



Universiteit
Leiden
The Netherlands

Metabolomics in community-acquired pneumonia: exploring metabolomics-based biomarkers for diagnosis and treatment response monitoring of community-acquired pneumonia

Hartog, I. den

Citation

Hartog, I. den. (2024, September 17). *Metabolomics in community-acquired pneumonia: exploring metabolomics-based biomarkers for diagnosis and treatment response monitoring of community-acquired pneumonia*. Retrieved from <https://hdl.handle.net/1887/4083598>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4083598>

Note: To cite this publication please use the final published version (if applicable).

Metabolomics in community-acquired pneumonia

Exploring metabolomics-based biomarkers for
diagnosis and treatment response monitoring of
community-acquired pneumonia

Ilona den Hartog

Manuscript

Cover design: DALL·E 2, I. den Hartog , A. Kamer

Layout: I. den Hartog

Printed by: Ridderprint BV

Copyright

© I. den Hartog, 2024

All rights reserved. No part of this thesis may be reproduced in any form or by any means without permission of the author.

Funding

This work is part of the research program ‘Metabolomic fingerprint biomarkers to guide antibiotic therapy and reduce resistance’ with project number 541001007, which is financed by ZonMW, the Netherlands Organization for Health Research and Development associated with the Dutch Research Council (NWO).

Metabolomics in community-acquired pneumonia

Exploring metabolomics-based biomarkers for
diagnosis and treatment response monitoring of
community-acquired pneumonia

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus, prof.dr.ir H. Bijl,
volgens het besluit van het college voor promoties
te verdedigen op dinsdag 17 september 2024
klokke 14:30 uur

door

Ilona den Hartog

geboren te Montfoort
in 1992

Promotores: Prof.dr. J.G.C. van Hasselt
Prof.dr. T. Hankemeier

Co-promotor: Dr. E.M.W. van de Garde (Universiteit Utrecht)

Promotiecommissie: Prof.dr. H. Irth
Prof.dr. E.C.M. de Lange
Prof.dr. M.G.J. de Boer (Leiden UMC)
Prof.dr. A. Geluk (Leiden UMC)
Dr. H. Endeman (Erasmus MC)

The research described in this thesis was performed at the Systems Pharmacology and Pharmacy division of the Leiden Academic Centre for Drug Research (LACDR), Leiden University, Leiden, The Netherlands.

Contents

1	Introduction	1
2	Metabolomic profiling of microbial disease etiology in community-acquired pneumonia	5
3	Differential metabolic host response to pathogens associated with community-acquired pneumonia	33
4	Longitudinal metabolite profiling of <i>Streptococcus pneumoniae</i>-associated community-acquired pneumonia	53
5	The Immunometabolic Atlas: a tool for design and interpretation of metabolomics studies in immunology	79
6	General discussion and summary	103
7	Nederlandse samenvatting	111
	Bibliography	115
	List of Abbreviations	127
	List of Publications	131
	List of Affiliations	133
	Dankwoord	135
	Curriculum Vitae	137

CHAPTER 1

Introduction

1.1 Community-acquired pneumonia

Community-acquired pneumonia (CAP) is a common infection of the lower respiratory tract that results in the hospitalization of approximately 1 in every 500 adults every year [1, 2]. CAP has the highest burden of mortality and morbidity in the elderly [3, 4] and is among the most common causes of sepsis in all age groups [5]. The most common cause of CAP is the bacterial pathogen *Streptococcus pneumoniae*, followed by *Mycoplasma pneumoniae*, and viral infections such as influenza [2, 6].

The clinical diagnosis of CAP is based on symptoms such as shortness of breath, cough, fever, and new-found focal chest signs, and can be confirmed by the presence of a visible lung infiltrate on a chest radiograph [3]. Several microbiological tests are available to establish the microbial etiological diagnosis of CAP, including pathogen culturing, antigen testing, and polymerase chain reaction (PCR)-based diagnostics [7]. These tests are not applied to every CAP patient but merely those with more severe symptoms or specific risk factors for worse outcomes. Microbial etiological diagnosis using current laboratory techniques can take up to 48 hours and remains inconclusive often. In over 50% of CAP patients, no causative pathogen is identified because a sputum sample cannot be obtained, or because microbiological testing yields no causative pathogen [6, 8].

1.2 Antibiotic treatment and antimicrobial resistance

Hospitalized patients with moderate to severe CAP typically receive empirical broad-spectrum antibiotic therapy to provide broad microbial coverage prior to further determination of potential causal pathogens. Empirical antibiotic treatment of CAP usually consists of a beta-lactam antibiotic, which in more severe cases can be combined with a macrolide or fluoroquinolone monotherapy [9]. If the causative pathogen and its susceptibility profiles have been determined, more targeted narrow-spectrum antibiotics can be selected, avoiding undesired effects on nonpathogenic commensal bacteria. Importantly, in a subset of patients receiving empirical antibiotic therapy,

viral pathogens such as influenza are responsible for the symptoms. In those patients, antibiotic treatment is not useful and should be avoided as antibiotic use promotes the emergence of antimicrobial resistance (AMR). In conclusion, the availability of rapid and well-performing diagnostic tools to guide microbiological etiological diagnosis of CAP is essential to ensure effective treatment with antibiotics while limiting the risks of promoting AMR.

1.3 Treatment response biomarkers

Biomarkers predictive of treatment response in individual (hospitalized) CAP patients are important to optimize the effectiveness of therapeutic strategies. Such longitudinal monitoring of the treatment response can guide decisions about adapting the antibiotic treatment if the therapy is working insufficiently or terminating the antibiotic treatment when the infection is successfully treated. Early termination of antibiotic treatment because the bacterial infection has been effectively treated likely reduces the risk for development of AMR [10, 11]. Monitoring the treatment response is commonly performed through a combination of monitoring the clinical symptoms, such as fever, and through measurement of biochemical markers reflecting inflammation, such as C-reactive protein (CRP) [12, 13]. While CRP is commonly used for this purpose, it has limitations in terms of its correlation to clinical outcomes such as length of hospitalization or mortality, its specificity towards infectious causes alone and not towards other non-infectious inflammatory conditions, and its correlation with the underlying disease progression dynamics. As such, an unmet need exists for additional biomarkers that can further improve effective monitoring of the clinical response in CAP patients to guide treatment strategies [14].

1.4 Metabolomics as platform for discovery of biomarkers

The field of metabolomics concerns the large-scale measurement of biomolecules, or metabolites, with a molecular weight of <2000 Dalton in tissues, cells, or body fluids [15]. Metabolomics allows for the identification of metabolites that are associated with a state of health or disease. Because the metabolome is closely related to the biochemical state of organisms, it is a relevant source of potential biomarkers for various diseases [16]. For diagnosis and treatment response monitoring in several infectious diseases, metabolomics approaches have been explored [17].

The measurement of metabolites as part of metabolomics workflows is performed using mass spectrometry (MS) or nuclear magnetic resonance (NMR) based techniques [17], where MS-based approaches are more commonly used due to improved sensitivity and selectivity as well as increased throughput [18]. Metabolomics studies can be designed as targeted or untargeted studies. Untargeted approaches are useful for hypothesis-free discovery of new metabolites associated to a certain health condition, while targeted metabolomics can help to quantify the levels of metabolites that are known to be

involved in certain biological processes [19]. Targeted metabolomics is therefore more relevant for the discovery of biomarkers to be used for diagnosis or treatment response monitoring in patients.

Currently, the microbial etiological diagnosis of CAP is established through tests focused on the causative pathogen. To identify new biomarkers for the etiological diagnosis of CAP, the specific host-response of the patient could also be an interesting target. The immunometabolome, which concerns the interplay between metabolism and immunology, could be a source of such biomarkers for viral or bacterial infections. The relevance of immunometabolomics has been shown in previous studies, for example, to separate tuberculosis patients from patients with community-acquired pneumonia and healthy controls [20], and to separate sepsis patients from emergency room controls [21]. Also, some small studies have shown the potential of metabolomics for pathogen identification. For example, patients with H1N1 influenza pneumonia could be discriminated from patients with bacterial pneumonia based on serum metabolite profiles of CAP patients admitted to the ICU [22], and CAP patients with *Streptococcus pneumoniae* could be differentiated from CAP patients with other bacterial and viral infections based on their urinary metabolite profile [23].

For monitoring the treatment response, metabolic biomarkers could have potential because the metabolome closely represents the current state of the patient, which may be highly dynamic. This has been shown, for example, in patients with chronic hepatitis B infections, of who the progression of the disease could be associated with increased concentrations of long-chain triglycerides, together with citrulline and ornithine [24]. Also, a study in CAP patients has demonstrated that the change in lysophosphatidylcholines (LPC) mirrors the transition from acute illness to recovery after the start of antibiotic treatment [25].

For biochemical and functional interpretation of metabolomics study results, published research on individual metabolites and the biological processes they are involved in, can be evaluated. In addition, several computational tools are available for biochemical and functional analysis [26, 27, 28]. However, determining the relationships between metabolism and immunological processes remains challenging because of the many possible interactions and sparse literature on these interactions. As such, there is a need for new methodologies to determine associations between metabolites and immune processes to aid in the biological interpretation of metabolomics data.

1.5 Scope of this thesis

For this thesis, our central hypothesis is that changes in the host metabolome associated with the immune response in patients with CAP may be a potential source for novel biomarkers. To this end, the overall aim of this thesis is to assess the potential utility of metabolomics-based biomarkers for the diagnosis and the monitoring of the treatment response in patients with CAP.

In **Chapter 2** we aimed to investigate if predictive metabolic biomarkers for the microbial etiological diagnosis of CAP could be identified. Serum samples from CAP patients with confirmed microbial etiologies: *S. pneumoniae*, atypical bacteria, or viral infections were analyzed using targeted mass-spectrometry-based metabolomics techniques. In **Chapter 3** we further studied the specific differences in the metabolic host response to distinct CAP-associated pathogens. We performed a systematic characterization of differential metabolite profiles for different pathogens, which can support evaluation of diagnostic performance and may contribute to insights in disease pathogenesis. In **Chapter 4** we aimed to characterize longitudinal metabolite profiles in 25 hospitalized CAP patients with *S. pneumoniae* to determine their potential relationship to disease severity and treatment response, and thereby their utility as treatment response biomarkers. Finally, in **Chapter 5** we describe the development of the Immunometabolic Atlas, which is a computational tool for the interpretation of metabolomics data to aid in the design and interpretation of metabolomics studies.

CHAPTER 2

Metabolomic profiling of microbial disease etiology in community-acquired pneumonia

Ilona den Hartog, Laura B. Zwep, Stefan M.T. Vestjens, Amy C. Harms, G. Paul Voorn, Dylan W. de Lange, Willem J.W. Bos, Thomas Hankemeier, Ewoudt M.W. van de Garde, J.G. Coen van Hasselt. Metabolomic profiling of microbial disease etiology in community-acquired pneumonia, *PLoS One* **16:6** (2021).

Abstract

Diagnosis of microbial disease etiology in community-acquired pneumonia (CAP) remains challenging. We undertook a large-scale metabolomics study of serum samples in hospitalized CAP patients to determine if host-response associated metabolites can enable diagnosis of microbial etiology, with a specific focus on discrimination between the major CAP pathogen groups *S. pneumoniae*, atypical bacteria, and respiratory viruses. Targeted metabolomic profiling of serum samples was performed for three groups of hospitalized CAP patients with confirmed microbial etiologies: *S. pneumoniae* (n=48), atypical bacteria (n=47), or viral infections (n=30). A wide range of 347 metabolites was targeted, including amines, acylcarnitines, organic acids, and lipids. Single discriminating metabolites were selected using Student's T-test and their predictive performance was analyzed using logistic regression. Elastic net regression models were employed to discover metabolite signatures with predictive value for discrimination between pathogen groups. Metabolites to discriminate *S. pneumoniae* or viral pathogens from the other groups showed poor predictive capability, whereas discrimination of atypical pathogens from the other groups was found to be possible. Classification of atypical pathogens using elastic net regression models was associated with a predictive performance of 61% sensitivity, 86% specificity, and an AUC of 0.81. Targeted profiling of the host metabolic response revealed metabolites that can support diagnosis of microbial etiology in CAP patients with atypical bacterial pathogens compared to patients with *S. pneumoniae* or viral infections.

2.1 Introduction

Community-acquired pneumonia (CAP) is a commonly occurring respiratory tract infection caused by bacterial or viral pathogens that can lead to severe disease, especially in elderly patients [4]. The predominant pathogens found in hospitalized CAP patients are *Streptococcus pneumoniae* and to a lesser extent, *Haemophilus influenzae*, *Legionella pneumophila*, and respiratory viruses [29, 30]. Patients hospitalized with severe CAP typically receive empirical antibiotic treatment with broad-spectrum antibiotics until the microbial etiology is determined [31, 8]. Current standard diagnostic methods for microbial identification are pathogen-targeted and include culturing, antigen testing, and molecular diagnostics such as PCR [8]. In over 60% of CAP patients, no causative pathogen can be identified with these pathogen-targeted diagnostic techniques [29, 32]. As a consequence, broad-spectrum antibiotics are over-used, which facilitates the emergence of antimicrobial resistance [1, 33]. To this end, a need exists to explore innovative methods to enhance the diagnostic performance for the detection of microbial pathogens in CAP.

Evaluation of differences in the host-response to CAP-associated pathogens may be an alternative approach to improve diagnosis [34]. There is growing evidence that the host, i.e. the patient, metabolic response to infections can be a relevant source of novel host immune response biomarkers to infections [35, 36]. Several small studies have reported differences in metabolite profiles in blood and urine samples in patients with different types of infections (Table 2.4) [37, 38, 23, 20, 39, 22, 40]. For instance, studies comparing metabolomic changes in CAP and tuberculosis (TB) patients show increased levels of plasma lipids and decreased levels of metabolites involved in cholesterol synthesis [37, 20]. A study comparing viral and bacterial respiratory tract infections showed that plasma metabolite profiles of patients with influenza A and bacterial pneumonia differed significantly [22]. In another study, urine samples of patients with a respiratory syncytial virus (RSV) or a bacterial respiratory tract infection showed differences in metabolite levels as well [40]. An important limitation of these studies is that the comparisons made cannot yet support the etiological diagnosis of CAP but merely focus on differences between diseases such as TB versus CAP. The studies that compared viral and bacterial causative pathogen groups of CAP used an untargeted metabolomics approach. While an untargeted approach is especially useful for the discovery of new features and hypothesis-free analysis, a targeted approach that can be fully quantified to clinical laboratory standards may be preferable for clinical implementation. Furthermore, these studies have the limitation that they focus on the comparison of pediatric patients while most hospitalized CAP patients are adults. No studies have evaluated differences in metabolite profiles of CAP patients comparing different microbial etiologies relevant for treatment of CAP, i.e. *S. pneumoniae*, atypical pathogens, and viral infections.

In the current study, we performed extensive targeted metabolomic profiling for three groups of hospitalized CAP patients with confirmed microbial etiologies of *S.*

pneumoniae, atypical bacteria, or viral infections. We aimed to determine whether host-response associated metabolites can enable diagnosis of microbial etiology, focusing on discrimination between the pathogen groups *S. pneumoniae*, atypical bacteria, and respiratory viruses in patients hospitalized with CAP.

2.2 Materials and methods

2.2.1 Study population

Serum samples were taken from 505 patients that were diagnosed with CAP in two previously conducted clinical studies that were executed between October 2004 and September 2010. [29, 30]. The samples were taken from CAP patients within 24 hours after hospital admission. In 57% of these patient samples, the causative pathogen could be identified using conventional diagnostic methods such as culturing, PCR, and urinary antigen tests. The most commonly found causative pathogen in these patients was *S. pneumoniae*, followed by atypical bacterial and viral pathogens. A minority of patients was diagnosed with other bacteria.

From the selection of patients in which a causative pathogen was identified, we excluded patients with mixed infections. Furthermore, we constructed three distinctive groups of patients with *Streptococcus pneumoniae*, atypical (*Coxiella burnetii*, *Chlamydophila psittaci*, *Legionella pneumophila* or *Mycoplasma pneumoniae*), or viral (influenza virus, herpes simplex virus (HSV), respiratory syncytial virus (RSV), parainfluenza virus, or another respiratory virus) infections. The number of available samples for the patient group with confirmed viral CAP infection was limited (n=31). The patients included in the *S. pneumoniae* and atypical bacterial groups were randomly drawn from the remaining study population in an iterating fashion until the bacterial groups were composed in such a way that three groups showed comparable means for sex and pneumonia severity index scores. This resulted in a group of 49 patients with *S. pneumoniae* and a group of 50 patients with atypical infections (Figure 2.1). No matching of individual samples was performed. An overview of patient characteristics is provided in Table 2.1 and Table 2.5. Patient characteristics that might be considered as possible covariates were: age, sex, nursing home resident, renal disease, congestive heart failure, CNS disease, malignancy, COPD, diabetes, altered mental status, respiratory rate, systolic blood pressure, temperature, pulse, pH, BUN, sodium, glucose, hematocrit, partial pressure of oxygen, pleural effusion on x-ray, duration of symptoms before admission, antibiotic treatment before admission. The analyses performed in this study were executed conform the informed consent given by the patients. The clinical data was anonymized before use.

2.2.2 Bioanalytical procedures

Serum samples were analyzed with five liquid chromatography methods and one gas chromatography, mass spectrometry-based, targeted, metabolomics method. The

Table 2.1 Patient characteristics per pathogen group.

	<i>S. pneumoniae</i> (n=48)	Atypical (n=47)	Viral (n=30)	P-value
Age (years)				
Mean (SD)	62.2 (18.9)	54.7 (14.6)	70.1 (16.4)	<0.01
Median [Min, Max]	63.5 [18.0, 98.0]	52.0 [26.0, 81.0]	74.0 [29.0, 95.0]	
Sex				
Male	22 (45.8%)	34 (72.3%)	21 (70.0%)	0.12
PSI score				
<50	9 (18.8%)	9 (19.1%)	2 (6.7%)	0.33
51-70	7 (14.6%)	13 (27.7%)	6 (20.0%)	
71-90	5 (10.4%)	10 (21.3%)	7 (23.3%)	
91-130	23 (47.9%)	12 (25.5%)	11 (36.7%)	
>131	4 (8.3%)	3 (6.4%)	4 (13.3%)	
Liver disease				
No	48 (100%)	47 (100%)	30 (100%)	
Kidney disease				
Yes	3 (6.2%)	1 (2.1%)	4 (13.3%)	0.30
Cardiovascular disease				
Yes	6 (12.5%)	5 (10.6%)	3 (10.0%)	0.93
CNS disease				
No	46 (95.8%)	44 (93.6%)	28 (93.3%)	0.66
Yes	1 (2.1%)	3 (6.4%)	2 (6.7%)	
Missing	1 (2.1%)	0 (0%)	0 (0%)	
Malignancy				
No	44 (91.7%)	46 (97.9%)	28 (93.3%)	0.66
Yes	3 (6.2%)	1 (2.1%)	2 (6.7%)	
Missing	1 (2.1%)	0 (0%)	0 (0%)	
COPD				
No	24 (50.0%)	44 (93.6%)	25 (83.3%)	0.16
Yes	9 (18.8%)	3 (6.4%)	5 (16.7%)	
Missing	15 (31.2%)	0 (0%)	0 (0%)	
Diabetes				
No	26 (54.2%)	45 (95.7%)	26 (86.7%)	0.17
Yes	7 (14.6%)	2 (4.3%)	4 (13.3%)	
Missing	15 (31.2%)	0 (0%)	0 (0%)	
Duration of symptoms before admission (days)				
Mean (SD)	4.06 (3.03)	5.83 (5.65)	4.70 (3.21)	0.33
Median [Min, Max]	3.50 [1.00, 14.0]	5.00 [1.00, 42.0]	4.00 [0.00, 14.0]	
Missing	16 (33.3%)	0 (0%)	0 (0%)	
Antibiotic treatment before admission				
No	27 (56.2%)	29 (61.7%)	23 (76.7%)	0.17
Yes	5 (10.4%)	18 (38.3%)	7 (23.3%)	
Missing	16 (33.3%)	0 (0%)	0 (0%)	
Corticosteroid use before admission				
No	29 (60.4%)	46 (97.9%)	29 (96.7%)	0.67
Yes	2 (4.2%)	1 (2.1%)	1 (3.3%)	
Missing	17 (35.4%)	0 (0%)	0 (0%)	

Data are presented as number (%) or mean (SD). Abbreviations: PSI: pneumonia severity index; CNS: central nervous system; COPD: chronic obstructive pulmonary disease.

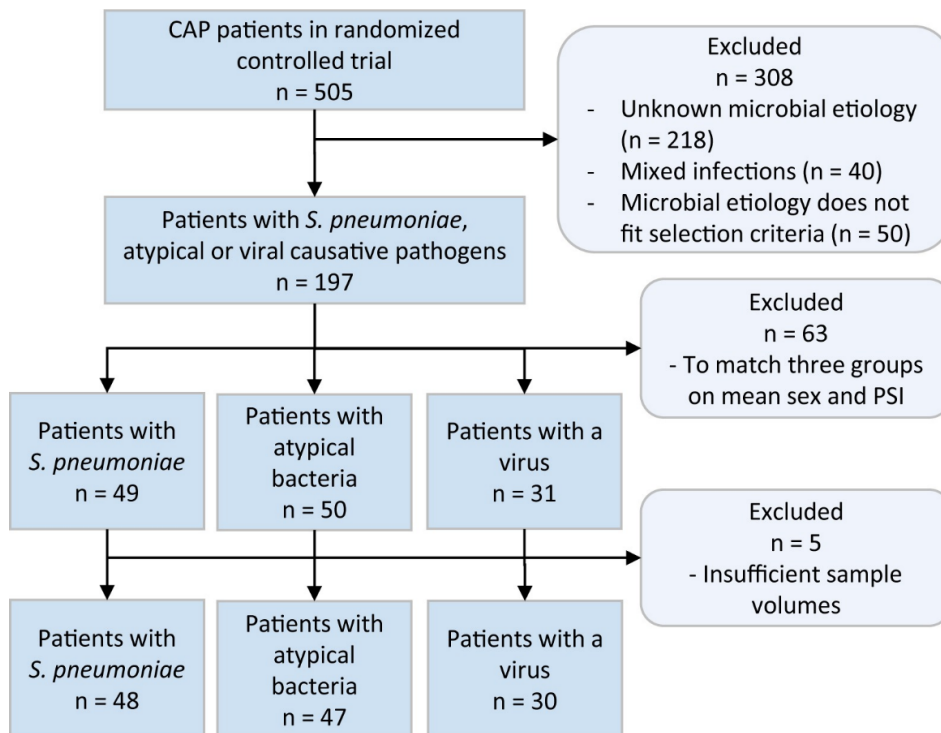


Figure 2.1 Flow chart of the formation of the three studied patient groups.

metabolomics profiling covered 596 metabolite targets from 25 metabolite classes, including amino acids, biogenic amines, acylcarnitines, organic acids, and multiple classes of lipids (Table 2.6). Levels of 374 unique metabolites were detected in the samples. The metabolomic profiling was performed within the Biomedical Metabolomics Facility of Leiden University in Leiden, The Netherlands. Details of the metabolomic analysis methods used are provided in section 2.5.

2.2.3 Data analysis

The data resulting from the metabolomic profiling was cleaned by removing patient samples with more than 10 missing metabolite values, for example, if results from one measurement platform were missing because of too low sample volumes, and by removing metabolites with missing patient samples, for example, because of a sample preparation error. The clean dataset consisted of 347 metabolite levels (Table 2.7) for 125 patients diagnosed with the microbial etiology *S. pneumoniae* (n=48), atypical (n=47), or viral (n=30). The pathogens identified in each group are shown in Table 2.2. The resulting metabolite levels were preprocessed by applying log transformation and

standardized to correct for heteroscedasticity. The preprocessed metabolomics dataset was visually inspected using a principal component analysis.

Data imputation was performed for patient characteristics that were to be evaluated as covariates in the statistical analysis and showed missingness in the data. Five times repeated imputation using predictive mean matching was performed with the ‘mice’ package for R to impute the patient data for the covariates with less than 25% missing data. Predictive mean matching is suitable for both numeric and binary covariates. Patient characteristics with >25% missing data were excluded from further analysis.

We performed logistic regression and elastic net regression modeling to determine if patients in one pathogen group could be discriminated from patients in the remaining two groups. Also, we aimed to determine which metabolites were important for prediction of the causative pathogen. In both methods, five-fold cross-validation was used to make the most efficient use of the available data for estimation of the predictive performance of the models and its associated metabolites [41]. Furthermore, the model generation was repeated 100 times to obtain robust estimates of the predictive performance of the models.

To identify single discriminative metabolites, Student’s T-tests with false discovery rate (FDR) multiple testing corrections were performed ($p < 0.05$). Then, significant metabolites and a combination of significant metabolites were modeled using logistic regression. Also, models containing covariates age and sex and all covariates were generated. The predictive logistic regression models were analyzed by comparison of their area under the curve (AUC), sensitivity, specificity, balanced error rate (BER), and receiver operating characteristic (ROC) curve.

Elastic net regression was performed to test if the predictive power of the metabolite data could be increased by including correlations between metabolites in addition to evaluating single metabolites. In elastic net regression, metabolites that have no explanatory power can be set to zero, as in a lasso regression, and metabolites that explain the same amount of variance can all be included with balanced coefficient sizes, as in a ridge regression [42].

To obtain robust estimates of the predictive performance of the elastic net model, hyperparameters were optimized in a five-fold nested-cross validation, where the hyperparameters were selected truly independent of the calculation of the predictive performance, as is schematically shown in Figure 2.2 [43]. In the inner cross-validation loop, the model optimization loop, optimal values for model hyperparameters α and λ were determined. In the outer cross-validation loop, the model performance loop, the optimal model for the training fold was built on the set hyperparameters α and λ (Figure 2.5). Hyperparameter selection was performed using the balanced error rate (BER), which can be calculated from the true- and false positive (TP, FP), and true- and false-negative rates (TN, FN, Equation 2.1). The BER accounts for different group sizes

per model and therefore gives an accurate picture of the performance of models in the model optimization and model performance loop.

$$\text{BER} = 0.5 * \left(\frac{\text{FP}}{\text{TN} + \text{FP}} + \frac{\text{FN}}{\text{FN} + \text{TP}} \right) \quad (2.1)$$

The overall predictive diagnostic performance was evaluated using sensitivity and specificity performance measures, generated from the confusion matrix that represents the number of samples falling into each possible outcome (Equation 2.2-2.3). The average sensitivity and specificity of all 500 generated models and its standard deviation were used to compare the assay performance to currently used methods.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2.2)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (2.3)$$

The relative contribution of metabolites to provide predictions of the expected pathogen group were quantified using the variable importance in prediction (VIP) score, expressed as a percentage. The VIP score was calculated per metabolite per fold or repeat as follows:

$$\text{VIP} (\%) = \frac{\beta_j}{\sum_{i=0}^p |\beta_i|} \cdot 100\% \quad (2.4)$$

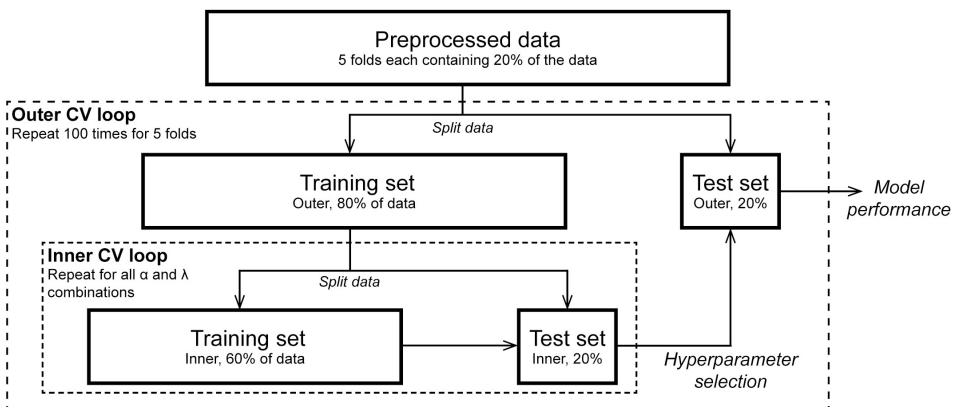
where β_j is the regression coefficient for fold j over the sum of all regression coefficient values in the model. Metabolites were arranged based on their mean VIP score over all folds and repeats. Metabolites with an absolute VIP $> 1\%$ were considered to be most important. Furthermore, to determine the need to include age and sex, or all covariates in the models we compared the BER for models with and without age and sex, or all covariates included. Finally, mean AUC values and ROC curves were calculated and generated to compare the performance of the elastic net models to the logistic regression models.

The scripts used for the statistical analyses were deposited in Github at <http://github.com/vanhasseltlab/MetabolomicsEtiologyCAP>.

Table 2.2 Distribution of causative microbial agents per pathogen group for statistical data analysis.

Causative pathogen	<i>S. pneumoniae</i> (n=48)	Atypical bacterial (n=47)	Viral (n=30)
<i>S. pneumoniae</i>	48 (100%)	0 (0%)	0 (0%)
<i>Legionella pneumophila</i>	0 (0%)	18 (38.3%)	0 (0%)
<i>Coxiella burnetii</i>	0 (0%)	17 (36.2%)	0 (0%)
<i>Chlamydophila psittaci</i>	0 (0%)	7 (14.9%)	0 (0%)
<i>Mycoplasma pneumoniae</i>	0 (0%)	5 (10.6%)	0 (0%)
Influenza virus	0 (0%)	0 (0%)	11 (36.7%)
HSV	0 (0%)	0 (0%)	6 (20.0%)
RSV	0 (0%)	0 (0%)	4 (13.3%)
Parainfluenza virus	0 (0%)	0 (0%)	3 (10.0%)
Other viruses	0 (0%)	0 (0%)	6 (20.0%)

Data are presented as number (%). Abbreviations: *S. pneumoniae*: *Streptococcus pneumoniae*; HSV: *herpes simplex virus*; RSV: *respiratory syncytial virus*.

**Figure 2.2** Schematic representation of stratified nested cross-validation for elastic net regression model optimization and performance [43]. Abbreviations: CV: cross-validation.

2.3 Results

2.3.1 Metabolomics profiling and exploratory analysis of metabolomics data

Metabolomics profiling was performed for 130 patients and 596 metabolite targets. Preprocessing of the metabolomics dataset resulted in a reduced dataset including 125 patients and 347 metabolites (Figure 2.1). The patient characteristics of these 125 patients are displayed in Table 2.1. The patients were diagnosed with the microbial etiology *S. pneumoniae* (n=48), atypical bacteria (n=47), or respiratory virus (n=30) (Table 2.2). A list of all targeted and detected metabolites and their identifiers can be found in Table 2.7. Unsupervised principal component analysis showed no clear separation between pathogen groups (Figure 2.6).

2.3.2 Single discriminating metabolites for pathogen groups

Three significant metabolites were found for the discrimination of atypical pathogens from *S. pneumoniae* and viral pathogens using a Student's T-test with FDR multiple testing correction ($p < 0.05$): glycylglycine, symmetric dimethylarginine (SDMA), and lysophosphatidylinositol (18:1) (LPI (18:1)). For the other comparisons, no significantly discriminating metabolites were found.

The significantly differentiating metabolites were included in logistic regression models to differentiate patients with atypical pathogens from patients suffering from CAP caused by *S. pneumoniae* or viral pathogens. The logistic regression models were evaluated based on their AUC, sensitivity, specificity, BER, and ROC curve after fivefold cross-validation with 100 repeats (Table 2.3, Figure 2.3). They show that logistic regression models of the individual metabolites glycylglycine, SDMA, and LPI(18:1) can differentiate atypical pathogens from *S. pneumoniae* and viral pathogens with AUCs between 0.70-0.72, sensitivities between 0.32-0.36, specificities between 0.83-0.85, and BERs of 0.39-0.41. A logistic regression model including all three significantly discriminating metabolites yields a more successful separation with an AUC of 0.78, sensitivity of 0.57, specificity of 0.83, and BER of 0.30. Addition of the covariates age and sex to the three metabolite model, slightly improved the predictive performance of the model resulting in a sensitivity of 0.63 and a specificity of 0.84. This model also showed the highest AUC (0.79) and lowest BER (0.26) of the tested logistic regression models. The addition of other covariates to the logistic regression model resulted in lower performance, probably due to overfitting of the model. The ROC curves emphasize the increased model performance upon the addition of more discriminating metabolites to the logistic regression model (Figure 2.3).

2.3.3 Predictive metabolites for diagnosis of CAP-associated pathogens

Elastic net models including multiple metabolites were fit to discriminate *S. pneumoniae*, atypical bacterial, and viral pathogens from the remaining two groups (e.g., *S. pneumoniae* versus atypical bacterial and viral pathogens). Elastic net models separating patients with atypical bacterial pathogens from patients with *S. pneumoniae* and viral infections resulted in a mean AUC of 0.81, a sensitivity of 0.61, a specificity of 0.86, and a BER of 0.26. Prediction of *S. pneumoniae* or viral infection etiologies showed lower predictive capabilities with AUC's of 0.74 and 0.63, high sensitivities of 0.83 and 0.89, but low specificities of 0.5 and 0.23, and BER's of 0.33 and 0.44, respectively (Table 2.3).

We included the covariates age and sex, and all covariates in the elastic net models to account for potential confounding effects. The addition of these covariates showed no improved performance of the elastic net models for differentiation of atypical pathogens or *S. pneumoniae* from the other groups. For the differentiation of viral pathogens from the other two pathogen groups, a slight performance improvement was seen upon the addition of the covariates age and sex resulting in an AUC of 0.63, a sensitivity of 0.89, a specificity of 0.23, and a BER of 0.44 (Table 2.3).

The ROC curves for the separation of atypical pathogens from *S. pneumoniae* and viral pathogens show that elastic net models perform better than the logistic regression models for single metabolites. However, the logistic regression model including the three significant metabolites and the covariates age and sex shows similar performance as the elastic net regression which included 100 metabolites on average (Figure 2.3).

2.3.4 Metabolite classes predictive for atypical bacterial pathogens

Focusing on the metabolites that have shown to be predictive for atypical bacterial pathogens, i.e., the only comparison with clinically relevant predictive performance, we identified 26 metabolites with an absolute VIP > 1% using elastic net regression (Figure 2.4). The metabolites originated from multiple metabolite classes. However, the classes of biogenic amines and lysophospholipids were well represented (4-5 metabolites per class), compared to the other classes. The number of metabolites included in the models varied across folds without a clear correlation to the BER. Commonly, models including all metabolites were favored, followed by models including 20-100 metabolites (Figure 2.7). We visualized the separation of the different pathogens in the atypical pathogen group using an unsupervised PCA analysis including all metabolites. The PCA plot indicated that no clear sub-group is present within the atypical group that would prominently drive the separation from the *S. pneumoniae* and viral infections (Figure 2.8).

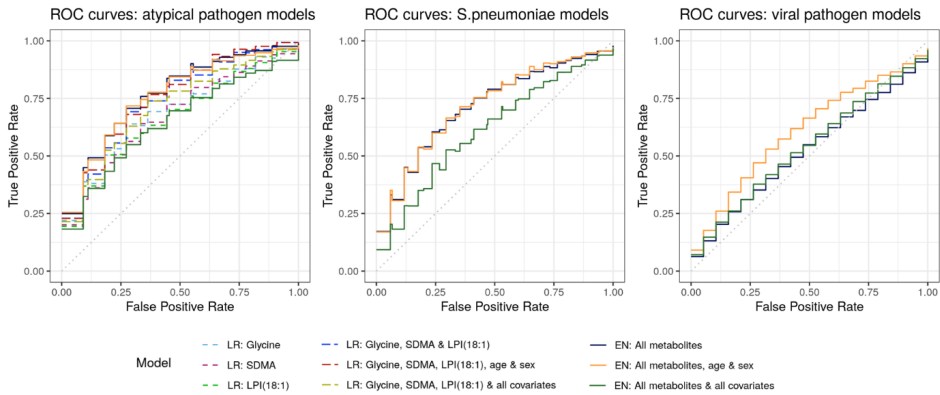


Figure 2.3 ROC curves of the results from logistic regression and elastic net regression models that were tested in five-fold cross-validation with 100 repeats for the comparisons: atypical versus *S. pneumoniae* and viral pathogens; *S. pneumoniae* pathogens versus atypical and viral pathogens; and viral versus *S. pneumoniae* and atypical pathogens. Abbreviations: LR: logistic regression, EN: elastic net regression, SDMA: symmetric dimethylarginine, LPI (18:1): lysophosphatidylinositol (18:1).

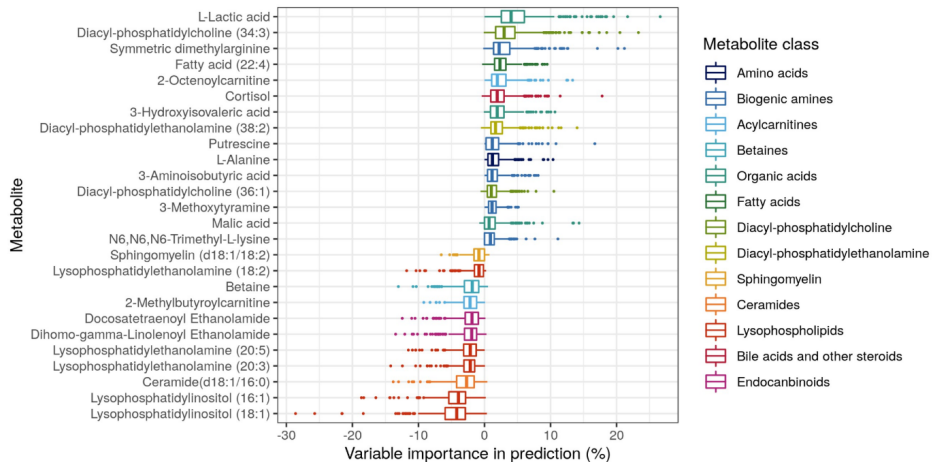


Figure 2.4 Variable importance of metabolites for the prediction of an atypical bacterial infection versus *S. pneumoniae* and viral infections. Only metabolites with an absolute mean percentage of influence > 1% are visualized.

Table 2.3 Results from the logistic regression and elastic net regression models that were tested in a fivefold cross-validation with 100 repeats. The table displays the performance of the models for the three comparisons: atypical versus *S. pneumoniae* and viral pathogens; *S. pneumoniae* pathogens versus atypical and viral pathogens; and viral versus *S. pneumoniae* and atypical pathogens. Logistic regression is only included for the comparison of atypical versus *S. pneumoniae* and viral pathogens because no significant single metabolites were found for the other comparisons. The performance is evaluated using the mean area under the curve (AUC), the mean sensitivity, the mean specificity, and the mean balanced error rate (BER) over all folds and repeats. All performances result from the test sets within the cross-validation. The best performing model per comparison and evaluation measure is displayed in bold and underlined.

Model	Variables	AUC	Sensitivity	Specificity	BER
<i>Atypical – (S. pneumoniae + viral)</i>					
LR	Glycylglycine	0.72 (0.094)	0.36 (0.14)	0.83 (0.110)	0.40 (0.084)
LR	SDMA	0.72 (0.093)	0.36 (0.15)	0.86 (0.100)	0.39 (0.082)
LR	LPI.18.1.	0.70 (0.099)	0.32 (0.14)	0.85 (0.100)	0.41 (0.082)
LR	Age + sex	0.71 (0.097)	0.39 (0.15)	0.85 (0.090)	0.38 (0.071)
LR	All covariates	0.65 (0.098)	0.52 (0.15)	0.68 (0.120)	0.40 (0.087)
LR	Glycylglycine + SDMA + LPI.18.1.	0.78 (0.094)	0.57 (0.16)	0.83 (0.100)	0.30 (0.090)
LR	Glycylglycine + SDMA + LPI.18.1. + age + sex	0.79 (0.089)	0.63 (0.16)	0.84 (0.095)	0.26 (0.085)
LR	Glycylglycine + SDMA + LPI.18.1. + all covariates	0.75 (0.097)	0.60 (0.16)	0.78 (0.110)	0.31 (0.093)
ENR	100 (82)	0.81 (0.087)	0.61 (0.18)	0.86 (0.092)	0.27 (0.094)
ENR	110 (91) incl. age & sex	0.80 (0.094)	0.61 (0.17)	0.84 (0.096)	0.28 (0.090)
ENR	270 (140) incl. all covariates	0.69 (0.100)	0.58 (0.17)	0.70 (0.120)	0.36 (0.098)
<i>S. pneumoniae – (atypical + viral)</i>					
ENR	210 (120)	0.74 (0.091)	0.83 (0.10)	0.50 (0.160)	0.33 (0.087)
ENR	240 (130) incl. age & sex	0.74 (0.095)	0.80 (0.10)	0.52 (0.160)	0.34 (0.084)
ENR	290 (120) incl. all covariates	0.63 (0.110)	0.69 (0.13)	0.51 (0.17)	0.40 (0.098)
<i>Viral – (S. pneumoniae + atypical)</i>					
ENR	170 (140)	0.54 (0.120)	0.88 (0.11)	0.16 (0.170)	0.48 (0.075)
ENR	130 (130) incl. age & sex	0.63 (0.130)	0.89 (0.08)	0.23 (0.160)	0.44 (0.082)
ENR	180 (160) incl. all covariates	0.56 (0.130)	0.79 (0.11)	0.31 (0.190)	0.45 (0.099)

Data are presented as mean (SD). Variables are presented as variable names or as the number of variables that are included in the model. Abbreviations: LR: Linear regression, ENR: Elastic net regression, SDMA: symmetric dimethylarginine, LPI (18:1): lysophosphatidylinositol (18:1), AUC: area under the curve, BER: balanced error rate.

2.4 Discussion

Targeted profiling of the host metabolic response revealed metabolites that can support the diagnosis of microbial etiology in CAP patients with atypical bacterial pathogens compared to patients with *S. pneumoniae* or viral infections. CAP patients suffering from *S. pneumoniae* and viral infection could not be as successfully discriminated from the other groups based on the metabolic host-response.

The currently used clinical assays still outperform the metabolomics host-response assays developed in this study. For atypical pathogens, the sensitivity of 63% and specificity of 86% reported in this study are lower than the current urinary antigen tests for detection of *Legionella pneumophila* which shows a sensitivity of approximately 70% and a specificity up to 96% [44]. For detection of *S. pneumoniae*, the 83% sensitivity reached with the metabolomics-based assay outperforms the current antigen tests that show 70% sensitivity. However, the specificity of the metabolomics-based assay is only 50% while antigen tests reach specificity up to 96% [45, 46]. PCR assays of nasopharyngeal swabs for viral pathogens show sensitivities of up to 96% for influenza viruses A and B [47]. Our viral metabolomics-based assay shows a good sensitivity of 89% as well. However, the specificity of this assay is with 23% very low. The expected clinical utility of the studied metabolite classes as host-response biomarkers for etiological diagnosis of CAP may therefore be considered limited.

The combination of the metabolites glycyglycine, SDMA, and LPI (18:1) and the covariates age and sex showed predictive capacities similar to elastic net models including 100 metabolites in the comparison of atypical pathogens versus *S. pneumoniae* and viral pathogens. This result suggests that a simple model might perform as well as a more complex elastic net model, which is an important finding when considering the use of these biomarkers for clinical diagnostic applications, e.g., where a limited set of 3 metabolites is preferable.

Glycyglycine, a biogenic amine, showed to be significantly contributing to the differentiation of atypical pathogens from the other pathogens, but was not often included in elastic net models. In contrast, SDMA and LPI (18:1) were often included in the elastic net models as was shown in the overview of the 26 most influential metabolites. Metabolites of the classes biogenic amines and lysophospholipids, to which SDMA and LPI (18:1) have been assigned, were most represented in the 26 most influential metabolites compared to other metabolite classes in the comparison of atypical versus *S. pneumoniae* and viral pathogens. A comparison of the most influential metabolites in this study to metabolites of interest reported in previous studies of metabolomics in CAP patients shows limited overlap. Major reasons for this could be that (i) not all studies measured the same set of metabolic classes; (ii) some other studies poorly controlled patient comparator groups; and (iii) difference in bioanalytical methodologies, e.g. the use of NMR or MS as analytical method with their respective (dis)advantages might provide different results [48]. For example, most lipids found to be predictive in this study have not been reported previously, most

likely because the applied bioanalytical methodologies did not allow their detection. However, some overlap was found between the most influential metabolites for the comparison of atypical versus *S. pneumoniae* and viral pathogens in this study, and the metabolites of interest from other metabolomics studies involving CAP patients. The amino acid alanine was found in multiple studies [23, 39, 22]. Ceramide (d18:1/16:0), two diacyl-phosphatidylcholines, and diacyl-phosphatidylethanolamine (38:2) were found in other studies as well, the latter in the form of choline and ethanolamine [20, 39, 40]. Lactic acid was identified by several other metabolomics studies to respiratory bacterial and viral infections [37, 23, 22]. Lactic acid levels are also known to rise in case of severe disease. However, because the three pathogen groups were balanced in terms of disease severity and, for example, did not show significant differences in pH levels, we hypothesize that the differences in lactate levels are, in this case, an effect of the pathogen-specific host-response to infection. The result showed that models including disease severity covariates do not perform better than models without these confounders, thus supporting this hypothesis. Finally, 3-hydroxyisovaleric acid and betaine have been reported in a previous study comparing viral and bacterial pneumonia [40]. The overlap in these findings may provide insights into common metabolic responses to pathogens involved in CAP.

Multiple biological processes besides infection can influence metabolic processes in patients. Inclusion of age and sex in the models did not improve the predictive performance of the elastic net models for atypical bacteria and *S. pneumoniae* but did improve the model for viral pathogens. The average age in the viral pathogen group was higher than in the other groups, which could explain this result. For the other comparisons, we see that a model including age and sex or more covariates does not outperform models without these possible confounders. This doesn't imply there is no metabolomic effect of age in the bacterial pathogen groups but implies that the separation between bacterial pathogen groups is more dependent on the metabolomic host-response to the infection than on the age-related metabolomic changes. In this study, we included patients with mild to severe CAP, reflecting the target patient population for which improvements in a diagnostic assay are required. However, the combination of samples from patients with different disease severities may negatively influence the predictive capabilities of the model because the effect from the causative pathogen on the host-metabolism may be less pronounced for less severe disease [49]. However, separating the patients into groups with comparable disease severity scores would decrease the power for statistical analysis. Furthermore, no standardization of sampling times and conditions was applied, e.g., patients had not fasted before blood sampling, which may influence the metabolite patterns found. Since variations in sampling conditions were unknown, we were unable to consider these in our analyses. However, we expect that the impact of not standardizing and correcting for these factors is limited because the noise in metabolite levels introduced by these factors is expected to be random with regard to the pathogen groups compared in this study. A standardized sampling approach could improve the sensitivity of the models to detect predictive metabolites because some noise is reduced. However, the specificity of the models with

respect to the prediction of specific pathogens would be unchanged, since no correlation with pathogen groups is likely.

The sample size of this study (n = 125) was relatively large compared to studies researching metabolomic differences between causative pathogens of CAP that included approximately 70 patients [22, 40]. The compared groups *S. pneumoniae*, atypical bacteria, and viruses were chosen because antibiotic treatment strategies differ between these three groups. Ideally, we would have further investigated differences within studied groups, e.g. to identify metabolic responses to specific pathogens within the atypical pathogens and viral infection groups. For example, it would be of interest to study *Legionella* species more in-depth because their intracellular growth might result in a differentiated host-response. However, this was considered not feasible in this study due to sample size restrictions. The heterogeneous pathogen population in the atypical bacterial and viral pathogen groups might have lowered the predictive performance of the metabolomic analysis. Studying the individual pathogens in bigger sample sizes might reveal more characteristic metabolite signatures. In this study, no control group was included because the goal of the study was to provide a faster and optimal diagnostic method and a guide for antibiotic treatment in hospitalized CAP patients. In further studies, it would be preferable to include patients with all causes of CAP, including the remaining microorganisms, which were excluded in the current study because of their low frequency, to enable a more comprehensive comparison with current clinical assays. In this study, CAP patients with unknown pathogens were excluded. In a follow-up study, the metabolite pattern of the patients with unknown causative pathogens could be compared to the metabolite patterns of the distinguished pathogen groups to gain more information about the metabolomic resemblance of the samples in which pathogens could and could not be identified using the conventional diagnostic techniques.

Metabolomics analysis resulted in some missing data because of sample preparation errors or the limited volume of the samples. Because the measurement platforms covered multiple metabolites within one pathway, metabolites with missing data could be removed without influencing the final results. Some patient samples had to be removed because of multiple missing metabolite levels, for example, if the results from a whole metabolomics platform were missing. Data imputation was not performed for the metabolomics data, because the wide range of patients included in the dataset did, in our opinion, not provide enough information for accurate data imputation.

In summary, this comprehensive analysis of the host metabolic response across multiple metabolic classes and based on a well-balanced study cohort of CAP patients has shown the possibility to identify atypical pathogens in CAP and limited utility of predicting *S. pneumoniae* and viral infection disease etiologies.

2.5 Supporting information

Details on metabolomic sample analysis

Batch design: Aliquoted samples were run in a randomized fashion in several batches together with quality control (QC) samples (every 10 samples), sample replicates (every 7 samples), internal standards (ISTDs), blanks, and calibration lines.

Quality control: Blank samples were used to determine the blank effect. Replicate samples were used to check the instrument for repeatability. In-house developed algorithms were applied using the pooled QC samples to compensate for shifts in the sensitivity of the mass spectrometer over the batches.

Reported results: After quality control correction the metabolites that complied with the acceptance criteria of a relative standard deviation of the quality control samples (RSD_{qc}) <15% were reported. The data was reported as relative response ratio (analyte signal area / ISTD area; unit free) of the metabolites after QC correction. Metabolites that did not comply with the acceptance criteria of the quality control, but have been included in the results present RSDs up to 30% and should be handled with caution.

Amine profiling: Amine profiling was performed according to the validated amine profiling analytical platform with minor optimization [50]. The amine platform covers amino acids and biogenic amines employing an Accq-Tag derivatization strategy adapted from the protocol supplied by Waters. 5,0 μL sample was spiked with an internal standard solution. Protein precipitation was performed by addition of MeOH and the sample was dried in a speedvac. The residue was reconstituted in borate buffer (pH 8.5) with AQC reagent. The prepared samples were transferred to autosampler vials and placed in an autosampler tray. The vials were cooled at 4°C upon injection. 1,0 μL prepared sample was injected in a UPLC-MS/MS system. Chromatographic separation was achieved by an Agilent 1290 Infinity II LC System on an Accq-Tag Ultra column (Waters) with a flow of 0.7 mL/min over an 11 min gradient. The UPLC was coupled to electrospray ionization on a triple quadrupole mass spectrometer (AB SCIEX Qtrap 6500). Analytes were detected in the positive ion mode and monitored in Multiple Reaction Monitoring (MRM) using nominal mass resolution. Acquired data was evaluated using MultiQuant Software for Quantitative Analysis (AB SCIEX, Version 3.0.2), by the integration of assigned MRM peaks and normalization using proper internal standards. For analysis of amino acids, their ¹³C¹⁵N-labeled analogs were used. For other amines, the closest-eluting internal standard was employed. After quality control correction 48 amines complied with the acceptance criteria of RSD_{qc} <15%. Additionally, 7 amines presented an RSD_{qc} between 15 and 30%. They are included in the results but these compounds should be considered with caution.

Acylcarnitine profiling: The acylcarnitine platform covers acylcarnitines as well as trimethylamine-N-oxide, choline, betaine, deoxycarnitine, and carnitine. 10 μL sample was spiked with an internal standard solution. Protein precipitation was performed by

addition of MeOH. The supernatant was transferred to an autosampler vial and placed into an autosampler. The vials were cooled at 10°C upon injection. 1.0 µL of the prepared sample was injected into a triple quadrupole mass spectrometer. Chromatographic separation was achieved by UPLC (Agilent 1290, San Jose, CA, USA) on an Accq-Tag Ultra column (Waters) with a flow of 0.7 mL/min over an 11 min gradient. The UPLC was coupled to electrospray ionization on a triple quadrupole mass spectrometer (Agilent 6460, San Jose, CA, USA). Analytes were detected in the positive ion mode and monitored in Multiple Reaction Monitoring (MRM) using nominal mass resolution. Acquired data was evaluated using Agilent MassHunter Quantitative Analysis software (Agilent, Version B.05.01), by integration of assigned MRM peaks and normalization using proper internal standards. The closest-eluting internal standard was employed. After quality control correction 24 acylcarnitines complied with the acceptance criteria of RSDqc <15%. Additionally, 4 acylcarnitines presented an RSDqc between 15 and 30%. They are included in the results but these compounds should be considered with caution.

Organic acid profiling: The organic acid platform covers 28 organic acids. 50 µL sample was spiked with an internal standard solution. Protein precipitation was performed by addition of MeOH. After centrifugation, the supernatant was transferred and the sample was dried using a speedvac. Then, two-step derivatization procedures were performed on-line: oximation using methoxyamine hydrochloride (MeOX, 15 mg/mL in pyridine) as the first reaction and silylation using N-Methyl-N-(trimethylsilyl)-trifluoroacetamide (MSTFA) as the second reaction. 1 µL of each sample was directly after its derivatization injected on GC-MS. Gas chromatography was performed on an Agilent Technologies 7890A equipped with an Agilent Technologies mass selective detector (MSD 5975C) and MultiPurpose Sampler (MPS, MXY016-02A, GERSTEL). Chromatographic separations were performed on an HP-5MS UI (5% Phenyl Methyl Silox), 30 m × 0.25 mm ID column with a film thickness of 25 µm, using helium as the carrier gas at a flow rate of 1,7 mL/min. A single-quadrupole mass spectrometer with electron impact ionization (EI, 70 eV) was used. The mass spectrometer was operated in SCAN mode mass range 50-500. Acquired data was evaluated using Agilent MassHunter Quantitative Analysis software (Agilent, Version B.05.01). After quality control correction and considering blank effects, 9 organic acid compounds complied with the acceptance criteria RSDqc <15% and blank effect <20%. 4 organic acids reported an RSDqc between 15 and 30% and should be considered with caution.

Negative lipid profiling: The negative lipid platform is a semi-target methodology for the identification of 30 fatty acids. 50 µL sample was spiked with 50 µL of an internal standard solution. Protein precipitation was performed by addition of 550 µL MeOH. After centrifugation, 600 µL supernatant was transferred and the sample was dried using a speedvac. The residue was reconstituted in 300 µL of isopropanol with 0,1% formic acid. The prepared samples were transferred to autosampler vials and placed in an autosampler tray. 8,0 µL of the prepared sample was injected into an LC-MS. The analysis was performed on an ACQUITY UPLC™ (Waters, the Netherlands) coupled to a high-resolution mass spectrometer with a Synapt G2 Q-TOF system (Waters, the

Netherlands) using reference lock mass correction. Lipids were detected in full scan in the negative ion mode. Chromatographic separation was achieved using an HSS T3 column (1.8 μm , 2.1 * 100 mm) with a flow of 0.4 mL/min over a 16-minute gradient. Acquired data was preprocessed using Targetlynx software (Masslynx, V4.1, SCN916). After quality control correction, 10 compounds complied with the acceptance criteria RSD_{qc} <15%. 6 compounds reported an RSD_{qc} between 15 and 30% and should be considered with caution.

Positive lipid profiling: The positive lipid platform covers 185 compounds including triglycerides (TGs, n=85) and non-triglycerides (non-TGs, n=100). 10 μL preprocessed sample was spiked with 1000 μL IPA containing internal standards and vortexed for 30 sec. Prepared samples were transferred to autosampler vials for LC-MS analysis. In total 2.5 μL prepared sample was injected for analysis. Chromatographic separation was achieved on an ACQUITY UPLC™ (Waters, Ettenleur, the Netherlands) with an HSS T3 column (1.8 μm , 2.1 * 100 mm) with a flow of 0.4 mL/min over a 16 min gradient. The lipid analysis is performed on a UPLC-ESI-Q-TOF (Agilent 6530, Jose, CA, USA) high-resolution mass spectrometer using reference mass correction. Lipids were detected in full scan in the positive ion mode. The raw data were preprocessed using Agilent MassHunter Quantitative Analysis software (Agilent, Version B.04.00). After quality control correction, 56 TGs and 39 non-TGs compounds complied with the acceptance criteria RSD_{qc}<15% and blank effect <40 %. 1 TG and 53 non-TGs reported an RSD_{qc} between 15 and 30% and should be considered with caution.

Signaling lipid profiling: The signaling lipids platform covers various isoprostane classes together with their respective prostaglandin isomers from different poly unsaturated fatty acids (PUFA), including n-6 and n-3 PUFAs such as dihomo- γ -linoleic acid (DGLA) and arachidonic acid (both n-6) and eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) (both n-3). Also included in this platform are endocannabinoids, bile acids, and signaling lipids from the sphingosine and sphinganine classes and their phosphorylated forms, as well as three classes of lysophosphatidic acids. The three lysophosphatidic acid classes include lysophosphatidic acids (LPAs), lysophosphatidylglycerol (LPG), lysophosphatidylinositol (LPI), lysophosphatidylserine (LPS), lysophosphatidylethanolamines (LPE), cyclic-phosphatidic acids(cLPA), and fatty acid all ranging from C14 to C22 chain length species. The signaling and peroxidized lipids platform is divided into two chromatographic methods: low and high pH. In the low pH method, isoprostanes, prostaglandins, nitro-fatty acids, lyso-sphingolipids, endocannabinoids, and bile acids are analyzed. The high pH method covers lyso-sphingolipids, lysophosphatidic acids, lysophosphatidylglycerol, lysophosphatidylinositol, lysophosphatidylserine, lysophosphatidylethanolamines, cyclic-phosphatidic acids, and fatty acid. Each sample was spiked with antioxidant and internal standard solution. The extraction of the compounds is performed via liquid-liquid extraction (LLE) with butanol and methyl tert-butyl ether (MTBE). After collection, the organic phase is concentrated by first drying followed by reconstituted in a smaller volume. After reconstitution, the extract is transferred into amber autosampler vials and used for high and low pH injection. A Shimadzu system, formed

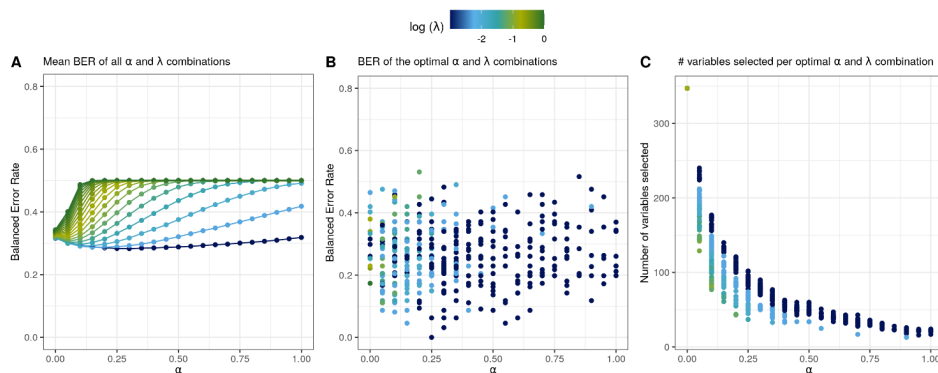


Figure 2.5 Optimization of α and λ in the inner cross-validation (CV) to reach a minimal balanced error rate (BER) in the outer CV. (A) Shows all α and λ values tested in inner CV against mean BER of the inner CV. (B) A plot of the optimal α and λ combinations chosen in the inner CV against their BER in the outer CV shows a variety of favorable α and λ concentrations. (C) A plot of the number of variables selected in the elastic net model in outer CV shows that with increasing alpha, the number of variables decreases as is expected in an elastic net model. The data shown in the figure is a result of the comparison Atypical – (*S. pneumoniae* + viral).

by three high-pressure pumps (LC-30AD), a controller (CBM-20Alite), an autosampler (SIL-30AC), and an oven (CTO-30A) from Shimadzu Benelux, was coupled online with an LCMS-8050 triple quadrupole mass spectrometer (Shimadzu) for high pH measurements. An LCMS-8060 triple quadrupole mass spectrometer (Shimadzu) was coupled to the Shimadzu system for low pH measurements. Both systems were operated using LabSolutions data acquisition software (Version 5.89, Shimadzu). The samples were analyzed by UPLC-MS/MS. An Acquity UPLC BEH C18 column (Waters) was used to measure the samples in the low pH method. For the high pH method, a Kinetex EVO column by Phenomenex was used. The triple quadrupole mass spectrometer was used in polarity switching mode and all analytes were monitored in dynamic Multiple Reaction Monitoring (dMRM). The acquired data was evaluated using LabSolutions Insight software (Version 3.1 SP1, Shimadzu), by integration of assigned MRM peaks and normalization using accordingly selected internal standards. When available, a deuterated version of the target compound was used as an internal standard. For the other compounds, the closest-eluting internal standard was employed. For low pH mode, after quality control correction, 46 metabolites complied with the acceptance criteria of RSD_{qc} <15% and blank effect <40%. 6 compounds reported an RSD_{qc} between 15 and 30% and should be considered with caution. For high pH mode, after quality control correction, 43 metabolites complied with the acceptance criteria of RSD_{qc} <15% and blank effect <40%. Additionally, 18 compounds reported an RSD_{qc} between 15 and 30% and should be considered with caution.

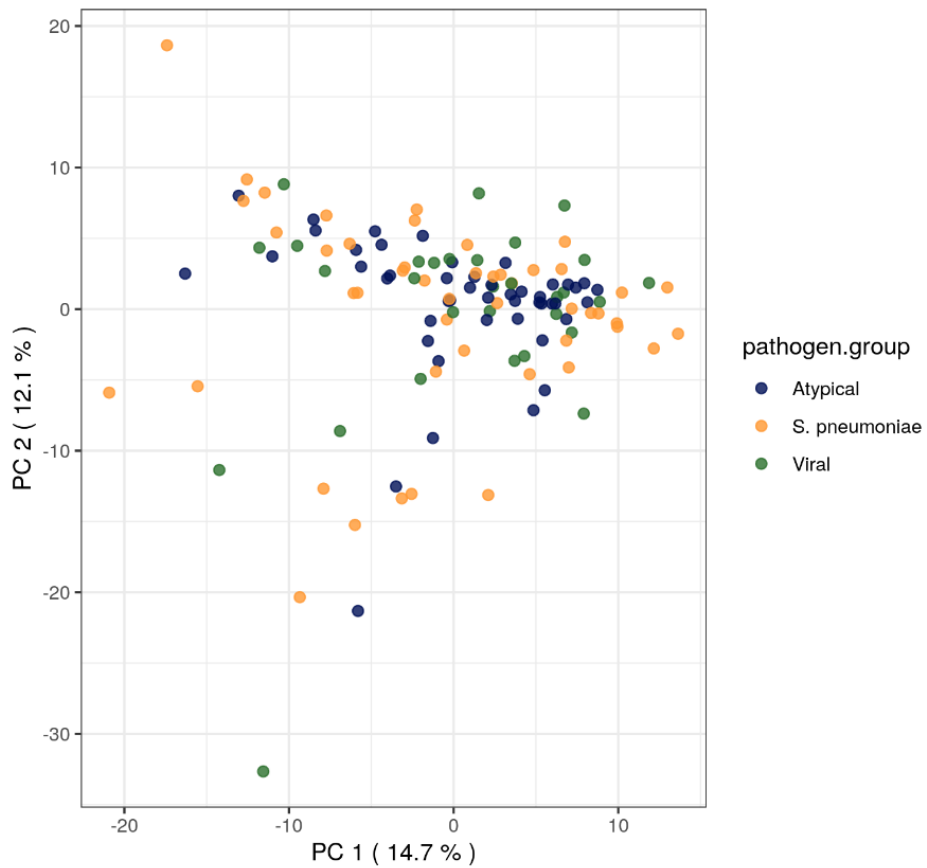


Figure 2.6 Unsupervised principal component analysis (PCA) plot of all pathogen groups.

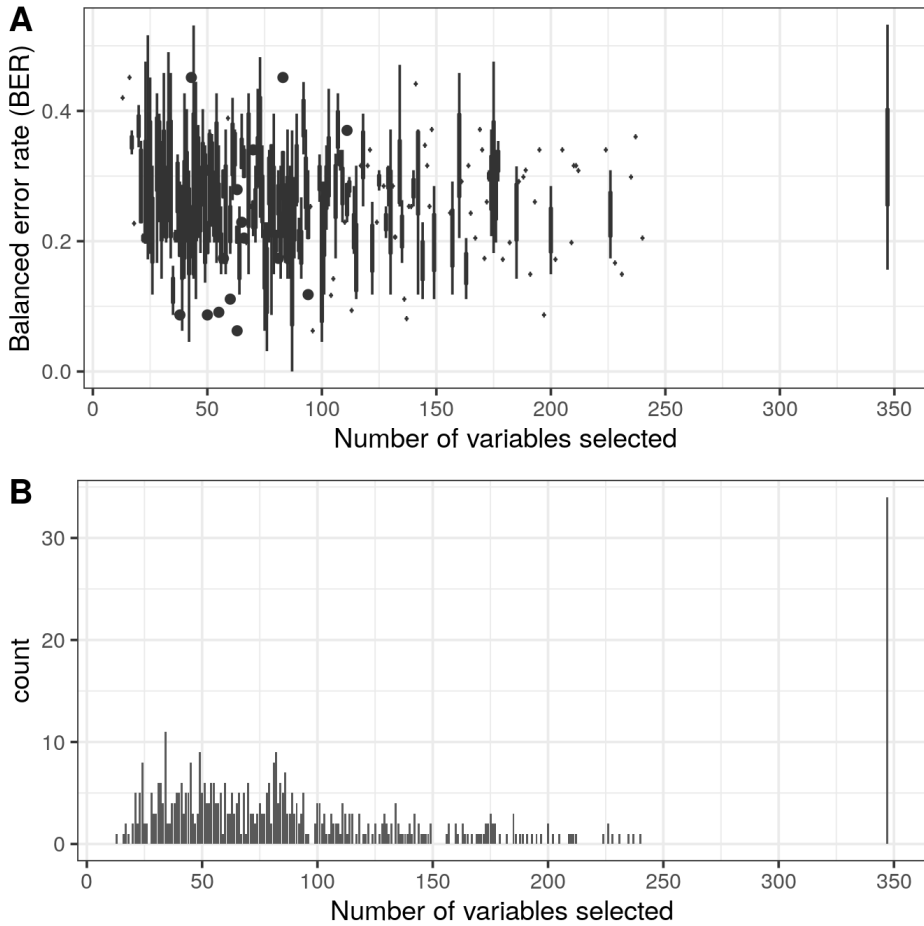


Figure 2.7 Boxplot of BER per number of variables selected shows no clear relation between the number of variables selected and model performance. (B) Histogram of the number of variables selected shows that a model with all metabolites included is favored, followed by models including 34, 49, 82, 24, or 45 metabolites. Both Figs contain the data of all folds and repeats ($n=500$) for the comparison between atypical versus *S. pneumoniae* and viral infections.

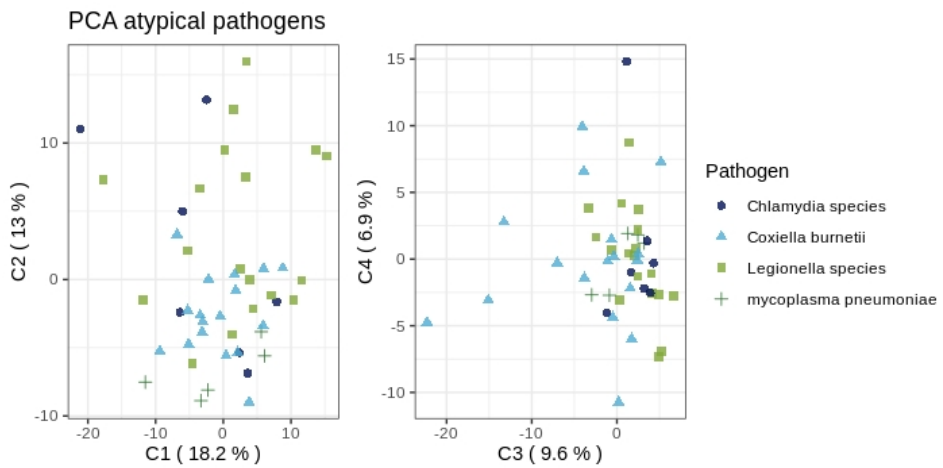


Figure 2.8 Principal component analysis (PCA) of the atypical pathogen group (log-transformed and standardized data) shows that there is no clear subgroup within the atypical group that would prominently drive the separation from the *S. pneumoniae* and viral infections.

Table 2.4 Summary of previous studies focusing on bacterial and viral respiratory tract infections and related metabolites.

Compared groups	Matrix	Analytical method	Significantly altered metabolites	Reference
30 CAP vs 30 HC	Plasma	NMR	Upregulated 1-methylhistidine	Zhou et al. (2015)
30 CAP vs 38 TB	Plasma	NMR	lactate, pyruvate, lipids, ketone bodies	Zhou et al. (2015)
11 pneumonia vs 11 HC (children)	Plasma	UPLC-TOF-MS	uric acid, hypoxanthine, glutamic acid	Laiakis et al. (2010)
11 pneumonia vs 11 HC (children)	Urine	UPLC-TOF-MS	uric acid, L-histidine	Laiakis et al. (2010)
47 pneumonia vs 47 HC	Urine	NMR	glucose, lactate, ketone bodies, amino acids [alanine, asparagine, isoleucine, leucine, lysine, serine, threonine, tryptophan, tyrosine, valine], carnitine, acetylcarnitine, hypoxanthine, fucose, myo-inositol, taurine, quinolate, adipate, dimethylamine, creatine, 2-oxoglutarate, fumarate	Slupsky et al. (2009)
30 CAP vs 46 TB	Plasma	UPLC-QTOF-MS	12(R)-hydroxyeicosatetraenoic acid, ceramide (d18:1/16:0), cholesterol sulfate, 4a-formyl-4b-methyl-5a-cholesta-8-en-3b-ol	Lau et al. (2015)
42 Influenza A vs 30 Bacterial CAP	Plasma	NMR, GC-MS	3-Methyl-2-Isovalerate, 3-Methyl-2-oxovalerate, 4-Hydroxybutyrate, Adipate, Alanine, Arabinonic acid, Asparagine, Aspartic Acid, Citrate, Citric acid, Fumerate, Histidine, Lysine, Methionine, Myoinositol, Phenylalanine, Serine, Threonic Acid, Threonine, Tyrosine, Uric acid, Urea	Banoei et al. (2017)
55 RVS vs 24 Bacterial pneumonia vs 37 HC (children)	Urine	NMR	3-Hydroxyisovalerate, 3-Indoxylsulfate, Acetoacetate, Betaine, Blue 1.06, Ethanolamine, Glutamate, N,N-Dimethylglycine, Pantothenate, Succinate, Tartrate, Uracil	Adamko et al. (2016)
			Downregulated lactate, ketone bodies amino acids[leucine, isoleucine, valine], 1-methylhistidine, glucose, nicotinate, GPC L-tryptophan, adenosine-diphosphate citrate, trigonelline, 1-methylnicotinamide, succinate, levoglucosan, 1-methylhistidine 2-amino Butanoic acid, Acetoacetate, Alkane, Benzoic acid, Beta-alanine, Carnitine, Dimethylamine, Formate, Glycine, Gulonic acid, Hexanoic acid, Leucine, Lactic acid, Pentadecane, Pyruvic acid, Quinic acid Hippurate, Serine, Threonine	

Abbreviations: CAP: community-acquired pneumonia; VAP: ventilator-associated pneumonia; HAP: hospital-acquired pneumonia; TB: tuberculosis; RSV: respiratory syncytial virus; HC: healthy control; NMR: nuclear magnetic resonance; UPLC: ultra-performance liquid chromatography; GC: gas chromatography; TOF: time-of-flight; QTOF: quadrupole time-of-flight; MS: mass spectrometry.

Table 2.5 Additional patient characteristics per pathogen group.

	<i>S. pneumoniae</i> (N=48)	Atypical (N=47)	Viral (N=30)	P-value
Race				
Other	1 (2.1%)	1 (2.1%)	0 (0%)	0.81
White	31 (64.6%)	46 (97.9%)	30 (100%)	
Missing	16 (33.3%)	0 (0%)	0 (0%)	
Nursing home resident				
No	46 (95.8%)	47 (100%)	25 (83.3%)	0.07
Yes	1 (2.1%)	0 (0%)	4 (13.3%)	
Missing	1 (2.1%)	0 (0%)	1 (3.3%)	
Altered mental status				
No	43 (89.6%)	42 (89.4%)	27 (90.0%)	0.85
Yes	3 (6.2%)	5 (10.6%)	3 (10.0%)	
Missing	2 (4.2%)	0 (0%)	0 (0%)	
Respiratory rate				
Mean (SD)	25.3 (6.64)	25.5 (6.44)	26.9 (7.32)	0.81
Median [Min, Max]	25.5 [12.0, 40.0]	26.0 [14.0, 40.0]	29.0 [12.0, 44.0]	
Missing	8 (16.7%)	8 (17.0%)	6 (20.0%)	
Systolic blood pressure				
Mean (SD)	131 (25.3)	133 (15.8)	137 (23.2)	0.81
Median [Min, Max]	130 [88.0, 226]	130 [99.0, 161]	135 [90.0, 186]	
Missing	1 (2.1%)	0 (0%)	1 (3.3%)	
temperature				
Mean (SD)	23.5 (8.41)	24.2 (11.5)	19.9 (9.40)	0.37
Median [Min, Max]	24.0 [6.00, 42.0]	24.0 [1.00, 41.0]	20.0 [3.00, 39.0]	
Missing	1 (2.1%)	0 (0%)	0 (0%)	
pulse				
Mean (SD)	104 (21.6)	96.6 (18.6)	94.3 (17.9)	0.28
Median [Min, Max]	109 [60.0, 144]	93.0 [50.0, 140]	96.0 [60.0, 120]	
Missing	1 (2.1%)	0 (0%)	0 (0%)	
pH				
Mean (SD)	12.7 (4.64)	14.5 (4.28)	12.0 (4.66)	0.31
Median [Min, Max]	14.0 [3.00, 21.0]	14.0 [3.00, 22.0]	13.0 [1.00, 19.0]	
Missing	8 (16.7%)	19 (40.4%)	5 (16.7%)	
BUN				
Mean (SD)	38.9 (24.9)	46.0 (21.6)	46.5 (24.1)	0.50
Median [Min, Max]	35.0 [2.00, 81.0]	48.0 [1.00, 84.0]	52.0 [4.00, 82.0]	
Missing	1 (2.1%)	0 (0%)	1 (3.3%)	
sodium				
Mean (SD)	132 (4.68)	131 (5.59)	136 (5.03)	0
Median [Min, Max]	132 [117, 141]	132 [119, 141]	136 [125, 152]	
Missing	1 (2.1%)	0 (0%)	0 (0%)	
glucose				
Mean (SD)	34.6 (13.9)	34.7 (14.0)	36.9 (17.6)	0.85
Median [Min, Max]	35.0 [1.00, 58.0]	35.0 [5.00, 59.0]	42.0 [2.00, 59.0]	
Missing	3 (6.2%)	0 (0%)	5 (16.7%)	
hematocrit				
Mean (SD)	11.0 (4.29)	11.5 (3.75)	10.5 (3.75)	0.81
Median [Min, Max]	11.0 [1.00, 19.0]	12.0 [1.00, 17.0]	11.5 [1.00, 16.0]	
Missing	2 (4.2%)	0 (0%)	2 (6.7%)	

	<i>S. pneumoniae</i> (N=48)	Atypical (N=47)	Viral (N=30)	P-value
Partial pressure of oxygen				
Mean (SD)	30.6 (16.5)	31.8 (17.7)	30.1 (16.3)	
Median [Min, Max]	34.5 [2.00, 56.0]	33.0 [1.00, 55.0]	37.0 [1.00, 50.0]	0.92
Missing	8 (16.7%)	19 (40.4%)	5 (16.7%)	
Pleural effusion on x ray				
No	39 (81.2%)	45 (95.7%)	25 (83.3%)	
Yes	8 (16.7%)	2 (4.3%)	5 (16.7%)	0.31
Missing	1 (2.1%)	0 (0%)	0 (0%)	

Data are presented as number (%) or mean (SD). Abbreviations: BUN: blood urea nitrogen.

Table 2.6 Overview of the number of metabolites included in the metabolomics platforms, measured in the samples and included in the data analysis.

Measurement platform	Number of metabolites included in platform	Number of metabolites measured in samples	Number of metabolites included in data analysis
Amines	74	55	55
Acylcarnitines	48	28	28
Organic acids	28	13	13
Negative lipids	30	16	16
Signaling lipids	231	113	91
Positive lipids	185	149	144
Total	596	374	347

Table 2.7 Information on measurement platforms used, metabolite classes targeted per platform, targeted metabolites, their abbreviations and names in R (if detected) and identifiers (if available).

This table is available as Excel file (S4 Table) on the website of the publisher at <https://doi.org/10.1371/journal.pone.0252378>.

Table 2.8 Metabolomics data after quality control.

This table is available as csv file (S5 Table) on the website of the publisher at <https://doi.org/10.1371/journal.pone.0252378>.

CHAPTER 3

Differential metabolic host response to pathogens associated with community-acquired pneumonia

Ilona den Hartog, Naama Karu, Laura B. Zwep, G. Paul Voorn, Ewoudt M.W. van de Garde, Thomas Hankemeier, J.G.C. van Hasselt. Differential metabolic host response to pathogens associated with community-acquired pneumonia. *Metabolism Open* **18** (2023).

Abstract

Background: Metabolic changes induced by the host immune response to pathogens found in patients with community-acquired pneumonia (CAP) may provide insight into its pathogenesis. In this study, we characterized differences in the host metabolic response to common CAP-associated pathogens.

Method: Targeted metabolomic profiling was performed on serum samples obtained from hospitalized CAP patients (n=119) at admission. We quantified 347 unique metabolites across multiple biochemical classes, including amines, acylcarnitines, and signaling lipids. We evaluated if unique associations between metabolite levels and specific CAP-associated pathogens could be identified.

Results: Several acylcarnitines were found to be elevated in *C. burnetii* and herpes simplex virus (HSV), and lowered in *M. pneumoniae* as compared to other pathogens. Phenylalanine and kynurenine were found elevated in *L. pneumophila* as compared to other pathogens. S-methylcysteine was elevated in patients with *M. pneumoniae*, and these patients also showed lowered cortisol levels in comparison to almost all other pathogens. For the herpes simplex virus, we observed a unique elevation of eicosanoids and several amines. Many lysophosphatidylcholines showed an altered profile in *C. burnetii* versus *S. pneumoniae*, *L. pneumophila*, and respiratory syncytial virus. Finally, phosphatidylcholines were negatively affected by the influenza virus in comparison to *S. pneumoniae*.

Conclusions: In this exploratory analysis, metabolites from different biochemical classes were found to be altered in serum samples from patients with different CAP-associated pathogens, which may be used for hypothesis generation in studies on differences in pathogen host response and pathogenesis of CAP.

3.1 Introduction

Community-acquired pneumonia (CAP) is a common infection of the lower respiratory tract, caused by bacterial or viral pathogens [51]. Causative pathogens that are commonly found in CAP patients include *Streptococcus pneumoniae*, respiratory viruses, *Haemophilus influenzae*, and, to a lesser extent *Mycoplasma pneumoniae* and *Legionella pneumophila* [51, 29, 30]. CAP is associated with high mortality and morbidity, especially in the elderly [3, 31], and with distinct differences in clinical outcomes between different CAP-associated pathogens [52, 53].

There is an increasingly recognized role of the patient-associated host metabolic response to infection and its association with overall innate immune system activation and clinical outcomes [17, 34, 54]. Previous studies have found distinct effects of this host metabolic response in COVID-19 and sepsis [55, 56, 57, 58]. We previously evaluated the discriminatory power of host-associated metabolites to support the microbial diagnosis of CAP and identified metabolic biomarkers to support the diagnosis of atypical bacterial pathogens [59].

In this study, we aimed to further characterize the differential metabolic host response associated with distinct CAP-associated pathogens, which may support understanding the role of metabolic changes in the pathogenesis associated with specific pathogens. To this end, we quantified 347 unique metabolites in 119 serum samples taken at hospital admission from patients with CAP and studied their association with the pathogens found in these patients.

3.2 Materials and methods

3.2.1 Study population

We analyzed previously collected serum samples obtained from 119 hospitalized patients with CAP, obtained as part of two previously conducted clinical studies executed between October 2004 and September 2010 and collected within 24 hours after hospital admission [29, 30]. For all patients, microbial disease etiology was confirmed using conventional diagnostic methods such as culturing, PCR, and urinary antigen tests. Inflammatory markers such as CRP and leukocyte counts were also available. Most patients (80%) were diagnosed with bacterial infections and the sample size per pathogen varied between 3 to 48 samples. Pathogen groups, a group of patients in which the same pathogen was found, including only one patient were excluded. None of the included patients had received treatment with corticosteroids. The patient distribution per pathogen showed no significant differences in sex and pneumonia severity index (PSI) score. However, the mean age differed significantly between pathogen groups ($p < 0.05$). All patients that participated in the previous clinical studies provided informed consent for the secondary use of their materials. Data were handled in an anonymized

way. An overview of the patient characteristics is provided in Table 3.1 and additional information is available in Table 3.2.

Table 3.1 Patient characteristics per CAP-associated pathogen.

Patient characteristic	<i>S. pneumoniae</i> (n=48)	<i>L. pneumophila</i> (n=18)	<i>C. burnetii</i> (n=17)	<i>C. psittaci</i> (n=7)	<i>M. pneumoniae</i> (n=5)	Influenza virus (n=11)	HS virus (n=6)	RS virus (n=4)	Parainfluenza virus (n=3)
Age (years) *									
Mean (SD)	62.2 (18.9)	62.9 (12.5)	47.4 (10.0)	62.6 (14.5)	38.8 (10.4)	67.0 (15.9)	71.3 (18.6)	69.3 (25.9)	83.3 (5.77)
Sex									
Male	22 (45.8%)	13 (72.2%)	14 (82.4%)	5 (71.4%)	2 (40.0%)	9 (81.8%)	5 (83.3%)	2 (50.0%)	2 (66.7%)
PSI score									
<50	9 (18.8%)	0 (0%)	5 (29.4%)	1 (14.3%)	3 (60.0%)	1 (9.1%)	0 (0%)	1 (25.0%)	0 (0%)
51-70	7 (14.6%)	3 (16.7%)	9 (52.9%)	0 (0%)	1 (20.0%)	3 (27.3%)	1 (16.7%)	0 (0%)	0 (0%)
71-90	5 (10.4%)	6 (33.3%)	1 (5.9%)	2 (28.6%)	1 (20.0%)	1 (9.1%)	1 (16.7%)	1 (25.0%)	0 (0%)
91-130	23 (47.9%)	7 (38.9%)	1 (5.9%)	4 (57.1%)	0 (0%)	6 (54.5%)	2 (33.3%)	1 (25.0%)	2 (66.7%)
131>	4 (8.3%)	2 (11.1%)	1 (5.9%)	0 (0%)	0 (0%)	0 (0%)	2 (33.3%)	1 (25.0%)	1 (33.3%)
Duration of symptoms before admission *									
Mean (SD)	4.06 (3.03)	5.11 (1.23)	5.06 (1.71)	4.14 (2.04)	13.4 (16.1)	4.45 (2.34)	5.17 (4.45)	6.50 (5.07)	4.00 (2.65)
CRP upon admission									
Mean (SD)	265 (166)	306 (90.8)	233 (115)	272 (110)	145 (86.0)	207 (87.9)	271 (121)	289 (135)	220 (142)
Leukocyte count upon admission*									
Mean (SD)	16.9 (5.99)	13.0 (4.24)	9.49 (1.56)	10.7 (5.03)	10.5 (3.29)	10.1 (4.93)	15.3 (4.89)	14.2 (7.94)	13.7 (9.30)

* Significant difference between pathogens (p -value < 0.05). Abbreviations: HS: herpes simplex; RS: respiratory syncytial; PSI: pneumonia severity index; CRP: C-reactive protein; SD: standard deviation.

3.2.2 Bioanalytical procedures

Metabolomic profiling was performed by the Biomedical Metabolomics Facility of Leiden University (BMFL), Leiden, The Netherlands. Details of the bio-analytical procedures are described elsewhere [59]. Briefly, the serum samples were quantitatively analyzed using five liquid- and one gas- chromatography methods coupled to mass spectrometry detectors, with a total coverage of 596 metabolite targets from 25 biochemical metabolite classes such as organic acids, amino acids, biogenic amines, acylcarnitines, and lipids. Following data integration and quality control steps, 374 unique metabolites were reported and underwent statistical analysis.

3.2.3 Data analysis

The metabolite levels were log-transformed and standardized prior to statistical analysis. Biological and biochemical relevant sums and ratios of metabolites (Table 3.3) were calculated from the raw metabolite data and were added as variables to the dataset, undergoing the same analysis together with metabolites.

Exploratory data analysis was performed using principal component analysis (PCA). To identify differences in metabolite levels between pathogens, we performed analysis of variance (ANOVA) with a false discovery rate (FDR) multiple testing correction. P-adjusted significant metabolites ($q < 0.1$) were tested with Tukey's posthoc test to identify which pathogen groups the metabolite was able to distinguish between. In addition to the significance threshold, we applied a cutoff of >20% fold change between the medians of the raw metabolite levels of each group.

All data analysis was performed using R. The scripts used for the data analysis are available on GitHub at <http://github.com/vanhasseltlab/MetabolicHostResponseToCAP>.

3.3 Results and discussion

3.3.1 Metabolic markers differentiate between pathogens

The metabolomics analysis of the patient serum samples yielded 347 quantifiable unique metabolites from different biochemical metabolite classes. No visible separation between samples obtained from patients infected by the various pathogens was apparent using PCA (Figure 3.6). Systematic analysis of individual metabolites for a specific pathogenic species yielded 64 unique metabolites that were significantly altered (ANOVA, $q < 0.1$ and $FC > 20\%$) as compared to other pathogenic species (Figure 3.1-3.5). All significant results are shown together in Figure 3.8 and Table 3.4. Large variation in group size for specific pathogenic species ($n=3-48$) and large within pathogen species group variation was present (Figure 3.7), limiting the ability to identify associations between metabolites and specific pathogenic species. Nonetheless, although often not statistically significant, trends between groups of metabolites and specific pathogenic

species were apparent, which is why we have maintained these in visualization for each metabolic class (Figure 3.1-3.4). The discriminating metabolites belong to different biochemical classes. The results are reported by biochemical class to allow focused biochemical discussion. A comprehensive qualitative overview of the main associations between pathogen species and metabolites identified has been summarized in Figure 3.5.

3.3.2 Carnitine and acylcarnitines

We identified significant differences in carnitines between serum samples of CAP patients that carried the pathogens *C. burnetii*, herpes simplex virus (HSV), or *M. pneumoniae* compared to CAP patients with other pathogens (Figure 3.1). The most consistent and statistically significant differences were found in patients with *C. burnetii* compared to patients with *L. pneumophila* or *S. pneumoniae* infections. Patients with *C. burnetii* showed 60-80% elevated long-chain acylcarnitines (LCAC), leading to a 2-fold lower ratio of free carnitine/total acylcarnitines. Changes in acylcarnitines can be related to mitochondrial energy production shifts between glycolysis and beta-oxidation of long-chain fatty acids [60, 61]. Naturally, such shifts depend on the glucose stores and the availability of carnitine as a limiting factor.

We observed increased levels of carnitine and short-chain acylcarnitines (SCAC) for *L. pneumophila* vs. *C. burnetii* and *M. pneumoniae*. Elevated carnitine and short-chain acylcarnitines suggest successful fatty acid beta-oxidation. Conversely, we also saw elevated mid-chain acylcarnitines (MCAC), most pronounced in patients with herpes simplex virus versus *C. burnetii*. MCAC could indicate incomplete beta-oxidation of long-chain fatty acids and may be related to mitochondrial dysfunction in the diseased state [62]. In contrast, LCAC were elevated mainly in patients with *C. burnetii* infection, suggesting a low rate of beta-oxidation rather than an incomplete process. It is unknown how well the levels of circulating acylcarnitines reflect the cellular levels, however, accumulation of long-chain acylcarnitines in lung cells indicated a defective beta-oxidation process, increasing oxidative stress and potentially leading to epithelial cell death [63]. Moreover, LCAC accumulation in the air-fluid interface of the lung can reduce lung function due to the inhibition of pulmonary surfactant [64].

It has been reported that impaired utilization of acylcarnitines for energy production can be caused by various viruses, including influenza and respiratory syncytial virus [63], although the latter two did not show any significant results in our study.

Overall, for acylcarnitines high within-pathogen group variance was present (Figure 3.7), which limited identifying further metabolite-pathogen associations. Acylcarnitine levels are associated with food consumption, certain medications such as valproic acid, and carnitine supplementation [65, 66, 67].

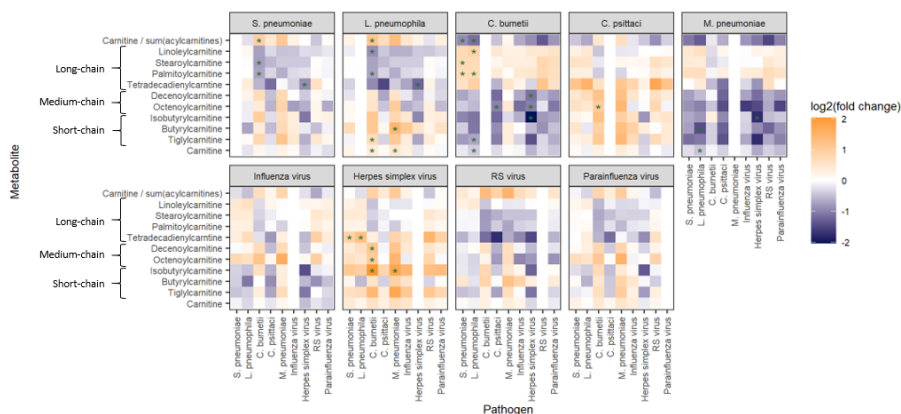


Figure 3.1 Fold changes of carnitine and acylcarnitines as median values per pathogen group compared to CAP patients with other pathogens. Significant differences ($q < 0.1$ and $FC > 20\%$) are marked with a green star.

3.3.3 Organic acids, amino acids, and related compounds

Significant differences in serum amino acids and derivatives were found in all groups of patients infected by CAP-associated pathogens (Figure 3.2). Several significantly different metabolites were only found in infections by specific pathogens. For example, patients with *L. pneumophila* had elevated kynurenine (1.5-1.7 fold) and phenylalanine (1.4-1.6 fold), while *M. pneumoniae* showed elevated S-methylcysteine (1.3-1.7 fold). Also, glutamate was elevated in *C. burnetii* compared to *L. pneumophila* (1.8 fold). Most of the stronger findings can be related to inflammation and oxidative stress and were associated with negative health outcomes in studies of various infections.

Elevated kynurenine was observed in patients with CAP and sepsis [68, 69], and severe COVID-19 [56, 70, 71]. Kynurenine increases due to enhanced IDO (Indoleamine-2,3-dioxygenase) conversion of tryptophan during the inflammatory response. IDO is expressed in the lungs and blood cells, and its activity and expression are induced by pro-inflammatory cytokines, viral proteins, and lipopolysaccharides [72, 73, 74]. Kynurenine acts as an immune suppressor via binding to the aryl-hydrocarbon receptor (AHR) [75], and also inhibits nitric oxide synthase (NOS), thereby increasing oxidative stress. Kynurenine also contributes to vasodilation and hypotension in sepsis [58, 76] and increases endothelial activation which leads to impaired microvascular reactivity [68, 77].

Elevated phenylalanine was associated with immune activation and negative outcome in patients with COVID-19 [55, 56], post-trauma, or sepsis [57, 58]. A possible explanation for elevated phenylalanine in the inflammatory environment is ROS-induced oxidation of the cofactor tetrahydrobiopterin (BH4) which is essential for the conversion

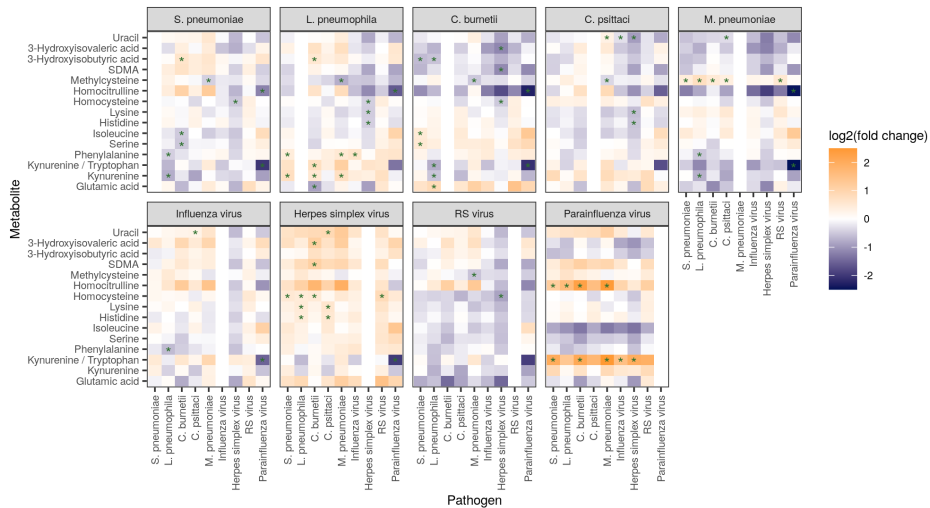


Figure 3.2 Fold changes of organic acids, amino acids, and derivatives as median values per pathogen group compared to CAP patients with other pathogens. Significant differences ($q < 0.1$ and $FC > 20\%$) are marked with a green star.

of phenylalanine to tyrosine and the downstream catecholamines by the enzyme phenylalanine 4-hydroxylase [57, 78].

Glutamate, a neurotransmitter in the CNS, also modulates the activity of immune cells via the NMDA receptor, and excessive production was related to damaged endothelial cells [79, 80, 81].

We found elevated uracil in patients with influenza and herpes simplex virus compared to patients with *C. psittaci* (1.2-1.9 fold). Patients with *S. pneumoniae* and *L. pneumophila* showed higher 3-hydroxyisobutyric acid (1.5-1.6 fold) and lower isoleucine (1.4 fold lower) compared to *C. burnetii*.

3.3.4 Signaling lipids

The signaling lipids covered in the applied platforms included oxylipins, endocannabinoids, their precursor fatty acids, and also cortisol. Metabolites that showed significant results were included in Figure 3.3.

Although we did not find significant differences in fatty acids in this study, the oxylipin profile of patients with herpes simplex virus was unique in comparison to other infections, showing significantly higher (all $FC > 1.8$) prostaglandin F2 alpha (PGF2a), 9,10,13-TriHOME (TriHOME), and thromboxane B2 (TXB2). Oxylipins are immune modulators produced by the peroxidation of C20 or C22 polyunsaturated fatty acids (PUFA). The omega-6 PUFA oxylipins exhibit mainly pro-inflammatory action

during the acute phase of inflammation. Some of these oxylipins also aid in the transition from innate to adaptive immune response. The arachidonic acid-produced PGF₂-alpha is a direct metabolite of PGE₂ and its increase aligns with the existing evidence that prostaglandin synthesis reactivates the herpes virus and increases its spread [82]. Although fluctuating along the course of inflammation, low levels of prostanoids (PGE₂ and TXB₂) are characteristic of sepsis non-survivors [83, 84] and COVID-19 [85]. TriHOMEs were reported to be dysregulated in respiratory diseases such as asthma and COPD [86, 87].

Of the covered endocannabinoids, several fatty acyl ethanolamides (EAs) were significantly different between patients with various infections. Docosatetraenoyl ethanolamide (DEA), which is the ethanolamide of the omega-6 adrenic acid (22:4) docosatetraenoic acid, was elevated in patients with *C. psittaci* compared to patients with *S. pneumoniae* or influenza virus (1.3-1.4 fold). Patients with the influenza virus showed the lowest levels of both DEA and dihomo-gamma-linolenoyl ethanolamide (DGLEA), the ethanolamide of the omega-6 (20:3) dihomo-gamma-linolenic acid. In comparison to other pathogens, patients with herpes simplex virus showed trends of elevated palmitoleoyl ethanolamide (POEA) and alpha-linolenoyl ethanolamide (aLEA), which are the ethanolamides of the omega-7 (16:1) palmitoleic acid, and the omega-3 (18:3) alpha-linolenic acid, respectively. A 2-fold significant elevation of POEA was measured in comparison to patients with *C. burnetii* and a 1.4-fold aLEA elevation in comparison to *M. pneumoniae*. Endocannabinoids do not only act as CNS modulators but are also peripheral immune mediators. Their pro- or anti-inflammatory activity varies and depends on their acyl group, the type of cells (immune; endothelial, etc.), and the receptors they bind to. In times of increased demand, for example during an acute immune response, and depending on diet and fat reserves, increased levels of endocannabinoids may reflect accelerated catabolism of cellular lipids, to serve as precursors for free fatty acids and oxylipins. It is also possible to see a shift towards the synthesis of fatty acyl EAs from fatty acids that occurs without medication [88], or due to treatment with corticosteroids and NSAIDs, that inhibit PLA₂ and COX enzymes [89, 90]. The last significant signaling lipid, cortisol, was lower in patients with *M. pneumoniae* compared to all other pathogens (2-3.3 fold lower), and lower for *C. burnetii* patients in comparison to *S. pneumoniae* and *L. pneumophila* patients (1.6 fold lower, Figure 3.8). Reduced excretion of cortisol indicates lower activation of the hypothalamic-pituitary-adrenal (HPA) axis in response to stress and innate immune activation following any infection. Cortisol stimulates gluconeogenesis and reduces inflammation, therefore its levels also affect (directly or indirectly) other metabolites measured in this study. In multiple studies on infections of the lower respiratory system, higher cortisol levels at hospital admission were predictive of mortality [91], indicating its relation to acute inflammation and disease severity.

3.3.5 Phospholipids and derivatives

Phospholipids form cell membranes and are hydrolyzed into lysophospholipids and long-chain fatty acids by various phospholipase A₂ (PLA₂) enzymes. The host's innate

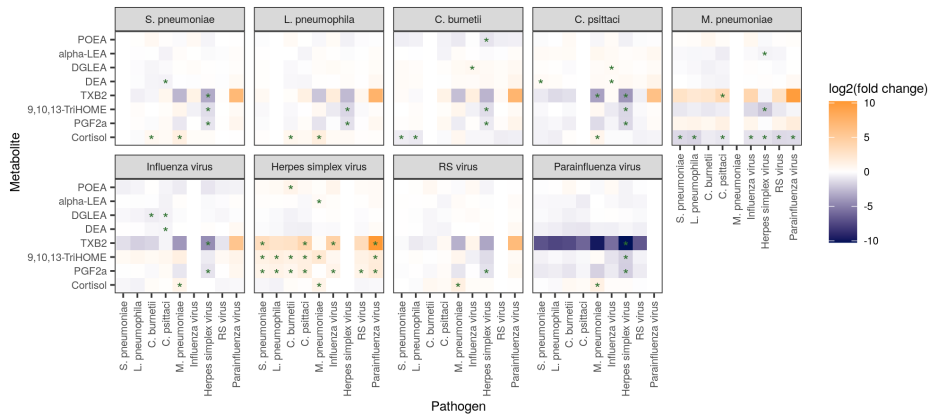


Figure 3.3 Fold changes of oxylipin, endocannabinoid, and cortisol as median values per pathogen group compared to CAP patients with other pathogens. Significant differences ($q < 0.1$ and $FC > 20\%$) are marked with a green star.

immune response to infection, as well as the pathogen itself, can induce the activity of PLA2 [92]. Patients with *S. pneumoniae* infection showed the lowest levels of any lysophospholipids (LPCs) compared to other pathogens, alongside the highest levels of intact phosphatidylcholines (PCs), suggesting reduced activity of PLA2. Such a profile was also characteristic of sepsis [91, 93], possibly due to the varying degree of expression and activities of different PLA2 enzymes. Many LPCs and the total LPC/PC ratio were significantly higher in *C. burnetii* compared to *S. pneumoniae*, *L. pneumophila*, and RS virus (1.5-4 fold, Figure 3.4). Bacteria can use host lipids as building blocks for bacterial membrane formation and as an energy source. Also, host lipids are crucial for the entry of bacteria, viruses or toxins into cells. For example, *C. burnetii* has been found to use phospholipase A to integrate lipids into its membrane. [94]. *L. pneumophila* increased its virulence by secretion of PLA2 that destructed alveolar surfactant phospholipids [95]. Conversely, many intact PCs showed significantly lower levels in patients with the influenza virus in comparison to those with *S. pneumoniae*. The M protein of viruses like the influenza virus can interact with lipids and phosphatidylcholine vesicles allowing membrane fusion [96, 97, 98, 99].

Regarding lysophosphatidyls with serine or inositol functional groups (LPS and LPI), patients with parainfluenza virus showed 3-5 fold higher LPS (22:6) compared to *S. pneumoniae*, *C. psittaci*, and herpes simplex virus. Patients with *L. pneumophila* infection showed 30-90% higher LPIs compared to *S. pneumoniae*, influenza, and herpes simplex virus.

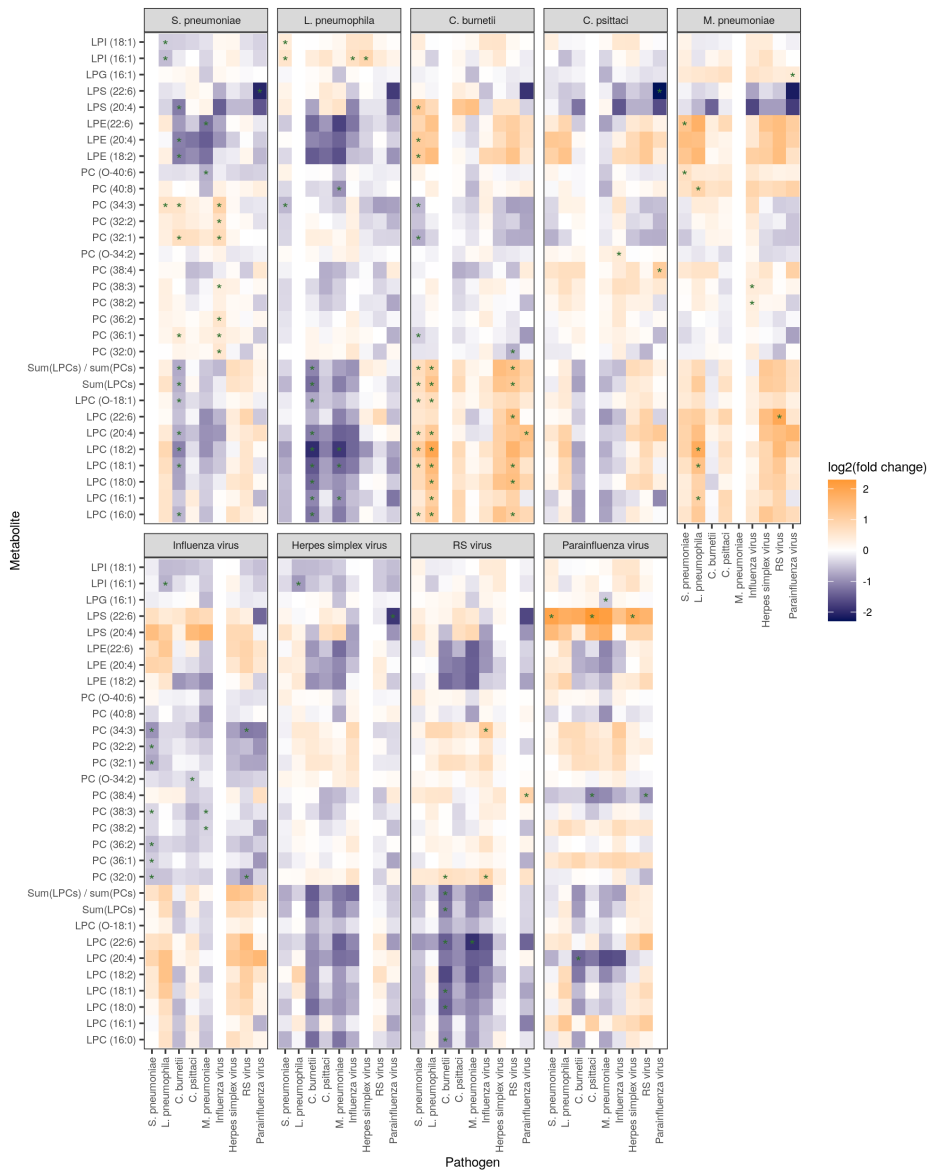


Figure 3.4 Fold changes of phospholipids including (lyso)phosphatidylcholines median values per pathogen group as median values per pathogen group compared to CAP patients with other pathogens. Significant differences ($q < 0.1$ and $FC > 20\%$) are marked with a green star.

3.3.6 Summary of metabolic findings per pathogen

The consolidation of the obtained results associates specific bacterial and viral CAP infections with metabolic perturbations related to inflammation, oxidative stress, lipids utilization, and energy production (Figure 3.5). Patients with *S. pneumoniae* infection showed high levels of the stress hormone cortisol that can be excreted as an HPA axis response to pro-inflammatory cytokines and hypoxia. Moreover, patients with *S. pneumoniae* infections consistently showed the highest PCs in comparison to all other pathogen groups, and very low lysophospholipids, suggesting lower activation of PLA2 enzymes, as described before in studies on sepsis [58, 91]. The lowest levels of LPCs were found in patients with *L. pneumophila* infection, who also showed the highest levels of cortisol, kynurenine, and phenylalanine, all characteristic of acute inflammatory response. Possibly indicating signs of oxidative stress and potentially also mitochondrial dysfunction [100], these patients had increased production of energy via beta-oxidation of fatty acids, evidenced by the elevated end-products of free carnitine and short-chain acylcarnitines. In comparison, patients with *C. burnetii* infection showed inhibition of fatty acid beta-oxidation, suggested by high levels of long-chain acylcarnitines, which are associated with increased ROS. These patients also had elevated glutamate accompanied by the highest levels of lysophospholipids, all can be linked to either inflammation, oxidative stress, or endothelial dysfunction [80, 81, 91, 98, 101]. Patients with *C. psittaci* infection possibly suffered from metabolic defects causing incomplete beta-oxidation of long-chain fatty acids, suggested by the elevated mid-chain fatty acids. This interrupted metabolism may be related to their relatively high levels of omega-6 fatty acyl ethanolamides, and high kynurenine, indicating acute inflammation. Despite the small sample size, patients with herpes simplex infection showed unique disruption in lipid metabolism. This included accumulation of long- and medium-chain acylcarnitines, elevated non-omega-6 fatty acyl ethanolamides, and very high levels of certain oxylipins, that possibly support the spread of the infection.

3.3.7 Study limitations and future perspectives

In this study, the sample size of patient groups with different CAP-associated pathogenic species was unbalanced and in several cases low (Table 3.1). Furthermore, a large within-group variance of metabolite levels was observed, which may impact our findings, especially in small sample size pathogen groups. Finally, the mean age of patients was unbalanced across the cohort, possibly acting as an additional confounder, since many aspects of metabolism change with age [102]. Overall, these factors limit the power to detect statistically significant metabolite-pathogen associations and require careful consideration of the relationships identified and further validation in a larger and more balanced study cohort. Nonetheless, several of the identified relationships are in line with expectations and other literature reports, supporting the biological relevance of our analysis.

Our study only considered metabolite levels at a single time point, even though patients will enter the clinical at different stages of their infection. The use of longitudinal

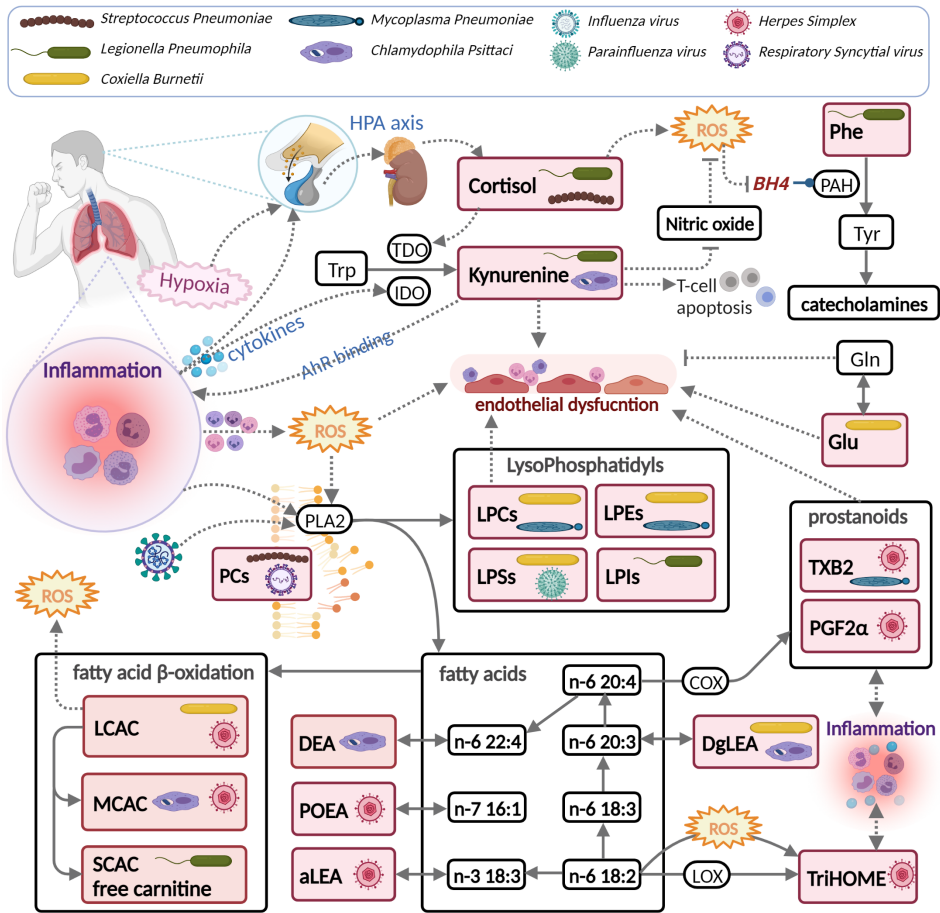


Figure 3.5 Overview of differential metabolic host response associations for community-acquired pneumonia-associated pathogens inferred from the current analysis. A pathogen symbol indicates that this patient group had the highest levels of metabolite(s) in the red box. Metabolites that were either not detected or detected without significant results are in white boxes. Enzymes are in ellipses, full lines are part of a metabolic pathway or transformation, and broken lines indicate an inducing or inhibiting effect. Abbreviations: see the abbreviation list. This figure was created with BioRender.com.

metabolite levels measured in patients represents an important next step to help to correct for such differences, and, to elucidate the relationship between changes in metabolite levels and biochemical or cellular biomarkers for immunological and/or inflammatory response markers, e.g. C-reactive protein levels, cytokines, and leukocyte cell counts [12].

Importantly, we have made available the full raw dataset underlying this study, which could be of interest for further bioinformatics analyses of the immune-metabolic responses reported [103].

3.4 Conclusion

In this study, we identified metabolites associated with the host response to specific CAP-associated pathogens, which is of relevance to further elucidate the pathogenesis of specific pathogens. We identified 64 metabolites that were significantly different between several pathogens. Specific bacterial and viral pathogens could be associated with metabolic perturbations that can be related to inflammation, oxidative stress, lipids utilization, and energy production. Further research should be conducted to validate our results and to further unravel the mechanisms of the metabolic host response to different pathogens causing CAP.

3.5 Supplementary data

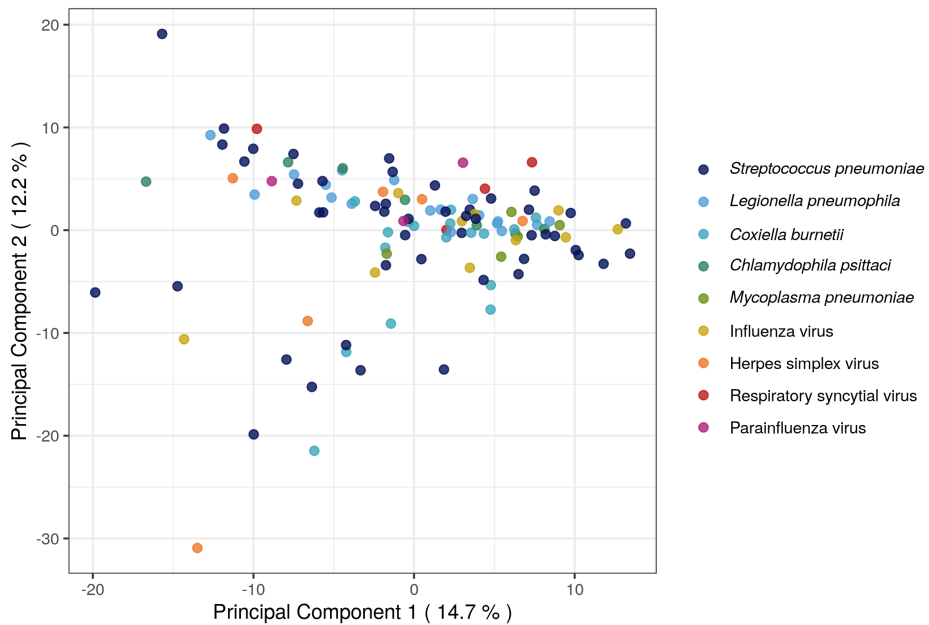


Figure 3.6 Principal component analysis (PCA) plot of all pathogens. The percentages denote the percentage of variance explained by each principal component. Data were log-transformed and standardized prior to analysis.

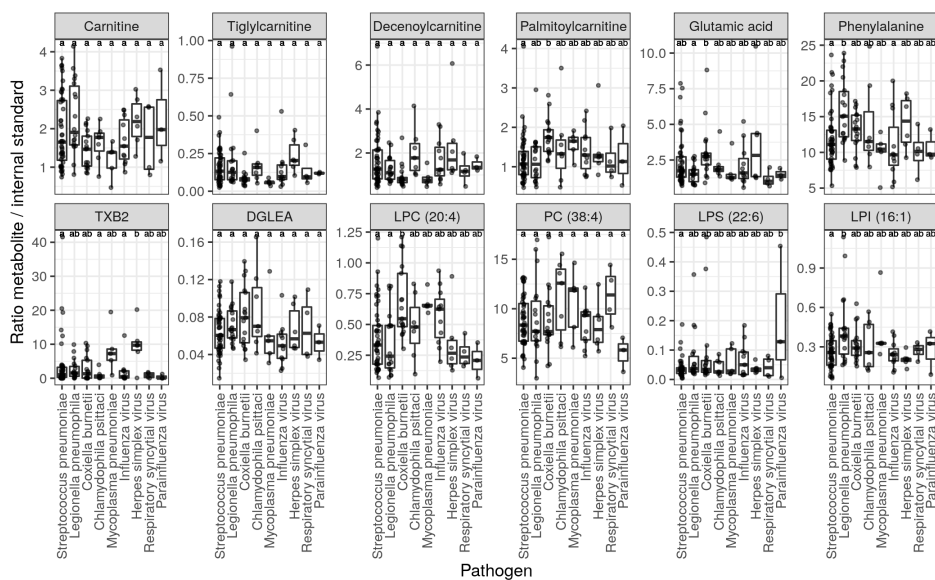


Figure 3.7 Box-scatter plots of selected pneumolite metabolites from different biochemical classes, that showed significant differences between pathogen patient groups. The box shows the lower quartile, the median, and the upper quartile of the metabolite ratios for each pathogen, with the whiskers denoting the minimum and maximum metabolite ratios within the interquartile range. Individual samples are shown as grey dots. Different letters above each bar represent significant differences between the groups.



Figure 3.8 Fold change of metabolite levels for all comparisons and all biochemical classes. Fold change values are shown only for metabolites with a significance of $q < 0.1$ & $FC > 20\%$.

Table 3.2 Extended patient characteristics

This table is available as pdf file (Table A.1) on the website of the publisher at <https://doi.org/10.1016/j.metop.2023.100239>

Table 3.3 Metabolite ratios and sums

Metabolite sum or ratio name in R	Metabolite sum or ratio formula
BCAA_sum	isoleucine + leucine + valine
TCA_cycle_sum	Citric acid + lactic acid + malic acid + fumaric acid
urea_cycle_sum	Citrulline + arginine + ornithine + fumaric acid
lc_Carnitines_sum	Myristoylcarnitine + Hexadecenoylcarnitine + Palmitoylcarnitine + Stearoylcarnitine + Dodecenoylcarnitine + Tetradecenoylcarnitine + Linoleylcarnitine + Oleoylcarnitine + Tetradecadienylcarnitine
mc_Carnitines_sum	Hexanoylcarnitine + Octanoylcarnitine + Octenoylcarnitine + Decanoylcarnitine + Lauroylcarnitine + Nonacylcarnitine + Pimelylcarnitine + Decenoylcarnitine
sc_Carnitines_sum	Acetylcarnitine + Propionylcarnitine + Isobutyrylcarnitine + Butyrylcarnitine + Tiglylcarnitine + Methylbutyrylcarnitine + Isovalerylcarnitine
Cer_sum	Cer(d18:1/22:1) + Cer. (d18:1/24:1. + Cer(d18:1/24:0) + Cer(d18:1/16:0) + Cer(d18:1/23:0) + Cer(d18:1/24:0)
SM_sum	Sphingomyelin (d18:1/14:0) + (d18:1/15:0) + (d18:1/16:0) + (d18:1/16:1) + (d18:1/17:0) + (d18:1/18:0) + (d18:1/18:1) + (d18:1/18:2) + (d18:1/20:0) + (d18:1/20:1) + (d18:1/21:0) + (d18:1/22:0) + (d18:1/22:1) + (d18:1/23:0) + (d18:1/ 23:1) + (d18:0/24:0) + (d18:0/24:1) + (d18:0/24:2) + (d18:0/25:0) + (d18:0/25:1)
LPC_sum	Lysophosphatidylcholine (14:0) + (16:0) + (16:1) + (18:0) + (18:1) + (18:2) + (18:3) + (20:4) + (20:5) + (22:6) + (O-16:1) + (O-18:1)
PC_sum	Diacyl-phosphatidylcholine (32:0) + (32:1) + (32:2) + (34:1) + (34:2) + (34:3) + (34:4) + (36:1) + (36:2) + (36:3) + (36:4) + (36:5) + (36:6) + (38:2) + (38:3) + (38:4) + (38:5) + (38:6) + (38:7) + (40:4) + (40:5) + (40:6) + (40:7) + (40:8) + (O-34:1) + (O-34:2) + (O-34:3) + (O-36:2) + (O-36:3) + (O-36:4) + (O-36:5) + (O-36:6) + (O-38:4) + (O-38:5) + (O-38:6) + (O-38:7) + (O-40:6) + (O-42:6) + (O-44:5)
HT5_Trp_ratio	Serotonine / Tryptophan
ADMA_Arg_ratio	ADMA / Arginine
SDMA_Arg_ratio	SDMA / Arginine
Carnitine_sum_lc_Carnitines_ratio	Carnitine / LCAC sum
Carnitine_sum_mc_Carnitines_ratio	Carnitine / MCAC sum
Carnitine_sum_sc_Carnitines_ratio	Carnitine / SCAC sum
DCA_CA_ratio	DCA / CA
FA_14.1_14.0	FA (14:1) / FA (14:0)
FA_16.1_16.0	FA (16:1) / FA(16:0)
Gln_Glu	Glutamine / Glutamic acid
Kyn_Trp	Kynurenine / Tryptophan
sum_BCAA_sum_Phe_Tyr_ratio	BCAA sum / (Phenylalanine + Tyrosine)
sum_CER_sum_SM_ratio	Cer sum / SM sum
sum_LPC_sum_PC_ratio	LPC sum / PC sum

Table 3.4 Significant metabolites for the host response to different CAP pathogens (q < 0.1, FC > 20%)

This table is available as Excel file (Table A.3) on the website of the publisher at <https://doi.org/10.1016/j.metop.2023.100239>

CHAPTER 4

Longitudinal metabolite profiling of *Streptococcus pneumoniae*-associated community-acquired pneumonia

Ilona den Hartog*, Laura B. Zwep*, Thomas Hankemeier, Jacqueline J. Meulman, Ewoudt M.W. van de Garde, J.G. Coen van Hasselt. Longitudinal metabolite profiling of *Streptococcus pneumoniae*-associated community-acquired pneumonia. *Metabolomics* (2024). (* Shared first authors)

Abstract

Introduction: Longitudinal biomarkers in patients with community-acquired pneumonia (CAP) may help in monitoring of disease progression and treatment response. The metabolic host response could be a potential source of such biomarkers since it closely associates with the current health status of the patient.

Objectives: In this study we performed longitudinal metabolite profiling in patients with CAP for a comprehensive range of metabolites.

Methods: Previously collected serum samples from CAP patients with confirmed *Streptococcus pneumoniae* infection (n=25) were used. Samples were collected at multiple time points, up to 30 days after admission. A wide range of metabolites was measured, including amines, acylcarnitines, organic acids, and lipids. The associations between metabolites and C-reactive protein (CRP), procalcitonin, CURB disease severity score at admission, and total length of stay were evaluated.

Results: Distinct longitudinal profiles of metabolite profiles were identified, including cholesteryl esters, diacyl-phosphatidylethanolamine, diacylglycerols, lysophosphatidylcholines, sphingomyelin, and triglycerides. Positive correlations were found between CRP and phosphatidylcholine (34:1) (cor=0.63) and negative correlations were found for CRP and nine lysophosphocholines (cor=-0.57 to -0.74). The CURB disease severity score was negatively associated with six metabolites, including acylcarnitines (tau=-0.64 to -0.58). Negative correlations were found between the length of stay and six triglycerides (TGs), especially TGs (60:3) and (58:2) (cor=-0.63 and -0.61).

Conclusion: The identified metabolites may provide insight into biological mechanisms underlying disease severity and may be of interest for exploration as potential treatment response monitoring biomarker.

4.1 Introduction

Community-acquired pneumonia (CAP) is a lower respiratory tract infection with a high incidence and is associated with the hospitalization of approximately one million adults per year [104]. The most common cause of CAP is *Streptococcus pneumoniae* [105]. In hospitalized CAP patients, there is a need to monitor the antibiotic treatment response to optimize the treatment strategy [106]. In addition, there is a need for guidance on decisions about earlier termination of antibiotic treatment to minimize the risk of antimicrobial resistance. Monitoring of treatment response is currently achieved through observation of clinical symptoms and with inflammatory markers such as C-reactive protein (CRP) and procalcitonin (PCT) [13, 12]. In particular, PCT is relevant for informing early treatment termination decisions but lacks predictive performance for CAP prognosis [107, 108]. Therefore, there is a need for biomarkers that give early insights into the clinical course of CAP.

Biomarkers that reflect the current physiological state of the patient have the potential to accurately monitor and predict the treatment response in CAP patients. Because the metabolome closely represents this physiological state, metabolomics-techniques may enable discovery of relevant novel biomarkers. Indeed, for CAP and sepsis, the potential for metabolomics-based biomarkers measured at a static time point has been demonstrated [109]. However, the longitudinal monitoring of metabolic changes within patients may allow for an improved characterization of treatment response [14]. For example, CAP patients show a change in lysophosphatidylcholines that mirrors the transition from acute illness to recovery after starting antibiotic treatment [25]. Further systematic characterization of longitudinal metabolic changes in CAP patients may thus be of relevance for identification of metabolic biomarkers that can predict and monitor the treatment response in these patients.

To this end, in this study, we aimed to comprehensively characterize the change of longitudinal metabolite profiling in hospitalized CAP patients with a confirmed *S. pneumoniae* infection using metabolomite profiling and evaluate how metabolic changes relate to disease severity based on CURB scores, established inflammation markers, and clinical treatment response quantified using the length of stay in the hospital.

4.2 Materials and methods

4.2.1 Patient cohort

We utilized serum samples collected at multiple time points during hospitalization from 25 hospitalized CAP patients with an *S. pneumoniae* infection. These samples were previously collected as part of a larger clinical study that was performed between November 2007 and September 2010 [29]. The causative pathogen was identified

using culturing or a urinary antigen test. We selected samples from patients with a confirmed *S. pneumoniae* infection. We excluded patients with a mixed infection involving additional pathogen(s) and one patient that died during the study period. Samples were collected at five time points: at the day of admission (day 0), and at days 1, 2, 4, and 30 after admission. CRP and creatinine were measured in the hospital at the same time points as the blood samples used for metabolite profiling obtained. Samples were stored at -80 degrees Celsius, and went through a maximum of 2 freeze-thaw cycles, so stable metabolites were preserved in the samples [110, 111]. Not all time points were available for each patient, resulting in 115 samples over the 25 patients. On the day of admission, disease severity was determined using the CURB score, which is a scoring system based on confusion, blood urea > 7 mmol/l, respiratory rate (RR) ≥ 30 /min; systolic BP < 90 mmHg or diastolic BP ≤ 60 mmHg [112]. A score of two or higher is classified as severe CAP.

4.2.2 Bio-analytical procedures

Serum samples were analyzed using five targeted LCMS methods and one targeted GCMS method by the Biomedical Metabolomics Facility of Leiden University, Leiden, The Netherlands, as described previously [59]. The metabolomics profiling covered 596 metabolite targets from 25 metabolite classes, including amino acids, biogenic amines, acylcarnitines, organic acids, and multiple classes of lipids. Details of the metabolomic analysis methods used are provided in section 4.5. A total of 369 unique metabolites was measured as relative levels, of which 6 metabolites were removed due to high missingness ($\geq 20\%$), resulting in 363 metabolites being evaluated in data analysis. Biochemically-selected sums and ratios of metabolites were calculated and added to the data (Table 4.2).

PCT was measured in the same serum samples used for the metabolite profiling analysis. PCT analysis was performed using the human procalcitonin CLIA kit from Abbexa (abx190129). Samples were measured in duplicate if sample volumes were sufficient (95% of samples).

4.2.3 Data analysis

The metabolite levels were scaled through log-transformation and standardization. To explore the variability of the high-dimensional metabolite profiling dataset, principal component analysis (PCA) was used. The PCA was used on the scaled metabolite profiling data over the different time points, with the metabolites as variables and each observation being a sample from a patient for a specific time point [113]. As part of the PCA, missing values were imputed through multiple imputation using expectation maximization (EM-PCA), which iteratively calculates the principal components and imputes the missing values [114].

To evaluate how much of the variation in the metabolites could be explained by the change over time, the first two principal components were related to time using a

polynomial regression model. The importance of the metabolites to explain the variation between the patients over time was evaluated by evaluating the squared variable loadings. Specifically, the squared variable loadings within and between biochemical metabolite classes were evaluated to study similarities within classes and see which biochemical classes vary more between the patients.

To characterize the metabolic time profiles and profiles of current inflammation markers for different patients, we estimated the correlations between the scaled metabolite levels and CRP, PCT and creatinine levels over time. Next, we evaluated which metabolites could be of interest for the prediction of the clinical course, by estimating the Kendall's Tau correlation between the scaled metabolite levels and a clinical disease severity marker, the CURB score [112] at hospital admission, and estimating Pearson correlation between the scaled metabolite levels and the outcome length of stay (LOS) in the hospital. Since the CURB and LOS are static values, while the metabolites changed over time, the correlations between these outcomes and the change in metabolite levels from baseline ($m_{t=k} - m_{t=0}$) at each time point (k) were calculated. Due to the large number of correlations calculated and the small sample size, the correlations were not tested for significance, to prevent multiple testing problems. In our analysis we focused on metabolites with the largest (positive or negative) correlation, as exploratory analysis. The metabolites with the largest correlations were further evaluated in literature research to assess their biological function.

All analyses were performed in R. The scripts used for the analyses were deposited on GitHub (<http://github.com/vanhasselmlab/LongitudinalMetabolomicsCAP/tree/manuscript>).

4.3 Results

4.3.1 Metabolite time profiles

Metabolite profiling was performed for 25 patients and resulted in 363 metabolite levels on five time points (Table 4.3). The patient characteristics are displayed in Table 4.1. Comorbidities present in patients included kidney disease ($n = 1$), cardiovascular disease ($n = 4$), malignancy ($n = 2$), COPD ($n=1$, $n_{\text{missing}} = 15$), diabetes ($n=3$, $n_{\text{missing}} = 15$). No patients were using corticosteroids before admission ($n_{\text{missing}} = 15$).

Metabolite profiles within all CAP patients shifted over time, as shown in the PCA over all time points (Figure 4.1). The close relationship between metabolite levels and time is reflected in the results from the polynomial regression model which indicated that 45% of the metabolite variation captured in these first two principal components could be explained by time. Due to the large age range, we tested whether age was a large explanatory factor for the metabolite differences between individuals, but did not find a significant contribution of age (section 4.5).

Table 4.1 Patient characteristics

Patient characteristics	CAP patients (n=25)
Age (years)	
Median [Min, Max]	67.0 [18.0, 98.0]
Sex	
Male	12 (48.0%)
Female	13 (52.0%)
CURB score	
Median [Min, Max]	1.00 [0, 3.00]
Duration of symptoms before admission (days)	
Median [Min, Max]	3.00 [1.00, 14.0]
Missing	15 (60.0%)
Antibiotic treatment before admission	
No	8 (32.0%)
Yes	2 (8.0%)
Missing	15 (60.0%)
Length of stay (days)	
Median [Min, Max]	7.50 [2.50, 24.5]

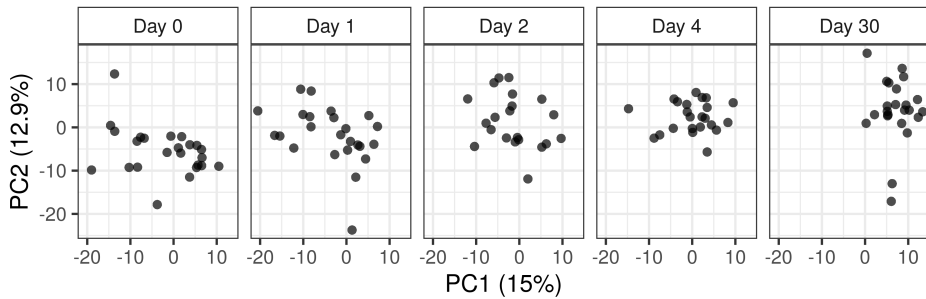


Figure 4.1 PCA scores for patient metabolite profiles over time. Every point represents the scores of an individual patient at a certain time point, in two dimensions based on the metabolite values. The panels show a trend over time of the metabolite profiles.

The metabolites that were targeted in this study were categorized into different biochemical classes. Metabolites from different biochemical classes showed distinct contributions to the total variation between the patients over time as was expressed in the variable loadings and directionality of the principal components (Figure 4.2). The squared PCA loadings represent the weight that the different metabolites in the biochemical class have in explaining the variation between patients over time. Of the variation in principal component one and two, 48% was explained by metabolites of the classes of cholesteryl esters, LPC's, sphingomyelins, diacylglycerols, and triglycerides (Figure 4.2A). The metabolites were categorized in classes based on their biochemistry and not based on their biological functions. The PCA indicate that metabolites that are categorized in the same class do not necessarily behave similarly (Figure 4.2B). For example, amino acids behave very differently from each other. Metabolites that do behave similarly in their biochemical class are for example triglycerides and sphingomyelins.

For each patient, the metabolic time profiles were shown as the two first components from the PCA (Figure 4.3, Figure 4.7). Generally, a shift from low to high principal component values was seen over time, corresponding to the shift in metabolite levels for the different metabolites (Figure 4.2B). The large variability in the time profiles indicates a large interpatient variability in metabolic levels and changes over time.

4.3.2 Inflammation marker associations

To explore associations between metabolite profiles and inflammation, the metabolite values were compared to currently used inflammation biomarkers. Correlations were found between CRP and PCT and several metabolites. For example, phosphocholine (PC) (34:1) showed a positive correlation with CRP ($cor = 0.63$). Several individual lysophosphocholines (LPCs) and the sum of all LPCs showed a negative correlation with CRP ($cor = -0.57$ to -0.74 , Figure 4.4A). PC (34:1) was found to decrease over time and several LPCs showed an increase over time, thereby mirroring the clinical disease progression (Figure 4.4B). Positive correlations with CRP and PCT were reported for the short-chain acylcarnitines (SCACs) tiglylcarnitine, 2-methylbutyrylcarnitine, and isovalerylcarnitine (cor with PCT = 0.61, 0.58, and 0.57; cor with CRP = 0.54, 0.64, and 0.51, respectively). Negative correlations were seen between the long-chain acylcarnitine (LCAC) stearoylcarnitine and CRP ($cor = 0.62$). This trend for decreasing SCACs over time is also represented by the positive correlation of CRP and PCT with the sum of all SCACs ($cor = 0.55$ and 0.53 , respectively).

Correlations between metabolite levels and creatinine, a marker of renal failure, were also identified. The same trends were seen for creatinine as for CRP and PCT (Figure 4.8). Strong positive correlations were observed between creatine and 1-Methylhistidine, SDMA, inositol, homoserine, methionine sulfone, and octanoylcarnitine ($cor > 0.7$)

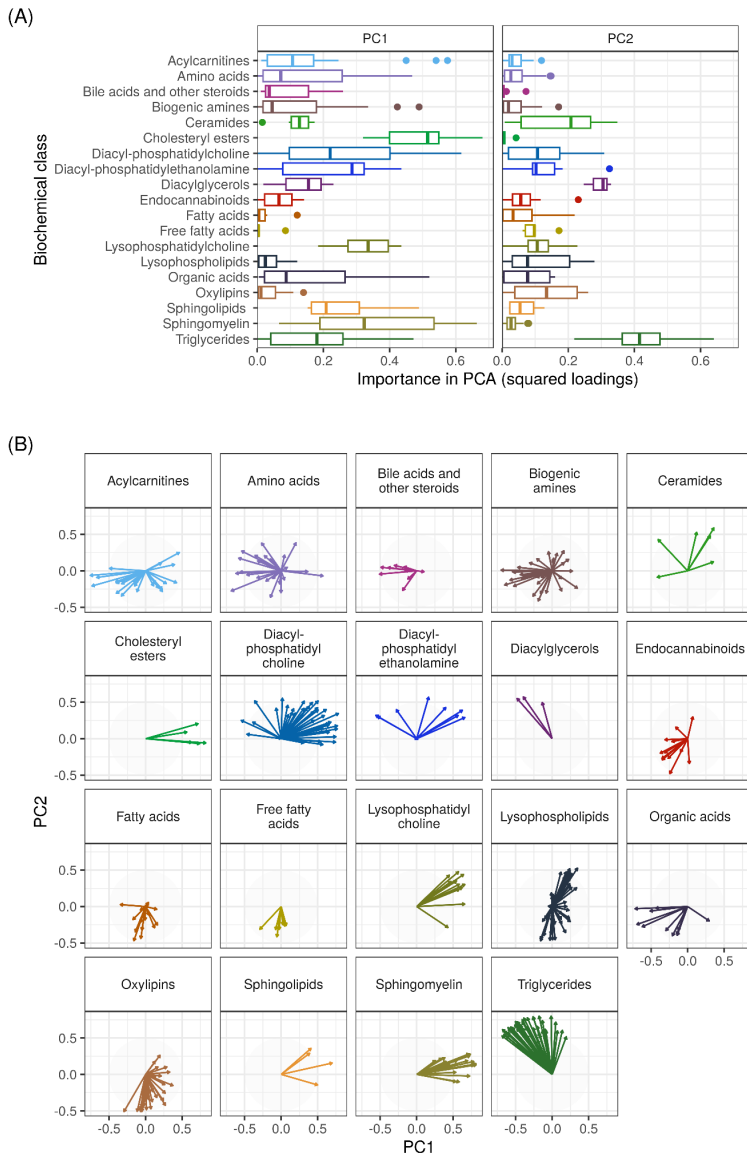


Figure 4.2 Metabolite contributions to the two dimensions of the PCA as variable loadings. a) The importance of each biochemical class for the different principal components (PCs), expressed by their squared metabolite loadings. Each box represents the squared loadings of the metabolites within a metabolic class. High squared loadings indicate a larger contribution to explaining the variation between patients. b) The loading plots for each biochemical metabolite class. The arrows indicate the importance (length) and direction of the metabolites in the principal component space. For example, high PC1 values correspond to high metabolite levels for metabolites with right pointing arrows, and low metabolite levels for metabolites with left pointing arrows. Arrows with a similar direction have similar metabolite patterns. Abbreviations: PC: principal component.

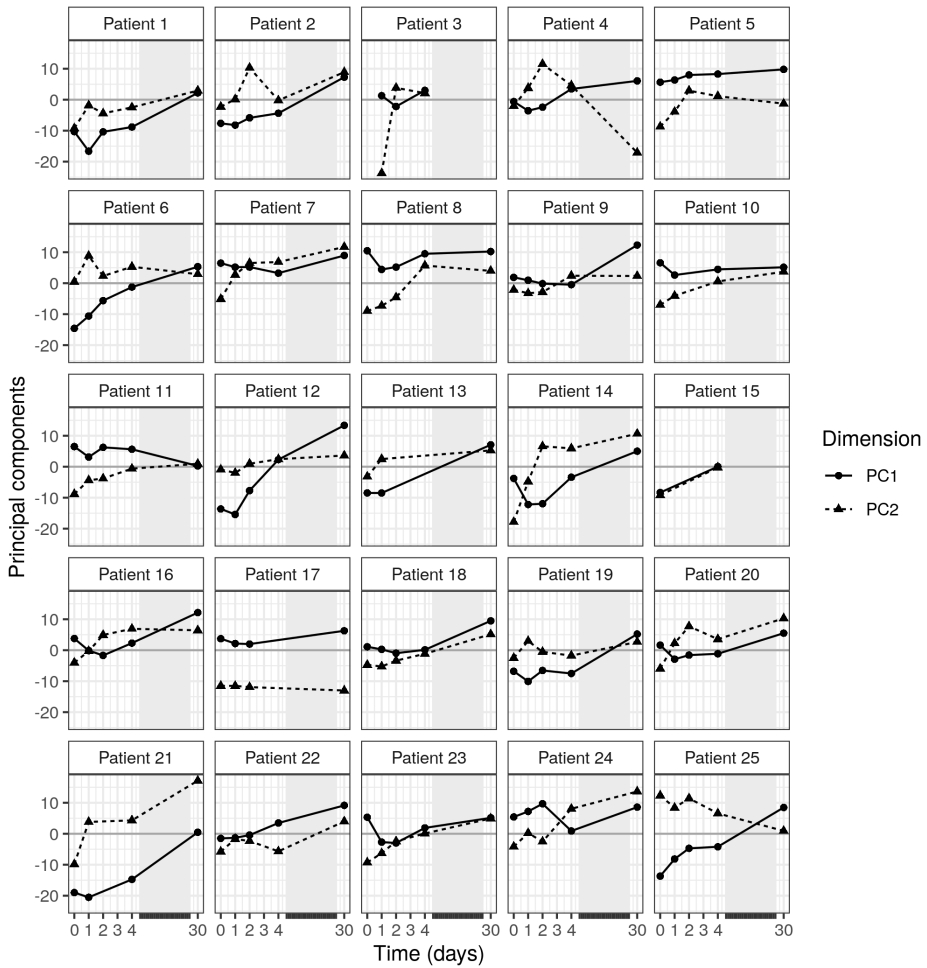


Figure 4.3 Individual time profiles over PC1 and PC2. The lines PC1 (solid) and PC2 (dashed), indicate the change in the corresponding principal component over time. Changes in PC values correspond to changes in metabolite levels according to their respective loadings. Abbreviations: PC: principal component. Abbreviations: PC: principal component.

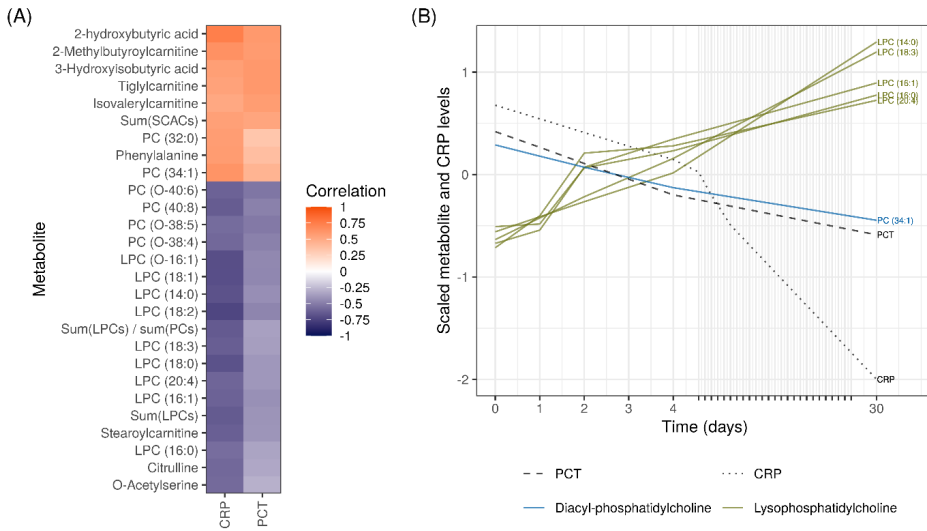


Figure 4.4 Correlations between inflammation markers CRP and PCT, and metabolites. A) The correlations between metabolites and CRP or PCT. Metabolites with a correlation >0.55 or <-0.55 for at least one marker are shown. A positive correlation (orange) indicates that a higher CRP or PCT level corresponds to an increase of that metabolite over time, while a negative correlation (blue) indicates a decrease over time for patients with a higher CRP or PCT level. B) Average CRP, PCT, PC (34:1), and LPC levels over time over all patients. Metabolite and CRP data were scaled. Abbreviations: see the abbreviation list.

4.3.3 Disease severity score associations

To identify possible metabolic biomarkers for indication of disease severity, associations between the CURB disease severity score at admission and the change in metabolite levels on from day 0 to days 1, 2, 4, and 30 were evaluated (Figure 4.8). Negative associations were found between the CURB score and the change of metabolite levels (m) between day 0 and day 30 ($m_{t=30} - m_{t=0}$) of tiglylcarnitine, isovalerylcarnitine, 3-hydroxyisovaeric acid, carnitine, N6,N6,N6-trimethyl-lysine, and isobutyryl carnitine ($\tau = -0.64$ to -0.58 , Figure 4.5). Patients with higher CURB scores showed decreasing levels of these metabolites.

4.3.4 Hospital length of stay associations

We evaluated the association between metabolites and clinical outcomes using the length of stay (LOS) as a potential surrogate endpoint. The strongest negative correlations to LOS were reported for the metabolite change over the first two days of admission ($m_{t=2} - m_{t=0}$, Figure 4.6), especially for the triglycerides (TGs) (60:3) and (58:2) ($\text{cor} = -0.63$ and -0.61 respectively). The correlations of these metabolites to LOS were much stronger than to CRP and PCT ($\text{cor} = -0.08$ and -0.25 respectively). Positive

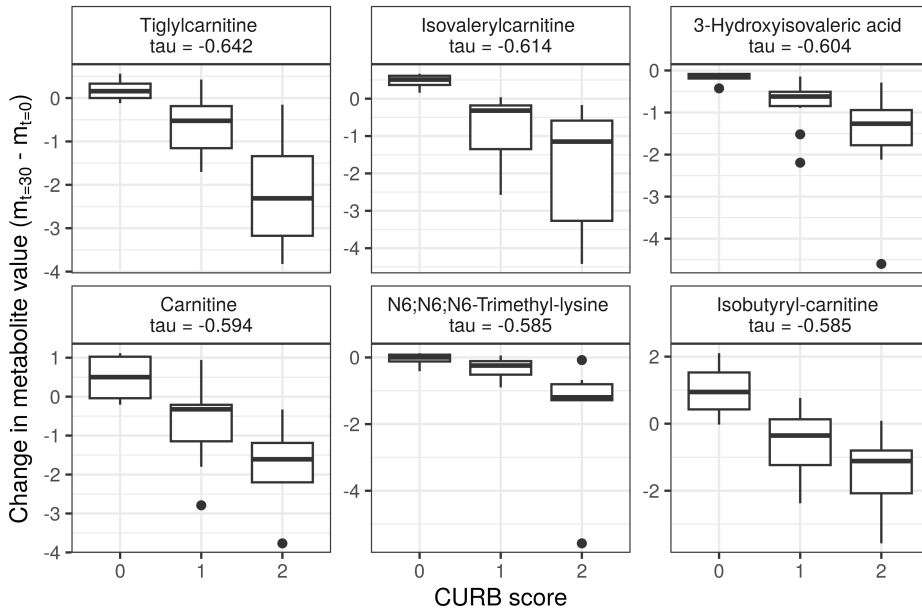


Figure 4.5 The correlation between the CURB score and six metabolites with highest associations. The change in metabolite level is the difference between the scaled metabolite level at day 30 and scaled metabolite level at admission (y-axis). These six metabolites all show a negative correlation with the CURB score (τ). This means, for patients with a CURB score of 0 the metabolite change between day 30 and day 0 is positive, so their metabolite levels were increasing over time. For patients with a CURB score of 2, the metabolite levels decreased over time.

correlations were most pronounced when analyzing the metabolite change from the day of admission to day 30 ($m_{t=30} - m_{t=0}$). In the case of fatty acid (FA) (22:1) the day after admission ($m_{t=1} - m_{t=0}$) was the most strongly positively correlated to the LOS ($\text{cor} = 0.58$).

4.4 Discussion

In this study, we characterized the dynamics of the serum metabolites in pneumococcal CAP patients. We found that a large part of the variation in the metabolite values was associated with time-varying changes in metabolites within the patients. We furthermore found that several groups of metabolites were found to correlate with inflammation markers, CURB score, and length of hospital stay. These findings both support the potential relevance of metabolite-based biomarkers to monitor the treatment response or disease progression in CAP.

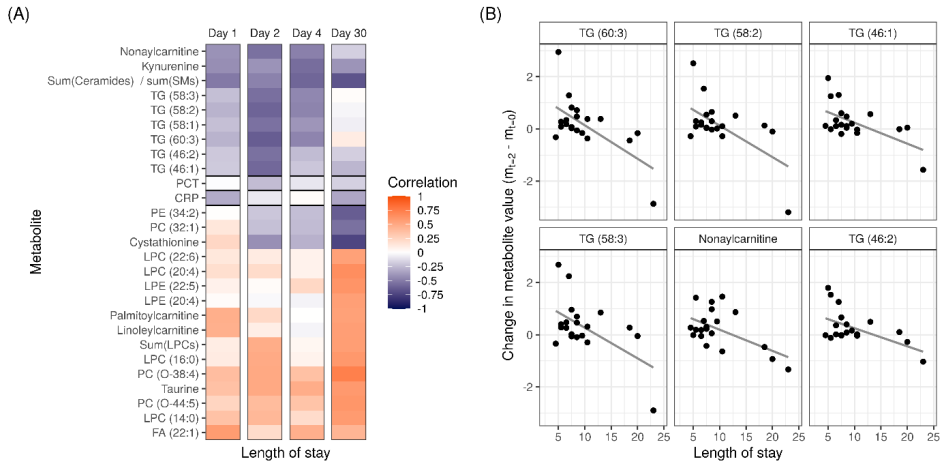


Figure 4.6 a) The correlations between the length of stay and metabolite change from baseline at days 1, 2, 4, and 30 after admission ($m_{t=k} - m_{t=0}$). CRP and PCT are added as a reference. A positive correlation (orange) indicates that a longer stay in the hospital corresponds to an increase of that metabolite over time, while a negative correlation (blue) indicates a decrease over time for patients with longer stay. b) Metabolite levels over time for individual patients for metabolites with large negative correlations ($\text{cor} < -0.55$) over the first two days after admission. Abbreviations: see the abbreviation list.

We found that length of stay in the hospital was negatively correlated with the triglycerides TG (60:3) and TG (58:2). Interestingly, these TGs are not highly correlated to CRP, PCT, or the CURB score, which suggests that they can explain a part of the variability of disease progression in patients not explained by established biomarkers for inflammation. We previously found that TGs do not contribute to the etiological prediction of pathogenic in CAP [59]; as such TGs may be of interest as potential biomarker beyond pneumococcal CAP studied in this analysis. Further studies should however consider the potential impact of diet on TGs, as a potential confounding factor [115].

Phosphatidylcholine (PC) (34:1) and lysophosphatidylcholines (LPCs) (14:0), (16:0), (16:1), (18:0), (18:1), (18:2), (18:3) and (20:4) correlated to inflammatory markers, which also corresponds to previous findings [25, 116]. PC (34:1), a ligand of nuclear receptor PPAR α 30, showed a positive correlation with CRP, which was previously associated with an anti-inflammatory response [117]. LPC 14:0 has been recently identified as a biomarker for disease severity in CAP patients [118]. Due the correlation with CRP, these metabolites could be of interest as treatment response biomarkers, also beyond pneumococcal CAP patients [34].

The CURB score was negatively associated with six metabolites, including several acylcarnitines. One of these acylcarnitines, tiglylcarnitine, has previously been found

to be increased in non-survivors of CAP and could be considered a marker for disease severity [116]. Isovalerylcarnitine and isobutyrylcarnitine have, to our knowledge, not been studied as disease severity marker before, but may show a comparable performance to tigylcarnitine as their direction on the first principal component is similar.

In this analysis we demonstrated which biochemical metabolite classes explain most of the variation in metabolite patterns between individuals and over time. Triglycerides and LPCs were important for explaining the variation over time in the principal component analysis (PCA) and correlated with LOS and inflammatory markers. Within the biochemical classes, not all metabolites showed similar patterns, indicating that metabolites in some biochemical classes behave similarly during the infection, while metabolites in other classes behave differently (Figure 4.2B). The amino acids behave very differently, which could be expected since they are involved in a wide variety of biological functions [119].

The longitudinal analysis of the metabolite profiling data enabled us to gain insight into acute and longer-term changes in the metabolome during the clinical course of CAP. Since patients are admitted to the hospital in different stages of the disease, interpretation of the metabolite profile at one time point can be challenging. The longitudinal metabolite profiles that were measured in this study give more information about the state of the patient and elucidate the effect of comorbidities and co-medications. The principle component analysis (Figure 4.3) showed large variability between different patients, indicating the importance of considering changes within patients, instead of evaluating the metabolite profile at one timepoint. We found that the differences in metabolite levels were largely explained by changes over time and were, therefore, related to the treatment response.

This study was conducted in a well-characterized set of CAP patients with *S. pneumoniae* infections. *S. pneumoniae* is a common cause of CAP, but other bacterial or viral pathogens can also be the cause of CAP. A previous study did not show significant differences in metabolic profiles between common causes of CAP [59]. The results of the current study may apply to CAP patients with these other causative pathogens, but this is still unsure because the previous study does not cover changes over time. Especially metabolites associated with length of stay should be validated in CAP cohorts with various causative pathogens, since they are not related to the general inflammatory response.

In further research, the addition of patients with other causes of CAP is of interest to compare metabolic time profiles for different treatment strategies based on the causative pathogen. Early recognition of a pathogen-drug mismatch using metabolite profiling could make antibiotic therapies more targeted and shorter. This study shows that mainly TGs, LPCs, PCs, and acylcarnitines are of interest for the disease severity and the length of stay for patients with CAP. By focusing on these metabolite classes, the number of metabolites that has to be measured for every patient can be reduced.

In conclusion, we find that that metabolomics-based biomarkers have potential for treatment response monitoring in CAP patients. The triglycerides found in this study could potentially complement the currently available biomarkers such as CRP and PCT as they yield additional information about the clinical course in these patients. This study furthermore supports the relevance for collecting longitudinal data to follow the highly dynamic metabolite profiles in patients, which can further enable the development of personalized treatment strategies.

4.5 Supplementary Information

Details on metabolite profiling methods

Batch design: Aliquoted samples were run in a randomized fashion in several batches together with quality control (QC) samples (every 10 samples), sample replicates (every 7 samples), internal standards (ISTDs), blanks, and calibration lines.

Quality control: Blank samples were used to determine the blank effect. Replicate samples were used to check the instrument for repeatability. In-house developed algorithms were applied using the pooled QC samples to compensate for shifts in the sensitivity of the mass spectrometer over the batches.

Reported results: After quality control correction the metabolites that complied with the acceptance criteria of a relative standard deviation of the quality control samples (RSD_{qc}) <15% were reported. The data was reported as relative response ratio (analyte signal area / ISTD area; unit free) of the metabolites after QC correction. Metabolites that did not comply with the acceptance criteria of the quality control, but have been included in the results present RSDs up to 30% and should be handled with caution.

Amine profiling : Amine profiling was performed according to the validated amine profiling analytical platform with minor optimization [50] . The amine platform covers amino acids and biogenic amines employing an Accq-Tag derivatization strategy adapted from the protocol supplied by Waters. 5,0 μ L sample was spiked with an internal standard solution. Protein precipitation was performed by addition of MeOH and the sample was dried in a speedvac. The residue was reconstituted in borate buffer (pH 8.5) with AQC reagent. The prepared samples were transferred to autosampler vials and placed in an autosampler tray. The vials were cooled at 4^o C upon injection. 1,0 μ L prepared sample was injected in a UPLC-MS/MS system. Chromatographic separation was achieved by an Agilent 1290 Infinity II LC System on an Accq-Tag Ultra column (Waters) with a flow of 0.7 mL/min over an 11 min gradient. The UPLC was coupled to electrospray ionization on a triple quadrupole mass spectrometer (AB SCIEX Qtrap 6500). Analytes were detected in the positive ion mode and monitored in Multiple Reaction Monitoring (MRM) using nominal mass resolution. Acquired data was evaluated using MultiQuant Software for Quantitative Analysis (AB SCIEX, Version 3.0.2), by the integration of assigned MRM peaks and normalization using proper

internal standards. For analysis of amino acids, their $^{13}\text{C}^{15}\text{N}$ -labeled analogs were used. For other amines, the closest-eluting internal standard was employed. After quality control correction the amines that complied with the acceptance criteria of $\text{RSD}_{\text{qc}} < 15\%$ were included in the results. Additionally, the amines that presented an RSD_{qc} between 15 and 30% were included in the results but these compounds should be considered with caution.

Acylcarnitine profiling: The acylcarnitine platform covers acylcarnitines as well as trimethylamine-N-oxide, choline, betaine, deoxycarnitine, and carnitine. 10 μL sample was spiked with an internal standard solution. Protein precipitation was performed by addition of MeOH. The supernatant was transferred to an autosampler vial and placed into an autosampler. The vials were cooled at 10°C upon injection. 1.0 μL of the prepared sample was injected into a triple quadrupole mass spectrometer. Chromatographic separation was achieved by UPLC (Agilent 1290, San Jose, CA, USA) on an Accq-Tag Ultra column (Waters) with a flow of 0.7 mL/min over an 11 min gradient. The UPLC was coupled to electrospray ionization on a triple quadrupole mass spectrometer (Agilent 6460, San Jose, CA, USA). Analytes were detected in the positive ion mode and monitored in Multiple Reaction Monitoring (MRM) using nominal mass resolution. Acquired data was evaluated using Agilent MassHunter Quantitative Analysis software (Agilent, Version B.05.01), by integration of assigned MRM peaks and normalization using proper internal standards. The closest-eluting internal standard was employed. After quality control correction the compounds that complied with the acceptance criteria of $\text{RSD}_{\text{qc}} < 15\%$ were included in the results. Additionally, the compounds that presented an RSD_{qc} between 15 and 30% were included in the results but these compounds should be considered with caution.

Organic acid profiling: The organic acid platform covers 28 organic acids. 50 μL sample was spiked with an internal standard solution. Protein precipitation was performed by addition of MeOH. After centrifugation, the supernatant was transferred and the sample was dried using a speedvac. Then, two-step derivatization procedures were performed on-line: oximation using methoxyamine hydrochloride (MeOX, 15 mg/mL in pyridine) as the first reaction and silylation using N-Methyl-N-(trimethylsilyl)-trifluoroacetamide (MSTFA) as the second reaction. 1 μL of each sample was directly after its derivatization injected on GC-MS. Gas chromatography was performed on an Agilent Technologies 7890A equipped with an Agilent Technologies mass selective detector (MSD 5975C) and MultiPurpose Sampler (MPS, MXY016-02A, GERSTEL). Chromatographic separations were performed on an HP-5MS UI (5% Phenyl Methyl Silox), 30 m \times 0.25 m ID column with a film thickness of 25 μm , using helium as the carrier gas at a flow rate of 1.7 mL/min. A single-quadrupole mass spectrometer with electron impact ionization (EI, 70 eV) was used. The mass spectrometer was operated in SCAN mode mass range 50-500. Acquired data was evaluated using Agilent MassHunter Quantitative Analysis software (Agilent, Version B.05.01). After quality control correction and considering blank effects, the organic acid compounds that complied with the acceptance criteria $\text{RSD}_{\text{qc}} < 15\%$ and blank effect

<20% were included in the results. Also, the organic acids that reported an RSD_{qc} between 15 and 30% were included and should be considered with caution.

Negative lipid profiling: The negative lipid platform is a semi-target methodology for the identification of 30 fatty acids. 50 μL sample was spiked with 50 μL of an internal standard solution. Protein precipitation was performed by addition of 550 μL MeOH. After centrifugation, 600 μL supernatant was transferred and the sample was dried using a speedvac. The residue was reconstituted in 300 μL of isopropanol with 0,1% formic acid. The prepared samples were transferred to autosampler vials and placed in an autosampler tray. 8,0 μL of the prepared sample was injected into an LC-MS. The analysis was performed on an ACQUITY UPLC™ (Waters, the Netherlands) coupled to a high-resolution mass spectrometer with a Synapt G2 Q-TOF system (Waters, the Netherlands) using reference lock mass correction. Lipids were detected in full scan in the negative ion mode. Chromatographic separation was achieved using an HSS T3 column (1.8 μm , 2.1 * 100 mm) with a flow of 0.4 mL/min over a 16-minute gradient. Acquired data was preprocessed using Targetlynx software (Masslynx, V4.1, SCN916). After quality control correction, the compounds that complied with the acceptance criteria RSD_{qc} <15% were included in the results. Additionally, the compounds that reported an RSD_{qc} between 15 and 30% were included in the results and should be considered with caution.

Positive lipid profiling: The positive lipid platform covers 185 compounds including triglycerides (TGs, n=85) and non-triglycerides (non-TGs, n=100). 10 μL preprocessed sample was spiked with 1000 μL IPA containing internal standards and vortexed for 30 sec. Prepared samples were transferred to autosampler vials for LC-MS analysis. In total 2.5 μL prepared sample was injected for analysis. Chromatographic separation was achieved on an ACQUITY UPLC™ (Waters, Ettenleur, the Netherlands) with an HSS T3 column (1.8 μm , 2.1 * 100 mm) with a flow of 0.4 mL/min over a 16 min gradient. The lipid analysis is performed on a UPLC-ESI-Q-TOF (Agilent 6530, Jose, CA, USA) high-resolution mass spectrometer using reference mass correction. Lipids were detected in full scan in the positive ion mode. The raw data were preprocessed using Agilent MassHunter Quantitative Analysis software (Agilent, Version B.04.00). After quality control correction, the TGs and non-TGs compounds that complied with the acceptance criteria RSD_{qc} <15% and blank effect <40 % were included in the results. The TG and non-TGs that reported an RSD_{qc} between 15 and 30% were also included and should be considered with caution.

Signaling lipid profiling: The signaling lipids platform covers various isoprostane classes together with their respective prostaglandin isomers from different poly unsaturated fatty acids (PUFA), including n-6 and n-3 PUFAs such as dihomo- γ -linoleic acid (DGLA) and arachidonic acid (both n-6) and eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) (both n-3). Also included in this platform are endocannabinoids, bile acids, and signaling lipids from the sphingosine and sphinganine classes and their phosphorylated forms, as well as three classes of lysophosphatidic acids. The three lysophosphatidic acid classes include lysophosphatidic acids (LPAs),

lysophosphatidylglycerol (LPG), lysophosphatidylinositol (LPI), lysophosphatidylserine (LPS), lysophosphatidylethanolamines (LPE), cyclic-phosphatidic acids(cLPA), and fatty acid all ranging from C14 to C22 chain length species. The signaling and peroxidized lipids platform is divided into two chromatographic methods: low and high pH. In the low pH method, isoprostanes, prostaglandins, nitro-fatty acids, lyso-sphingolipids, endocannabinoids, and bile acids are analyzed. The high pH method covers lyso-sphingolipids, lysophosphatidic acids, lysophosphatidylglycerol, lysophosphatidylinositol, lysophosphatidylserine, lysophosphatidylethanolamines, cyclic-phosphatidic acids, and fatty acid. Each sample was spiked with antioxidant and internal standard solution. The extraction of the compounds is performed via liquid-liquid extraction (LLE) with butanol and methyl tert-butyl ether (MTBE). After collection, the organic phase is concentrated by first drying followed by reconstituted in a smaller volume. After reconstitution, the extract is transferred into amber autosampler vials and used for high and low pH injection. A Shimadzu system, formed by three high-pressure pumps (LC-30AD), a controller (CBM-20Alite), an autosampler (SIL-30AC), and an oven (CTO-30A) from Shimadzu Benelux, was coupled online with an LCMS-8050 triple quadrupole mass spectrometer (Shimadzu) for high pH measurements. An LCMS-8060 triple quadrupole mass spectrometer (Shimadzu) was coupled to the Shimadzu system for low pH measurements. Both systems were operated using LabSolutions data acquisition software (Version 5.89, Shimadzu). The samples were analyzed by UPLC-MS/MS. An Acquity UPLC BEH C18 column (Waters) was used to measure the samples in the low pH method. For the high pH method, a Kinetex EVO column by Phenomenex was used. The triple quadrupole mass spectrometer was used in polarity switching mode and all analytes were monitored in dynamic Multiple Reaction Monitoring (dMRM). The acquired data was evaluated using LabSolutions Insight software (Version 3.1 SP1, Shimadzu), by integration of assigned MRM peaks and normalization using accordingly selected internal standards. When available, a deuterated version of the target compound was used as an internal standard. For the other compounds, the closest-eluting internal standard was employed. For low pH mode, after quality control correction, the metabolites that complied with the acceptance criteria of RSD_{qc} <15% and blank effect <40% were included in the results. Additionally, the compounds that reported an RSD_{qc} between 15 and 30% were included in the results and should be considered with caution. For high pH mode, after quality control correction, the metabolites that complied with the acceptance criteria of RSD_{qc} <15% and blank effect <40% were included in the results. Additionally, the compounds that reported an RSD_{qc} between 15 and 30% were included in the results and should be considered with caution.

Testing the influence of age on metabolite profiles

To test whether the age is a factor to take into account in the correlation analysis between the change in metabolite values and the CURB score and hospitalization time, we tested the whether the interindividual variance of the metabolite profiling was explained by age, to decide whether age should be a confounder in the analysis. To test this, we did an anova test to compare a mixed effect model on the principal components, which

represent the metabolite profiles in a lower dimension, in two models: one with only a patient specific random effect and one model that included both a patient specific random effect and a parameter for age. The anova was done two times, with the first and second principle component scores as outcomes respectively. The code and outcomes of the anova are shown below, where `subject.id` denotes the patient and `age` is the age variable. The p-values for principle component 1 and 2 were 0.18 and 0.09 respectively and did not indicate a significant improvement of the model including age over the model not including age, which motivated the correlation analysis without adding age as confounder or using it to stratify the analysis. The small sample size could be a reason for not finding significance, but this is also a reason for not stratifying the analysis.

```
lmer_age_pc1 <- lmer(PC1 ~ age + (1|subject.id),
                    data = pca_data, REML = F)
lmer_pc1 <- lmer(PC1 ~ (1|subject.id),
                data = pca_data, REML = F)
anova(lmer_pc1, lmer_age_pc1)

              npar    AIC    BIC logLik deviance Chisq Df Pr(>Chisq)
lmer_pc1          3 768.60 776.83 -381.30   762.60
lmer_age_pc1      4 768.82 779.80 -380.41   760.82 1.7827  1    0.1818

lmer_age_pc2 <- lmer(PC2 ~ age + (1|subject.id),
                    data = pca_data, REML = F)
lmer_pc2 <- lmer(PC2 ~ (1|subject.id),
                data = pca_data, REML = F)
anova(lmer_pc2, lmer_age_pc2)

              npar    AIC    BIC logLik deviance Chisq Df Pr(>Chisq)
lmer_pc2          3 770.96 779.19 -382.48   764.96
lmer_age_pc2      4 770.10 781.08 -381.05   762.10 2.8589  1    0.09087
```

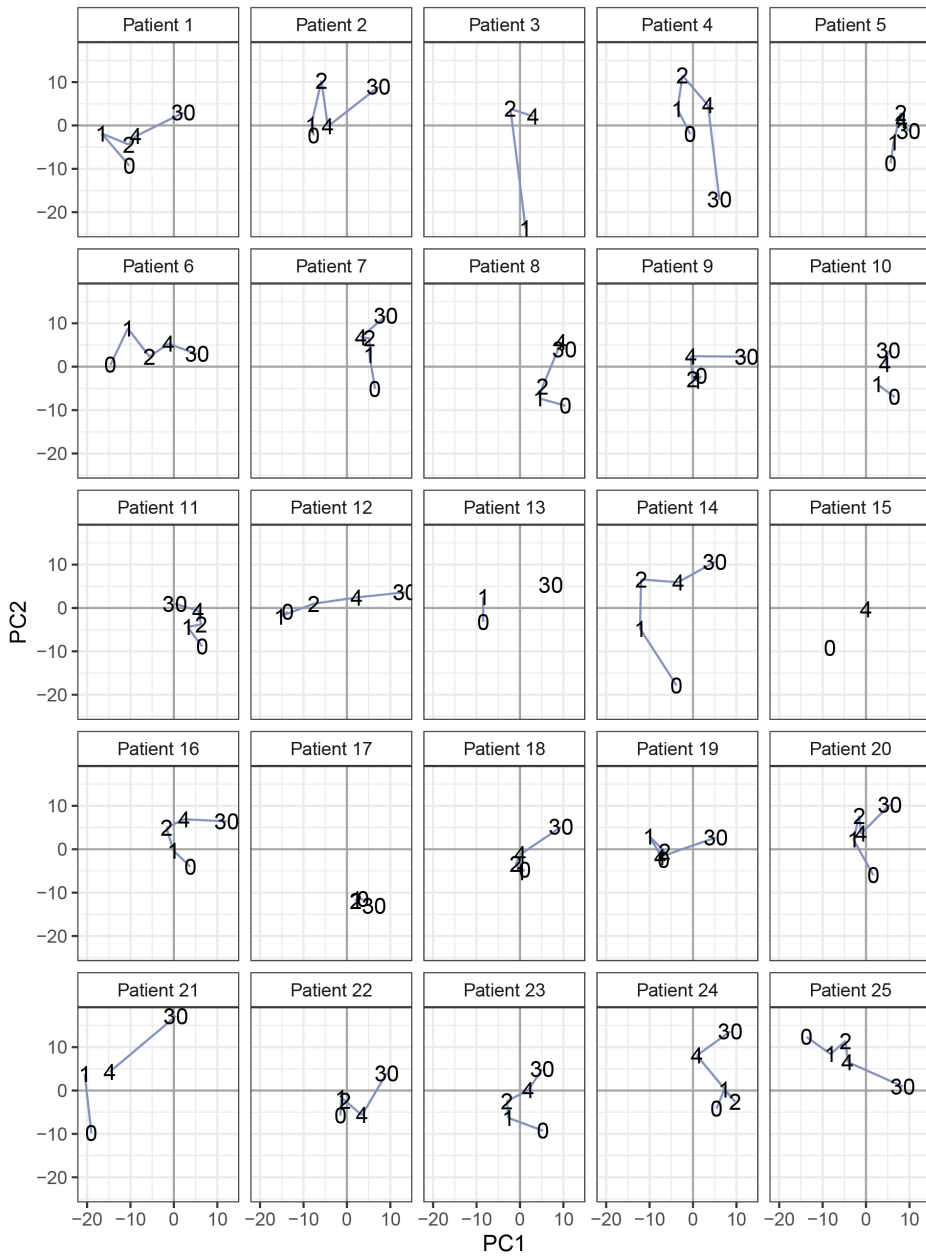


Figure 4.7 PCA score plots for each patient. For each patient, the time points are labelled and connected with lines. Abbreviations: PC: principal component

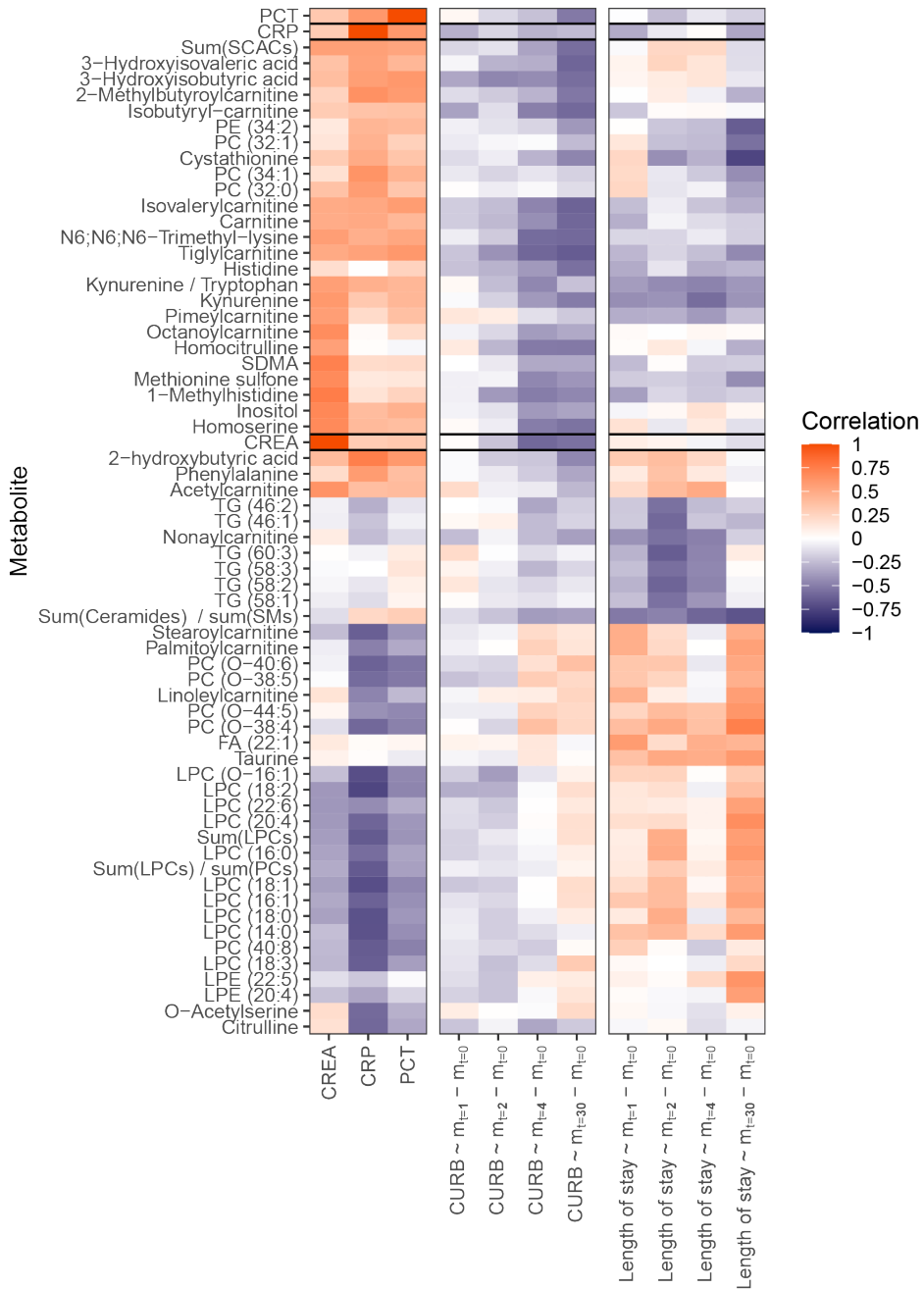


Figure 4.8 The correlations between metabolites and creatinine, CRP, and PCT over time; and the correlations of the CURB score and length of stay with a change of the metabolites between day k and day 0, where the change in metabolite levels is denoted by $mt=k$ $mt=0$.

Table 4.2 Metabolite ratios and sums

Metabolite sum or ratio name in R	Metabolite sum or ratio formula
BCAA_sum	isoleucine + leucine + valine
TCA_cycle_sum	Citric acid + lactic acid + malic acid + fumaric acid
urea_cycle_sum	Citrulline + arginine + ornithine + fumaric acid
lc_Carnitines_sum	Myristoylcarnitine + Hexadecenoylcarnitine + Palmitoylcarnitine + Stearoylcarnitine + Dodecenoylcarnitine + Tetradecenoylcarnitine + Linoleylcarnitine + Oleoylcarnitine + Tetradecadienylcarnitine
mc_Carnitines_sum	Hexanoylcarnitine + Octanoylcarnitine + Octenoylcarnitine + Decanoylcarnitine + Lauroylcarnitine + Nonanoylcarnitine + Pimeylcarnitine + Decenoylcarnitine
sc_Carnitines_sum	Acetylcarnitine + Propionylcarnitine + Isobutyrylcarnitine + Butyrylcarnitine + Tiglylcarnitine + Methylbutyrylcarnitine + Isovalerylcarnitine
Cer_sum	Cer(d18:1/22:1) + Cer. (d18:1/24:1. + Cer(d18:1/24:0) + Cer(d18:1/16:0) + Cer(d18:1/23:0) + Cer(d18:1/24:0)
SM_sum	Sphingomyelin (d18:1/14:0) + (d18:1/15:0) + (d18:1/16:0) + (d18:1/16:1) + (d18:1/17:0) + (d18:1/18:0) + (d18:1/18:1) + (d18:1/18:2) + (d18:1/20:0) + (d18:1/20:1) + (d18:1/21:0) + (d18:1/22:0) + (d18:1/22:1) + (d18:1/23:0) + (d18:1/ 23:1) + (d18:0/24:0) + (d18:0/24:1) + (d18:0/24:2) + (d18:0/25:0) + (d18:0/25:1)
LPC_sum	Lysophosphatidylcholine (14:0) + (16:0) + (16:1) + (18:0) + (18:1) + (18:2) + (18:3) + (20:4) + (20:5) + (22:6) + (O-16:1) + (O-18:1)
PC_sum	Diacyl-phosphatidylcholine (32:0) + (32:1) + (32:2) + (34:1) + (34:2) + (34:3) + (34:4) + (36:1) + (36:2) + (36:3) + (36:4) + (36:5) + (36:6) + (38:2) + (38:3) + (38:4) + (38:5) + (38:6) + (38:7) + (40:4) + (40:5) + (40:6) + (40:7) + (40:8) + (O-34:1) + (O-34:2) + (O-34:3) + (O-36:2) + (O-36:3) + (O-36:4) + (O-36:5) + (O-36:6) + (O-38:4) + (O-38:5) + (O-38:6) + (O-38:7) + (O-40:6) + (O-42:6) + (O-44:5)
HT5_Trp_ratio	Serotonine / Tryptophan
ADMA_Arg_ratio	ADMA / Arginine
SDMA_Arg_ratio	SDMA / Arginine
Carnitine_sum_lc_Carnitines_ratio	Carnitine / LCAC sum
Carnitine_sum_mc_Carnitines_ratio	Carnitine / MCAC sum
Carnitine_sum_sc_Carnitines_ratio	Carnitine / SCAC sum
DCA_CA_ratio	DCA / CA
FA_14.1_14.0	FA (14:1) / FA (14:0)
FA_16.1_16.0	FA (16:1) / FA(16:0)
Gln_Glu	Glutamine / Glutamic acid
Kyn_Trp	Kynurenine / Tryptophan
sum_BCAA_sum_Phe_Tyr_ratio	BCAA sum / (Phenylalanine + Tyrosine)
sum_CER_sum_SM_ratio	Cer sum / SM sum
sum_LPC_sum_PC_ratio	LPC sum / PC sum

Table 4.3 Metabolite ratios and sums

Biochemical class	Metabolite	Biochemical class	Metabolite
Acylcarnitines	Acetylcarnitine	Betaines	Betaine
Acylcarnitines	Butyrylcarnitine	Bile acids and other steroids	Cholic acid
Acylcarnitines	Decanoylcarnitine	Bile acids and other steroids	Cortisol
Acylcarnitines	Decenoylcarnitine	Bile acids and other steroids	Deoxycholic acid
Acylcarnitines	Dodecanoylcarnitine	Bile acids and other steroids	GCA
Acylcarnitines	9-Hexadecenoylcarnitine	Bile acids and other steroids	GCDCA
Acylcarnitines	Hexanoylcarnitine	Bile acids and other steroids	GDCA
Acylcarnitines	Isobutyryl-carnitine	Bile acids and other steroids	GLCA
Acylcarnitines	Isovalerylcarnitine	Bile acids and other steroids	GUOCA
Acylcarnitines	Lauroylcarnitine	Bile acids and other steroids	TCA
Acylcarnitines	Linoleylcarnitine	Biogenic amines	ADMA
Acylcarnitines	Myristoylcarnitine	Biogenic amines	Anserine
Acylcarnitines	Nonacylcarnitine	Biogenic amines	Beta-Alanine
Acylcarnitines	2-Octenoylcarnitine	Biogenic amines	Cystathionine
Acylcarnitines	Octanoylcarnitine	Biogenic amines	3-Aminoisobutyric acid
Acylcarnitines	Oleylcarnitine	Biogenic amines	Ethanolamine
Acylcarnitines	Palmitoylcarnitine	Biogenic amines	N2-gamma-Glutamylglutamine
Acylcarnitines	Pimeylcarnitine	Biogenic amines	gamma-Glutamylalanine
Acylcarnitines	Propionylcarnitine	Biogenic amines	Glutathione
Acylcarnitines	Stearoylcarnitine	Biogenic amines	Glycylglycine
Acylcarnitines	Tetradecadienylcarnitine	Biogenic amines	Glycylproline
Acylcarnitines	Tetradecanoylcarnitine	Biogenic amines	Homocitrulline
Acylcarnitines	Tiglylcarnitine	Biogenic amines	Homocysteine
Acylcarnitines	2-Methylbutyrylcarnitine	Biogenic amines	5-Hydroxylysine
Amino acids	Citrulline	Biogenic amines	Aminoadipic acid
Amino acids	Cysteine	Biogenic amines	Alpha-aminobutyric acid
Amino acids	Glycine	Biogenic amines	Homoserine
Amino acids	4-Hydroxyproline	Biogenic amines	Kynurenine
Amino acids	Alanine	Biogenic amines	Methionine sulfoxide
Amino acids	Arginine	Biogenic amines	Methionine sulfone
Amino acids	Asparagine	Biogenic amines	N6, N6, N6-Trimethyl-lysine
Amino acids	Aspartic acid	Biogenic amines	O-Acetylserrine
Amino acids	Glutamic acid	Biogenic amines	Putrescine
Amino acids	Glutamine	Biogenic amines	Methylcysteine
Amino acids	Histidine	Biogenic amines	Saccharopine
Amino acids	Isoleucine	Biogenic amines	Sarcosine
Amino acids	Leucine	Biogenic amines	SDMA
Amino acids	Lysine	Biogenic amines	Serotonin
Amino acids	Methionine	Biogenic amines	Taurine
Amino acids	Phenylalanine	Biogenic amines	1-Methylhistidine
Amino acids	Proline	Biogenic amines	3-Methoxytyramine
Amino acids	Serine	Biogenic amines	5-Aminolevulinic acid
Amino acids	Threonine	Carnitines	Carnitine
Amino acids	Tryptophan	Ceramides	Ceramide (d18:0/24:0)
Amino acids	Tyrosine	Ceramides	Ceramide (d18:1/16:0)
Amino acids	Valine	Ceramides	Ceramide (d18:1/22:1)
Amino acids	Ornithine	Ceramides	Ceramide (d18:1/23:0)
		Ceramides	Ceramide (d18:1/24:0)
		Ceramides	Ceramide (d18:1/24:1)

Biochemical class	Metabolite	Biochemical class	Metabolite
Cholesteryl esters	CE (18:3)	Diacyl-phosphatidylethanolamine	PE (34:2)
Cholesteryl esters	CE (18:2)	Diacyl-phosphatidylethanolamine	PE (36:3)
Cholesteryl esters	CE (18:1)	Diacyl-phosphatidylethanolamine	PE (36:4)
Cholesteryl esters	CE (20:5)	Diacyl-phosphatidylethanolamine	PE (38:2)
Cholesteryl esters	CE (22:6)	Diacyl-phosphatidylethanolamine	PE (38:4)
Cholines	Choline	Diacyl-phosphatidylethanolamine	PE (38:6)
Diacylglycerols	DG (36:2)	Diacyl-phosphatidylethanolamine	PE (O-36:5)
Diacylglycerols	DG (36:3)	Diacyl-phosphatidylethanolamine	PE (O-38:5)
Diacylglycerols	DG (36:4)	Diacyl-phosphatidylethanolamine	PE (O-38:7)
Diacyl-phosphatidylcholine	PC (32:0)	Endocannabinoids	alpha-LEA
Diacyl-phosphatidylcholine	PC (32:1)	Endocannabinoids	AEA
Diacyl-phosphatidylcholine	PC (32:2)	Endocannabinoids	DEA
Diacyl-phosphatidylcholine	PC (34:1)	Endocannabinoids	DGLEA
Diacyl-phosphatidylcholine	PC (34:2)	Endocannabinoids	DHEA
Diacyl-phosphatidylcholine	PC (34:3)	Endocannabinoids	LEA
Diacyl-phosphatidylcholine	PC (34:4)	Endocannabinoids	O-AEA
Diacyl-phosphatidylcholine	PC (36:1)	Endocannabinoids	PEA
Diacyl-phosphatidylcholine	PC (36:2)	Endocannabinoids	POEA
Diacyl-phosphatidylcholine	PC (36:3)	Endocannabinoids	SEA
Diacyl-phosphatidylcholine	PC (36:4)	Endocannabinoids	1-/2-Arachidonoyl Glycerol (20:4)
Diacyl-phosphatidylcholine	PC (36:5)	Endocannabinoids	1-Linoleoyl Glycerol (18:2)
Diacyl-phosphatidylcholine	PC (36:6)	Fatty acids	FA (14:0)
Diacyl-phosphatidylcholine	PC (38:2)	Fatty acids	FA (14:1)
Diacyl-phosphatidylcholine	PC (38:3)	Fatty acids	FA (15:0)
Diacyl-phosphatidylcholine	PC (38:4)	Fatty acids	FA (16:0)
Diacyl-phosphatidylcholine	PC (38:5)	Fatty acids	FA (16:1)
Diacyl-phosphatidylcholine	PC (38:6)	Fatty acids	FA (17:0)
Diacyl-phosphatidylcholine	PC (38:7)	Fatty acids	FA (17:1)
Diacyl-phosphatidylcholine	PC (40:4)	Fatty acids	FA (18:1)
Diacyl-phosphatidylcholine	PC (40:5)	Fatty acids	FA (20:0)
Diacyl-phosphatidylcholine	PC (40:6)	Fatty acids	FA (20:1)
Diacyl-phosphatidylcholine	PC (40:7)	Fatty acids	FA (20:2)
Diacyl-phosphatidylcholine	PC (40:8)	Fatty acids	FA (22:1)
Diacyl-phosphatidylcholine	PC (O-34:1)	Fatty acids	FA (22:4)
Diacyl-phosphatidylcholine	PC (O-34:2)	Fatty acids	FA (22:5)-w6
Diacyl-phosphatidylcholine	PC (O-34:3)	Fatty acids	FA (22:6)
Diacyl-phosphatidylcholine	PC (O-36:2)	Fatty acids	FA (24:1)
Diacyl-phosphatidylcholine	PC (O-36:3)	Free fatty acids	FA (18:1)
Diacyl-phosphatidylcholine	PC (O-36:4)	Free fatty acids	FA (18:2)
Diacyl-phosphatidylcholine	PC (O-36:5)	Free fatty acids	FA (20:5)
Diacyl-phosphatidylcholine	PC (O-36:6)	Free fatty acids	FA (22:4)-w6
Diacyl-phosphatidylcholine	PC (O-38:4)	Free fatty acids	FA (22:5)-w3
Diacyl-phosphatidylcholine	PC (O-38:5)	Free fatty acids	FA (22:5)-w6
Diacyl-phosphatidylcholine	PC (O-38:6)	Free fatty acids	FA (22:6)
Diacyl-phosphatidylcholine	PC (O-38:7)		
Diacyl-phosphatidylcholine	PC (O-40:6)		
Diacyl-phosphatidylcholine	PC (O-42:6)		
Diacyl-phosphatidylcholine	PC (O-44:5)		

Biochemical class	Metabolite	Biochemical class	Metabolite
Lysophosphatidylcholine	LPC (14:0)	Lysophospholipids	LPS (18:0)
Lysophosphatidylcholine	LPC (16:0)	Lysophospholipids	LPS (18:1)
Lysophosphatidylcholine	LPC (16:1)	Lysophospholipids	LPS (20:4)
Lysophosphatidylcholine	LPC (18:0)	Lysophospholipids	LPS (22:4)
Lysophosphatidylcholine	LPC (18:1)	Lysophospholipids	LPS (22:6)
Lysophosphatidylcholine	LPC (18:2)	Nitro-Fatty Acids	10-NO ₂ (OA)NO ₂ (OA)
Lysophosphatidylcholine	LPC (18:3)	Organic acids	Inositol
Lysophosphatidylcholine	LPC (20:4)	Organic acids	2-hydroxybutyric acid
Lysophosphatidylcholine	LPC (20:5)	Organic acids	Citric acid
Lysophosphatidylcholine	LPC (22:6)	Organic acids	Glutamic acid
Lysophosphatidylcholine	LPC (O-16:1)	Organic acids	Lactic acid
Lysophosphatidylcholine	LPC (O-18:1)	Organic acids	Malic acid
Lysophosphatidylethanolamines	LPE(22:6)	Organic acids	Fumaric acid
Lysophospholipids	LPA (14:0)	Organic acids	Pyroglutamic acid
Lysophospholipids	LPA (16:0)	Organic acids	3-Hydroxybutyric acid
Lysophospholipids	LPA (16:1)	Organic acids	Aspartic acid
Lysophospholipids	LPA (18:0)	Organic acids	3-Hydroxyisobutyric acid
Lysophospholipids	LPA (18:1)	Organic acids	3-Hydroxyisovaleric acid
Lysophospholipids	LPA (18:2)	Organic acids	Uracil
Lysophospholipids	LPA (20:3)	Oxylipins	PGF _{2a}
Lysophospholipids	LPA (20:4)	Oxylipins	TXB ₂
Lysophospholipids	LPA (20:5)	Oxylipins	10-HDoHE
Lysophospholipids	LPA (22:4)	Oxylipins	11,12-DiHETrE
Lysophospholipids	LPA (22:6)	Oxylipins	11-HETE
Lysophospholipids	LPE (14:0)	Oxylipins	12,13-DiHODE
Lysophospholipids	LPE (16:0)	Oxylipins	12,13-DiHOME
Lysophospholipids	LPE (16:1)	Oxylipins	12-HETE
Lysophospholipids	LPE (18:0)	Oxylipins	12S-HEPE
Lysophospholipids	LPE (18:1)	Oxylipins	12S-HHTrE
Lysophospholipids	LPE (18:2)	Oxylipins	13-HODE
Lysophospholipids	LPE (20:3)	Oxylipins	14,15-DiHETrE
Lysophospholipids	LPE (20:4)	Oxylipins	14-HDoHE
Lysophospholipids	LPE (20:5)	Oxylipins	15-HETE
Lysophospholipids	LPE (22:4)	Oxylipins	15S-HETrE
Lysophospholipids	LPE (22:5)	Oxylipins	16-HDoHE
Lysophospholipids	LPE (22:6)	Oxylipins	17,18-DiHETE
Lysophospholipids	LPG (16:0)	Oxylipins	19,20-EpDPE
Lysophospholipids	LPG (16:1)	Oxylipins	19,20-DiHDPA
Lysophospholipids	LPG (18:0)	Oxylipins	20-HETE
Lysophospholipids	LPG (18:1)	Oxylipins	5,6-DiHETrE
Lysophospholipids	LPG (18:2)	Oxylipins	5-HETE
Lysophospholipids	LPG (20:3)	Oxylipins	5S-HEPE
Lysophospholipids	LPG (20:4)	Oxylipins	8,9-DiHETrE
Lysophospholipids	LPI (16:0)	Oxylipins	8-HETE
Lysophospholipids	LPI (16:1)	Oxylipins	9,10,13-TriHOME
Lysophospholipids	LPI (18:0)	Oxylipins	9,10-DiHOME
Lysophospholipids	LPI (18:1)	Oxylipins	9,12,13-TriHOME
Lysophospholipids	LPI (20:4)	Oxylipins	9-HODE
Lysophospholipids	LPI (22:4)	Oxylipins	9-HOTrE
Lysophospholipids	LPS (16:0)	Platelet activating factor	PAF(18:2)-adduct

Biochemical class	Metabolite	Biochemical class	Metabolite
Sphingolipids	S1P (16:1)	Triglycerides	TG (54:1)
Sphingolipids	S1P (18:0)	Triglycerides	TG (54:2)
Sphingolipids	S1P (18:1)	Triglycerides	TG (54:3)
Sphingolipids	S1P (18:2)	Triglycerides	TG (54:4)
Sphingomyelin	SM (d18:1/14:0)	Triglycerides	TG (54:5)
Sphingomyelin	SM (d18:1/15:0)	Triglycerides	TG (54:6)
Sphingomyelin	SM (d18:1/16:0)	Triglycerides	TG (54:7)
Sphingomyelin	SM (d18:1/16:1)	Triglycerides	TG (55:2)
Sphingomyelin	SM (d18:1/17:0)	Triglycerides	TG (55:3)
Sphingomyelin	SM (d18:1/18:0)	Triglycerides	TG (56:2)
Sphingomyelin	SM (d18:1/18:1)	Triglycerides	TG (56:3)
Sphingomyelin	SM (d18:1/18:2)	Triglycerides	TG (56:4)
Sphingomyelin	SM (d18:1/20:0)	Triglycerides	TG (56:5)
Sphingomyelin	SM (d18:1/20:1)	Triglycerides	TG (56:6)
Sphingomyelin	SM (d18:1/21:0)	Triglycerides	TG (56:7)
Sphingomyelin	SM (d18:1/22:0)	Triglycerides	TG (55:1)
Sphingomyelin	SM (d18:1/22:1)	Triglycerides	TG (58:1)
Sphingomyelin	SM (d18:1/23:0)	Triglycerides	TG (58:10)
Sphingomyelin	SM (d18:1/ 23:1)	Triglycerides	TG (58:2)
Sphingomyelin	SM (d18:0/24:0)	Triglycerides	TG (58:3)
Sphingomyelin	SM (d18:0/24:1)	Triglycerides	TG (58:4)
Sphingomyelin	SM (d18:0/24:2)	Triglycerides	TG (58:5)
Sphingomyelin	SM (d18:0/25:0)	Triglycerides	TG (58:6)
Sphingomyelin	SM (d18:0/25:1)	Triglycerides	TG (58:8)
Triglycerides	TG (44:2)	Triglycerides	TG (58:9)
Triglycerides	TG (46:1)	Triglycerides	TG (60:2)
Triglycerides	TG (46:2)	Triglycerides	TG (60:3)
Triglycerides	TG (48:0)	Trimethylamine-N-oxides	TMAO
Triglycerides	TG (48:1)		
Triglycerides	TG (48:2)		
Triglycerides	TG (48:3)		
Triglycerides	TG (O-50:2)		
Triglycerides	TG (50:1)		
Triglycerides	TG (50:2)		
Triglycerides	TG (50:3)		
Triglycerides	TG (50:4)		
Triglycerides	TG (51:1)		
Triglycerides	TG (51:2)		
Triglycerides	TG (51:3)		
Triglycerides	TG (51:4)		
Triglycerides	TG (52:1)		
Triglycerides	TG (52:2)		
Triglycerides	TG (52:3)		
Triglycerides	TG (52:4)		
Triglycerides	TG (52:5)		
Triglycerides	TG (52:6)		
Triglycerides	TG (53:1)		

CHAPTER 5

The Immunometabolic Atlas: a tool for design and interpretation of metabolomics studies in immunology

Pascal Maas*, Ilona den Hartog*, Alida Kindt, Sonja Boman, Thomas Hankemeier, Coen van Hasselt. The Immunometabolic Atlas: A tool for design and interpretation of metabolomics studies in immunology. *PLoS One* 17:5 (2022). (* Shared first authors)

Abstract

Immunometabolism, which concerns the interplay between metabolism and the immune system, is increasingly recognized as a potential source of novel drug targets and biomarkers. In this context, the use of metabolomics to identify metabolic characteristics associated with specific functional immune response processes is of value. Currently, there is a lack of tools to determine known associations between metabolites and immune processes. Consequently, interpretation of metabolites in metabolomics studies in terms of their role in the immune system, or selection of the most relevant metabolite classes to include in metabolomics studies, is challenging. Here, we describe the Immunometabolic Atlas (IMA), a public web application and library of R functions to infer immune processes associated with specific metabolites and vice versa. The IMA derives metabolite-immune process associations utilizing a protein-metabolite network analysis algorithm that associates immune system-associated annotated proteins in Gene Ontology to metabolites. We evaluated IMA inferred metabolite-immune system associations using a text mining strategy, identifying substantial overlap, but also demonstrating a significant chemical space of immune system-associated metabolites that should be confirmed experimentally. Overall, the IMA facilitates the interpretation and design of immunometabolomics studies by the association of metabolites to specific immune processes.

5.1 Introduction

Immunometabolism, or the interplay of immunology and metabolism, has received increasing interest because of its role in the function and regulation of immune system processes in health and disease. Metabolites with *e.g.*, pro- or anti-inflammatory functions may be of interest as biomarkers or drug targets for inflammation and immune system-associated pathologies such as infection, cancer, and various auto-immune diseases [120, 121, 14]. Significant knowledge gaps related to the relationship between metabolism and immune function remain to be elucidated. To this end, metabolomics technologies can facilitate the identification and quantification of metabolites in relation to the immune system in experimental models and clinical studies.

For biochemical and functional interpretation of metabolomics study results, different computational tools can be used: biochemical pathways analysis can be executed using tools such as MetaboAnalyst or KEGG, and for functional analysis, STITCH can be employed [27, 26, 28]. However, inferring the relationship of metabolites with immune system processes remains challenging. In contrast, for the analysis of genes, gene expression, and proteins, such biological interpretation is straightforward through the use of high-quality annotated ontologies such as Gene Ontology [122, 123].

For hypothesis-driven metabolomics studies that require absolute quantification of measured metabolites, targeted metabolomics methods are preferred over untargeted metabolomics methods. However, targeted mass spectrometry-based metabolomics studies measure by design only a subset of metabolites and metabolite classes at once. Guidance in the selection of the most relevant subset of metabolites for the immune process of interest is therefore of relevance. However, tools to facilitate the design of targeted metabolomics studies by pre-selection of metabolites of interest are lacking.

To address the current hurdles of hypothesis generation and biological interpretation of metabolomics studies, we developed the Immunometabolic Atlas (IMA). The IMA enables inference of immune system associated functions, and vice versa, to determine relevant metabolites with specific immune system processes. We infer metabolite-immune process associations utilizing a protein-metabolite network analysis algorithm that associates immune system-associated annotated proteins (Figure 5.1), leveraging protein-metabolite interaction databases [124, 125] and protein annotations of immune system processes in Gene Ontology (GO). We then characterize the global metabolite-immune process coverage and perform validation through text mining-derived immune system associations. The application of the IMA is demonstrated in a case study and is made available as an R package and public web application.

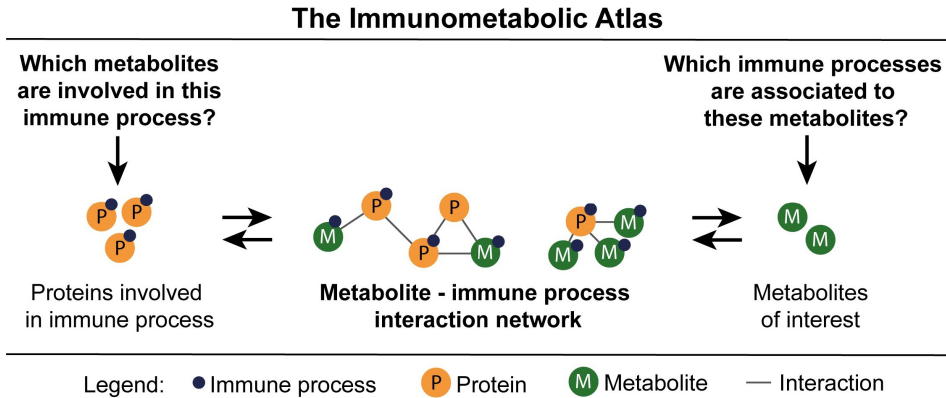


Figure 5.1 A conceptual overview of the Immunometabolic Atlas (IMA). The IMA provides associations between metabolites and immune processes of interest through the generation and evaluation of a protein-metabolite interaction network.

5.2 Methods

5.2.1 Assembly of immune process-metabolite interaction network

We constructed a database that contains associations between specific immune process terms, proteins, and metabolites through the integration of publicly available databases (Figure 5.2A). Through the integration of these resources, we constructed an interaction network to associate metabolites with immune processes. In the following paragraphs, the development of the immune process-metabolite interaction network is described.

5.2.1.1 Immune processes

Immune processes were retrieved as GO terms from Gene Ontology. The associated gene names that were descendants of “Immune System Process” (GO:0002376) were acquired using the EBI QuickGo application programming interface (API, version 2021-05-24) [122, 123, 126].

5.2.1.2 Proteins and protein-immune process associations

Human proteins (Swiss-Prot) were retrieved from the UniProt database [127]. The requested UniProt data included: entry (UniProt identifiers), protein name, cofactors, EC number, transporter protein (TCDB), Ensembl transcript, and GO immune processes. The reference to GO immune processes in the UniProt data was used to identify immune system-related proteins.

5.2.1.3 *Metabolites*

Metabolite names and associated metadata were obtained from the Human Metabolome Database (HMDB, version 4.0). We only included metabolites that were known to either have a biological role and/or were part of a naturally occurring process to exclude any synthetic drugs. We also excluded any inorganic compounds. The retrieved HMDB data included: name, class, superclass, accession (HMDB identifiers), ChEBI ID, UniProt ID, biospecimen, cellular locations, and metabolic pathways.

5.2.1.4 *Protein-protein interactions*

Protein-protein interactions were obtained from STRING's functional protein association networks version 11.0 [124]. Ensembl transcripts from the UniProt data were converted to Ensembl Protein IDs using the Ensembl API [128]. Subsequently, STRING was parsed using these IDs to extract protein-protein interactions.

5.2.1.5 *Metabolite-protein interactions*

The UniProt identifiers in the HMDB data were used to connect the metabolites to the proteins in the UniProt data, obtaining metabolite-protein interactions. Proteins without immune system-related GO terms were excluded from further analysis.

5.2.1.6 *Metabolite-metabolite interactions*

Metabolite-metabolite interactions for the obtained metabolites from HMDB were retrieved using the Rhea-Annotated reactions database (RheaDB, release 118) [125]. We cross-referenced HMDB with ChEBI to extract interactions stored in Rhea. We applied an all-versus-all method, where each reactant-product combination results in an individual interaction.

5.2.1.7 *Building the interaction network*

To construct the interaction network for each immune process extracted from GO, first, proteins involved in the immune processes were identified. Then, protein-metabolite, protein-protein, and metabolite-metabolite interactions were added to the network. To build an interaction network for metabolites of interest, proteins associated with the metabolites of interest were identified. Related protein-metabolite, protein-protein, and metabolite-metabolite interactions were then added to the network (Figure 5.2B). For metabolites with only metabolite-metabolite interactions, no interaction network can be constructed, because at least one protein-metabolite interaction is necessary to inherit immune processes.

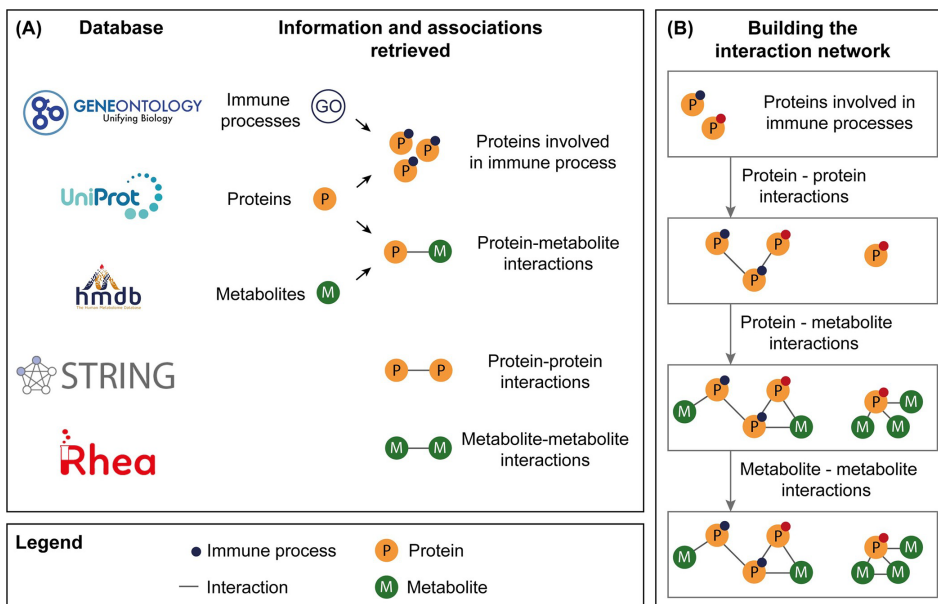


Figure 5.2 Overview of (A) information and associations retrieved from available databases, and (B) the study flow to build protein-metabolite interaction networks to associate metabolites and immune processes of interest. First, the proteins that are associated with the immune processes of interest are added to the network. Then, protein-protein, protein-metabolite, and metabolite-metabolite interactions are added to the network to generate the final interaction network for the immune processes of interest.

5.2.1.8 Inheritance of immune processes by metabolites

To associate metabolites to immune processes, an inheritance methodology was applied (Figure 5.3B). In the default, first-order inheritance method, metabolites inherit the immune processes of the directly neighboring proteins only. For second-order and third-order inheritance, metabolites inherit both the immune processes of their direct neighboring proteins and the first neighbors of that protein, two or three interaction steps away, respectively. The preferred inheritance order can be defined by the user.

5.2.2 Evaluation of network-inferred metabolite and immune processes

5.2.2.1 Overrepresentation analysis

We test for the overrepresentation of metabolites and immune processes in the interaction network using Fisher's exact test with multiple testing correction, using the IMA metabolites and immune processes as background. Based on this, we rank

by p-value to identify the most significant metabolites or immune processes associated with either an immune process or metabolite set. The p-value for Fisher's exact test was computed as follows (Equation 5.1)

$$p - \text{value Fisher's exact test} = \frac{(a+b)! (c+d)! (a+c)! (b+d)!}{a!b!c!d! (a+b+c+d)!} \quad (5.1)$$

Here, for metabolite-based overrepresentation analysis, a is the number of associations of a specific metabolite to a specific immune process in the interaction network (via multiple proteins), b is the number of associations of other immune processes to the specific metabolite in the network, c is the total number of associations of the specific metabolite to the specific immune process in the database minus the number of associations of the specific metabolite to the specific immune process in the network, and d is the total number of immune process associations to the specific metabolite in the database minus the number of associations of other immune processes to the specific metabolite in the network. For immune process-based overrepresentation analysis, a is the number of appearances of a specific immune process in the interaction network, b is the number of the immune process appearances in the IMA database, c is the number of other immune processes in the network, and d is the number of other immune processes appearances in the IMA database.

5.2.2.2 Metabolite centrality

We calculated metabolite centrality to determine the position of a metabolite in the interaction network. A metabolite could be on the edge of a network with minimal interactions, in the center of a network with a lot of interactions, or somewhere in between. The centrality was calculated as harmonic closeness, which is a distance-based centrality metric that is suitable for disconnected graphs, in contrast to classical closeness. A high harmonic closeness value indicates a central position of the metabolite in the network. For node i , the harmonic closeness is calculated by taking the sum of all reciprocals of distance d to other node j (Equation 5.2). The centrality was determined for each metabolite in the network separately.

$$\text{Harmonic closeness}(i) = \sum_{j \neq i} \frac{1}{d_{i,j}} \quad (5.2)$$

5.2.2.3 Metabolite precision

Metabolite precision was computed to quantify how specific a metabolite is for a certain immune process. The precision score was computed for each metabolite-immune process association to allow discrimination between metabolites that are contributing either to a single process or to multiple processes and between common and rare metabolites that have comparable centrality scores. The precision of a metabolite for

an immune process of interest is determined by the ratio of the metabolite associations with the process of interest, compared to all its associations remaining in the IMA database (Equation 5.3). A high metabolite precision value indicates that most of the immune process associations the metabolite could have, according to the IMA database, are present in the immune process network.

$$\text{Precision score}(i, j) = \frac{N_{i,j}}{N_i \cdot V_j} \quad (5.3)$$

Here, for metabolite i in process j , with N being the number of interacting nodes of metabolite i and V the number of nodes in process j . The precision score is corrected by the number of nodes in the process.

5.2.3 Evaluation of IMA metabolite-immune process association performance through text mining

To evaluate the evidence available for metabolite-immune process associations identified by the IMA, an external validation dataset was created using text mining. We selected papers including one or more metabolites and immune processes that were also present in the IMA database using the EuropePMC API on 6 March 2021 [129]. We included EuropePMC-listed journal articles in which a metabolite and immune process term from the IMA database was detected in the abstract, methods, results, supplement, figures, and/or tables. Introduction and discussion sections were excluded since comparisons to results of other studies are often made in these sections, possibly leading to biased text mining results. Also, papers were only included if they were related to humans.

The text mining resulted in a list of PubMed identifiers (PMIDs) which were used to find associations between metabolites and immune processes. These associations were included in the quantitative text mining validation dataset. Metabolite-immune process associations with only one occurrence in the text mining dataset were removed to limit false positives. We excluded the superclass lipids and lipid-like molecules as defined within HMDB from the validation because the complex nomenclature of these metabolites made text mining unfeasible.

We characterized the IMA database by cross-referencing metabolites and processes with the text mining database. Metabolites were grouped according to their presence or absence in the IMA database. Furthermore, we evaluated the quality of the IMA database by calculating the specificity, sensitivity, precision, accuracy, and F_1 -score (Equation 5.4-5.8). The F_1 -score focuses on the positive predictions and leaves out the True Negatives. The F_1 -score represents the performance of the IMA better than other evaluation measures because it evaluates how well associations are made instead of how well associations are excluded.

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}} \quad (5.4)$$

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (5.5)$$

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad (5.6)$$

$$\text{Accuracy} = \frac{\text{True Negatives} + \text{True Positives}}{\text{True Negatives} + \text{True Positives} + \text{False Negatives} + \text{False Positives}} \quad (5.7)$$

$$F_1 - \text{score} = \frac{\text{True Positives}}{\text{True Positives} + \frac{1}{2}(\text{False Positives} + \text{False Negatives})} \quad (5.8)$$

5.2.4 R package and Shiny application

We implemented the IMA in the R package IMA, which facilitates users to create various graph-based analyses. The package includes an interactive R shiny application that allows for a user-friendly interpretation of our interaction database. The app adds extensions that are useful for additional analyses, including metadata from HMDB and UniProt, and allows networks to be built using either one or multiple immune processes, or by one or multiple metabolites. If one or multiple immune processes are used as input, all connected metabolites that are in the Immunometabolic Atlas database will be included in the graphical network. The app also features two additional versions of interaction datasets, which allows users to determine the strictness of the app. These datasets include proteins that are unrelated to the immune system but do have an interaction with an immune system-related protein. The first dataset includes neighbors of immune system-related proteins, whereas the second dataset includes the second neighbors of an immune system-related protein. The package and all other scripts used for analysis are available in our Github repository <https://github.com/vanhasseltlab/IMAtlas>.

5.3 Results

5.3.1 Development of the IMA database and metabolite-immune process algorithm

The IMA database includes all child processes of the immune system process (GO:0002376) and contains 97 525 metabolites, 3 101 proteins, 1 712 immune processes,

664 metabolite-metabolite interactions, 172 291 protein-metabolite interactions, and 411 286 protein-protein interactions (Table 5.3).

We associated immune processes and metabolites in a stepwise process (Figure 5.3). Immune processes were assigned to metabolites using a first, second, or third-order inheritance strategy (Figure 5.3B). By default, first-order inheritance of immune processes is used, in which metabolites only inherit immune processes from their directly interacting protein. To determine if a metabolite of interest plays a central role in the metabolite-immune process interaction network, a centrality score was calculated (Figure 5.3C). Metabolites with a high centrality score are typically located in a central point in the network and have multiple interactions with surrounding metabolites and proteins, while metabolites with a low centrality score are less closely connected to other metabolites or proteins in the network and are typically located towards the edges of a network. To indicate how specific a metabolite is for a certain immune process, the precision score was calculated (Figure 5.3D). The precision of a metabolite for an immune process of interest is determined by the ratio of the metabolite associations with the process of interest, compared with all its associations remaining in the database. Metabolites with a high precision score are typically committed to a smaller number of immune processes. To rank metabolites and immune processes in the interaction network, we calculated a p-value that signifies the overrepresentation of the metabolites and immune processes in the network in comparison to the ones in the database using Fisher's exact test (Figure 5.3E). This resulted in a performance table with significance values for every metabolite and immune process within the network. The significance value for overrepresentation of the immune process for a specific metabolite is indicative of the strength of metabolite-immune process association. The network-based significance value that indicates the overrepresentation of an immune process within the entire network indicates the importance of the collection of metabolites for the immune process. In summary, the metabolites and immune processes in the network are ranked based on their metabolite centrality, precision and p-value, and the immune process p-value (Figure 5.3A).

5.3.2 Overview of metabolism-immune response associations

To provide an overview of IMA-inferred metabolite-immune response process associations, we categorized GO terms according to main high-level immune response processes as defined in the standard textbook Janeway's Immunobiology [130]. We determined for each of these immune response processes the biochemical metabolite superclasses of identified metabolites (Figure 5.4). We found significant differences in metabolite classes associated with unique immune response processes (Figure 5.4A). The average number of immune processes per protein is in the same order of magnitude for all superclasses except for benzenoids (Table 5.4). Metabolites of the superclass of lipids and lipid-like molecules, here referred to as 'lipids', were abundantly present with 90 280 occurrences (92.6%) but interacted with a relatively small portion of proteins (5.1%). Excluding lipids from the analysis shows a lower average number of metabolites that were associated with a specific immune process. There were 18,995 unique metabolites

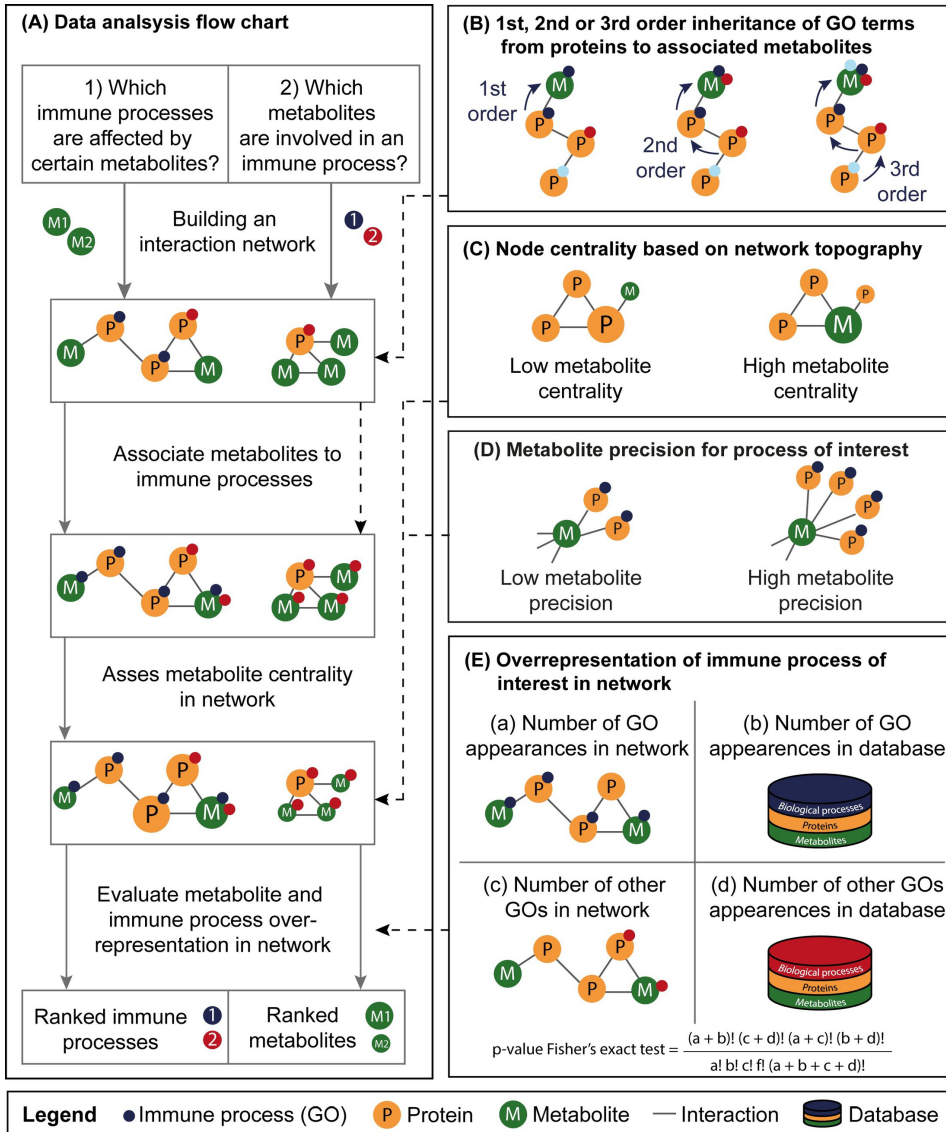


Figure 5.3 Overview of the IMA network-based interaction analysis. (A) Flowchart of the data analysis of an interaction network using a first-order inheritance strategy. (B) The metabolite of interest can inherit immune processes directly (1st order, default) or indirectly (2nd or 3rd order) from neighboring proteins based on the order of inheritance chosen by the user. (C) The centrality of a protein or metabolite in a graphical network is determined using the harmonic closeness score. This topology-based score is the highest for metabolites with multiple connections in a central point in the network. (D) The precision of a metabolite for a process of interest is determined by the ratio of its interactions within the process compared to all its interactions and represents the commitment of a metabolite to the process of interest. (E) The overrepresentation of an immune process (GO term) in an interaction network is evaluated using a Fisher's Exact Test with FDR multiple testing correction. The resulting significance levels can be used to rank immune processes in a network.

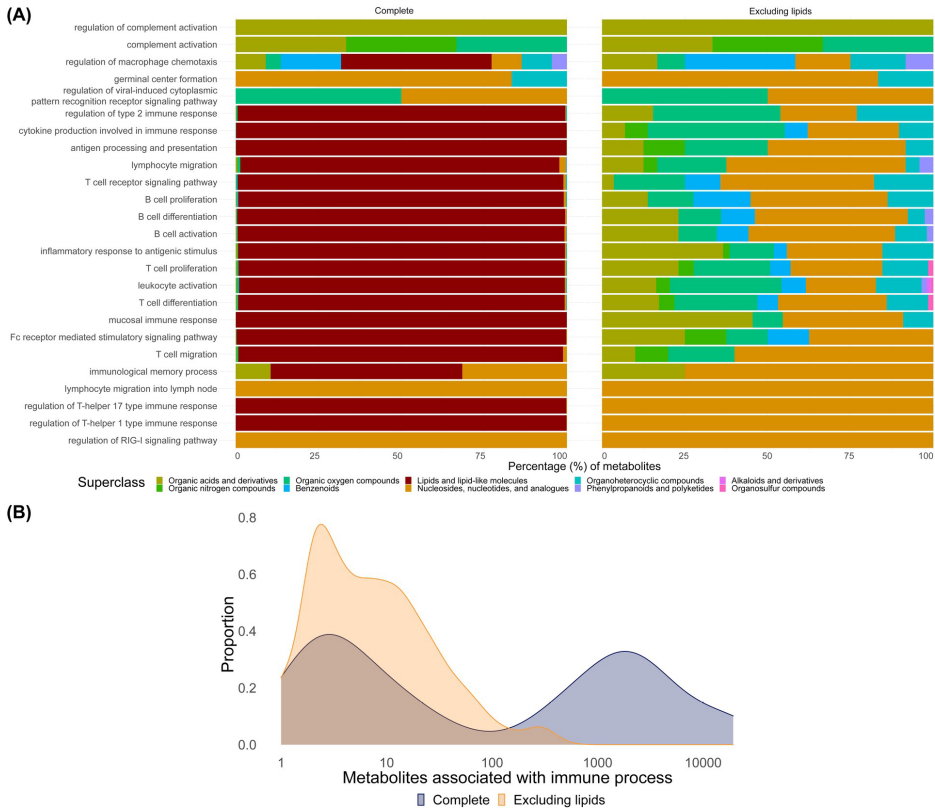


Figure 5.4 Overview of immune process and metabolite associations. (A) Distribution of the biochemical metabolite classes identified for common immune processes in the IMA classified according to the standard textbook Janeway’s Immunobiology, either including (left) or excluding (right) lipids. (B) Distribution of the number of metabolites associated with specific immune processes inferred from the IMA, using first-order inheritance, either for excluding lipids (orange) or including all metabolites including lipids (blue).

associated with the main immune processes when lipids were included and 342 when they were excluded. A large variation of the number of metabolites associated with immune processes was present (Figure 5.4B). Excluding lipids results in a shift from many metabolites to smaller numbers of metabolites that are associated with an immune process. The exclusion of the superclass of lipids and lipid-like molecules from these results excludes several metabolite classes including fatty acyls, glycerophospholipids, and prenol lipids.

5.3.3 Validation of the metabolite-immune process associations

The methodology was validated by comparing the results from an interaction analysis of all metabolites and immune processes in the IMA database to metabolite-immune

process associations found in literature for the same metabolites and immune processes. The immune system process interaction network was built using 1st order immune process inheritance and resulted in 432 metabolites associated with 767 immune processes.

Associations of metabolites and immune processes related to the immune system process in literature were collected using a text mining approach. We identified 1 046 metabolites that were associated with 565 immune system processes (Table 5.5). The overlap between associations found by the IMA and found in literature was 31.5% (290 metabolites in 398 processes, Figure 5.5).

Of all associations found in text mining and by the IMA, 58.7% of the metabolites involved in metabolite-immune process associations were only found in literature and not by the IMA ($n = 614$ metabolites). Of these, 398 metabolites were lacking any interaction according to the IMA database and therefore remained undetected using the IMA methodology. The 216 remaining metabolites with known interactions could be classified as having either only metabolite-metabolite interactions and/or protein-protein interactions. Metabolites that were only interacting with other metabolites, and not with proteins, could not be detected because immune processes are only inherited through proteins in the current IMA methodology. Metabolites that were only interacting with proteins that were not in the immune system process (according to GO), were also not included in the IMA database. Finally, 13.6% of the metabolites that inherited an immune system process were only found using the IMA and not in literature ($n = 142$). Of these, 55 metabolites were identified in literature but were lacking a link to the immune system. The remaining 87 metabolites were not found in any immune-related studies through text mining.

The metabolite-immune process associations found in literature were considered as the gold standard for the evaluation of the performance of different orders of immune process inheritance. By default, the inheritance of processes was done through direct protein interactions (first-order), but inheritance through indirect protein interactions was also evaluated (second and third-order). Therefore, specificity, sensitivity, precision, F_1 -score, and accuracy were calculated (Table 5.1). All methods of inheritance yielded high specificity and accuracy values, indicating that the IMA is strict in linking processes to metabolites. Relatively low values for precision and sensitivity were reported, indicating discrepancies between the associations found in literature and made by the IMA. Comparing direct- and indirect inheritance showed a higher precision for direct inheritance, while indirect inheritance showed higher sensitivity. Since the IMA gives high numbers of True Negatives, also the F_1 -score, which does not take the number of true negatives into account, was calculated to quantify the difference between the methods of inheritance. The F_1 -score favoured direct inheritance.

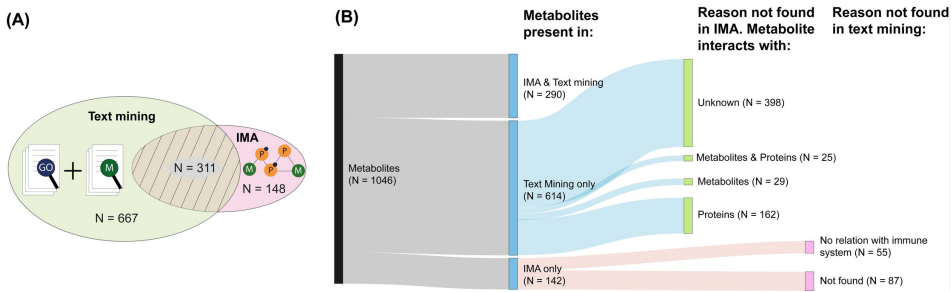


Figure 5.5 Comparison of metabolites in IMA database and text mining dataset for validation. (A) Text mining was used to identify co-occurrences of GO-terms and metabolites present in the IMA database. Metabolites obtained from co-occurrences were compared with associations found using IMA. (B) Sankey diagram of metabolites found in both IMA and text mining. 290 of 1 046 metabolites found are present in both IMA and text mining datasets. The portion of non-overlapping metabolites can be explained based on exclusion criteria for the atlas. 614 metabolites were only found in literature, of which 398 have no known interaction to any proteins or metabolites in the IMA. There were interactions found of the metabolites to other metabolites ($n=81$), metabolites and proteins (42) or proteins (160), but these did not have an association to an immune process in the IMA database. Of the 142 metabolites that were only found in the atlas, 87 could not be found in literature and 55 were found in literature but missed any connection to an immune process.

Table 5.1 Absolute performance measure results of first, second, and third-order inheritance.

Order of inheritance	Specificity	Sensitivity	Precision	Accuracy	F1-score
First order	0.99	0.11	0.29	0.97	0.16
Second order	0.85	0.43	0.07	0.84	0.12
Third order	0.73	0.49	0.04	0.73	0.08

5.3.4 Identification of possible biomarkers using network-based interaction analysis

To identify which metabolites could be of interest for a specific immune process, we reported the position of a metabolite in the network (centrality) and the exclusivity of the metabolite for a certain immune process (precision, Figure 5.6B-C). Metabolites with high centrality and precision scores might be of interest as biomarkers for the associated immune process. Therefore, all metabolite-immune process associations made by the IMA were analyzed on centrality and precision. Only statistically significant metabolite-immune process associations after FDR multiple testing correction were included in the results ($p < 0.05$, Figure 5.6A, Table 5.6). To identify specific metabolic biomarkers, we selected metabolite-immune process associations that were above the set threshold of the mean plus two standard deviations for both centrality and precision. After analysis of the associations in the current IMA association dataset, 48 metabolites were found to be of interest as a potential biomarker in 47 immune processes. Several metabolites were found to be involved in antigen processing via MHC class 1B, including

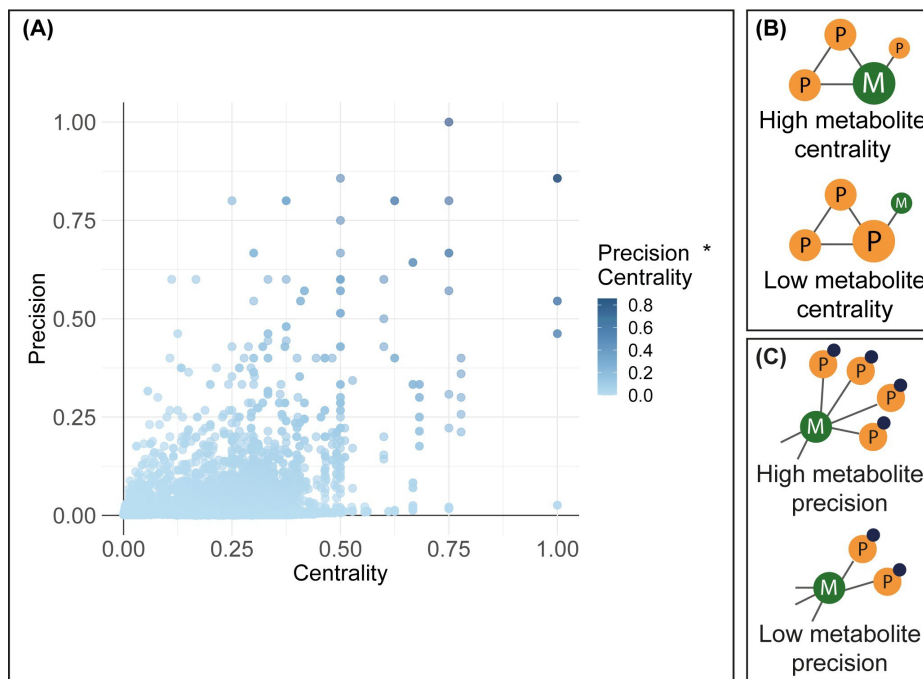


Figure 5.6 Precision-centrality plot of the significant metabolite-immune process associations in the IMA excluding lipids. (A) For each metabolite in each immune process, the centrality and precision were calculated and normalized to the network size. Metabolites with high centrality and precision scores might be of interest as biomarkers for the associated immune process (B) The centrality represents the position of a metabolite in the network. (C) The precision represents the exclusivity of the metabolite for a certain immune process.

sphingosine, sphinganine, and dihydroceramide. Furthermore, we identified strong relationships between several pyruvic acids and positive regulation of prostaglandin secretion.

5.3.5 Positive regulation of T cell-mediated immunity

As an example, the interplay between metabolites and proteins for the immune process of positive regulation of T cell-mediated immunity is demonstrated (GO:0002711, Figure 5.7). The interaction network that was built for this process shows one big cluster of proteins and metabolites, and some unconnected proteins (not shown). Unconnected proteins may interact with non-immune-related proteins which are not considered in the current IMA methodology and indicate the current knowledge gap. Of the 8 metabolites in the network, 5 were found to be significant for the immune process ($p < 0.05$, Table 5.2). These molecules are highly related as they interact with the same proteins. This results in the same centrality value for each molecule; however, the precision varies as they can interact with proteins in other processes. The exception here is 3-Dehydrospinganine,

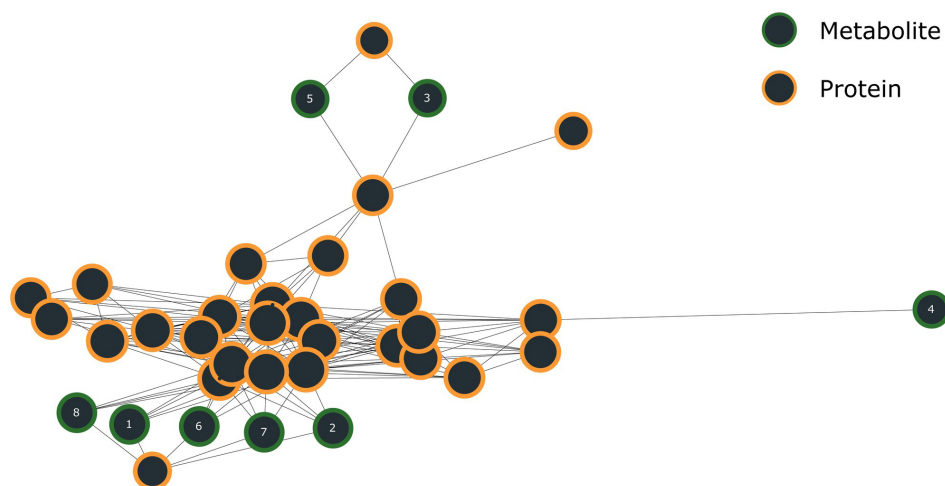


Figure 5.7 The interaction network of positive regulation of T cell-mediated immunity (GO:0002711) shows the interplay between metabolites (green) and proteins (orange).

Table 5.2 Metabolites associated with positive regulation of T cell-mediated immunity. Fisher's exact test with FDR multiple testing correction was used to calculate p-values, while centrality and precision values are indicators of the importance of the metabolites in this process. The metabolite number in the table corresponds to the number in the interaction network in Figure 5.7.

Metabolite	Metabolite superclass	Centrality	Precision	P-value	Metabolite number in network
3-Dehydrosphinganine	Organic oxygen compounds	0.30	0.63	<0.001	6
Phytosphingosine	Organic nitrogen compounds	0.30	0.56	<0.001	7
Sphinganine	Organic nitrogen compounds	0.30	0.45	<0.001	2
Sphingosine	Organic nitrogen compounds	0.30	0.39	<0.001	1
Dihydroceramide	Organicacids and derivatives	0.30	0.31	<0.001	8
S-Adenosylmethionine	Nucleosides, nucleotides, and analogues	0.22	0.07	1.00	4
ADP	Nucleosides, nucleotides, and analogues	0.22	0.01	0.58	5
ATP	Nucleosides, nucleotides, and analogues	0.22	0.01	0.58	3

which only interacts with proteins involved in positive regulation of T cell-mediated immunity, resulting in a precision of 1.00.

Several significant metabolites form a primary component for sphingolipids. Sphingolipids are membrane lipids that function as ligands for sphingosine-1-phosphate receptors (S1PR) and are especially associated with the determination of T cell phenotypes [131]. Previous studies have shown that deficiency in S1PR can cause failure in mature T cells leaving the thymus [132]. Finally, it has been shown to be an important factor for coordinating adaptive immune responses through the S1P₁-Akt-mTOR pathway [133].

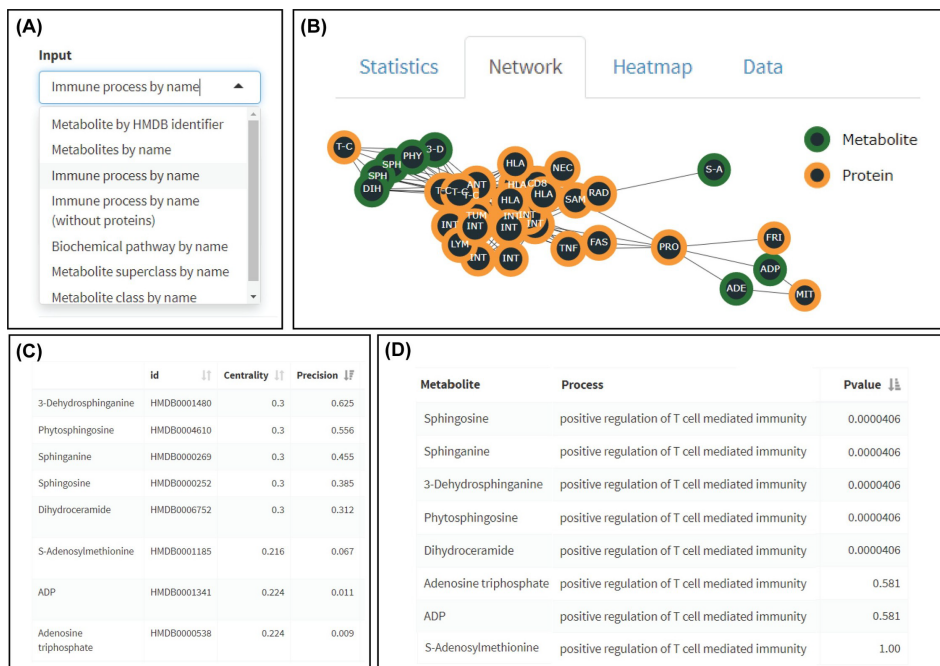


Figure 5.8 IMA Rshiny application concept. (A) The application allows users to enter processes, metabolites, proteins, or identifiers from a list to produce networks and calculate statistics. This flexibility of input possibilities enables to obtain metabolites from immune processes and vice versa. (B) Several visualizations have been included to visualize associations between metabolites and processes. Here, we zoomed in on the connected part of the interaction network of T cell-mediated immunity. (C, D) Example outputs for metabolites associated with positive regulation of T cell-mediated immunity including centrality, precision, and p-values.

5.3.6 IMatlas R package and R shiny application

We have implemented the IMA as an R package and R shiny module. The IMA supports several search modes to facilitate the construction of networks, using either immune processes or metabolites as input (Figure 5.8). An interaction network is built and evaluation metrics such as metabolite centrality and p-value are calculated. The IMA supports bulk input of HMDB identifiers or GO terms to produce graphs that can be adjusted using several thresholds using the settings panel. For example, confidence thresholds used by STRING for protein-protein interactions can be increased to include only very well-curated interactions. Other features include generating neighborhood graphs of a given set of metabolites and the ability to search using (super)classes and/or biochemical pathways. In summary, the application contains useful features to construct network graphs for non-programmatic applications.

5.4 Discussion

We describe the development of the Immunometabolic Atlas (IMA), which leverages protein-metabolite interaction analysis to identify metabolites associated with immune processes, and vice versa, and which can be used to interpret and design metabolomics studies.

The IMA is based on metabolites included on the Human Metabolome Database (HMDB), which is a large, comprehensive, and well-annotated database of metabolites found in humans, and is more complete than alternative human metabolite databases [134]. HMDB contains a large number of lipid metabolites, which were found to be associated with many immune processes. Lipids are highly biologically relevant in various biological functions as is also extensively studied within the field of lipidomics [135, 136]. In this study, the superclass of lipids and lipid-like molecules was excluded from the validation and the example shown because the complex nomenclature of these metabolites made text mining unfeasible. However, inclusion of the superclass lipids and lipid-like molecules is available for researchers using the IMA. Not all lipids were removed by excluding this superclass. For example, sphingolipids were included in the example in Figure 5.7. Depending on the method of classification that is used, some lipid metabolites will be classified as such, and some will be classified further into other categories. The method of classification of metabolites by HMDB could be debated but is considered to be out of the scope of this study.

Association between metabolites and immune processes was based on the inferred protein-metabolite network, where proteins were associated with GO-associated immune processes either through direct or higher-order inheritance based on protein-protein interactions present. The rationale for this strategy is based on the assumption that metabolites and proteins commonly interact: e.g., as enzyme-substrate or co-factor [137]. Of course, the majority of proteins are not limited to a single biological process. Evaluation of association strength of metabolite-immune processes through classical over-representation analysis is important to identify those metabolites or immune processes of primary interest [105]. Of note, the absolute p-value obtained from the over-representation analysis should be interpreted with caution because in the IMA not the whole metabolome for all human biological processes are used as a background, but only the metabolites and immune system processes in the IMA database.

To further interpret metabolites for their value as specific and selective biomarkers or drug targets we have included computation of precision and centrality network-structure inferred metrics. Centrality concerns the position of a metabolite in a network and has been proposed before [138]. The precision score assesses how specific the metabolite is for the associated process within the network. These metrics may help to identify metabolites of increased interest as a biomarker for the specific immune process, e.g. because they are more likely to have a specific function, as inferred from the underlying network structure.

We evaluated immune response-metabolite associations through a comparison of literature text mining derived metabolite-immune process associations. We found that a substantial part (67%) of the IMA associations overlapped with the associations found in literature. Overall, the text mining approach identified a two times higher number of associations than the IMA. We expect that this is related to the nature of methodology used to identify associations, because of its intrinsic high likelihood of identifying false-positive associations, which we attempted to reduce through applying several filtering steps. Ultimately we found that the IMA yields a specificity of 73-99% and sensitivity of 11-49% depending on the inheritance method used, which indicates our method shows sufficient performance to be used as a tool for hypothesis generation or to guide metabolomics study design.

A similar tool for the functional interpretation of metabolomics study results is STITCH [28], which is a database incorporating known and predicted interactions between metabolites and proteins. STITCH assigns processes to metabolites based on direct interactions and a clustering-based algorithm. STITCH does not include topological measurements, whereas in the IMA this is applied for easier interpretation of larger networks. In contrast to the IMA, associations between GO biological processes and metabolites can only be made in the direction of metabolites to processes, but not from (immune) processes to metabolites.

Limitations of the current IMA include the lack of directionality of associations in the protein-metabolite network, which could help in identifying biochemical interactions that are most relevant and plausible. In addition, incorporation of data on cell-type-specific as well as (sub-) cellular locations of metabolites or metabolite-protein associations may help in refining metabolite-immune system associations inferred, in particular, because of the complex and multi-cellular nature of the immune response.

We conclude that the developed IMA can be a relevant tool to guide researchers in the field of immunometabolomics in the interpretation of immune-metabolomics data from experiments or clinical studies and to guide the design of prospective metabolomics studies in the field of immunology, which we facilitate by making our tool available both as R package and user-friendly web-application. Finally, we expect that the conceptual approach and developed algorithms for inferring metabolite-immune process associations through protein-metabolite interaction networks can be expanded towards complete biological ontologies, and is not just limited to immune processes.

5.5 Supporting information

Table 5.3 Contents of IMA database.

Type of data	Number included in the IMA
Immune processes	1 712
Proteins	3 101
Metabolites	97 525
Immune process-protein associations	69 641
Protein-protein interactions	411 286
Protein-metabolite interactions	172 291
Metabolite-metabolite interactions	664

Table 5.4 Summary of superclass characteristics in the IMA

Superclass	Number of metabolites	Number of proteins	Number of unique immune processes	Metabolite / protein ratio	Number of immune processes per protein	Number of unique immune processes per protein
Organic acids and derivatives	714	138	386	5	57	3
Organic nitrogen compounds	134	29	148	5	74	5
Organic oxygen compounds	1278	162	431	8	84	3
Benzenoids	966	24	166	40	116	7
Lipids and lipid-like molecules	90309	164	467	551	95	3
Nucleosides, nucleotides, and analogues	152	436	639	0	78	1
Organoheterocyclic compounds	1786	72	291	25	52	4
Phenylpropanoids and polyketides	1915	9	74	213	62	8
Alkaloids and derivatives	113	1	16	113	16	16
Organosulfur compounds	174	9	77	19	42	9

Table 5.5 Text mining results listing papers that mention both metabolite and immune process that is present in the IMA database.

This table is available as zip file (S3 Table) on the website of the publisher at <https://doi.org/10.1371/journal.pone.0268408>

Table 5.6 All significant metabolite-immune process associations in the IMA.

This table is available as zip file (S4 Table) on the website of the publisher at <https://doi.org/10.1371/journal.pone.0268408>

CHAPTER 6

General discussion and summary

Community-acquired pneumonia (CAP) is a lower respiratory tract infection with a high incidence [1, 2]. Hospitalized patients with moderate to severe CAP typically receive empirical broad-spectrum antibiotic therapy, that can be switched to targeted therapy when the microbial etiology is determined. To optimize patient care and to reduce the risk for development of antimicrobial resistance (AMR) in treatment of CAP, there is a need for additional biomarkers to support microbial diagnosis and treatment monitoring of CAP. Metabolomics is a key technology of relevance as a potential source of new biomarkers for CAP. In this thesis, in **Chapter 1**, we first introduced current challenges in diagnosis and treatment of CAP, the basic concepts in the use of metabolomics for biomarkers discovery, and our central hypothesis that changes in the host metabolome in patients with CAP may be a potential source for novel biomarkers. We then described our studies to assess the potential utility of metabolomics-based biomarkers for diagnostic purposes (**Chapter 2 & Chapter 3**) and for the monitoring of the treatment response (**Chapter 4**) in patients with CAP. Finally, we described the development of a computational tool which can help in the design and analysis of metabolomics studies investigating the host immune response (**Chapter 5**)

6.1 Diagnosis of CAP

Currently used approaches for the etiological diagnosis of pathogens in CAP are based on several techniques, including culturing, antigen testing, and PCR [44, 45, 46, 47]. The adoption of PCR-based assays has expanded over the past few years. Due to the high sensitivity and short turnaround times, PCR point-of-care assays have great value for the diagnosis of CAP. Using multiplex PCR respiratory panels, a large number of potential pathogens can be detected [139, 140]. Respiratory PCR assays are typically performed on sputum, which can be a limitation for CAP patients, who are often unable to produce a sputum sample. Nasal or nasopharyngeal swab PCR tests that could overcome this challenge have been proven effective in the diagnosis of Methicillin-resistant

Staphylococcus aureus (MRSA) and COVID-19 [141, 142]. However, the identified pathogens in these PCR tests may not be conclusively the cause of CAP because the sample does not originate from the lower respiratory tract [143]. Blood-based diagnostic assays investigated in this thesis that address these limitations are therefore of potential relevance for diagnosis or prognosis of CAP. Another potential limitation of PCR-based assays relates to their targeted nature, i.e., pathogens for which the associated target sequences are not included, can also not be detected. Potentially, assays which consider the host immune response to specific pathogens could address this limitation.

Evaluation of metabolic biomarkers for etiological diagnosis of CAP

In **Chapter 2**, we aimed to determine if predictive metabolomics-derived biomarkers could be identified to discriminate between key pathogens or pathogen groups in CAP. To this end, we performed extensive metabolomics profiling of serum samples from CAP patients collected at the time of hospitalization. Specifically, we assessed whether patients with a confirmed infection of *Streptococcus pneumoniae*, atypical bacterial pathogens, or viral pathogens could be identified using metabolomics-based biomarkers. The choice for these pathogen groups was made because the identification of either one of these groups drives clinical treatment decision-making. The increased number of patients per group also enhanced the statistical power to detect potential biomarkers.

In our analyses, no metabolites were found to discriminate *Streptococcus pneumoniae* or viral pathogens from the other pathogen groups. However, patients with atypical CAP pathogens, such as *Coxiella burnetii*, *Chlamydophila psittaci*, *Legionella pneumophila*, and *Mycoplasma pneumoniae*, could be discriminated from patients with *Streptococcus pneumoniae* or viral infections using a predictive model. This model included three metabolites: glycyglycine, symmetric dimethylarginine, and lysophosphatidylinositol (18:1). The model showed a predictive performance of 63% sensitivity and 84% specificity. This performance is not superior to the predictive performance of established clinical assays based on culturing, antigen testing, or PCR [44, 45, 46, 47]. Potentially, the use of the identified metabolite-based biomarkers could, however, be further explored to complement established clinical assays, e.g., through combined consideration of both pathogen- and host-response associated characteristics.

Uncovering metabolic differences to CAP-associated pathogens

Whereas in Chapter 2 we focused on groups in pathogens, in **Chapter 3** we further characterized the differential metabolic host response to distinct CAP-associated pathogens. The goal of the analysis was to enhance our understanding of host (or patient)-associated metabolic changes in the pathogenesis of specific pathogens. To this end, we studied associations between metabolite levels and CAP-associated pathogens. While in chapter 2 we did not find metabolites that could discriminate patients with *S. pneumoniae* from all other patients, in chapter 3 we found that patients with *Streptococcus pneumoniae* infections showed high levels of the stress hormone cortisol, high phosphatidylcholines (PCs), and low lysophospholipids (LPCs) in comparison to all

other pathogens, all of which are associated to inflammation [92]. From these results, we can conclude that the pathogens within the viral and atypical pathogen group in chapter 2, are so different from each other, that we should study them individually.

This need for individual assessment of the different pathogens is confirmed by our findings in the group that we defined in chapter 2 as atypical pathogens. In patients with *Legionella pneumophila* infections, we found the lowest levels of LPCs, the highest levels of cortisol, kynurenine, and phenylalanine, and elevated free carnitine and short-chain acylcarnitines in comparison to the other pathogen groups, and which could be related to inflammation [95] and oxidative stress [100]. In patients with a *Coxiella burnetii* infection, high levels of long-chain acylcarnitines and LPCs, and elevated glutamate compared to other pathogen groups were observed, which can be linked to either inflammation, oxidative stress, or endothelial dysfunction [91, 81, 80, 101, 98].

Within the group of viral pathogens, we found a singular profile for the herpes simplex virus. Despite the small sample size, patients with herpes simplex infection showed a unique disruption in their lipid metabolism. An important limitation of this analysis was the limited sample size per group, the differences in patient age, and the substantial within-group variance of metabolite levels, which could have confounded the results. As such, we view these results as preliminary findings which require further confirmation, but which can still drive hypothesis generation and inform the design of follow-up studies. Also, a larger number of individuals per group would increase the power and change of finding significant differences.

Clinical utility of metabolomics-based biomarkers for etiological diagnosis of CAP

Based on our findings in **Chapters 2 and 3** we conclude that the added value of metabolomics-based blood-based assays for the etiological diagnosis of CAP may be limited. This is because the predictive performance for key pathogens in CAP is so far inferior or, at best similar to current diagnostic assays (**Chapter 2**), and with limitations in pathogen-specificity (**Chapter 3**). However, the information yielded in metabolomics assays could potentially be used to complement information gained from established clinical assays, i.e., through combination of diagnostic assays for host- and pathogen-associated characteristics and/or blood- and sputum-based matrices [144], such as recently has been explored in sepsis patients where combined host and pathogen data improved diagnostic sensitivity to 97-100% [145].

Alternative omics-based biomarkers

Over the last few years, other molecular profiling or “omics” technologies have been explored for the diagnosis of CAP. For example, recent studies suggest potential value of metagenomics next-generation sequencing [146], transcriptomics [147], and proteomics [148] approaches to further enhance the detection rate in pneumonia. Ultimately, we expect that diagnostic strategies which combine several of these technologies and thereby combining predictive biochemical host- and pathogen-associated features may

be most effective in the identification of predictive biomarkers for CAP, and which may be combined in combinatory diagnostic biomarker panels for CAP.

6.2 Monitoring of treatment response in CAP

Monitoring the course of disease progression and treatment response in CAP is essential to inform optimal clinical treatment strategies, including the use of effective antimicrobial therapies. To this end, patients are currently monitored based on their clinical symptoms, such as fever. In addition, inflammatory biomarkers in blood, such as C-reactive protein (CRP) or procalcitonin (PCT), are commonly measured longitudinally in patients to determine whether an antimicrobial treatment strategy appears to be effective and to guide when antimicrobial treatment can be terminated, which is relevant to control the risk of AMR emergence. However, current biomarkers have key limitations, i.e., even though CRP is commonly used, it is not specific to infection, and its kinetics are delayed in relationship to the underlying infection, and as such do not directly reflect the current state of the patient [12].

Longitudinal metabolic biomarkers for treatment response monitoring of CAP patients

In **Chapter 4**, we explored the potential of longitudinal metabolomics-based biomarkers for treatment response monitoring in CAP patients. To this end, we measured metabolite profiles in 25 CAP patients with a confirmed *S. pneumoniae* infection during treatment. We aimed to comprehensively characterize the change in longitudinal metabolite profiles of these patients and investigated associations with disease severity, inflammation markers, and treatment response outcomes, quantified using the length of hospital stay. We found that a large part of the variation in the metabolite values could be explained due to the changes over time within the patients. Several groups of metabolites were found to correlate with inflammation markers, CURB score, and length of hospital stay. The results showed that the inflammation marker C-reactive protein (CRP) correlated positively with phosphatidylcholine and negatively with nine LPCs. The CURB disease severity score was negatively associated with six metabolites, including three acylcarnitines. Length of stay correlated negatively with six triglycerides (TGs), and especially with TG (60:3) or TG (58:2). Since these TGs are not highly correlated to CRP, PCT, or the CURB score, they explain a part of the variability of the disease progression that is not captured by conventional treatment response biomarkers and are therefore of interest as new additional biomarkers associated with length of hospital stay.

To further evaluate the potential of TGs, LPCs, and PCs for treatment response monitoring, a prospective follow-up study is warranted to further characterize the predictive value of these biomarkers and their relationship with disease severity (CURB) and treatment outcomes such as length of stay (LOS). For such a study, the collection of additional frequent blood samples until hospital discharge would be optimal. In this study we have made a deliberate choice to only focus on *S. pneumoniae*-associated infections as the predominant CAP-associated pathogen, and to ensure that a controlled

analysis of the longitudinal profiles could be performed. However, in follow-up studies, assessment of the causative pathogens will be essential to determine whether the identified biomarkers are generalizable to other pathogens besides *S. pneumoniae*.

Beyond the analysis described in this chapter, only very limited research has so far been done investigating longitudinal, treatment response, metabolic biomarkers in CAP. In this context, previous studies investigating procalcitonin in CAP and/or sepsis are the most prominent [149, 108]. When considering the discovery of novel biomarker candidates, several recent studies in CAP patients have focused on prognostic biomarkers for treatment outcome or disease severity, such as studies for serum surviving and Cysteine-rich 61 [150, 151]. However, such prognostic biomarkers which have been determined only at the start of treatment cannot be directly applied as a clinical decision-making tool to adjust treatment strategy during the course of treatment.

6.2 Improving metabolomics study design and interpretation with the Immunometabolic Atlas

The biological interpretation of metabolomics data can be challenging because it is often unclear how certain metabolic biochemical changes associate with immunobiological or inflammatory processes. In addition, during the design of metabolomics studies, choices often need to be made regarding the specific chemical classes of metabolites to be measured, i.e., those which are likely associated with the (immuno-) biological process of interest. There are currently not many computational tools available to guide this association between known metabolic pathways and immune-biological processes.

In **Chapter 5**, we describe the development of the Immunometabolic Atlas (IMA), which can support the interpretation and design of metabolomics studies. In the IMA, we integrated information on metabolites, metabolite-protein interactions and immunobiological processes available in large-scale public databases. Based on these databases, we established a metabolite-immune process interaction network with over 1.4 million metabolite-immune process associations. Through a web interface, this network can be used to infer immunobiological processes from metabolites and vice versa. A current limitation for such analyses remains the lack of uniform metabolite identifiers for many metabolite classes. For example, for 60% of the metabolites that were measured as part of this thesis, a Human Metabolite Database (HMDB) identifier was unavailable. This was in particular these case for lipids, where existing efforts to provide identifiers for lipids [152] should be integrated with platforms such as HMDB. Implementing and integration of such identifiers is crucial to further enable FAIR (re-)use of metabolomics data. Recently, recommendations for metabolite annotations have been made to overcome the inconsistencies in metabolite nomenclature [153]. Furthermore, a new methodology is developed to discover new protein-metabolite interactions that enhances the biological interpretation of metabolomics data [154].

6.3 Discovery and application of clinical metabolomics-based biomarkers

High-quality clinical data

The analyses in this thesis have been performed using patient samples which were collected during previously conducted clinical studies involving hospitalized patients with moderate to severe CAP, which required hospitalization [30, 29]. An important characteristic of these clinical studies was the availability of a confirmed microbial diagnosis, as well as detailed information on patient characteristics, disease severity, antibiotic therapy, inflammation markers measured during treatment, and length of stay in the hospital. As such, these clinical studies provided a unique and relevant set of samples to study the role of metabolomics-based biomarkers in CAP.

Metadata and cofounders

This thesis confirms the importance of complete and extensive data on patient characteristics, treatments administered, and comorbidities, which may all influence metabolite profiles and, therefore, could affect the ability to successfully detect novel biomarkers and should be included to correct for possible confounders in metabolomics data analysis. In our analyses, details such as sex, age, disease severity, comorbidities, and durations of symptoms before admission were already collected as part of the study design. However, several other factors were incomplete or lacking, such as saturation, fever, respiratory rate, food/fasting details, antibiotic dose, and the specific time of drawing a blood sample, and can influence the metabolic profile [155, 156, 157]. In this context, and because of the complexity of design and execution of prospective clinical trials, the role and further establishment of biobanks for patients with CAP or associated respiratory infections with extensive collection of all relevant metadata will be an essential step to aid in future biomarker discovery studies. This is especially a challenge when the number per patient group is as small as in this thesis.

Control groups and longitudinal data

In this thesis we did not incorporate samples from controls, i.e., individuals unaffected by an infection. This has impacted our ability to compare observed metabolite levels in patients with infections to metabolite levels in healthy individuals or in hospitalized patients without active infection. When considering variability in metabolite levels within patients, we did have samples available at 30 days post-admission for a subset of included patients, which can be considered as a within-patient controls. Nonetheless, establishing metabolome-wide large scale public data repositories for such controls could be helpful in future metabolomics-based biomarker discovery studies. However, this would require that metabolomics data can be compared between studies.

In this thesis we have shown that longitudinal metabolite profiles can give more information about the state of the patient and elucidate the effect of comorbidities and

co-medications. Therefore, the design of a new clinical study should include sampling for metabolomics at more than one time point. Since patients are admitted to the hospital in different stages of the disease, diagnostic testing based on the absolute metabolite concentrations at one time point can be challenging. The variable baseline metabolite levels upon admission and the dynamical metabolite profile differs per patient and complicates the development and analysis of longitudinal and diagnostic biomarkers and emphasizes the need for longitudinal biomarkers.

Absolute metabolite levels

In this thesis, metabolite levels were measured as metabolite peak ratio to its internal standard, as is common practice in the metabolomics field. However, for clinical application, metabolite levels should be measured as absolute concentrations so a threshold can be determined that can be easily interpreted by a physician. Measurement of absolute concentrations was not yet feasible for the hundreds of metabolites that were measured in the research in this thesis, as this would require calibration lines for all metabolites. However, if a subset of five to ten metabolites was measured, i.e., in the context of a dedicated clinical metabolomics-derived assay, this would be an achievable and important step. In addition, methods should be developed so that also 100's of metabolites can be reported as absolute concentrations.

Clinical metabolomics assay development

To measure a small number of metabolites with different chemical properties, such as polarity and charge, new analytical methodologies should be developed. Recent research has shown it to be possible to integrate the measurement of metabolites with different chemistries in one methodological platform, and this methodology should be developed further for clinical use [158]. In conclusion, to obtain the integration of metabolomics-based assays in the clinical laboratory, the number of measured metabolites should be minimized to allow fast processing of samples and easy interpretation of the results for the clinician. On the other hand, if a larger number of metabolites are quantified several diseases could be diagnosed (or excluded) in one analysis.

6.4 Conclusion

We have shown that metabolomics-based biomarkers have potential to monitor the treatment response in CAP patients. An important advantage of such biomarkers is their reflection of the host response to infection. Pathogen-specific metabolic responses described may be able to support discrimination between individual pathogens or pathogen groups based on the host response but should be further researched to explore their potential. Ultimately, metabolomics-based biomarkers could in the future be relevant to complement existing diagnostic tools and biomarkers for diagnosis and

treatment monitoring of CAP. For that larger studies that follow the above given recommendations are required.

CHAPTER 7

Nederlandse samenvatting

Community-acquired pneumonia (CAP), ofwel een thuis opgelopen longontsteking, is een veel voorkomende infectie van de onderste luchtwegen. CAP kan zowel door een virus als een bacteriële ziekteverwekker veroorzaakt worden. CAP patiënten met een ernstig ziekteverloop worden doorgaans in het ziekenhuis opgenomen. In dat geval krijgen ze meestal initieel een breed-spectrum antibioticum voorgeschreven. Tegelijkertijd wordt er met behulp van microbiologisch onderzoek geprobeerd de specifieke ziekteverwekker te identificeren zodat met een gericht, smaller spectrum antibioticum uitbehandeld kan worden. Als de oorzaak viraal blijkt te zijn, kan overwogen worden om de antibioticatherapie te stoppen. Huidige diagnostische tests voor bacteriële ziekteverwekkers hebben vaak een relatief lange doorlooptijd (>48 uur) waardoor het lang duurt voordat de therapie op de patient aangepast kan worden. Bovendien blijft in ongeveer de helft van de CAP-patiënten de ziekteverwekker onbekend. Nadat de antibioticatherapie is gestart, is het van belang om vroegtijdig te kunnen meten of de behandeling aanslaat en wanneer deze weer gestaakt kan worden. Gebruik van antibiotica is direct gerelateerd aan het risico op het ontstaan van antimicrobiële resistentie (AMR). Het is daarom van groot belang om antibiotica zo gericht mogelijk te gebruiken en te staken als de oorzaak viraal is of als de bacteriële infectie volledig is onderdrukt. Snelle en voorspellende diagnostische technieken zijn essentieel, zowel voor effectieve antibioticatherapie als om het risico op het ontstaan van AMR te beperken.

Metabolomics is de grootschalige meting van kleine biomoleculen (metabolieten) in weefsels, cellen of lichaamsvloeistoffen, zoals bloed. Deze lichaamseigen metabolieten spelen een rol in de regulering van allerlei processen in het lichaam, waaronder bij de reactie van het lichaam op een infectie. In dit proefschrift hebben wij onderzocht of er (combinaties van) metabolieten bestaan die specifiek gerelateerd kunnen worden aan patiënten met een CAP infectie. Hierbij hebben wij gekeken naar relaties van metabolietprofielen met verschillende soorten ziekteverwekkers, met de ernst van de ziekte en met het ziekteverloop. Onze specifieke focus lag op het bepalen of er klinische meerwaarde zou kunnen bestaan voor het gebruik van dergelijke nieuwe metabolomics-gebaseerde biomarkers in de diagnostiek en behandeling van CAP.

In **Hoofdstuk 2** hebben we onderzocht of bepaalde metabolieten onderscheid kunnen maken tussen verschillende groepen CAP-patiënten, afhankelijk van de ziekteverwekker. Hierbij hebben wij gekeken naar metabolietprofielen in bloedmonsters van CAP-patiënten die waren opgenomen in het ziekenhuis. We hebben patiënten met een bacteriële *Streptococcus pneumoniae* (pneumokok) infectie, een atypische bacteriële infectie, of een virale infectie, met elkaar vergeleken. We kozen deze drie groepen omdat de antibiotische behandeling per groep verschilt. Wij vonden dat het niet goed mogelijk is om pneumokokken of virale infecties te onderscheiden op basis van de onderzochte metabolietprofielen. Wel bleek het mogelijk om met een set van drie metabolieten atypische bacteriële infecties te onderscheiden zijn van andere groepen. De gevoeligheid (63%) en specificiteit (84%) van deze metabolieten, de voorspellende waarde, bleek echter niet beter dan deze van bestaande diagnostische tests zoals kweek, antigeentesten of PCR. Desondanks zou het kunnen dat deze metabolieten in aanvulling op bestaande technieken de identificatie van het type ziekteverwekker kunnen verbeteren, omdat deze specifiek kijkt naar de metabole respons van een patiënt.

In **Hoofdstuk 3** hebben we onderzocht of er met behulp van metabolietprofielen onderscheid kan worden gemaakt tussen individuele ziekteverwekkers. We hebben metabolietprofielen geanalyseerd in bloedmonsters van CAP-patiënten met als doel ons begrip van metabole veranderingen bij patiënten in relatie tot specifieke ziekteverwekkers te vergroten. In deze analyse vonden we met name in patiënten met een pneumokokkeninfectie, *Legionella pneumophila* en *Coxiella burnetii* infecties een sterkere associatie met specifieke metabolieten. Hierbij bleken met name de niveaus van fosfatidylcholines, lysofosfolipiden, cortisol, kynurenine, fenylalanine, vrije carnitine en korte-keten-acylcarnitines een onderscheidende rol te spelen. Deze metabolieten zijn deels eerder al in verband gebracht met ontstekingsprocessen, oxidatieve stress, en/of endotheelcelfunctie. Bij patiënten met het herpes simplex-virus vonden we daarnaast een unieke verstoring in het lipidenmetabolisme. Het aantal patiënten per ziekteverwekker was beperkt in deze studie en vormt een belangrijke beperking. Bovendien waren er aanzienlijke verschillen in de leeftijd van de patiënten en in variatie van metabolietniveaus binnen de groepen ziekteverwekkers. De gevonden associaties tussen specifieke metabolieten en ziekteverwekkers moeten dus als exploratief worden beschouwd, en kunnen mogelijk als basis van hypothesevorming in vervolgstudies gebruikt worden.

Het kunnen monitoren van het ziekteverloop en de behandelrespons is essentieel voor een optimale behandelstrategie. In **Hoofdstuk 4** hebben wij gekeken naar veranderingen in metabolietprofielen gedurende het ziekteverloop van CAP-patiënten. Momenteel gebeurt dit door het meten van klinische symptomen zoals koorts en door het bepalen van ontstekingsmarkers in het bloed gedurende de behandeling, zoals C-reactief proteïne (CRP). Een biomarker zoals CRP is niet specifiek voor de infectie en reageert relatief traag op veranderingen in de ernst van de infectie, waardoor ze de actuele toestand van de patiënt niet nauwkeurig weerspiegelen. In dit hoofdstuk hebben we onderzocht of er metabolieten zijn die mogelijk kunnen bijdragen aan het monitoren van de behandelingseffecten bij CAP-patiënten. We hebben hiervoor gekeken naar

metabolietenprofielen van 25 CAP-patiënten met een streptokokkeninfectie, waarbij wij op verschillende momenten tijdens en na de behandeling de metabolietprofielen in bloedmonsters hebben bepaald. We hebben onderzocht of veranderingen in deze over de tijd variërende metabolietprofielen gerelateerd kunnen worden aan scores voor de ernst van ziekte (de CURB score), andere ontstekingsmarkers, en aan de opnameduur in het ziekenhuis. Wij vonden verschillende groepen metabolieten die een correlatie met ontstekingsmarkers, ernst van de ziekte en/of opnameduur lieten zien. Met name een aantal triglyceriden bleken mogelijk relevant als longitudinale biomarker voor het bepalen van behandelrespons.

Op dit moment is er nog relatief weinig bekend over de rol en functie van metabolieten in de immuunrespons bij infecties. Het kiezen van de meest relevante groepen van metabolieten in een metabolomics studie, en de biologische interpretatie van gevonden metabolieten in metabolomics studies, is daarom ingewikkeld. In **Hoofdstuk 5** hebben wij een computationele tool ontwikkeld om deze uitdagingen te ondersteunen: de Immunometabolic Atlas (IMA). De IMA werkt door middel van het combineren van verschillende databases met informatie over metabolieten, eiwitten die aan deze metabolieten gerelateerd zijn en eiwitten die gerelateerd zijn aan biologische processen. Door de kennis van onderlinge connecties tussen eiwitten en metabolieten te benutten, wordt het mogelijk om (mogelijke) verbanden tussen metabolieten en immuno-biologische processen te identificeren. Het uiteindelijk ontwikkelde interactienetwerk tussen metabolieten en eiwitten had meer dan 1.4 miljoen van dergelijke verbindingen. Met behulp van de publiek beschikbare IMA kunnen onderzoekers in het vakgebied van immunometabolisme geholpen worden in zowel het opzetten van nieuwe metabolomics studies als de analyse daarvan.

Samenvattend hebben wij in dit proefschrift laten zien dat metabolietprofielen potentie hebben als mogelijke biomarkers bij CAP, met name voor het monitoren van behandelrespons. Hoewel de gevonden metabolietprofielen mogelijk voorspellende waarde hebben bij de behandeling van CAP, is verder onderzoek in grotere patiëntcohorten nodig om hun potentiële rol verder te kunnen vaststellen. Een belangrijk voordeel van metaboliet-gebaseerde biomarkers is dat zij de reactie van de patiënt op een infectie weerspiegelen, in plaats van de eigenschappen van de ziekteverwekker zelf. Mogelijk zijn metabolietgebaseerde biomarkers met name in aanvulling op bestaande diagnostische technieken van toegevoegde waarde bij de behandeling van CAP.

Bibliography

1. Bjarnason, A. *et al.* Incidence, etiology, and outcomes of community-acquired pneumonia: A population-based study. *Open Forum Infectious Diseases* **5** (2018).
2. Jain, S. *et al.* Community-Acquired Pneumonia Requiring Hospitalization among U.S. Adults. *New England Journal of Medicine* **373**, 415–427 (2015).
3. Prina, E., Ranzani, O. T. & Torres, A. Community-acquired pneumonia. *The Lancet* **386**, 1097–1108 (2015).
4. Kothe, H. *et al.* Outcome of community-acquired pneumonia: Influence of age, residence status and antimicrobial treatment. *European Respiratory Journal* **32**, 139–146 (2008).
5. Angus, D. C. *et al.* Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. *Critical care medicine* **29**, 1303–1310 (2001).
6. Wiersinga, W. J. *et al.* Dutch Guidelines on the Management of Community-Acquired Pneumonia in Adults. *SWAB* (2011).
7. Aliberti, S., Dela Cruz, C. S., Amati, F., Sotgiu, G. & Restrepo, M. I. Community-acquired pneumonia. *The Lancet* **398**, 906–919 (2021).
8. Wiersinga, W. J. *et al.* Management of Community-Acquired Pneumonia in Adults: 2016 Guideline Update From The Dutch Working Party on Antibiotic Policy (SWAB) and Dutch Association of Chest Physicians (NVALT). *The Netherlands journal of medicine* **76** (2016).
9. Lee, J. S., Giesler, D. L., Gellad, W. F. & Fine, M. J. Antibiotic Therapy for Adults Hospitalized With Community-Acquired Pneumonia: A Systematic Review. *JAMA* **315**, 593–602 (2016).
10. Avdic, E. *et al.* Impact of an Antimicrobial Stewardship Intervention on Shortening the Duration of Therapy for Community-Acquired Pneumonia. *Clinical Infectious Diseases* **54**, 1581–1587 (2012).
11. Viasus, D., Vecino-Moreno, M., De La Hoz, J. M. & Carratalà, J. Antibiotic stewardship in community-acquired pneumonia. *Expert Review of Anti-infective Therapy* **15**, 351–359 (2017).
12. Aulin, L. B. *et al.* Biomarker-Guided Individualization of Antibiotic Therapy. *Clinical Pharmacology & Therapeutics* **110**, 346–360 (2021).
13. Karakioulaki, M. & Stolz, D. Biomarkers in Pneumonia—Beyond Procalcitonin. *International Journal of Molecular Sciences* **20** (2019).

-
14. Kohler, I. *et al.* Integrating clinical metabolomics-based biomarker discovery and clinical pharmacology to enable precision medicine. *European Journal of Pharmaceutical Sciences* **109**, S15–S21 (2017).
 15. Gao, P. & Xu, G. Mass-spectrometry-based microbial metabolomics: recent developments and applications. *Anal Bioanal Chem* **407**, 669–680 (2015).
 16. Orešič, M., Vidal-Puig, A. & Hänninen, V. Metabolomic approaches to phenotype characterization and applications to complex diseases. *Expert Review of Molecular Diagnostics* **6**, 575–585 (2006).
 17. Tounta, V., Liu, Y., Cheyne, A. & Larrouy-Maumus, G. Metabolomics in infectious diseases and drug discovery. *Molecular Omics* **17**, 376–393 (2021).
 18. Aretz, I. & Meierhofer, D. Advantages and Pitfalls of Mass Spectrometry Based Metabolome Profiling in Systems Biology. *International Journal of Molecular Sciences* **17** (2016).
 19. Bingol, K. Recent advances in targeted and untargeted metabolomics by NMR and MS/NMR methods. *High-Throughput* **7**, 9 (2018).
 20. Lau, S. K. P. *et al.* Metabolomic Profiling of Plasma from Patients with Tuberculosis by Use of Untargeted Mass Spectrometry Reveals Novel Biomarkers for Diagnosis. *Journal of clinical microbiology* **53** (ed Land, G. A.) 3750–9 (2015).
 21. Kauppi, A. M. *et al.* Metabolites in Blood for Prediction of Bacteremic Sepsis in the Emergency Room. *PLOS ONE* **11** (2016).
 22. Banoei, M. M. *et al.* Plasma metabolomics for the diagnosis and prognosis of H1N1 influenza pneumonia. *Critical care* **21**, 97 (2017).
 23. Slupsky, C. M. *et al.* Pneumococcal Pneumonia: Potential for Diagnosis through a Urinary Metabolic Profile. *Journal of Proteome Research* **8**, 5550–5558 (2009).
 24. Schoeman, J. C. *et al.* Metabolic characterization of the natural progression of chronic hepatitis B. *Genome Medicine* **8**, 64 (2016).
 25. Müller, D. C. *et al.* Phospholipid levels in blood during community-acquired pneumonia. *PLOS ONE* **14** (2019).
 26. Chong, J., Wishart, D. S. & Xia, J. Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. *Current Protocols in Bioinformatics* **68** (2019).
 27. Kanehisa, M. & Goto, S. *KEGG: Kyoto Encyclopedia of Genes and Genomes* 2000.
 28. Szklarczyk, D. *et al.* STITCH 5: Augmenting protein-chemical interaction networks with tissue and affinity data. *Nucleic Acids Research* **44**, D380–D384 (2016).
 29. Meijvis, S. C. A. *et al.* Dexamethasone and length of hospital stay in patients with community-acquired pneumonia: A randomised, double-blind, placebo-controlled trial. *The Lancet* **377**, 2023–2030 (2011).

30. Endeman, H. *et al.* Clinical features predicting failure of pathogen identification in patients with community acquired pneumonia. *Scandinavian Journal of Infectious Diseases* **40**, 715–720 (2008).
31. Wunderink, R. G. & Waterer, G. W. Community-Acquired Pneumonia. *New England Journal of Medicine* **370**, 543–551 (2014).
32. Postma, D. F. *et al.* Antibiotic Treatment Strategies for Community-Acquired Pneumonia in Adults. *New England Journal of Medicine* **372**, 1312–1323 (2015).
33. WHO. *Antimicrobial resistance: global report on surveillance* tech. rep. (2014).
34. Saleh, M. A. A., Van De Garde, E. M. W. & Van Hasselt, C. J. G. Host-response biomarkers for the diagnosis of bacterial respiratory tract infections. *Clinical Chemistry and Laboratory Medicine* **57** (2019).
35. Pearce, E. L. & Pearce, E. J. Metabolic pathways in immune cell activation and quiescence. *Immunity* **38**, 633–643 (2013).
36. Khovidhunkit, W. *et al.* Effects of infection and inflammation on lipid and lipoprotein metabolism: Mechanisms and consequences to the host. *Journal of Lipid Research* **45**, 1169–1196 (2004).
37. Zhou, A. *et al.* Metabolomics specificity of tuberculosis plasma revealed by (1)H NMR spectroscopy. *Tuberculosis* **95**, 294–302 (2015).
38. Laiakis, E. C., Morris, G. A. J., Fornace, A. J., Howie, S. R. C. & Howie, S. R. C. Metabolomic Analysis in Severe Childhood Pneumonia in The Gambia, West Africa: Findings from a Pilot Study. *PloS one* **5** (2010).
39. Antcliffe, D., Jiménez, B., Veselkov, K., Holmes, E. & Gordon, A. C. Metabolic Profiling in Patients with Pneumonia on Intensive Care. *EBioMedicine* **18**, 244–253 (2017).
40. Adamko, D. J., Saude, E., Bear, M., Regush, S. & Robinson, J. L. Urine metabolomic profiling of children with respiratory tract infections in the emergency department: a pilot study. *BMC infectious diseases* **16**, 439 (2016).
41. Varma, S. & Simon, R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics* **7**, 91 (2006).
42. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**, 301–320 (2005).
43. Statnikov, A., Tsamardinos, I., Dosbayev, Y. & Aliferis, C. F. GEMS: A system for automated cancer diagnosis and biomarker discovery from microarray gene expression data. *International Journal of Medical Informatics* **74**, 491–503 (2005).
44. Yzerman, E. P. *et al.* Sensitivity of three urinary antigen tests associated with clinical severity in a large outbreak of Legionnaires' disease in the Netherlands. *Journal of Clinical Microbiology* **40**, 3232–3236 (2002).

-
45. Gutierrez, F. *et al.* Evaluation of the Immunochromatographic Binax NOW Assay for Detection of *Streptococcus pneumoniae* Urinary Antigen in a Prospective Study of Community-Acquired Pneumonia in Spain. *Clinical Infectious Diseases* **36**, 286–292 (2003).
 46. Sordé, R. *et al.* Current and potential usefulness of pneumococcal urinary antigen detection in hospitalized patients with community-acquired pneumonia to guide antimicrobial therapy. *Archives of Internal Medicine* **171**, 166–172 (2011).
 47. Van Elden, L. J., Nijhuis, M., Schipper, P., Schuurman, R. & Van Loon, A. M. Simultaneous detection of influenza viruses A and B using real-time quantitative PCR. *Journal of Clinical Microbiology* **39**, 196–200 (2001).
 48. Emwas, A. H. *et al.* Nmr spectroscopy for metabolomics research. *Metabolites* **9** (2019).
 49. Ning, P. *et al.* Metabolic profiles in community-acquired pneumonia: developing assessment tools for disease severity. *Critical Care* **22**, 130 (2018).
 50. Noga, M. J. *et al.* Metabolomics of cerebrospinal fluid reveals changes in the central nervous system metabolism in a rat model of multiple sclerosis. *Metabolomics* **8**, 253–263 (2012).
 51. Torres, A. *et al.* Pneumonia. *Nature Reviews Disease Primers* **7**, 1–28 (2021).
 52. Falguera, M. *et al.* Predictive factors, microbiology and outcome of patients with parapneumonic effusion. *European Respiratory Journal* **38**, 1173–1179 (2011).
 53. Sureja, B. R., Govani, K. J., Sureja, J. B. & Sureja, R. B. An Assessment of Clinical Profile & Outcome in Patients with Community Acquired Pneumonia (CAP). *National Journal of Community Medicine* **12**, 221–224 (2021).
 54. Kok, M., Maton, L., van der Peet, M., Hankemeier, T. & van Hasselt, J. G. Unraveling antimicrobial resistance using metabolomics. *Drug discovery today* **27**, 1774–1783 (2022).
 55. Páez-Franco, J. C. *et al.* Metabolomics analysis reveals a modified amino acid metabolism that correlates with altered oxygen homeostasis in COVID-19 patients. *Scientific reports* **11**, 1–12 (2021).
 56. Overmyer, K. A. *et al.* Large-Scale Multi-omic Analysis of COVID-19 Severity. *Cell Systems* **12**, 23–40 (2021).
 57. Ploder, M. *et al.* Serum phenylalanine in patients post trauma and with sepsis correlate to neopterin concentrations. *Amino acids* **35**, 303–307 (2008).
 58. Wang, J., Sun, Y., Teng, S. & Li, K. Prediction of sepsis mortality using metabolite biomarkers in the blood: a meta-analysis of death-related pathways and prospective validation. *BMC Medicine* **18**, 83 (2020).
 59. Den Hartog, I. *et al.* Metabolomic profiling of microbial disease etiology in community-acquired pneumonia. *PloS one* **16** (2021).
 60. Marcovina, S. M. *et al.* Translating the basic knowledge of mitochondrial functions to metabolic therapy: Role of L-carnitine. *Translational Research* **161** (2013).

61. Hoppel, C. *The role of carnitine in normal and altered fatty acid metabolism* 2003.
62. Langley, R. J. *et al.* Integrative "omic" analysis of experimental bacteremia identifies a metabolic signature that distinguishes human sepsis from systemic inflammatory response syndromes. *American Journal of Respiratory and Critical Care Medicine* **190**, 445–455 (2014).
63. Ayres, J. S. A metabolic handbook for the COVID-19 pandemic. *Nature Metabolism* **2**, 572–585 (2020).
64. Otsubo, C. *et al.* Long-chain Acylcarnitines Reduce Lung Function by Inhibiting Pulmonary Surfactant. *Journal of Biological Chemistry* **290**, 23897–23904 (2015).
65. Rousseau, M. *et al.* Associations Between Dietary Protein Sources, Plasma BCAA and Short-Chain Acylcarnitine Levels in Adults. *Nutrients* **11**, 173 (2019).
66. Moreno, F. A., Macey, H. & Schreiber, B. Carnitine Levels in Valproic Acid-Treated Psychiatric Patients: A Cross-Sectional Study. *The Journal of Clinical Psychiatry* **66**, 2314 (2005).
67. Mynatt, R. L. Carnitine and type 2 diabetes. *Diabetes/Metabolism Research and Reviews* **25**, S45–S49 (2009).
68. Darcy, C. J. *et al.* An observational cohort study of the kynurenine to tryptophan ratio in sepsis: association with impaired immune and microvascular function. *PloS one* **6** (2011).
69. Suzuki, Y. *et al.* Serum activity of indoleamine 2,3-dioxygenase predicts prognosis of community-acquired pneumonia. *Journal of Infection* **63** (2011).
70. Arshad, H. *et al.* Decreased plasma phospholipid concentrations and increased acid sphingomyelinase activity are accurate biomarkers for community-acquired pneumonia. *Journal of Translational Medicine* **17**, 1–18 (2019).
71. Lawler, N. G. *et al.* Systemic Perturbations in Amine and Kynurenine Metabolism Associated with Acute SARS-CoV-2 Infection and Inflammatory Cytokine Responses. *Journal of proteome research* (2021).
72. Takikawa, O. Biochemical and medical aspects of the indoleamine 2, 3-dioxygenase-initiated L-tryptophan metabolism. *Biochemical and biophysical research communications* **338**, 12–19 (2005).
73. Orabona, C., Pallotta, M. T. & Grohmann, U. Different partners, opposite outcomes: a new perspective of the immunobiology of indoleamine 2, 3-dioxygenase. *Molecular medicine* **18**, 834–842 (2012).
74. Jones, S. P. *et al.* Expression of the Kynurenine Pathway in Human Peripheral Blood Mononuclear Cells: Implications for Inflammatory and Neurodegenerative Disease. *PloS one* **10** (2015).
75. Mezrich, J. D. *et al.* An interaction between kynurenine and the aryl hydrocarbon receptor can generate regulatory T cells. *The Journal of Immunology* **185**, 3190–3198 (2010).

-
76. Changsirivathanathamrong, D. *et al.* Tryptophan metabolism to kynurenine is a potential novel contributor to hypotension in human sepsis. *Critical care medicine* **39**, 2678–2683 (2011).
 77. Leisman, D. E., Deutschman, C. S. & Legrand, M. Facing COVID-19 in the ICU: vascular dysfunction, thrombosis, and dysregulated inflammation. *Intensive Care Medicine* **46**, 1105–1108 (2020).
 78. Neurauter, G. *et al.* Chronic immune stimulation correlates with reduced phenylalanine turnover. *Current drug metabolism* **9**, 622–627 (2008).
 79. Pacheco, R., Gallart, T., Lluís, C. & Franco, R. Role of glutamate on T-cell mediated immunity. *Journal of Neuroimmunology* **185**, 9–19 (2007).
 80. Nassar, T. *et al.* tPA regulates pulmonary vascular activity through NMDA receptors. *American Journal of Physiology - Lung Cellular and Molecular Physiology* **301**, 307–314 (2011).
 81. Collard, C. D. *et al.* Neutrophil-derived Glutamate Regulates Vascular Endothelial Barrier Function. *Journal of Biological Chemistry* **277**, 14801–14811 (2002).
 82. Baker, D. A., Thomas, J., Epstein, J., Possilico, D. & Stone, M. L. The effect of prostaglandins on the multiplication and cell-to-cell spread of herpes simplex virus type 2 in vitro. *American Journal of Obstetrics and Gynecology* **144** (1982).
 83. Dalli, J. *et al.* Human sepsis eicosanoid and pro-resolving lipid mediator temporal profiles: correlations with survival and clinical outcomes. *Critical care medicine* **45**, 58 (2017).
 84. Bruegel, M. *et al.* Sepsis-associated changes of the arachidonic acid metabolism and their diagnostic potential in septic patients. *Critical Care Medicine* **40**, 1478–1486 (2012).
 85. Schwarz, B. *et al.* Severe SARS-CoV-2 infection in humans is defined by a shift in the serum lipidome resulting in dysregulation of eicosanoid immune mediators. *Journal of immunology* **206**, 329 (2021).
 86. Lundström, S. L. *et al.* Asthmatics Exhibit Altered Oxylin Profiles Compared to Healthy Individuals after Subway Air Exposure. *PLOS ONE* **6**, e23864 (2011).
 87. Balgoma, D. *et al.* Linoleic acid-derived lipid mediators increase in a female-dominated subphenotype of COPD. *European Respiratory Journal* **47**, 1645–1656 (2016).
 88. Dennis, E. A. & Norris, P. C. Eicosanoid storm in infection and inflammation. *Nature Reviews Immunology* **15**, 511–523 (2015).
 89. Malcher-Lopes, R., Franco, A. & Tasker, J. G. Glucocorticoids shift arachidonic acid metabolism toward endocannabinoid synthesis: A non-genomic anti-inflammatory switch. *European Journal of Pharmacology* **583**, 322–339 (2008).
 90. Chen, J. S. *et al.* Nonsteroidal Anti-inflammatory Drugs Dampen the Cytokine and Antibody Response to SARS-CoV-2 Infection. *Journal of Virology* **95** (2021).

91. Nickler, M. *et al.* Systematic review regarding metabolic profiling for improved pathophysiological understanding of disease and outcome prediction in respiratory infections. *Respiratory Research* **16**, 125 (2015).
92. Hurley, B. P. & McCormick, B. A. Multiple roles of phospholipase A2 during lung infection and inflammation. *Infection and Immunity* **76**, 2259–2272 (2008).
93. Mecatti, G. C. *et al.* Lipidomic Profiling of Plasma and Erythrocytes From Septic Patients Reveals Potential Biomarker Candidates. *Biomarker insights* **13** (2018).
94. Vromman, F. & Subtil, A. Exploitation of host lipids by bacteria. *Current Opinion in Microbiology* **17**, 38–45 (2014).
95. Flieger, A. *et al.* Phospholipase A secreted by *Legionella pneumophila* destroys alveolar surfactant phospholipids. *FEMS Microbiology Letters* **188**, 129–133 (2000).
96. Gregoriades, A. Interaction of influenza M protein with viral lipid and phosphatidylcholine vesicles. *Journal of Virology* **36**, 470–479 (1980).
97. Kawasaki, K. & ichi Ohnishi, S. Membrane fusion of influenza virus with phosphatidylcholine liposomes containing viral receptors. *Biochemical and Biophysical Research Communications* **186**, 378–384 (1992).
98. Vijay, R. *et al.* Critical role of phospholipase A2 group IID in age-related susceptibility to severe acute respiratory syndrome–CoV infection. *The Journal of Experimental Medicine* **212**, 1851 (2015).
99. Müller, C. *et al.* Inhibition of Cytosolic Phospholipase A2 α Impairs an Early Step of Coronavirus Replication in Cell Culture. *Journal of Virology* **92**, 1463–1480 (2018).
100. Houten, S. M. & Wanders, R. J. A general introduction to the biochemistry of mitochondrial fatty acid β -oxidation. *Journal of Inherited Metabolic Disease* **33**, 469–477 (2010).
101. Newsholme, P. Why Is L-Glutamine Metabolism Important to Cells of the Immune System in Health, Postinjury, Surgery or Infection? *The Journal of Nutrition* **131**, 2515–2522 (2001).
102. Lawton, K. A. *et al.* Analysis of the adult human plasma metabolome. *Pharmacogenomics* **9**, 383–397 (2008).
103. Maas, P. *et al.* The Immunometabolic Atlas: A tool for design and interpretation of metabolomics studies in immunology. *PLOS ONE* **17** (2022).
104. Battleman, D. S., Callahan, M. & Thaler, H. T. Rapid Antibiotic Delivery and Appropriate Antibiotic Selection Reduce Length of Hospital Stay of Patients With Community-Acquired Pneumonia: Link Between Quality of Care and Resource Utilization. *Archives of Internal Medicine* **162**, 682–688 (2002).
105. Marco-Ramell, A. *et al.* Evaluation and comparison of bioinformatic tools for the enrichment analysis of metabolomics data. *BMC Bioinformatics* **19**, 1 (2018).
106. Pletz, M. W. *et al.* Unmet needs in pneumonia research: a comprehensive approach by the CAPNETZ study group. *Respiratory Research* **23**, 1–15 (2022).

-
107. Guo, S., Mao, X. & Liang, M. The moderate predictive value of serial serum CRP and PCT levels for the prognosis of hospitalized community-acquired pneumonia. *Respiratory Research* **19**, 1–9 (2018).
 108. De Jong, E. *et al.* Efficacy and safety of procalcitonin guidance in reducing the duration of antibiotic treatment in critically ill patients: A randomised, controlled, open-label trial. *The Lancet Infectious Diseases* **16**, 819–827 (2016).
 109. Seymour, C. W. *et al.* Metabolomics in pneumonia and sepsis: An analysis of the GenIMS cohort study. *Intensive Care Medicine* **39**, 1423–1434 (2013).
 110. Breier, M. *et al.* Targeted metabolomics identifies reliable and stable metabolites in human serum and plasma samples. *PLoS ONE* **9** (2014).
 111. Goodman, K. *et al.* Assessment of the effects of repeated freeze thawing and extended bench top processing of plasma samples using untargeted metabolomics. *Metabolomics* **17** (2021).
 112. Neill, A. M. *et al.* Community acquired pneumonia: aetiology and usefulness of severity criteria on admission. *Thorax* **51**, 1010–1016 (1996).
 113. Van Der Ham, T., Meulman, J. J., Van Strien, D. C. & Van Engeland, H. Empirically based subgrouping of eating disorders in adolescents: A longitudinal perspective. *British Journal of Psychiatry* **170**, 363–368 (1997).
 114. Josse, J., Pagès, J. & Husson, F. Multiple imputation in principal component analysis. *Advances in Data Analysis and Classification* **5**, 231–246 (2011).
 115. Parks, E. J. Effect of Dietary Carbohydrate on Triglyceride Metabolism in Humans. *The Journal of Nutrition* **131**, 2772–2774 (2001).
 116. Banoei, M. M. *et al.* Plasma lipid profiling for the prognosis of 90-day mortality, in-hospital mortality, ICU admission, and severity in bacterial community-acquired pneumonia (CAP). *Critical Care* **24** (2020).
 117. Colombo, S. *et al.* Phospholipidome of endothelial cells shows a different adaptation response upon oxidative, glycativ and lipoxidative stress. *Scientific Reports* **8** (2018).
 118. Nan, W. *et al.* Myristoyl lysophosphatidylcholine is a biomarker and potential therapeutic target for community-acquired pneumonia. *Redox Biology* **58**, 102556 (2022).
 119. Wu, G. Amino acids: Metabolism, functions, and nutrition. *Amino Acids* **37**, 1–17 (2009).
 120. O'Neill, L. A., Kishton, R. J. & Rathmell, J. A guide to immunometabolism for immunologists. *Nature Reviews Immunology* **16**, 553–565 (2016).
 121. Cheng, S.-C. C., Joosten, L. A. B. & Netea, M. G. The interplay between central metabolism and innate immune responses. *Cytokine and Growth Factor Reviews* **25**, 707–713 (2014).
 122. Ashburner, M. *et al.* Gene ontology: Tool for the unification of biology. *Nature Genetics* **25**, 25–29 (2000).

123. Gene Ontology Consortium. The Gene Ontology resource: enriching a GOLD mine. *Nucleic Acids Research* **49** (2021).
124. Szklarczyk, D. *et al.* STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research* **47**, D607–D613 (2019).
125. Lombardot, T. *et al.* Updates in Rhea: SPARQLing biochemical reaction data. *Nucleic Acids Research* **47**, D596–D600 (2019).
126. Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research* **47**, W636–W641 (2019).
127. Bateman, A. *et al.* UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research* **49**, D480–D489 (2021).
128. Yates, A. D. *et al.* Ensembl 2020. *Nucleic Acids Research* **48**, D682–D688 (2020).
129. Ferguson, C. *et al.* Europe PMC in 2020. *Nucleic Acids Research* **49**, D1507–D1514 (2021).
130. Murphy, K. & Weaver, C. *Janeway's immunobiology* 213–252 (Garland Science, Taylor & Francis Group, LLC, 2017).
131. Garris, C. S., Blaho, V. A., Hla, T. & Han, M. H. Sphingosine-1-phosphate receptor 1 signalling in T cells: trafficking and beyond. *Immunology* **142**, 347 (2014).
132. Matloubian, M. *et al.* Lymphocyte egress from thymus and peripheral lymphoid organs is dependent on S1P receptor 1. *Nature* **427**, 355–360 (2004).
133. Liu, G. *et al.* S1P1 receptor overrides regulatory T cell-mediated immune suppression through Akt-mTOR. *Nature immunology* **10**, 769 (2009).
134. Noronha, A. *et al.* The Virtual Metabolic Human database: integrating human and gut microbiome metabolism with nutrition and disease. *Nucleic Acids Research* **47** (2019).
135. Hubler, M. J. & Kennedy, A. J. Role of lipids in the metabolism and activation of immune cells. *Journal of Nutritional Biochemistry* **34**, 1–7 (2016).
136. Lovewell, R. R., Sasseti, C. M. & VanderVen, B. C. Chewing the fat: Lipid metabolism and homeostasis during M. tuberculosis infection. *Current Opinion in Microbiology* **29**, 30–36 (2016).
137. Wang, Z. *et al.* Detection of Metabolite-Protein Interactions in Complex Biological Samples by High-Resolution Relaxometry: Toward Interactomics by NMR. *Journal of the American Chemical Society* **11** (2021).
138. Ashtiani, M. *et al.* A systematic survey of centrality measures for protein-protein interaction networks. *BMC Systems Biology* **12**, 1–17 (2018).
139. Gilbert, D. N. *et al.* Enhanced Detection of Community-Acquired Pneumonia Pathogens With the BioFire® Pneumonia FilmArray® Panel. *Diagnostic Microbiology and Infectious Disease* **99**, 115246 (2021).

-
140. Poole, S. *et al.* Molecular point-of-care testing for lower respiratory tract pathogens improves safe antibiotic de-escalation in patients with pneumonia in the ICU: results of a randomised controlled trial. *Journal of Infection* **85**, 625–633 (2022).
 141. Sindelar, M. A., Zepeski, A. E., Lawler, B. J., Johnston, S. D. & Faine, B. A. MRSA nasal swab PCR to de-escalate antibiotics in the emergency department. *The American Journal of Emergency Medicine* **55**, 133–137 (2022).
 142. Tsang, N. N. Y. *et al.* Diagnostic performance of different sampling approaches for SARS-CoV-2 RT-PCR testing: a systematic review and meta-analysis. *The Lancet Infectious Diseases* **21**, 1233–1245 (2021).
 143. Bouzid, D. *et al.* Agreement of respiratory viruses' detection between nasopharyngeal swab and bronchoalveolar lavage in adults admitted for pneumonia: a retrospective study. *Clinical Microbiology and Infection* (2023).
 144. Pepe, M. S. & Thompson, M. L. Combining diagnostic test results to increase accuracy. *Biostatistics* **1**, 123–140 (2000).
 145. Gant, V. & Singer, M. Combining pathogen and host metagenomics for a better sepsis diagnostic. *Nature Microbiology* **7**, 1713–1714 (2022).
 146. Xie, F. *et al.* Clinical metagenomics assessments improve diagnosis and outcomes in community-acquired pneumonia. *BMC Infectious Diseases* **21** (2021).
 147. Brands, X. *et al.* An epigenetic and transcriptomic signature of immune tolerance in human monocytes through multi-omics integration. *Genome Medicine* **13**, 1–17 (2021).
 148. Schuurman, A. R. *et al.* Integrated single-cell analysis unveils diverging immune features of covid-19, influenza, and other community-acquired pneumonia. *eLife* **10** (2021).
 149. Wootton, D. G. *et al.* A longitudinal modelling study estimates acute symptoms of community acquired pneumonia recover to baseline by 10 days. *European Respiratory Journal* **49**, 1602170 (2017).
 150. Cheng, J. Y. *et al.* Longitudinal associations of serum survivin with the severity and prognosis of community-acquired pneumonia patients. *Respiratory Investigation* **61**, 84–94 (2023).
 151. Yao, M. X. *et al.* Cross-sectional and longitudinal associations of serum Cysteine-rich 61 with severity and prognosis among community-acquired pneumonia patients in China. *Frontiers in Medicine* **9** (2022).
 152. Aimo, L. *et al.* The SwissLipids knowledgebase for lipid biology. *Bioinformatics* **31**, 2860–2866 (2015).
 153. Koistinen, V. *et al.* Towards a Rosetta stone for metabolomics: recommendations to overcome inconsistent metabolite nomenclature. *Nature Metabolism* **5**, 351–354 (2023).
 154. Hicks, K. G. *et al.* Protein-metabolite interactomics of carbohydrate metabolism reveal regulation of lactate dehydrogenase. *Science* **379**, 996–1003 (2023).

155. Rubio-Aliaga, I. *et al.* Metabolomics of prolonged fasting in humans reveals new catabolic markers. *Metabolomics* **7**, 375–387 (2011).
156. Krug, S. *et al.* The dynamic range of the human metabolome revealed by challenges. *The FASEB Journal* **26**, 2607–2619 (2012).
157. Morris, C. *et al.* The relationship between BMI and metabolomic profiles: a focus on amino acids. *Proceedings of the Nutrition Society* **71**, 634–638 (2012).
158. Van Der Laan, T. *et al.* High-Throughput Fractionation Coupled to Mass Spectrometry for Improved Quantitation in Metabolomics. *Analytical Chemistry* **92**, 14330–14338 (2020).

List of Abbreviations

ADP	Adenosine diphosphate
AHR	Aryl-hydrocarbon receptor
aLEA	Alpha-linolenoyl ethanolamide
AMR	Antimicrobial resistance
ANOVA	Analysis of variance
API	Application programming interface
ATP	Adenosine triphosphate
AUC	Area under the curve
BER	Balanced error rate
BH4	Tetrahydrobiopterin
BMI	Body mass index
BP	Blood pressure
BUN	Blood urea nitrogen
<i>C. burnetii</i>	<i>Coxiella burnetii</i>
<i>C. psittaci</i>	<i>Chlamydophila psittaci</i>
CAP	Community-acquired pneumonia
ChEBI	Chemical Entities of Biological Interest database
CNS	Central nervous system
COPD	Chronic obstructive pulmonary disease
cor	Correlation
COX	Cyclooxygenase
CRP	C-reactive protein

CURB	Confusion, uremia, respiratory rate, blood pressure
CV	Cross validation
DEA	Docosatetraenoyl ethanolamide
DGLEA	Dihomo-gamma-linolenoyl ethanolamide
EA	Ethanolamine
EBI	European Bioinformatics Institute
EM-PCA	Expectation maximization - principal component analysis
EN	Elastic net regression
FA	Fatty acids
FAIR	Findable, accessible, interoperable, reusable
FC	Fold change
FDR	False discovery rate
FN	False negative
FP	False positive
Gln	Glutamine
Glu	Glutamic acid
GO	Gene Ontology
HMDB	Human Metabolome Database
HPA axis	Hypothalamic-pituitary-adrenal axis
HSV	Herpes simplex virus
ICU	Intensive care unit
ID	Identifier
IDO	Indoleamine-2,3-dioxygenase
IMA	Immunometabolic Atlas
<i>L. pneumophila</i>	<i>Legionella pneumophila</i>
LCAC	Long-chain acylcarnitine
LOS	Length of stay
LOX	Lipoxygenase
LPC	Lysophosphatidylcholine
LPE	Lysophosphatidylethanolamine
LPI	Lysophosphatidylinositol
LPS	Lysophosphatidylserine
LR	Logistic regression
<i>M. pneumoniae</i>	<i>Mycoplasma pneumoniae</i>
Max	Maximum
MCAC	Medium-chain acylcarnitine
Min	Minimum
MS	Mass spectrometry
n	Number
NMDA	N-methyl-D-aspartate
NMR	Nuclear magnetic resonance
NOS	Nitric oxide synthase
NSAID	Non-steroidal anti-inflammatory drugs
p-value	Probability value
PC	Phosphatidylcholine

PCA	Principal component analysis
PCR	Polymerase chain reaction
PCT	Procalcitonin
PE	Phosphatidylethanolamine
PGE2	Prostaglandin E2
PGF2a	Prostaglandin F2 alpha
Phe	Phenylalanine
PLA2	Phospholipase A2
PMID	PubMed identifier
POEA	Palmitoleoyl ethanolamide
PSI	Pneumonia Severity Index
PUFA	Polyunsaturated fatty acids
q	P-adjusted significance
RheaDB	Rhea-Annotated reactions database
ROC	receiver operating characteristic
ROS	Reactive Oxygen Species
RR	Respiratory rate
RSV / RS virus	Respiratory syncytial virus
<i>S. pneumoniae</i>	<i>Streptococcus pneumoniae</i>
SCAC	Short-chain acylcarnitine
SD	Standard deviation
SDMA	Symmetric dimethylarginine
SM	Sphingomyelin
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins database
S1PR	Sphingosine-1-phosphate receptor
TB	Tuberculosis
TDO	Tryptophan-2,3-dioxygenase
TG	Triglyceride
TN	True negative
TP	True positive
TriHOME	9,10,13-TriHOME
Trp	Tryptophan
TXB2	Thromboxane B2
Tyr	Tyrosine
VIP	Variable importance in prediction

List of Publications

- Ilona den Hartog*, Laura B. Zwep*, Thomas Hankemeier, Jacqueline J. Meulman, Ewoudt M.W. van de Garde, J.G. Coen van Hasselt. Longitudinal metabolomic profiling of *Streptococcus pneumoniae*-associated community-acquired pneumonia. *Metabolomics* (2024). (* Shared first authors)
- Ilona den Hartog, Naama Karu, Laura B. Zwep, G. Paul Voorn, Ewoudt M.W. van de Garde, Thomas Hankemeier, J.G.C. van Hasselt. Differential metabolic host response to pathogens associated with community-acquired pneumonia. *Metabolism Open* **18** (2023).
- Pascal Maas*, Ilona den Hartog*, Alida Kindt, Sonja Boman, Thomas Hankemeier, Coen van Hasselt. The Immunometabolic Atlas: A tool for design and interpretation of metabolomics studies in immunology. *PLoS One* **17:5** (2022). (* Shared first authors)
- Ilona den Hartog, Laura B. Zwep, Stefan M.T. Vestjens, Amy C. Harms, G. Paul Voorn, Dylan W. de Lange, Willem J.W. Bos, Thomas Hankemeier, Ewoudt M.W. van de Garde, J.G. Coen van Hasselt. Metabolomic profiling of microbial disease etiology in community-acquired pneumonia, *PLoS One* **16: 6** (2021).
- Pichet Praveschotinunt, Noémie-Manuelle Dorval Courchesne, Ilona den Hartog, Chaochen Lu, Jessica J. Kim, Peter Q. Nguyen, Neel S. Joshi. Tracking of Engineered Bacteria In Vivo Using Nonstandard Amino Acid Incorporation. *ACS Synthetic Biology* **7:6**. (2018).
- Remco Arts, Ilona den Hartog, Stefan E. Zijlema, Vito Thijssen, Stan H. E. van der Beelen, Maarten Merckx. Detection of Antibodies in Blood Plasma Using Bioluminescent Sensor Proteins and a Smartphone. *Analytical Chemistry* **88:8** (2016).

List of Affiliations

Sonja Boman

Division of Systems Pharmacology & Pharmacy, Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands.

Willem J.W. Bos

Department of Internal Medicine, Leiden University Medical Center, Leiden, The Netherlands.

Ewoudt M.W. van de Garde

Division of Pharmacoepidemiology and Clinical Pharmacology, Department of Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands.
Department of Clinical Pharmacy, St. Antonius Hospital, Nieuwegein, The Netherlands.

Thomas Hankemeier

Metabolomics and Analytics Centre, Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands.

Amy C. Harms

Metabolomics and Analytics Centre, Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands.

J.G. Coen van Hasselt

Division of Systems Pharmacology & Pharmacy, Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands.

Naama Karu

Metabolomics and Analytics Centre, Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands.

Alida Kindt

Metabolomics and Analytics Centre, Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands.

Dylan W. de Lange

Intensive Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands.

National Poison Information Center, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands.

Pascal Maas

Metabolomics and Analytics Centre, Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands.

Jacqueline J. Meulman

LUXs Data Science, Leiden, The Netherlands.

Department of Statistics, Stanford University, Stanford, CA, United States.

Stefan M.T. Vestjens

Department of Medical Microbiology and Immunology, St. Antonius Hospital,
Nieuwegein, The Netherlands.

G. Paul Voorn

Department of Medical Microbiology and Immunology, St. Antonius Hospital,
Nieuwegein, The Netherlands.

Laura B. Zwep

Division of Systems Pharmacology & Pharmacy, Leiden Academic Centre for Drug
Research, Leiden University, Leiden, The Netherlands.

Dankwoord

Dit proefschrift is het resultaat van de samenwerking tussen vele mensen. Aan de start van dit onderzoek stonden mijn promotoren en co-promotor. De metabolomics analyse van de patiëntsamples is mogelijk gemaakt door medewerkers van het St. Antonius Ziekenhuis te Nieuwegein en het Biomedical Metabolomics Facility te Leiden. Ander experimenteel werk en de data analyse is gedaan in samenwerking met studenten en medewerkers van de afdeling Systems Pharmacology & Pharmacy van het LACDR aan de Universiteit Leiden, in het bijzonder met de Van Hasselt groep. De gepubliceerde wetenschappelijke artikelen zijn tot stand gekomen in samenwerking met collega's van verschillende universiteiten en medisch centra.

Op meer informele wijze waren mijn familie, schoonfamilie en vrienden, met in het bijzonder mijn partner Stefan, zeer belangrijk bij het tot stand komen van dit proefschrift.

Curriculum Vitae

Ilona den Hartog was born in Montfoort, the Netherlands, in 1992. She attended high school at the Kalsbeek College Woerden and received her VWO diploma in 2010. She obtained her propaedeutic diploma in Human Movement Sciences at the VU Amsterdam in 2011. She received her Bachelor's and Master's degree in Biomedical Engineering at the Eindhoven University of Technology, graduating in 2017.

During her Bachelor's internship, Ilona studied the detection of antibodies in blood plasma using bioluminescent sensor proteins for point-of-care diagnostics. In her Master's, she specialized in Protein Engineering in the group of Maarten Merkx. Her thesis focused on DNA-based antibody detection using rolling circle amplification to reach single molecule sensitivity. During her Master's she visited Harvard University, Boston, USA for a four month research externship at the Wyss Institute for Biologically Inspired Engineering in the group of Neel Joshi. During her externship she worked on improving genetically engineered bacteria that incorporate non-standard amino acids in the extracellular matrix of biofilms, marking specific environments and facilitating drug targeting to disease.

In 2017, Ilona started her PhD research on metabolomics in community-acquired pneumonia at the Leiden Academic Centre for Drug Research (LACDR) under supervision of Coen van Hasselt, Thomas Hankemeier, and Ewoudt van de Garde. Since October 2022, Ilona works as a technical author at Etteplan in Amersfoort.

