.washington.edu/GVS/) using data from the International HapMap Project [27] and SeattleSNPs [24] as indicated. A weighted Genetic Risk Score (GRS) was calculated for SPTD by PLINK for every participant, stratified by race/ethnicity, using SNPs that were associated with IPD status at p<0.05. The GRS is simply a sum across SNPs of the number of risk alleles (0, 1, or 2) at that SNP, multiplied by the odds ratio of the association. Participants with incomplete genotype data at any SNP used in the GRS were excluded from analysis. Logistic regression with continuous GRS as the independent variable, was used to evaluate the joint effects of associated genetic variants with IPD status.

The primary analysis was based on ABCs data for IPD case status. We also evaluated other clinical outcomes of pneumococcal bacteremia, pneumonia, or meningitis, defined by S. pneumoniae cultured from blood only with no other site-specific diagnosis, isolation from pleural fluid or a clinician diagnosis of pneumonia, and isolation from cerebrospinal fluid or a clinician diagnosis of meningitis, respectively. Because of the small numbers of pneumonia and meningitis cases, we used Fisher’s exact test for these analyses. Association results were also plotted graphically using Synthesis-View.[33]
Table S2 tagSNPs in candidate genes found associated with IPD. For European-Americans (182 Cases and 361 Controls), and African-Americans (53 Cases and 113 Controls), Allelic (2×2) and genotypic (2×3) models are used to calculate allelic, heterozygote and homozygote OR, 95% confidence intervals (CI) and minor allele frequency (MAF). Variants are ordered by decreasing significance of the allelic p-value.

Table S3 tagSNPs in candidate genes found associated with bacteremia. For European-Americans (120 Cases and 361 Controls) with 213 SNPs passing QC (HWE >0.0001, MAF >0.01, Genotyping efficiency >90%) and African-Americans (39 Cases and 113 Controls) with 281 SNPs passing QC. Allelic (2×2) and genotypic (2×3) models are used to calculate allelic, heterozygote and homozygote OR, 95% confidence intervals (CI) and minor allele frequency (MAF). Variants are ordered by decreasing significance of the allelic p-value.

References

1. Whitney CG, Farley MM, Hadler J, Harrison LH, Bennett NM, et al. (2003) Decline in invasive pneumococcal disease after the introduction of protein-poly saccharide conjugate vaccine. N Engl J Med 348: 1377–1374.

2. Robinson KA, Baughman W, Rothrock G, Barrett NL, Pass M, et al. (2001) Epidemiology of invasive Streptococcus pneumoniae infections in the United States, 1995-1998: Opportunities for prevention in the conjugate vaccine era. JAMA 285: 1729-1735.

3. Centers for Disease Control and Prevention (2005) Direct and indirect effects of routine vaccination of children with 7-valent pneumococcal conjugate vaccine on incidence of invasive pneumococcal disease—United States, 1998-2003. MMWR Morb Mortal Wkly Rep 54: 893-897.

4. Nuesa JP, Butler JC, Farley MM, Harrison LH, McGeer A, et al. (2000) Cigarette smoking and invasive pneumococcal disease. Active Bacterial Core Surveillance Team. N Engl J Med 342: 681-689.

5. Waterer GW, Wunderink RG (2005) Genetic susceptibility to pneumonia. Clin Chest Med 26: 29–38.

6. Crawford DC, Zimmer SM, Morin CA, Messonnier NE, Lynfield R, et al. (2004) Integrating genomics with surveillance for invasive bacterial diseases. Emerg Infect Dis 11: 1138-1140. PMCID: PMC2600343.

7. Joublat R, Kadioglu A, Iannelli F, Pozzi G, Eggleton P, et al. (2004) Binding and agglutination of Streptococcus pneumoniae by human surfactant protein D (SP-D) vary between strains but SP-D fails to enhance killing by neutrophils. Infect Immun 72: 709-716.

8. Forbes LR, Haczku A (2010) SP-D and regulation of the pulmonary innate immune system in allergic airway changes. Clin Exp Allergy 40: 547–562.

9. VandeVier RW, Ogden CA, Fadok VA, Hoffmann FR, Brown KK, et al. (2005) Role of surfactant proteins A, D, and C1q in the clearance of apoptotic cells in vivo and in vitro: calreticulin and CD91 as dual function surveillance molecules to suppress or enhance inflammation. J Cell Physiol 207: 414-426.

10. Iwaki D, Mitsuzawa H, Sano H, Takahashi H, et al. (2002) Murakami S, Iwaki D, Mitsuzawa H, Sano H, Takahashi H, et al. (2002) Decreasing significance of the allelic p-value.

11. Liu CF, Rivere M, Huang HJ, Pozu G, Wang JY (2010) Surfactant protein D inhibits mit-reduced alveolar macrophage and dendritic cell activations through TLR signalling and DC-SIGN expression. Clin Exp Allergy 40: 111–122.

12. Chapman SJ, Khor CC, Vanberg FO, Fordham A, Waley A, et al. (2007) MipRalpha genetic polymorphisms and invasive pneumococcal disease. Am J Respir Crit Care Med 176: 101-107.

13. Chapman SJ, Khor CC, Vanberg FO, Rautanen A, Segal S, et al. (2010) NFkB1 genetic polymorphisms and susceptibility to pneumococcal disease in European and African populations. Genes Immun 11: 319–325.

14. Shoshin NJ, Chapman SJ, Vanberg FO, Dunne A, Murphy C, et al. (2007) A Mal functional variant is associated with protection against invasive pneumo- coccal disease, bacteremia, malaria and tuberculosis. Nat Genet 39: 523-528.

15. Ivanov II, Periman A, Chen J, et al. (2009) RIP1-KAP1 and RIP3 form a complex to promote innate immune responses to necrotic cells. Science 326: 1487–1491.

16. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754–1760.

17. Mikkelsen TS,imbright OM, et al. (2011) Broad direct lineage analysis of mouse development reveals distinct regional identities and the role of maternal genome activation in lineage specification. Dev Cell 20: 864-877.

18. Ares SR, Pacheco RJ, et al. (2011) Conversion of B cells to antibody-secreting plasma cells by RIP1-Tag2 cells. Science 332: 1260–1265.

19. Handel MG, et al. (2011) The canonical NF-κB pathway is essential for myeloid cell survival and host defense against infection. Cell 146: 63–76.

20. Willment JA, et al. (2011) NF-κB and NAD+ metabolism in cancer cells. Nat Rev Cancer 11: 334–346.

21. Xiong Y, et al. (2011) Cooperation between RIP1 and nuclear factor κB (NF-κB) controls adaptive immunity and autoimmunity. J Exp Med 208: 1359–1366.

22. Koike K, et al. (2011) NF-κB mediates survival of hematopoietic stem cells. J Exp Med 208: 1367–1380.

23. Li Q, et al. (2011) NF-κB pathway is critical for the pathogenesis of pulmonary fibrosis. J Biol Chem 286: 35159–35168.

24. Araki A, et al. (2011) NF-κB signaling is required for lung innate immune responses against pneumococcal infection. J Exp Med 208: 1909–1922.

25. Togawa F, et al. (2011) NF-κB signaling is required for lung innate immune responses against pneumococcal infection. J Exp Med 208: 1909–1922.

26. Schuurman PA, et al. (2011) Role of neutrophils in the repair and remodeling of the lung. Am J Respir Crit Care Med 183: 708-718.

27. Togawa F, et al. (2011) NF-κB signaling is required for lung innate immune responses against pneumococcal infection. J Exp Med 208: 1909–1922.

28. Schuurman PA, et al. (2011) Role of neutrophils in the repair and remodeling of the lung. Am J Respir Crit Care Med 183: 708-718.

29. Togawa F, et al. (2011) NF-κB signaling is required for lung innate immune responses against pneumococcal infection. J Exp Med 208: 1909–1922.

30. Schuurman PA, et al. (2011) Role of neutrophils in the repair and remodeling of the lung. Am J Respir Crit Care Med 183: 708-718.

31. Togawa F, et al. (2011) NF-κB signaling is required for lung innate immune responses against pneumococcal infection. J Exp Med 208: 1909–1922.
31. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945–959.
32. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559–575.
33. Pendergrass SA, Dukleth SM, Crawford DC, Ritchie MD (2010) AnalysisView: visualization and interpretation of SNP association results for multi-cohort, multi-phenotype data and meta-analysis in BioDataMin. 3: 10.
34. Johansson L, Rytkonen A, Bergman P, Albiguer B, Kalliomaki H, et al. (2003) CD46 in meningococcal disease. Science 301: 373–375.
35. Seya T, Hirano A, Matsumoto M, Nomura M, Ueda S (1999) Human membrane cofactor protein (MCP, CD46): multiple isoforms and functions. Int J Biochem Cell Biol 31: 1255–1260.
36. Mold C, Nakayama S, Holzer TJ, Gewurz H, Du Clos TW (1981) C-reactive protein is protective against Streptococcus pneumoniae infection in mice. J Exp Med 154: 1703–1708.
37. Mold C, Redic-Pole B, Du Clos TW (2002) Protection from Streptococcus pneumoniae infection by C-reactive protein and natural antibody requires complement but not Fc gamma receptors. J Immunol 168: 6375–6381.
38. Yother J, Volanakis JE, Briles DE (1982) Human C-reactive protein is protective against fatal Streptococcus pneumoniae infection in mice. J Immunol 128: 2374–2376.
39. van der Pol WL, Huizinga TW, Vidarsson G, van der Linden MW, Jansen MD, et al. (2005) C-reactive protein is protective against pulmonary infection with Streptococcus pneumoniae, but not to a soluble pneumococcal conjugate vaccine. J Immunol 177: 6044–6051.
40. Ling E, Feldman G, Dagan R, Mizrachi-Nebenzahl Y (2003) Cytokine mRNA expression during pneumococcal meningitis. J Neuroimmunol 168: 21–33.
41. Paul R, Angele B, Sporer B, Pfister HW, Koedel U (2004) Inflammatory cytokines and risk of invasive pneumococcal disease: a case-control study. Lancet 363: 1569–1573.
42. Albiguer B, Sandgren A, Katsaragyi H, Meyer-Hoffert U, Beier K, et al. (2005) Myeloid differentiation factor 88-dependent signalling controls bacterial growth during colonization and systemic pneumococcal disease in mice. Cell Microbiol 7: 1603–1615.
43. Hartshorn KL, Crouch E, White MR, Colamussi ML, Kakkanatt A, et al. (1998) Pulmonary surfactant protein A and D enhance neutrophil uptake of bacteria. Am J Physiol 274: L587–96.
44. Lin Z, Pearson C, Pinchilli V, Pietrangelo SM, Luo J, et al. (2000) Polymorphisms of human SPA-SP-B and SP-D genes: association of SP-B Thr151Ile with ARDS. Clin Genet 58: 181–191.
45. Emonts M, Veenhoven RH, Wiertsema SP, Houwing-Duistermaat JJ, PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559–575.
46. Fanger XM, Schroder S, Hoelt A, Stuber F (1999) Comparison of two polymorphisms of the interleukin-1 gene family: interleukin-1 receptor antagonist polymorphism contributes to susceptibility to severe sepsis. Crit Care Med 27: 1330–1334.
47. Chen Q, Sen G, Snapper CM (2006) Endogenous IL-1R1 signaling is critical for cognate CD4+ T cell help for induction of in vivo type 1 and type 2 anti-poly saccharide and anti-protein IgG responses to intact Streptococcus pneumoniae, but not to a soluble pneumococcal conjugate vaccine. J Immunol 177: 6044–6051.
48. Emonts M, Verhoeven RH, Wiertsema SP, Hoving-Duijtermajt JA, Walraven V, et al. (2007) Genetic polymorphisms in immunoresponse genes TNFA, IL6, IL10, and TLR4 are associated with recurrent acute otitis media. Pediatrics 120: 814–823.
49. Schubert B, Rauffhake C, Erren M, Schotte H, Kipp F, et al. (2002) Effect of the interleukin-6 promoter polymorphism (-174 G/C) on the incidence and outcome of sepsis. Crit Care Med 30: 32–37.
50. Schmeck B, Zahleen J, Moog K, van Laak V, Huber S, et al. (2004) Streptococcus pneumoniae-induced p38 MAPK-dependent phosphorylation of RelA at the interleukin-8 promoter. J Biol Chem 279: 53241–53247.
51. Gordon SB, Jarman ER, Kanyanda S, French N, Pridmore AC, et al. (2005) Reduced interleukin-1 response to Streptococcus pneumoniae by alveolar macrophages from adults with HIV/AIDS. AIDS 19: 1197–1206.
52. Westendorp RG, Langemaa J, Huizinga TW, Elouali AH, Verweij CL, et al. (1997) Genetic influence on cytokine production and fatal meningococcal disease. Lancet 349: 170–173.
53. Schaaf BM, Boechne K, Emaaharhi H, Seitzer U, Kothe H, et al. (2003) Pneumococcal septic shock is associated with the interleukin-10-1082 geno promoter polymorphism. Am J Respir Crit Care Med 168: 476–480.
54. Gallagher PM, Lowe G, Fitzgerald T, Bella A, Green E, et al. (2003) Association of IL-10 polymorphism with severity of illness in community acquired pneumonia. Thorax 58: 134–136.
55. Lang E, Feldman G, Dagan R, Mizrachi-Nebenzahl Y (2003) Cytokine mRNA expression in pneumococcal carriage, pneumonia, and sepsis in young mice. J Infect Dis 188: 1752–1756.
56. Lynch JM, Briles DE, Metager DW (2003) Increased protection against pneumococcal disease by mucosal administration of conjugate vaccine plus interleukin-12. Infect Immun 71: 4780–4788.
57. Laue FN, Branger J, Horquin S, Speelman P, van Deventer SJ, et al. (2002) IL-18 improves the early antimicrobial host response to pneumococcal pneumonia. J Immunol 168: 372–378.
58. Paterson GK, Blue CE, Mitchell TJ (2005) Role of interleukin-18 in experimental infections with Streptococcus pneumoniae. J Med Microbiol 54: 323–326.
59. Zwiebneburg PJ, van der Pol T, Horquin S, Akira S, Takeda K, et al. (2003) Interleukin-18 gene-deficient mice show enhanced defense and reduced inflammation during pneumococcal meningococcal meningitis. J Neuroimmunol 138: 31–37.
60. Sterber F, Petersen M, Bokelmann F, Schade U (1996) A genomic polymorphism within the tumor necrosis factor locus influences plasma tumor necrosis factor-alpha concentrations and outcome of patients with severe sepsis. Crit Care Med 24: 381–384.
61. Mira JP, Carieu I, Graff F, Delclaux C, Losser MR, et al. (1999) Association of TNF2, a TNF-alpha promoter polymorphism, with septic shock susceptibility and mortality: a multicenter study. JAMA 282: 561–569.
62. McGuire W, Hill AV, Allopp CE, Greenwood BM, Kwaizikoowski D (1994) Variations in the TNF-alpha promoter region associated with susceptibility to cerebral malaria. Nature 371: 508–510.
63. Bernard GR (2003) Drotrecogin alfa (activated) (recombinant human activated protein C) for the treatment of severe sepsis. Crit Care Med 31: S65–93.
64. Bernard GR, Vincent JL, Laterre PF, LaRosa SP, Dhainaut JF, et al. (2001) Efficacy and safety of recombinant activated human protein C for severe sepsis. New Engl J Med 344: 699–709.
65. Dunkel DR, Gerard NP, Gerard C, Ilsaanap-Hekkila I, Tuomainen EI (1995) Streptococcus pneumoniae: anchor to activated human cells by the receptor for platelet-activating factor. Nature 377: 435–438.
66. Grashoff G, Binder A, Muller M, Zohrer B, Resch B, et al. (2005) 4G/5G promoter polymorphism in the plasminogen-activator-inhibitor-1 gene in children with systemic meningococcaemia. Eur J Pediatr 164: 486–490.
67. Westendorp RG, Hottinger JJ, Slagboom PE (1999) Variation in plasminogen-activator-inhibitor-1 gene and risk of meningococcal septic shock. Lancet 354: 561–563.
68. Jeunisens A, Wuyts G, Kasran A, Ramdani-Murli S, Blanken A, et al. (2004) The human antibody response to pneumococcal capsular polysaccharides is dependent on the CD14/CD18/human leukocyte receptor interaction. Eur J Immunol 34: 830–836.
69. Bourdeijn M, Jeunisens A, Wuyts M, Moens I, Boon L, et al. (2005) Blockade of CTLA-4 (CD152) enhances the murine antibody response to pneumococcal capsular polysaccharides. J leukoc Biol.
70. Matteo-Bello G, Liles WC, Frevert CW, Dhaniroyd S, Ballman K, et al. (2005) Blockade of the Fas/FasL system improves pneumococcal clearance from the lungs without preventing dissemination of bacteria to the spleen. J Infect Dis 191: 596–606.
71. Paul R, Angele B, Spooner P, Plister HW, Koedel U (2004) Inflammatory response during bacterial meningitis is unchanged in Fas- and Fas ligand-deficient mice. J Neuroimmunol 152: 7–8.