Machine learning-driven identification of serotype-independent pneumococcal vaccine candidates using samples from human infection challenge studies

Cheliotis, K.S.; Gonzalez-Dias, P.; German, E.L.; Gonçalves, A.N.A.; Mitsi, E.; Nikolaou, E.; ... ; Ferreira, D.M.

Vaccine 75 (2026) 128280

Contents lists available at ScienceDirect

Vaccine

journal homepage: www.elsevier.com/locate/vaccine

Machine learning-driven identification of serotype-independent pneumococcal vaccine candidates using samples from human infection challenge studies

Katerina S. Cheliotis [a,b,1], Patricia Gonzalez-Dias [a,c,o,1], Esther L. German [a,d], André N.A. Gonçalves [a,c,o], Elena Mitsi [a,c,o], Elissavet Nikolaou [a,e,f], Sherin Pojar [a,g], Eliane N. Miyaji [h], Rafaella Tostes [h], Jesús Reiné [a,c,o], Andrea M. Collins [a,i], Helder I. Nakaya [j,k], Stephen B. Gordon [a,l], Ying-Jie Lu [m], Shaun H. Pennington [a], Andrew J. Pollard [c,o], Richard Malley [m], Simon P. Jochems [n], Britta Urban [a,c,o,2], Carla Solórzano [a,c,o,2,*], Daniela M. Ferreira [a,c,*,2]

[a] Department of Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, United Kingdom
[b] City St George's, University of London, London, United Kingdom
[c] Oxford Vaccine Group, Department of Paediatrics, University of Oxford, Oxford, United Kingdom
[d] London School of Hygiene and Tropical Medicine, London, United Kingdom
[e] Infection, Immunity and Global Health, Murdoch Children's Research Institute, Melbourne, Victoria, Australia
[f] Department of Microbiology and Immunology, The University of Melbourne at the Peter Doherty Institute for Infection and Immunity, Melbourne, Victoria, Australia
[g] BioGrad Limited, Liverpool, United Kingdom
[h] Laboratório de Bacteriologia, Instituto Butantan, São Paulo, Brazil
[i] Liverpool University Foundation Hospitals Trust, Liverpool, United Kingdom
[j] Department of Clinical and Toxicological Analyses, School of Pharmaceutical Sciences, University of São Paulo, São Paulo, Brazil, 05508
[k] Hospital Israelita Albert Einstein, São Paulo, Brazil
[l] Centre for Inflammation Research, University of Edinburgh, Edinburgh, United Kingdom
[m] Boston Children's Hospital, Harvard, Medical School, MA, USA
[n] Leiden University Center for Infectious Diseases (LUCID), Leiden University Medical Center, Leiden, Netherlands
[o] NIHR Oxford Biomedical Research Centre, Oxford, United Kingdom

ARTICLE INFO

Keywords:
Streptococcus pneumoniae
Controlled human infection model
Correlates of protection
Serotype-independent vaccine
Machine learning
Vaccine antigen discovery
Systems vaccinology
Immune responses

ABSTRACT

Identifying conserved, immunogenic proteins that confer protection against *Streptococcus pneumoniae* (*pneumococcus*) colonization could enable development of serotype-independent vaccines.

In our controlled human infection model, no individual IgG or cytokine/chemokine response correlated significantly with protection against colonization with pneumococcus, suggesting that effective immunity reflects a coordinated, multi-antigen response. To capture these complex patterns, we trained independent Random Forest models on humoral and cellular datasets. The humoral model identified IgG responses to PdB, SP1069, and SP0899 as predictive of protection. The cellular model revealed that MCP-1 responses to SP1069 and SP0899, and IL-17A production in response to SP0648-3, were associated with protection. Elevated baseline IFN-γ, RANTES, and anti-protein IgG levels were linked to reduced colonization density.

We highlight SP1069 and SP0899 as potential serotype-independent vaccine candidates and demonstrate the utility of machine learning to identify immune correlates of protection.

* Corresponding authors at: Oxford Vaccine Group, Department of Paediatrics, University of Oxford, Oxford, United Kingdom.
E-mail address: carla.solorzanogonzalez@paediatrics.ox.ac.uk (C. Solórzano).
[1] Joint first authors.
[2] Joint senior authors.

https://doi.org/10.1016/j.vaccine.2026.128280
Received 12 November 2025; Received in revised form 5 January 2026; Accepted 23 January 2026
Available online 31 January 2026

1. Introduction

Pneumonia is the leading infectious cause of death in infants worldwide, with more than 800,000 annual deaths among children under 5 years of age globally [1–3]. The most common aetiological agent of pneumonia is the bacterium *Streptococcus pneumoniae* (pneumococcus). Nasopharyngeal colonization with pneumococcus, as well as being a prerequisite for invasive disease, is a source of transmission in the community [4].

The pneumococcal conjugate vaccine (PCV) generates population herd protection by reducing vaccine-type nasopharyngeal carriage and transmission, leading to marked declines in invasive pneumococcal disease across both vaccinated and unvaccinated age groups [5]. Vaccine-induced humoral responses to the pneumococcal capsule polysaccharide protect against pneumococcal carriage acquisition and disease caused by vaccine specific serotypes [6–10]. However, current vaccines – pneumococcal polysaccharide vaccine (PPV23) and PCVs – offer protection against a limited number of the >100 pneumococcal serotypes [11], leading to serotype replacement of non-vaccine types [12,13]. In addition, despite being included in PCVs, serotype 3 continues to circulate and cause a high disease burden particularly in older adults [14–16]. PCVs are also complex and costly to produce, which can present a barrier to vaccine access in resource-limited countries, particularly those graduating from Gavi support [17].

Pneumococcal protein antigens that are universally expressed across various serotypes may induce both humoral and cellular immune responses, offering broad protection. These antigens could potentially be introduced as stand-alone vaccines or be included as conjugate proteins in current polysaccharide-based vaccines [18].

Currently the regulatory approval pathway for next-generation pneumococcal vaccines in infants relies on the non-inferiority of IgG levels to capsular polysaccharide using a pre-defined threshold of 0.35μg/ml [19,20]. In adults, demonstration of non-inferior or superior opsonophagocytic antibody responses to each serotype compared to existing licensed vaccines is required [21]. Therefore, the regulatory pathway for capsular polysaccharide independent vaccines is not defined. The identification of serotype-independent cellular and humoral correlates of protection against pneumococcal infection could assist with this framework and help to accelerate development of serotype-independent vaccines.

Systems-biology and artificial intelligence can be used to significantly enhance our understanding of immunity and aid in the identification of correlates of protection. Embedding the understanding of individual components and their interactions, beyond canonical antibody binding and neutralization, may accelerate the design of more successful and long-lasting vaccines [22].

In this study, we leveraged a controlled human infection model (CHIM) to investigate immune responses to pneumococcal protein antigens that are conserved across serotypes, aiming to identify correlates of serotype-independent protection and to understand how immune response alters colonization dynamics in the nasopharynx. To enable high-throughput assessment of protein-specific IgG responses, we established a Luminex assay targeting 75 pneumococcal protein antigens. Given the well-established role of IL-17A in mediating protection against pneumococcus in murine models [23–27], we also evaluated cellular responses to IL-17A in protein-stimulated peripheral blood mononuclear cells (PBMCs). In parallel, we employed a multiplex cytokine assay to profile broader cytokine responses, with MCP-1 (CCL2), RANTES (CCL5), and IFN-γ selected for more detailed analysis.

2. Methods

2.1. Study design

The overall study design is shown in Fig. 1. Briefly, at baseline (day -4 and day -5), participants were assessed before intranasal inoculation (day 0) with *S. pneumoniae* (Spn). Carriage acquisition was monitored until day 14 for Spn15B and day 29 for Spn6B.

Immune responses were evaluated using Luminex assays, which measured IgG antibody responses to 75 pneumococcal antigens in 72

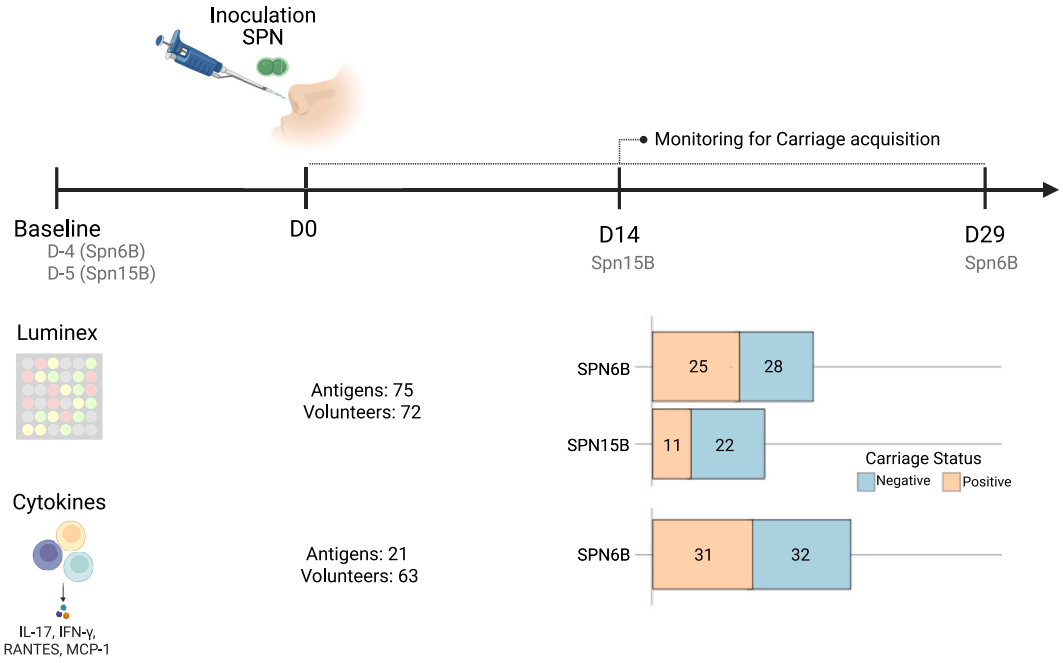

Fig. 1. Study design. Serum samples were collected at baseline (day -4 for Spn6B and day -5 for Spn15B) prior to experimental inoculation with *Streptococcus pneumoniae* (Spn) on day 0 (D0). Participants were monitored for pneumococcal carriage acquisition up to day 29 post-challenge (sampling on days 2, 6, 9, 14, and 29 for Spn6B; and days 2, 7, and 14 for Spn15B). Antibody responses were assessed using a Luminex assay, evaluating 75 pneumococcal antigens across 72 volunteers. Cytokine and chemokine responses (IL-17A, IFN-γ, RANTES [CCL5], and MCP-1 [CCL2]) were measured by ELISA following PBMC stimulation with pneumococcal protein antigens (21 antigens, 63 volunteers). Bar plots indicate the number of participants who acquired pneumococcal carriage (orange) versus those who remained carriage-negative (blue) for each serotype. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

volunteers, and cytokine profiling, which detected 21 cellular products (including IL-17A, IFN-γ, RANTES, and MCP-1) in 63 volunteers.

Carriage acquisition results are presented as bar graphs, depicting the number of individuals testing positive or negative for Spn6B and Spn15B at different time points. These results indicate the proportion of participants who acquired pneumococcal carriage following inoculation.

2.2. Participant cohort inoculated with serotype 6B pneumococcus

Blood samples were taken from healthy non-smoking participants between the ages of 18 and 50 who were intranasally inoculated with 80,000 CFU in 0.1 ml solution of *S. pneumoniae* serotype 6B (Spn6B) strain BHN418 per naris (full sequence available GenBank: ASHP00000000). Samples were taken prior to inoculation (baseline). Cohorts were recruited as part of EudraCT (2014-004634-26), a study to evaluate the effect of live attenuated influenza vaccine (LAIV) on experimental human pneumococcal colonization (EHPC) [28,29]. All participants gave written informed consent, with approval from the North West NHS Research Ethics Committee (14/NW/1460). Nasal wash samples were taken on days 2, 6, 9, 14 and 29 post-challenge with Spn6B. Colonization was determined by classical microbiology, as described previously [25], and the presence of Spn6B in nasal wash samples taken at pre-defined time-points between inoculation and 29 days post-inoculation. No participants had previously received any pneumococcal vaccine; nevertheless, previous exposition is expected.

2.3. Participant cohort inoculated with serotype 15B pneumococcus

Healthy, non-smoking participants aged 18–50 years were intranasally inoculated with 80,000 CFU *S. pneumoniae* serotype 15B (Spn15B) per naris (clinical isolate 15B P1262. European Nucleotide Archive accession number: ERS2632437). Spn15B is included in the PPV23 formulation but not in the 13-valent PCV formulation. All participants gave informed written consent, ethical approval was obtained from the National Health Service Research Ethics Committee, Liverpool East (15/NW/0931). Serum samples were taken at baseline (pre-inoculation) (Trial identifier: ISRCTN 68323432; 20,815). Colonization was determined by classical microbiology, as previously described [30], and the presence of Spn15B in nasal wash samples taken at pre-defined time-points between inoculation and 14 days post-inoculation. Nasal wash samples were taken on days 2, 7 and 14 post-challenge with Spn15B. No participants in either cohort (exposed to Spn6B or Spn15B) had previously received any pneumococcal vaccine; nevertheless, previous exposure is expected [31,32].

2.4. Protein antigens

The proteins used in the assay are given in Supplementary Table 1. The sequences encoding the alpha-helical N-terminal domain of PspA1, PspC6, PspC9, were cloned using DNA from strain BHN418 (serotype 6B pneumococcus). The sequence encoding the N-terminal domain of PspA4 was cloned from strain 255/00 (serotype 14, GenBank accession number EF649969). Recombinant proteins, including pneumolysin toxoid B (PdB), a pneumolysin mutant with a tryptophan-to-phenylalanine substitution at position 433 and 0.1% haemolytic activity of wild-type pneumolysin [33,34], were produced at Butantan Institute laboratories as described previously [35]. The remaining 70 proteins in the library were produced at Boston Children's Hospital as described [24], having been selected by in silico analysis for high conservation (>90% identity across 42 *pneumococcus* genomes) and predicted surface exposure.

2.5. Luminex assay

All 75 pneumococcal protein antigens were conjugated to magnetic microspheres at a concentration of 5 μg per million microspheres using the xMAP Antibody Coupling Kit as per the manufacturer's instructions. The assay was validated for the detection of anti-protein IgG as described previously [36,37]. Full methods are described in detail in Supplementary Text 1. A stock of pooled sera taken 29 days post-challenge and diluted 1:50 in assay buffer was used as an internal positive control across all plates and assay buffer was used as a negative control to detect background median fluorescence intensity (MFI) values.

2.6. Luminex data cleaning and statistical analysis

Raw Luminex data were cleaned prior to analysis through the following workflow. The coefficient of variation (CV) of replicates was calculated for all analytes. If any analyte had greater than 25% CV and both replicates had greater than 50 microspheres per well, the analyte was removed from analysis. If any paired samples had less than 50 microspheres of a given analyte in each well, the analyte was removed from analysis. The determinant of 50 microspheres is based on recommendations by the Luminex Corporation that a recovery of less than 50 microspheres renders a result unreliable as well as previous work showing that lower bead counts elicit higher CV values [38]. All statistical analysis was carried out in RStudio (version 1.0.153). Statistical significance was determined via a Mann-Whitney or Wilcoxon test for unpaired and paired groups, respectively, followed by Benjamini Hochberg correction for multiple comparisons to reduce the false positive rate. For the machine learning analysis, any proteins with greater than 25% of values missing and any participants with more than 50% of values missing were removed from the analysis.

2.7. Stimulation of human peripheral blood mononuclear cells (PBMCs) and cytokine/chemokine detection

Baseline PBMCs from healthy participants inoculated with Spn6B were thawed with 50 μg/ml DNAse I (MilliporeSigma) in prewarmed RPMI containing 10% FBS and washed once in media including DNAse I. 1×10^6 PBMCs/ml from an initial cohort of 12 carriage-negative and 11 carriage-positive participants were stimulated with 8 μg/ml of each protein from the complete library and 5 μg/ml of heat-inactivated Spn6B for 7 days at 37 °C under 5% CO_2. The plates were then centrifuged to pellet the cells, after which the supernatants were collected to measure IL-17A concentration using human ELISA kits (Invitrogen). In parallel, supernatants for all proteins from carriage-negative and supernatants from all proteins from carriage-positive participants from this initial cohort were combined and assessed for other cytokine and chemokine concentrations using a 30-plex magnetic human Luminex cytokine kit (Thermo Fisher Scientific) and acquired on an LX200. A log 2-fold change between carriage-negative and positive was calculated and a Wilcoxon paired test was used to identify cytokines/chemokines associated with a specific carriage status.

Twenty-one antigens from a 70-protein antigen library were selected for further analysis. The antigens were selected based on (i) their ability to induce high IL-17A responses in carriage-negative individuals (ii) negative correlation between high IL-17A levels and low pneumococcal density in participants who were experimentally colonized (iii) their ability to induce high expression of cytokines/chemokines at baseline prior to challenge (iv) available published data in which the antigens have shown protection against challenge after immunization in pre-clinical models of infection.

PBMCs from an additional 22 carriage-negative and 21 carriage-positive participants were stimulated with the selected 21 antigens and cytokine/chemokine concentrations were measured to validate the results obtained with the initial cohort. Human ELISA kits were used to measure concentrations of IL-17A (Invitrogen), MCP-1 (CCL-2) (BD OptEIA™), IFN-γ (BD OptEIA™), and RANTES (CCL-5) (Biolegend) in the collected supernatants following the manufacturers

recommendations. Non-stimulated (control) and protein-stimulated supernatants were diluted as follows: 1:2 for IL-17A, 1:10 for RANTES, 1:200 for IFN-γ, and 1:3200 for MCP-1.

2.8. ELISA data cleaning and statistical analysis

As detailed above, any proteins with greater than 25% of values missing and any participants with more than 50% of values missing were removed from the analysis. Statistical significance was determined via a Mann-Whitney test for unpaired groups, followed by Benjamini Hochberg correction for multiple comparisons.

2.9. Area under the curve (AUC) calculation

Area under the curve (AUC) was used to calculate overall colonizing density of *S. pneumoniae* in the nasopharynx of experimentally colonized participants. AUC was calculated by trapezoidal rule from log 10 transformed density results at 2, 6, 16, 22, 27 and 36 days or after challenge on day 2, 7 or 14 with interpolation of missing data assuming a linear change in density between known values. The AUC of pneumococcal colonizing density at each timepoint was derived and summarized using the above descriptive statistics and analysed using a generalized linear model with a single factor of treatment.

A Spearman test was performed to evaluate the correlation between the IgG and cytokine/chemokine levels elicited in response to each protein and the AUC values obtained between days 2 to 14 for each carriage-positive participant using the *cor.test* function in the software R. The analysis was performed using antibody response and cellular response from 34 and 27 participants, respectively. Correlations with *p*-value $\leq$0.05 were considered significant.

2.10. AUC threshold definition

Retrospectively, we calculated the AUC of nasopharyngeal pneumococcal density for each colonized participant between days 2 to 14 post-challenge. For participants inoculated with Spn6B (n = 128) or Spn15B (n = 9), the distributions of AUC values were used to define colonization density categories. Specifically, participants were divided into “Low,” “Mid,” and “High” density groups corresponding to the lower, middle, and upper tertiles ($\leq$33rd, 34–66th, and $\geq$67th percentiles, respectively) of the AUC distribution within each serotype cohort. Thresholds therefore differed between Spn6B and Spn15B due to differences in absolute density profiles. All data used for threshold definition were obtained from our prior CHIM studies employing identical experimental protocols, and no external datasets were included. Supplementary Fig. 1 illustrates the AUC distributions, with vertical dashed lines indicating tertile cutoffs and each dot representing an individual participant (Supplementary Fig. 1).

2.11. Machine learning analysis

Missing values in numerical variables were imputed using the “mice” R package [39]. Numeric features were standardized using z-score normalization. The Random Forest algorithm, an ensemble machine learning method that constructs multiple decision trees using random feature subsets and aggregates their outputs to enhance predictive accuracy and minimize overfitting [40], implemented in the R package “randomForest”. The model was trained with mtry = 1 and ntree = 500, while other hyperparameters were set to their default values. This algorithm was employed to identify baseline protein-specific antibodies and cytokines/chemokines produced in response to PBMC stimulation with protein antigens that best predicted carriage status post-challenge. Unlike univariate analysis, which assesses each feature independently, machine learning techniques such as Random Forest can uncover complex, multivariate relationships among features, improving predictive accuracy.

Feature importance was assessed using the mean decrease in Gini index, a metric that quantifies how much a given feature contributes to reducing impurity (i.e., improving class separation) within the decision trees. A higher mean decrease in Gini index indicates a stronger contribution to classification accuracy.

To identify the most relevant features and optimize model performance, we applied an iterative feature selection process based on feature importance. Initially, we trained the Random Forest model using all available features and evaluated their importance based on the mean decrease in Gini index. We then progressively removed the least informative features, retraining the model at each iteration and assessing its predictive performance. This process was repeated several times, testing progressively smaller feature sets (e.g., 50, 40, etc.), until the optimal configuration was identified, the smallest number of features yielding the best model performance, minimizing error, and reducing the risk of overfitting. Model evaluation was conducted using 10-fold cross-validation.

3. Results

3.1. Participant cohorts

Two cohorts of healthy, non-smoking adults aged 18–50 years participated in experimental human pneumococcal colonization (EHPC) studies. One cohort was intranasally inoculated with pneumococcal serotype 6B (Spn6B; strain BHN418), and the other with serotype 15B (Spn15B; clinical isolate P1262). Participants were followed for 29 days (Spn6B) or 14 days (Spn15B) to monitor pneumococcal carriage. None had previously received pneumococcal vaccination, although prior natural exposure to *S. pneumoniae* was expected.

Baseline blood samples were collected before inoculation for assessment of pre-existing antibody levels, and nasal wash samples were obtained at multiple time points post-inoculation to determine colonization. Immune responses were evaluated by Luminex-based IgG profiling and cytokine/chemokine assays, allowing assessment of both humoral and cellular immune responses following challenge.

3.2. Protection against experimental pneumococcal carriage is not driven by single antigenic IgG responses, suggesting combined immune signatures

A Luminex-based multiplex assay using a library of 75 highly conserved pneumococcal proteins (Supplementary Table 1), was successfully optimized and validated [37]. These data enabled a comprehensive, high-dimensional assessment of baseline humoral responses prior to experimental pneumococcal challenge with Spn6B or Spn15B.

To explore potential immune correlates of protection, we first examined associations between baseline serum IgG levels and subsequent carriage acquisition. While several antigens showed nominal associations, none remained statistically significant after Benjamini–Hochberg correction for multiple testing, which controls for false discovery (Fig. 2). This lack of single-antigen predictors led to multivariate and machine learning approaches to identify combinations of immune features associated with protection.

3.3. Screening of antigens using human PBMCs and cellular responses

A library of pneumococcal protein antigens was used to stimulate PBMCs collected from participants prior to experimental challenge with Spn6B (Supplementary Fig. 2). Due to the limited number of PBMCs obtained from each participant, we could not screen all protein antigens in all participants. To prioritize antigens for further analysis, we assessed IL-17A production, as this cytokine has previously been identified as a marker of protection against pneumococcal colonization in mice [23,41]. IL-17A secretion following ex-vivo protein stimulation was observed for all proteins at 7 days of culture.

To identify other cellular effector proteins that could associate with a

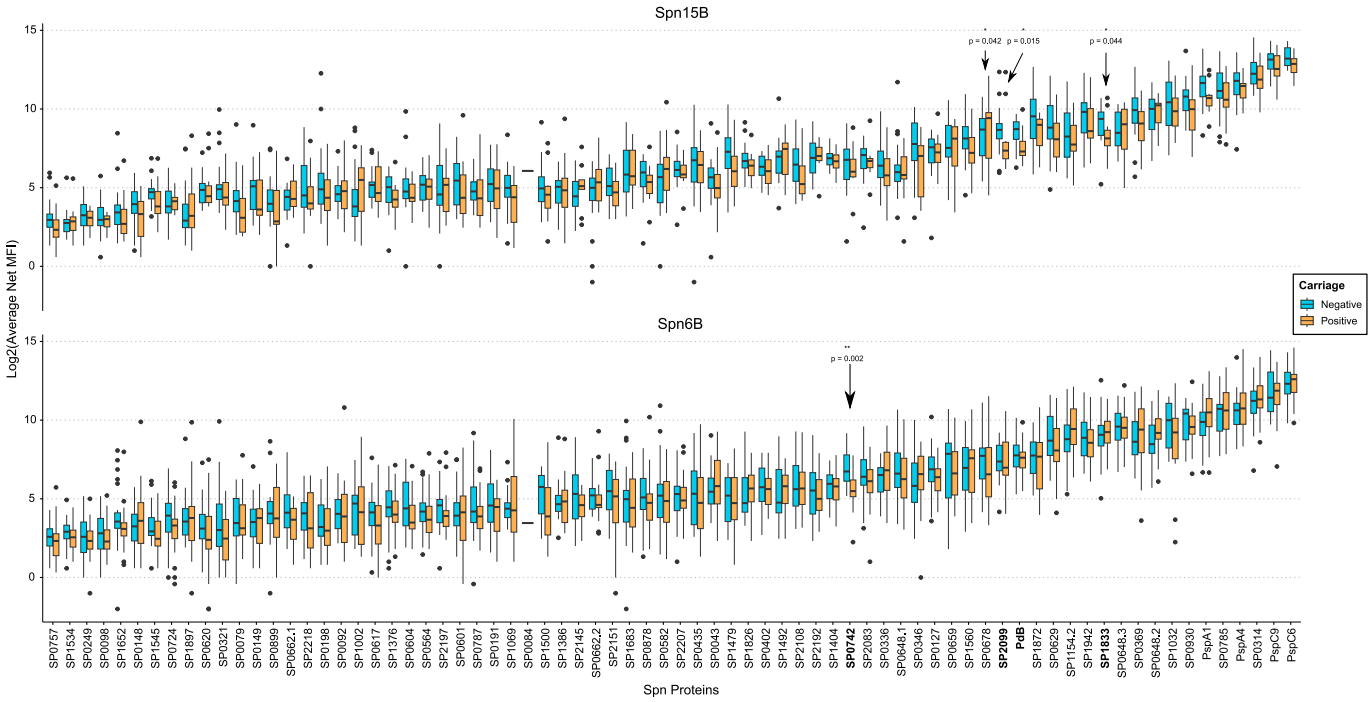

Fig. 2. Levels of anti-protein IgG at baseline in participant cohorts subsequently challenged with Spn6B or 15B. Baseline levels of anti-protein serum IgG in healthy adults aged 18–50, measured as average net median fluorescence intensity (MFI). Bars indicate interquartile range. Spn15B carriage-negative $n = 15$, Spn15B carriage-positive $n = 9$; Spn6B carriage-negative $n = 18$, Spn6B carriage-positive $n = 21$. Carriage status was defined as *binary*, participants were classified as carriage-positive if *S. pneumoniae* was detected at any sampling time point post-challenge, and carriage negative if it is not detected. Responses with *p*-value $\leq$0.05 prior to Benjamini-Hochberg correction for multiple comparisons are shown; this significance was lost following correction.

protective or susceptible profile in the CHIM, we pooled supernatants from PBMC cultures following individual protein stimulation from carriage-negative and carriage-positive participants and compared levels of 30 cytokines and chemokines using 30-plex Luminex in these two groups (Supplementary Fig. 3). MCP-1 (CCL-2) was selected as a marker of protection based on its well-documented role in monocyte recruitment and the clearance of pneumococcal carriage [28]. RANTES (CCL5) and IFN-γ were selected due to their elevated levels in carriage-positive individuals and their known involvement in T cell-mediated immune responses [42].

We used the data obtained from the initial IL-17A screening of susceptible and protected profiles (Supplementary Fig. 2), together with available data on immunization with those antigens and protection against carriage in mice [24], to down-select a limited number of antigens (Supplementary Table 2) for PBMC stimulation assays. PBMCs from a further 22 carriage-negative and 21 Spn6B carriage-positive participants were stimulated with the down-selected 21 antigens and IL-17A, RANTES, MCP-1 and IFN-γ responses were measured.

3.4. *Protection against experimental pneumococcal carriage acquisition with Spn6B is not driven by single antigenic cytokine/chemokine responses, suggesting combined immune signatures*

Comprehensive profiling of baseline protein-specific cytokine and chemokine responses was performed to investigate their association with protection against experimental pneumococcal colonization with Spn6B (Fig. 3). This analysis revealed several cytokines and chemokines with nominal differences ($p \leq 0.05$) between participants who became colonized and those who remained protected. However, after correction for multiple comparisons using the Benjamini–Hochberg method, none of these individual responses remained statistically significant. These results demonstrate that while baseline cytokine and chemokine activity differs modestly between groups, protection is unlikely to be driven by single mediators and may instead depend on coordinated multi-analyte immune signatures.

3.5. *Random Forest analysis identifies combined antibody and cytokine responses associated with protection against experimental pneumococcal carriage*

To explore potential multi-analyte immune signatures of protection, we applied a Random Forest classification algorithm [40] to baseline protein-specific antibody (IgG) and cytokine/chemokine datasets independently. The models were trained on multiple baseline immune features to predict susceptibility versus protection following experimental pneumococcal challenge. Among antibody responses, the combined baseline IgG levels to PdB, SP0899, and SP1069 emerged as the most informative predictors, with an out-of-bag (OOB) error rate of 36.05% (Supplementary Table 3). Similarly, for cytokine and chemokine responses, IL-17A production in response to SP0648-3 and MCP-1 production in response to SP1069 and SP0899 were identified as the most relevant features, with an OOB error rate of 38.98% (Supplementary Table 4). The relative importance of these IgG antibody and cytokine and chemokine predictors is shown in Supplementary Fig. 4.

Notably, SP0899 and SP1069 were shared features across both humoral and cellular response models, consistently ranking among the top predictors of protection (Fig. 4). Participants protected from experimental colonization exhibited higher baseline IgG responses to SP1069 and SP0899, and higher MCP-1 production following stimulation with these same antigens ($p = 0.035$ and $p = 0.049$, respectively; Fig. 4). Together, these findings indicate that integrated patterns of antibody and cytokine activity, rather than isolated immune measures, best predict protection to pneumococcal colonization.

3.6. *Correlation between antibody and cytokine/chemokine responses to machine learning–selected protein antigens*

To investigate the relationship between humoral and cellular

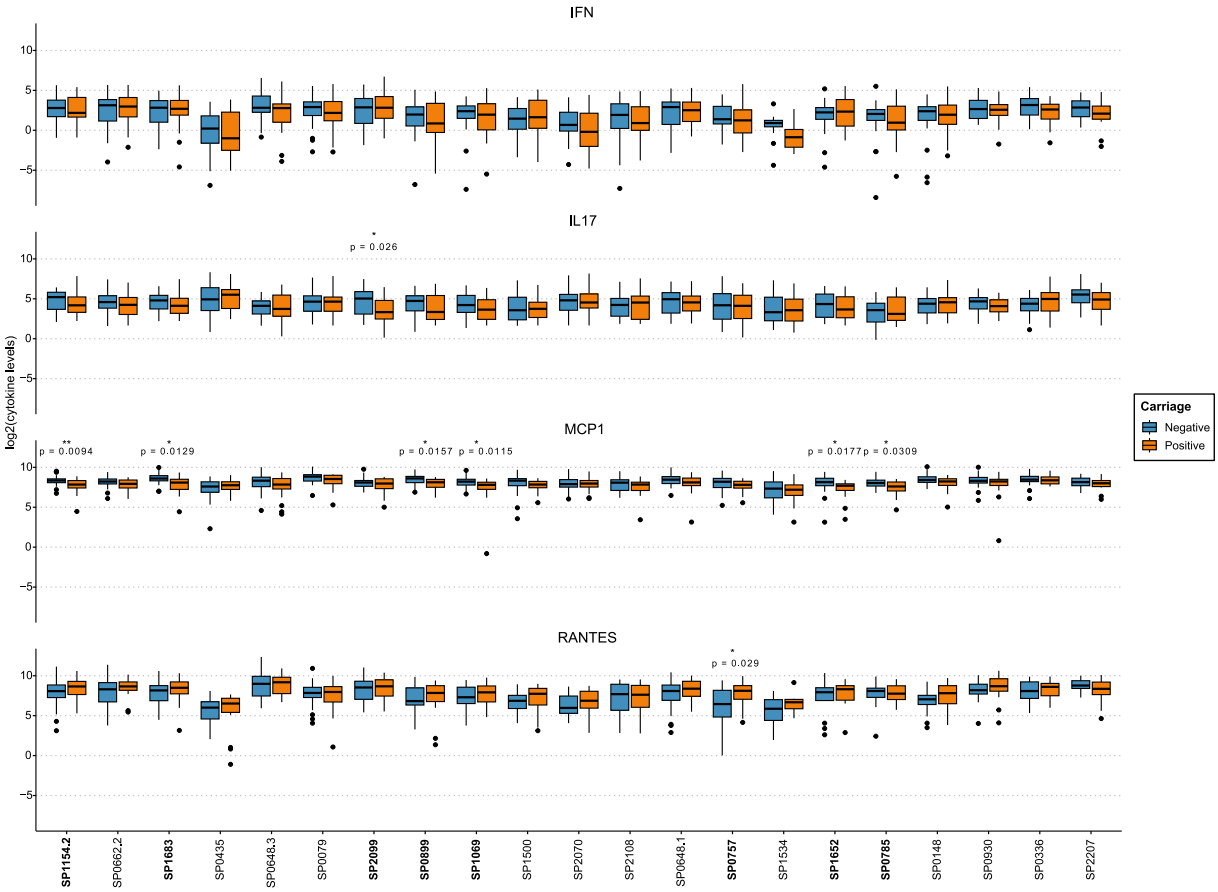

Fig. 3. Cytokine/chemokine responses to pneumococcal proteins. Baseline levels of cytokines/chemokines produced by PBMCs stimulated with pneumococcal proteins in healthy adults aged 18–50. Bars indicate interquartile range. Spn6B carriage-negative (blue) $n = 32$, Spn6B carriage-positive (orange) $n = 31$. Responses with p-value ≤ 0.05 prior to Benjamini-Hochberg correction for multiple comparisons are shown; this significance was lost following correction. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

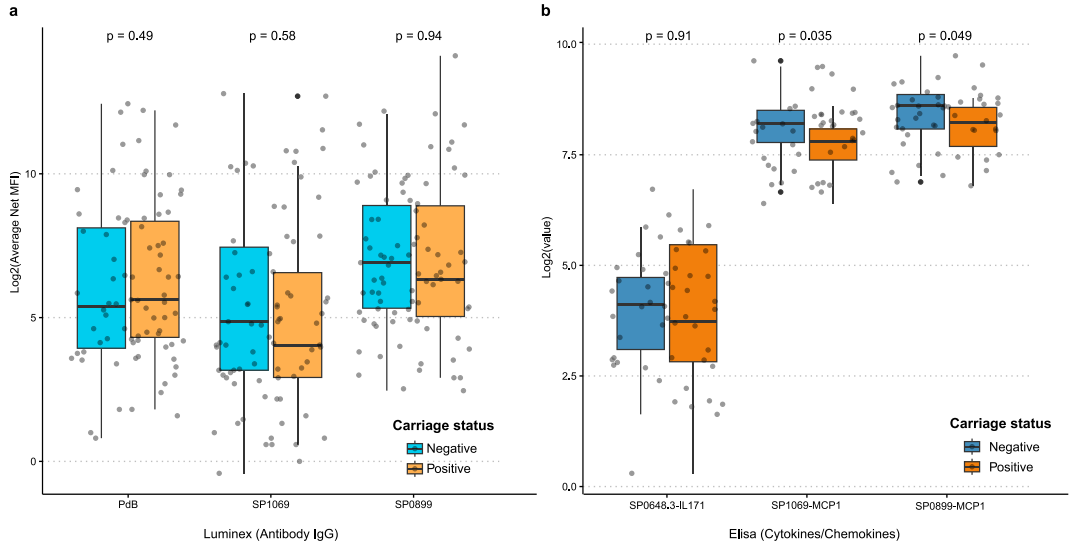

Fig. 4. Humoral and cellular responses to the machine learning best ranked proteins antigens. Boxplots showing (A) Log2 transformed average net MFI for Spn carriage-positive (orange; $n = 36$) and Spn carriage-negative (blue; $n = 50$) participants and (B) Log2 transformed cytokine/chemokine concentration for Spn carriage-positive (orange; n = 31) and Spn carriage-negative (blue; n = 32) participants. Each dot represents data from an individual participant. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

immune responses, we examined correlations between baseline antibody and cytokine/chemokine levels for the proteins identified by machine learning as predictive of colonization outcome. Among participants who became colonized, antibody and cytokine responses to

SP0648-3 were correlated across datasets ($\rho = 0.614$; $p = 0.009$), whereas SP1069 ($\rho = 0.378$; $p = 0.203$) and SP0899 ($\rho = -0.242$; $p = 0.425$) showed no correlation. No significant correlations were observed among participants who remained uncolonized (SP0648-3: $\rho = -0.076$; $p = 0.781$; SP1069: $\rho = -0.422$; $p = 0.133$; SP0899: $\rho = -0.098$; $p = 0.781$).

The lack of strong correlations suggests that antibody and cytokine responses capture distinct, nonredundant aspects of immunity, and their independent selection by the Random Forest model highlights complementary contributions to protection.

3.7. Correlation between protein-responses and AUC

To investigate which baseline protein-specific antibody and cytokine/chemokine responses associate with the magnitude of bacterial colonization, we correlated these immune measures with the area under the curve (AUC) of nasopharyngeal bacterial density (Fig. 5). This AUC-based analysis complements the binary carriage classification, because low AUC values may reflect partial immune control or an immunizing exposure rather than full susceptibility. Although no correlations remained significant after Benjamini–Hochberg correction, several nominal associations ($p \leq 0.05$) were observed (Fig. 5; Supplementary Tables 5 and 6). Notably, higher baseline IgG and selected cytokine responses to multiple pneumococcal proteins tended to associate with lower colonization AUC, consistent with the machine-learning results that implicate combined humoral and cellular signatures in limiting bacterial density (Fig. 6; Supplementary Fig. 4).

Participants were stratified by colonization density and duration (AUC) (Supplementary Fig. 5), and differences in RANTES, MCP-1, and IFN-γ production in response to protein stimuli (SP1154-2, SP0648-3, SP0079, and SP1683) were evident across AUC-defined groups (Supplementary Fig. 5B).

While not statistically significant after correction, higher baseline antibody and cytokine/chemokine responses to multiple pneumococcal proteins tended to associate with lower colonization density (Fig. 6). Elevated IgG responses were observed to 32 proteins (SP0321, SP0564, SP0346, SP1545, SP1032, SP0617, SP0620, SP0678, SP1479, SP1492, SP0079, SP0402, SP0629, SP0757, SP2145, SP0662-1, SP1683, SP0878, SP0582, SP1376, SP0724, SP1500, SP2083, SP0098, SP0127, SP0604, SP0648-1, SP2207, SP0742, SP0314, SP0648-2, PspA1) among participants who were either protected against experimental carriage acquisition, or in those who were experimentally colonized at a relatively low bacterial density (Fig. 6A). We also observed that cytokine and chemokine responses were elevated in participants who were either protected against experimental carriage acquisition or in the low AUC density group, followed by the mid AUC density and high AUC density groups – IFN-γ and RANTES responses were particularly elevated (Fig. 6B). Cytokine and chemokine levels in carriage-negative participants were generally not higher than those observed in the low AUC group. From the selected proteins, only IFN-γ and RANTES responses to SP1154-2 (IFN-γ $p = 0.034$; RANTES $p = 0.024$) and SP0079 (IFN-γ $p = 0.012$; RANTES $p = 0.005$) and the RANTES response to SP0648-1 ($p = 0.027$) have shown a $p \leq 0.05$ before Benjamini-Hochberg correction (Fig. 6B).

Overall, IgG responses to pneumococcal proteins were highest in carriage-negative participants and progressively lower with increasing colonization density, consistent with a protective role for antibodies. In contrast, cytokine and chemokine responses did not consistently follow this pattern; several, including IFN-γ and RANTES, were elevated in the low-AUC group, suggesting a role in the control of bacterial density rather than complete prevention of colonization. Consistent with the machine learning analysis, SP1069 and SP0899—identified as key predictive features for carriage status—showed similar trends: IFN-γ responses to SP1069 were highest in carriage-negative or low-density participants, while MCP-1 responses to SP0899 varied only modestly across groups. Together, these results indicate that baseline humoral responses are primarily associated with protection from colonization, whereas cellular cytokine activity contributes to limiting bacterial load, reflecting partial immune control rather than a simple protected/susceptible distinction.

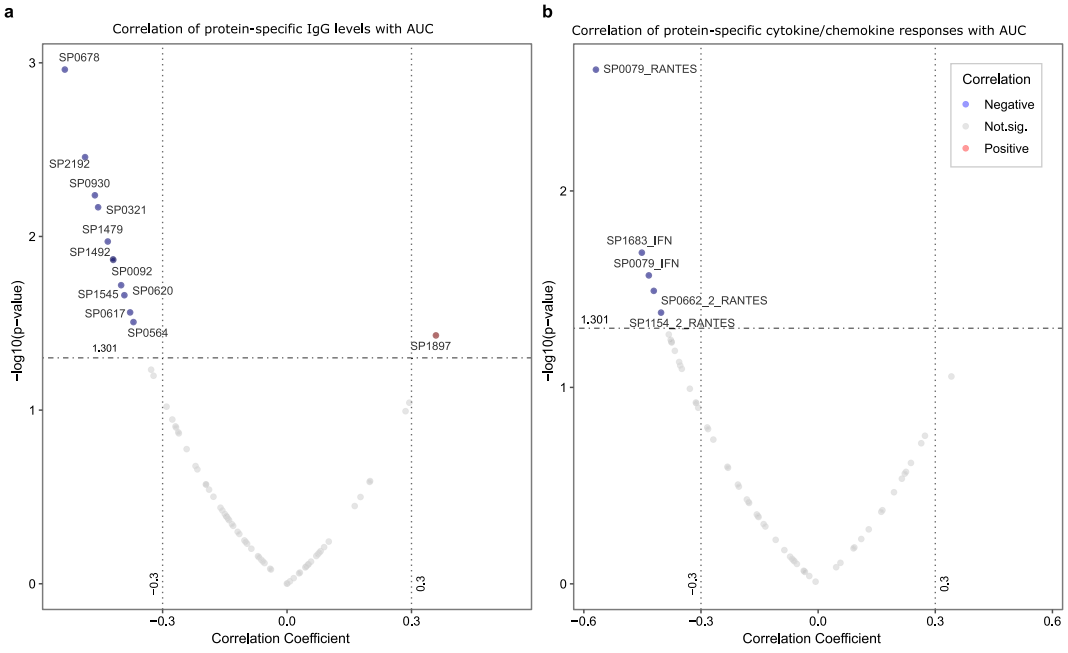

Fig. 5. Correlation of protein-specific IgG levels and cytokine/chemokine responses with colonizing density (AUC). Correlation results obtained between (A) antibody ($n = 34$) and (B) cytokine/chemokine ($n = 27$) levels for each protein and the AUC values obtained at the interval between days 2 to 14 post-challenge for each Spn6B carriage-positive participant. Proteins with p-value ≤0.05 (denoted by horizontal dashed line) were highlighted, blue dots represent protein responses negatively correlated with AUC, while red dots represent the positive correlations. Following Benjamini Hochberg correction for multiple comparisons, all significance was lost. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

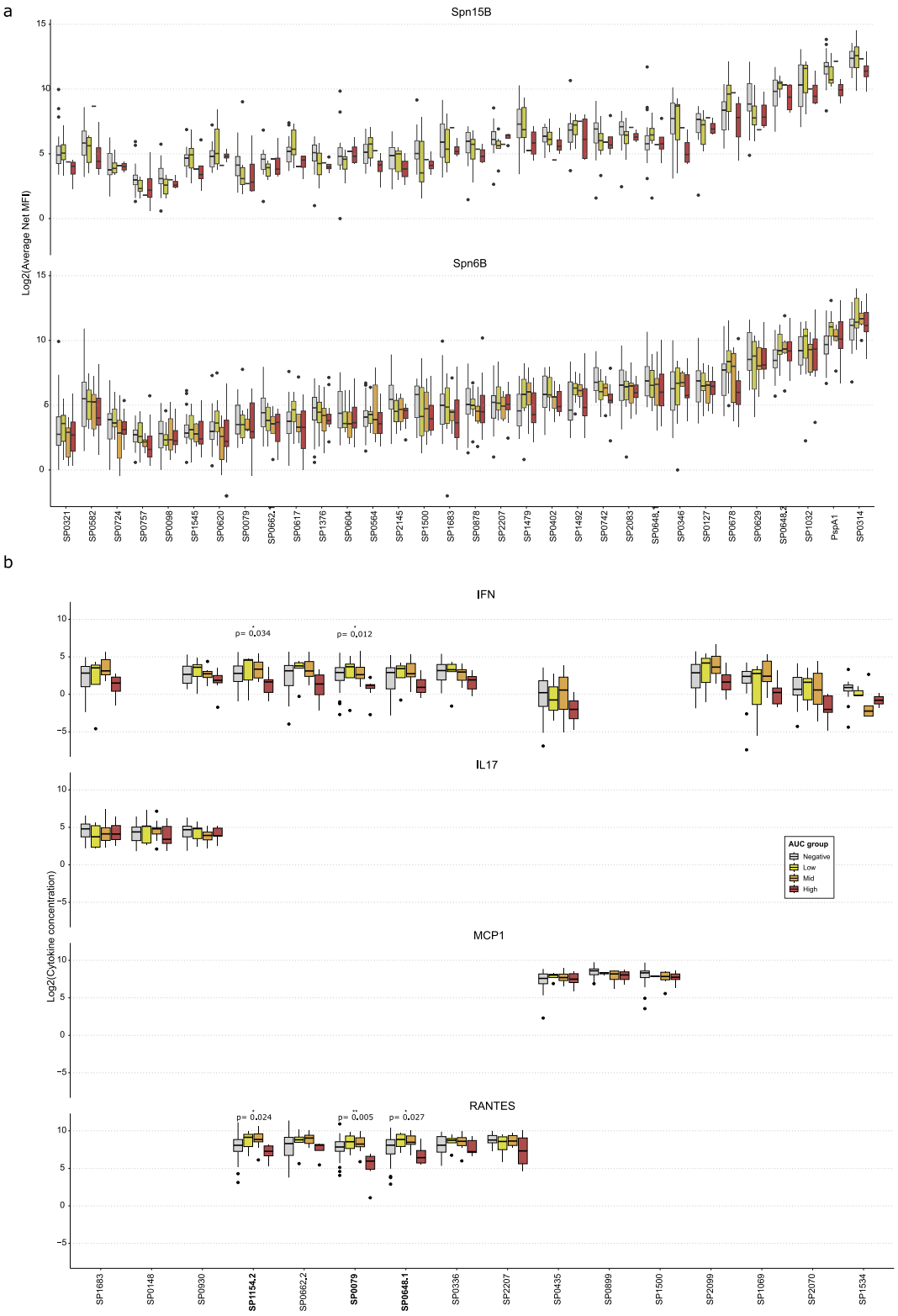

Fig. 6. Humoral and cellular responses across AUC groups for selected proteins showing trend-specific patterns. (A) Baseline levels of anti-protein serum IgG in healthy adults aged 18–50, measured as average net median fluorescence intensity (MFI). Bars indicate interquartile range. Spn15B carriage-negative n = 15, Spn15B carriage-positive $n = 13$, low AUC $n = 8$, mid n = 1 and high $n = 4$; Spn6B carriage-negative n = 18, Spn6B carriage-positive n = 21, low AUC group n = 4, mid $n = 6$ and high $n = 11$. (B) Baseline levels of selected cytokines/chemokines stimulated by pneumococcal proteins in PBMCs from healthy adults aged 18–50. Carriage-negative - grey; low AUC - yellow, mid AUC - orange; high AUC - red. Bars indicate interquartile range. Spn6B carriage-negative n = 18, Spn6B carriage-positive n = 27, low AUC group n = 4, mid $n = 14$ and high $n = 9$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

Machine learning and systems biology are increasingly employed in vaccine development to identify and design vaccine candidates as well as to identify correlates of successful immunization, predict individual vaccine responses, and correlates of protection when a disease/infection endpoint is possible [30,43]. As the number of measurable biological parameters expands, traditional univariate analyses are inappropriate, increasing the risk of type II errors (failing to detect true associations). In contrast, machine learning enables a more comprehensive and rigorous analysis of complex, high-dimensional datasets, offering deeper insights into the multifactorial nature of immune responses. In this study, we leveraged machine learning to identify immune features predictive of protection against nasal colonization. Our findings highlight the Random Forest algorithm's utility in identifying proteins linked to protection against experimental pneumococcal carriage, demonstrating its potential to accelerate vaccine development.

The ultimate aim of pneumococcal vaccination may not be sterilising immunity and complete elimination of nasopharyngeal colonization, a natural immunizing event [30,42,44,45], but rather reducing colonization density to prevent disease progression or transmission to others in the population, particularly during viral co-infections [46]. Suppressing bacterial density through protein-based vaccines could minimize transmission and reinforce immunity while maintaining mucosal exposure necessary for lung immunity [47].

Consistent with prior studies [30,44,48], baseline systemic IgG response to individual pneumococcal proteins did not correlate with protection against experimental colonization by serotype 6B or 15B pneumococcus. Similarly, baseline cell-mediated protein-specific responses in blood were not predictive of protection in our human cohort, despite evidence of Th17-mediated immunity against pneumococcal colonization in murine models [24].

Using machine learning, we identified two overlapping groups of antigens associated with reduced susceptibility to pneumococcal colonization: SP1069, SP0899, and PdB, which elicited IgG responses, and SP1069, SP0899, and SP0648-3, which stimulated protein-specific cellular responses. Notably, SP1069 and SP0899 induced both humoral and cellular immunity, highlighting their potential as dual-acting antigens. Although the precise functions of these proteins remain incompletely characterized, immunization studies in mice have shown that SP1069, SP0899, or SP0648-3 as individual protein antigens when co-administered with cholera toxin has led to a significant reduction in pneumococcal colonization compared with the control groups [24]. Co-administration of multiple antigens with a cholera toxin adjuvant has been shown to enhance the protective efficacy in murine models [24].

These findings suggest that protein-based vaccines incorporating multiple antigens may provide broader, serotype-independent protection, particularly if conjugated to capsular polysaccharides in next-generation PCV formulations [49,50]. Such strategies may also confer the added benefit of inducing cellular immunity alongside antibody-mediated protection.

In humans, protection against pneumococcal disease requires both humoral and cellular immunity [23,51–56]. Similarly, murine models indicate the necessity of both immunity types for nasopharyngeal colonization defence [24,27,51–53,57]. Using our CHIM, we previously reported that colonization-induced MCP-1 (CCL2) levels are associated with monocyte recruitment to the nasal mucosa and subsequent clearance of colonization [28]. These results also support previous findings from mouse models showing that CCL2 signalling and monocyte recruitment are key mediators of pneumococcal carriage clearance [26]. Elevated MCP-1 levels in protected participants against experimental acquisition of Spn6B and its involvement in monocyte and neutrophil recruitment underline its significance in pneumococcal clearance [28,58,59].

In our cohort, baseline anti-protein IgG and cytokine responses correlated with lower colonization density, suggesting roles for both humoral and cellular immunity in controlling carriage and preventing disease. Consistent with prior findings [60–63], IFN-γ and RANTES secretion correlated with reduced colonization density, likely enhancing innate immune cell recruitment and activity [64–66]. These results emphasize the importance of designing protein vaccines that elicit both antibody and cellular responses for effective protection.

While promising, this work has limitations. Firstly, we acknowledge limited sample sizes, which may have reduced our power to detect significant effects through our statistical analysis. Despite limited sample size, relevant effects were found using the random forest model. Out-of-bag error rates (~36–39%) indicate moderate predictive power, suggesting that further optimization, such as increasing cohort size or incorporating mucosal immune parameters, may be needed to improve model accuracy. Furthermore, because pneumococcal colonizing density of Spn15B was substantially lower than that of Spn6B, it would have been inappropriate to use the AUC thresholds established from the larger Spn6B cohort ($n = 128$). We therefore defined a cohort-specific AUC threshold for Spn15B to enable qualitative comparison of carriage dynamics between serotypes. As no prior datasets exist for Spn15B, this threshold could not be validated externally, and conclusions from the Spn15B AUC analysis should be interpreted cautiously due to the small sample size ($n = 9$).

Alternative analytical approaches could complement the present work. Time-to-event analyses of carriage loss may be informative for clearance kinetics; however, the study was designed to identify baseline immune features associated with acquisition and colonization density, and the limited cohort size and sampling frequency constrained the applicability of survival-based models. Shapley-based explanation methods provide local interpretability of machine learning models, but in small, correlated immunological datasets they may offer limited additional insight beyond global importance measures. Accordingly, permutation-based variable importance was prioritized for robustness and interpretability, while these approaches remain well suited for future studies with larger cohorts.

Immune responses to protein antigens in systemic circulation may differ from those in the nasal mucosa, and the study's small sample size, comprising mainly healthy young adults, may not reflect high-risk demographics. Validation in older adults, who represent a high-risk demographic, will be important [44,67]. Immune correlates may vary across populations with differing pneumococcal burden. Our studies in Malawi [10,68] will enable validation across diverse populations and environmental contexts. Further investigation into antigen-specific nasal T-cell responses and sex-stratified immune differences could provide additional insights [36].

As immunology and vaccinology increasingly rely on complex, high-dimensional datasets, predictive computational models offer substantial advantages over traditional analyses by uncovering intricate relationships between immune responses and protection. The integration of advanced computational approaches with human challenge models, which uniquely allow sampling immediately before and after defined pathogen exposure, promises to transform how we define protective immunity and accelerate rational vaccine design.

Despite the remarkable success of vaccines, which are estimated to have saved more than 150 million lives over the past 50 years [69], major challenges persist. Unequal access to vaccines and the biological complexity of difficult-to-target pathogens continue to slow progress in reducing the global burden of infectious diseases. A key scientific challenge is the identification of causal correlates of protection for high-frequency colonizing pathogens such as *Streptococcus pneumoniae*, *Staphylococcus aureus*, and *Escherichia coli*, where repeated antigen exposures at mucosal surfaces shape layered humoral and cellular immune responses. Addressing these gaps will require integrating advanced computational approaches, experimental human models, and global public health strategies to translate immunological insights into next-generation vaccines.

Authors contribution

Conceived the study - DMF, SPJ, CS, BU, RM.
Data acquisition - KSC, CS, RT, JR, ELG, EM, EN, YJL, SJP, ENM.
Generated the ELISA data - CS.
Performed the machine learning analysis - PGD.
Performed data analysis - KSC, PGD, CS.
Scientific discussion - AJP, SBG, HIN, SHP, ANAG, AMC, SPJ.
Project definition and discussion - KSC, PGD, CS, DMF, BU.
Provided protein library - RM, YJL, EM, RT.

PGD, KSC and CS jointly interpreted the results and wrote the manuscript with contributions from all co-authors. All authors read and approved the final manuscript.

CRediT authorship contribution statement

Katerina S. Cheliotis: Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Patricia Gonzalez-Dias:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Esther L. German:** Writing – review & editing, Methodology, Data curation, Conceptualization. **André N.A. Gonçalves:** Writing – review & editing, Visualization, Formal analysis. **Elena Mitsi:** Writing – review & editing, Data curation. **Elissavet Nikolaou:** Writing – review & editing, Data curation. **Sherin Pojar:** Writing – review & editing, Data curation. **Eliane N. Miyaji:** Writing – review & editing, Methodology, Investigation, Data curation. **Rafaella Tostes:** Writing – review & editing, Methodology, Data curation. **Jesús Reiné:** Writing – review & editing, Methodology, Data curation. **Andrea M. Collins:** Writing – review & editing, Investigation. **Helder I. Nakaya:** Writing – review & editing, Methodology, Investigation. **Stephen B. Gordon:** Investigation. **Ying-Jie Lu:** Writing – review & editing, Investigation, Data curation. **Shaun H. Pennington:** Writing – review & editing, Methodology, Investigation. **Andrew J. Pollard:** Writing – review & editing, Investigation. **Richard Malley:** Writing – review & editing, Investigation, Conceptualization. **Simon P. Jochems:** Writing – review & editing, Funding acquisition, Data curation, Conceptualization. **Britta Urban:** Writing – review & editing, Supervision, Conceptualization. **Carla Solórzano:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Daniela M. Ferreira:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: RM is a consultant to GSK and Merck, is a named inventor and patent holder on vaccine technologies, a member of the board of directors at Corner Therapeutics and the scientific advisory boards of Amplitude Therapeutics, Limmatech and Vitrivax. Y-JL is a consultant to GSK and is a named inventor and patent holder on vaccine technologies. All the other authors have declared that no conflicts of interest exist.

Acknowledgments

The authors thank all participants in the EHPC studies, as well as the Clinical Research Network North West Coast and the NIHR Comprehensive Local Research Network North West Coast for their continued support. We also extend our gratitude to the study clinical team members for their assistance in sample collection. This work by funded by grants from the Bill & Melinda Gates Foundation (OPP1117728) (to DMF), Medical Research Council (MR/M011569/1) (to SBG and DMF), Robert Austrian Award (to SPJ) and the Bacterial Vaccines (BactiVac) Network funded by the GCRF Networks in Vaccines Research and Development which was co-funded by the MRC and BBSRC (to DMF and CS). KSC was supported by the MRC Doctoral Training Programme (Award 1964515).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.vaccine.2026.128280.

Data availability

Data are available upon reasonable request by email directed to the corresponding authors at daniela.ferreira@paediatrics.ox.ac.uk and carla.solorzanogonzalez@paediatrics.ox.ac.uk

References

[1] GBD 2019 Antimicrobial Resistance Collaborators. Global mortality associated with 33 bacterial pathogens in 2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2022;400:2221–48.
[2] Wahl B, et al. Burden of *Streptococcus pneumoniae* and *Haemophilus influenzae* type b disease in children in the era of conjugate vaccines: global, regional, and national estimates for 2000-15. Lancet Glob Health 2018;6:e744–57.
[3] McAllister DA, et al. Global, regional, and national estimates of pneumonia morbidity and mortality in children younger than 5 years between 2000 and 2015: a systematic analysis. Lancet Glob Health 2019;7:e47–57.
[4] Weiser JN, Ferreira DM, Paton JC. *Streptococcus pneumoniae*: transmission, colonization and invasion. Nat Rev Microbiol 2018;16:355–67.
[5] Ladhani SN, Andrews N, Ramsay ME. Summary of evidence to reduce the two-dose infant priming schedule to a single dose of the 13-valent pneumococcal conjugate vaccine in the national immunisation programme in the UK. Lancet Infect Dis 2021;21:e93–102.
[6] Black S, et al. Efficacy, safety and immunogenicity of heptavalent pneumococcal conjugate vaccine in children. Northern California Kaiser Permanente Vaccine Study Center Group. Pediatr Infect Dis J 2000;19:187–95.
[7] Ota MOC, et al. Pneumococcal antibody concentrations of subjects in communities fully or partially vaccinated with a seven-valent pneumococcal conjugate vaccine. PloS One 2012;7:e42997.
[8] Collins AM, et al. First human challenge testing of a pneumococcal vaccine. Double-blind randomized controlled trial. Am J Respir Crit Care Med 2015;192:853–8.
[9] Nagel J, et al. The association between antibody levels before and after 7-valent pneumococcal conjugate vaccine immunization and subsequent pneumococcal infection in chronic arthritis patients. Arthritis Res Ther 2015;17:124.
[10] Bar-Zeev N, et al. Impact and effectiveness of 13-valent pneumococcal conjugate vaccine on population incidence of vaccine and non-vaccine serotype invasive pneumococcal disease in Blantyre, Malawi, 2006-18: prospective observational time-series and case-control studies. Lancet Glob Health 2021;9:e989–98.
[11] Singh A, Dutta AK. Pneumococcal vaccines - how many serotypes are enough? Indian J Pediatr 2018;85:47–52.
[12] Weinberger DM, Malley R, Lipsitch M. Serotype replacement in disease after pneumococcal vaccination. Lancet 2011;378:1962–73.
[13] Olaya-Abril A, Jiménez-Munguía I, Gómez-Gascón L, Obando I, Rodríguez-Ortega MJ. Identification of potential new protein vaccine candidates through pan-surfomic analysis of pneumococcal clinical isolates from adults. PloS One 2013;8:e70365.
[14] Tiley KS, et al. Nasopharyngeal carriage of pneumococcus in children in England up to 10 years after 13-valent pneumococcal conjugate vaccine introduction: persistence of serotypes 3 and 19A and emergence of 7C. J Infect Dis 2023;227:610–21.
[15] Pick H, et al. Pneumococcal serotype trends, surveillance and risk factors in UK adult pneumonia, 2013-18. Thorax 2020;75:38–49.
[16] Ladhani SN, et al. Rapid increase in non-vaccine serotypes causing invasive pneumococcal disease in England and Wales, 2000-17: a prospective national observational cohort study. Lancet Infect Dis 2018;18:441–51.
[17] Chen C, et al. Effect and cost-effectiveness of pneumococcal conjugate vaccination: a global modelling analysis. Lancet Glob Health 2019;7:e58–67.
[18] Lagousi T, Basdeki P, Routsias J, Spoulou V. Novel protein-based pneumococcal vaccines: assessing the use of distinct protein fragments instead of full-length proteins as vaccine antigens. Vaccines (Basel) 2019;7.
[19] Introduction. Recommendations to assure the quality, safety and efficacy of pneumococcal conjugate vaccines.
[20] Ryman J, et al. Predicting vaccine effectiveness against invasive pneumococcal disease in children using immunogenicity data. npj Vaccines 2022;7:140.
[21] Grant LR, et al. Persistence of IgG antibody following routine infant immunization with the 7-valent pneumococcal conjugate vaccine. Pediatr Infect Dis J 2015;34:e138–42.

[22] Britto C, Alter G. The next frontier in vaccine design: blending immune correlates of protection into rational vaccine design. Curr Opin Immunol 2022;78:102234.
[23] Lu Y-J, et al. Interleukin-17A mediates acquired immunity to pneumococcal colonization. PLoS Pathog 2008;4:e1000159.
[24] Lu Y-J, et al. Screening for Th17-dependent pneumococcal vaccine antigens: comparison of murine and human cellular immune responses. Infect Immun 2018;86.
[25] Moffitt KL, et al. T(H)17-based vaccine design for prevention of *Streptococcus pneumoniae* colonization. Cell Host Microbe 2011;9:158–65.
[26] Zhang Z, Clarke TB, Weiser JN. Cellular effectors mediating Th17-dependent clearance of pneumococcal colonization in mice. J Clin Invest 2009;119:1899–909.
[27] Moffitt KL, Malley R, Lu Y-J. Identification of protective pneumococcal T(H)17 antigens from the soluble fraction of a killed whole cell vaccine. PloS One 2012;7:e43445.
[28] Jochems SP, et al. Inflammation induced by influenza virus impairs human innate immune control of pneumococcus. Nat Immunol 2018;19:1299–308.
[29] Rylance J, et al. Two randomized trials of the effect of live attenuated influenza vaccine on pneumococcal colonization. Am J Respir Crit Care Med 2019;199:1160–3.
[30] Ferreira DM, et al. Controlled human infection and rechallenge with *Streptococcus pneumoniae* reveals the protective efficacy of carriage in healthy adults. Am J Respir Crit Care Med 2013;187:855–64.
[31] Bogaert D, De Groot R, Hermans PWM. *Streptococcus pneumoniae* colonization: the key to pneumococcal disease. Lancet Infect Dis 2004;4:144–54.
[32] Weiser JN. The pneumococcus: why a commensal misbehaves. J Mol Med 2010;88:97–102.
[33] Paton JC, et al. Purification and immunogenicity of genetically obtained pneumolysin toxoids and their conjugation to *Streptococcus pneumoniae* type 19F polysaccharide. Infect Immun 1991;59:2297–304.
[34] Ferreira DM, et al. DNA vaccines expressing pneumococcal surface protein A (PspA) elicit protection levels comparable to recombinant protein. J Med Microbiol 2006;55:375–8.
[35] Darrieux M, et al. Fusion proteins containing family 1 and family 2 PspA fragments elicit protection against *Streptococcus pneumoniae* that correlates with antibody-mediated enhancement of complement deposition. Infect Immun 2007;75:5930–8.
[36] Cheliotis KS, et al. Influence of sex, season and environmental air quality on experimental human pneumococcal carriage acquisition: a retrospective cohort analysis. ERJ Open Research 2022;8.
[37] Identifying Novel Pneumococcal Pneumonia Vaccine Candidates - LSTM Online Archive. https://archive.lstmed.ac.uk/21171/.
[38] Bjornstal O, et al. The impact of decreased bead count to determine concentrations of amyloid beta1-42, total-tau, and phosphorylated-tau181 in human cerebrospinal fluid using xMAP technology. J Pharm Sci 2011;100:4655–63.
[39] van Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. J Stat Softw 2011;45(3):1–67.
[40] Breiman L. Random forests45. Springer Science and Business Media LLC; 2001. p. 5–32.
[41] Keech CA, et al. A phase 1 randomized, placebo-controlled, observer-blinded trial to evaluate the safety and immunogenicity of inactivated *Streptococcus pneumoniae* whole-cell vaccine in adults. Pediatr Infect Dis J 2020;39:345–51.
[42] Mitsi E, et al. Nasal pneumococcal density is associated with microaspiration and heightened human alveolar macrophage responsiveness to bacterial pathogens. Am J Respir Crit Care Med 2020;201:335–47.
[43] Oberg AL, Kennedy RB, Li P, Ovsyannikova IG, Poland GA. Systems biology approaches to new vaccine development. Curr Opin Immunol 2011;23:436–43.
[44] Adler H, et al. Experimental human pneumococcal colonization in older adults is feasible and safe, not immunogenic. Am J Respir Crit Care Med 2021;203:604–13.
[45] German EL, et al. Protective effect of PCV vaccine against experimental pneumococcal challenge in adults is primarily mediated by controlling colonization density. Vaccine 2019;37:3953–6.
[46] Carr OJJ, et al. Nasopharyngeal pneumococcal colonization density is associated with severe pneumonia in young children in the Lao People's Democratic Republic. J Infect Dis 2022;225:1266–73.
[47] Khan MN, Xu Q, Pichichero ME. Protection against *Streptococcus pneumoniae* invasive pathogenesis by a protein-based vaccine is achieved by suppression of nasopharyngeal bacterial density during influenza a virus coinfection. Infect Immun 2017;85.
[48] Araujo AP, et al. Serum levels of anti-PspA and anti-PspC IgG decrease with age and do not correlate with susceptibility to experimental human pneumococcal colonization. PloS One 2021;16:e0247056.
[49] Chichili GR, et al. Phase 1/2 study of a novel 24-valent pneumococcal vaccine in healthy adults aged 18 to 64 years and in older adults aged 65 to 85 years. Vaccine 2022;40:4190–8.
[50] Kaplonek P, et al. A semisynthetic glycoconjugate provides expanded cross-serotype protection against *Streptococcus pneumoniae*. Vaccine 2022;40:1038–46.
[51] McCool TL, Weiser JN. Limited role of antibody in clearance of *Streptococcus pneumoniae* in a murine model of colonization. Infect Immun 2004;72:5807–13.
[52] Malley R, et al. CD4+ T cells mediate antibody-independent acquired immunity to pneumococcal colonization. Proc Natl Acad Sci U S A 2005;102:4848–53.
[53] van Rossum AMC, Lysenko ES, Weiser JN. Host and bacterial factors contributing to the clearance of colonization by *Streptococcus pneumoniae* in a murine model. Infect Immun 2005;73:7718–26.
[54] Cohen JM, et al. Protective contributions against invasive *Streptococcus pneumoniae* pneumonia of antibody and Th17-cell responses to nasopharyngeal colonization. PloS One 2011;6:e25558.
[55] Wilson R, et al. Protection against *Streptococcus pneumoniae* lung infection after nasopharyngeal colonization requires both humoral and cellular immune responses. Mucosal Immunol 2015;8:627–39.
[56] Wilson R, et al. Naturally acquired human immunity to pneumococcus is dependent on antibody to protein antigens. PLoS Pathog 2017;13:e1006137.
[57] Trzciński K, et al. Effect of serotype on pneumococcal competition in a mouse colonization model. MBio 2015;6:e00902–15.
[58] Nikolaou E, et al. Experimental human challenge defines distinct pneumococcal kinetic profiles and mucosal responses between colonized and non-colonized adults. MBio 2021;12.
[59] Carr MW, Roth SJ, Luther E, Rose SS, Springer TA. Monocyte chemoattractant protein 1 acts as a T-lymphocyte chemoattractant. Proc Natl Acad Sci U S A 1994;91:3652–6.
[60] Trzciński K, et al. Protection against nasopharyngeal colonization by *Streptococcus pneumoniae* is mediated by antigen-specific CD4+ T cells. Infect Immun 2008;76:2678–84.
[61] Hoe E, et al. Reduced IL-17A secretion is associated with high levels of pneumococcal nasopharyngeal carriage in Fijian children. PloS One 2015;10:e0129199.
[62] Vuononvirta J, Peltola V, Ilonen J, Mertsola J, He Q. The gene polymorphism of IL-17 G-152A is associated with increased colonization of *Streptococcus pneumoniae* in young Finnish children. Pediatr Infect Dis J 2015;34:928–32.
[63] Lu Y-J, et al. GMP-grade pneumococcal whole-cell vaccine injected subcutaneously protects mice from nasopharyngeal colonization and fatal aspiration-sepsis. Vaccine 2010;28:7468–75.
[64] Soudja SM, et al. Memory-T-cell-derived interferon-γ instructs potent innate cell activation for protective immunity. Immunity 2014;40:974–88.
[65] Mancuso RI, et al. Impaired expression of CXCL5 and matrix metalloproteinases in the lungs of mice with high susceptibility to *Streptococcus pneumoniae* infection. Immun Inflamm Dis 2018;6:128–42.
[66] Shenoy AT, et al. Lung CD4+ resident memory T cells remodel epithelial responses to accelerate neutrophil recruitment during pneumonia. Mucosal Immunol 2020;13:334–43.
[67] Urban BC, et al. Inflammation of the nasal mucosa is associated with susceptibility to experimental pneumococcal challenge in older adults. Mucosal Immunol 2024;17:973–89.
[68] Morton B, et al. A feasibility study of controlled human infection with *Streptococcus pneumoniae* in Malawi. EBioMedicine 2021;72:103579.
[69] Shattock AJ, et al. Contribution of vaccination to improved survival and health: modelling 50 years of the Expanded Programme on Immunization. Lancet 2024;403:2307–16.