**A robust mRNA signature obtained via Recursive Ensemble Feature Selection predicts the responsiveness of omalizumab in moderate-to-severe asthma**

**S. Kidwai**1**, P. Barbiero**3**, I. Meijerman**1**, A. Tonda**4**, P.Perez-Pardo**1 **P. Lio ́** 3**, A. H. Maitland-van der Zee**5 **D. L. Oberski**2**, A.D Kraneveld**1 and **A. Lopez-Rincon** 1,2

1Division of Pharmacology,Utrecht Institute for Pharmaceutical Science, Faculty of Science,Utrecht University,Universiteitsweg 99, 3584 CG Utrecht,the Netherlands  
2Department of Data Science, University Medical Center Utrecht  
3Department of Computer Science and Technology, University of Cambridge, 15 JJ Thomson Avenue, Cambridge CB3 0FD, United Kindgdom  
4UMR 518 MIA, INRAE, Universite Paris-Saclay, Paris, France  
5Department of Pulmonary Medicine, Amsterdam UMC, University of Amsterdam, Amsterdam, Netherlands

corresponding author: A. Lopez-Rincon [a.lopezrincon@uu.nl](mailto:a.lopezrincon@uu.nl)

# Acknowledgements

We would like to thank students Daphne S. Roozendaal, Hilde M. Spierenburg, Astha L. Holm and Renee Metcalf for their contributions to the work (i.e., data collection) reported in the manuscript

# Abstract

**Background**: Not being well controlled by therapy with inhaled corticosteroids and long-acting β2 agonist bronchodilators is a major concern for severe-asthma patients. Current treatment option for these patients is the use of biologicals such as anti-IgE treatment, omalizumab, as add-on therapy. Despite the accepted use of omalizumab, patients do not always benefit from it. Therefore, there is a need to identify reliable biomarkers as predictors of omalizumab response.

**Methods**: Two novel computational algorithms, machine-learning based Recursive Ensemble Feature Selection (REFS) and rule-based algorithm Logic Explainable Networks (LEN ) were used on open accessible mRNA expression data from moderate-to-severe asthma patients to identify genes as predictors of omalizumab response

**Results**: With REFS, the number of features were reduced from 28,402 genes to 5 genes while obtaining a cross-validated accuracy of 0.975. The 5 responsiveness predictive genes encode for the following proteins: Coiled-coil domain- containing protein 113 (CCDC113), Solute Carrier Family 26 Member 8 (SLC26A), Protein Phosphatase 1 Regulatory Subunit 3D (PPP1R3D), C-Type lectin Domain Family 4 member C (CLEC4C) and LOC100131780 (not annotated). The LEN algorithm found 4 identical genes with REFS: CCDC113 ,SLC26A8 PPP1R3D and LOC100131780. Literature research showed that the 4 identified responsiveness predicting genes are associated with: mucosal immunity, cell metabolism, and airway remodeling.

**Conclusion and clinical relevance**: Both computational methods show 4 identical genes as predictors of omalizumab response in moderate-to-severe asthma patients. The obtained high accuracy indicates that our approach has potential for clinical settings. Future studies in relevant cohort data should validate our computational approach.

Keywords: anti-IgE, asthma, biomarker, machine-learning, omalizumab

# Introduction

Asthma is a common chronic respiratory disease, affecting more than 300 million individuals worldwide.1 Around 50-70 % of the total asthmatic population represent individuals with moderate-to-severe asthma.2, 3 This form of asthma is difficult to treat, resulting in frequent asthma exacerbations, hospitalizations, and even death.4, 5 Patients diagnosed with moderate-to-severe asthma cannot always control their disease with long acting inhaled β 2 agonists and corticosteroids.6 As an add-on therapy for these patients, a biological such as omalizumab7 can be prescribed. This anti-IgE agent specifically binds free IgE and prevents IgE binding to the high affinity IgE receptor FcεRI expressed on immune cells such as mast cells, basophils and antigen-presenting cells.8 It reduces allergic airway inflammation by decreasing IgE levels in blood, and by consequent downregulation of the expression of FcεRI on airway inflammatory cells and lowers eosinophil counts.9–11

Despite omalizumab’s widely accepted use, moderate-to-severe asthma patients do not always benefit from this anti-IgE therapy. Because there are also other biologicals such as anti-IL-5 (mepolizumab and reslizumab) 12, anti-IL-5R (benralizumab) 13 and anti-IL-4R (dupilimab) 14 available prediction of treatment response will support choosing the best biological for the individual patient. This will reduce the burden of disease and health care cost that arise due to the use of improper or suboptimal medications. However, predictors of response to omalizumab are still limited.15–17 Therefore, it is of clinical significance to identify reliable biomarkers as predictors of omalizumab response that can match individual moderate-to-severe asthma patients with the most effective medication.

Finding accurate biomarkers for omalizumab response is difficult as these biomarker studies are diverse and complex. Various studies have explored potential biomarkers to predict omalizumab responsiveness such as genes 18 or clinical measurements (e.g., total serum IgE, FEV1).19–21 But the available biomarkers are limited in number and accuracy. Consequently, there is no consensus on the best set of biomarkers of omalizumab therapy for those who do benefit from the drug.

Machine learning (ML), a branch of artificial intelligence, provides a set of techniques that can potentially deal with complex omics data to find the right biomarkers. Typically, omics datasets contain thousands of features (such as genetic variations, mRNA expression, miRNAs expression or metabolites concentrations) in a relative low number of samples, making it difficult to analyse using standard univariate techniques.22, 23

Using ML gene cluster analysis, Upchurch et al. (2020)6 identified transcriptional differences between omalizumab responders and non-responders. The study showed 3867 genes which were differentially expressed between healthy controls, responders, and non-responders. However, this set of response-predictive genes is too large to be effectively used in a clinical setting. With Recursive Ensemble Feature Selection (REFS) 22 a more accurate and robust gene signature can be found over the gene cluster analysis method that was used by Upchurch. Feature selection (REFS) method shows unbiased gene interactions in contrast to clustering of over/under expressed genes dependent on annotated pathway functions or a single algorithm.24,25 REFS allows identification of compact gene signatures. Logic explained networks (LEN), an advanced computational algorithm, gives in contrast to other algorithms 26 an interpretable result by providing an explanation for the possible relationships between the features. It simplifies how to read mRNA expression as a set of logical rules which may serve to take clinical action/decision. Measuring mRNA expression in patients’ whole blood could give a direction for omalizumab therapy responsiveness based on LEN rules.

In this study, we use two ML techniques, REFS and LEN, to reduce the number of features to a more compact size which is effective for use in a clinical setting.24, 25

We implemented the two-state-of-art computational algorithm in ML i.e., the REFS algorithm and LEN algorithm on mRNA expression profiles in whole blood of moderate-to-severe asthma patients. We used the same samples as used by Upchurch et al. (2020)6 to find predictors of omalizumab response. Single Response-predictive genes were identified through ML-based REFS and validated with 10-fold cross-validation. 24, 25 In addition, we used LEN 26 which identifies gene groups and their interrelationship as predictors of omalizumab response. The LEN algorithm provides rules on sets of mRNA expression compared to healthy controls. The biological context and pathway annotation of REFS and LEN overlapping omalizumab response-predictive genes were validated through literature research to gain a better understanding of mechanisms involved in omalizumab responsiveness.

# Methods

**2.1 Patient Samples and Data**

Data on individual moderate-to-severe adult asthma patients,’ whole blood mRNA profiles and clinical information was retrieved from the publicly available database ‘Gene Expression Omnibus’ (GEO) with accession code GSE134544. This dataset was published by Upchurch et al., (2020)6

The GSE134544 database has 239 blood mRNA samples of 40 moderate-to-severe adult asthmatic patients, and 17 non-asthmatic healthy controls. Of the 40 patients, 30 were defined as responders and 10 as non-responders. For the LEN analysis, besides the whole blood mRNA expression of patients, 17 healthy controls were also included. Patients were prescribed omalizumab (Xolair®, Genentech) and were dosed as per the manufacturer’s dosing table (according to serum IgE and body weight).

The blood transcriptome included all blood cell types. For every patient, the sample contained 28,402 gene expression levels. Whole blood mRNA expression profiles measured 1 week before start of treatment of the 40 moderate-to-severe asthmatics were used for the REFS analysis.

Patients were not eligible for the study if they were pregnant, under the age of 18 or recently on omalizumab. mRNA expression was measured with Illumina HumanHT-12 V4.0 beadchip platform, containing genes for approximately 20,000 transcripts.

**2.2 Definition of disease severity and responders/non-responders**

Upchurch et al., (2020)6 selected asthmatic patients who had uncontrolled asthma despite treatment with inhaled corticosteroids (ICS) and/or long-acting β-agonists (LABA). Patients and non-asthmatic healthy controls6 were recruited under protocols approved by the Institutional Review Board.

According to the Global Initiative for Asthma guideline, 66 moderate asthma patients are well-controlled with low or medium dose ICS and/or LABA, and severe asthma patients were defined as those that remain uncontrolled despite optimized treatment with high dose ICS-LABA, or that require high dose ICS-LABA to prevent asthma from becoming uncontrolled.

Disease severity was based on a combination of asthma control test (ACT), low lung function (FEV1 < 80%) and symptom frequency, including total number of days with symptoms per week and of night-time awakenings per week. 6

Omalizumab responsiveness was defined by Upchurch et al (2020) 6 as improvements in asthma control, with non-responders’ not able to achieve asthma control. Uncontrolled asthma was defined as a combination of factors including ACT score (<19), asthma related symptoms in number of days in previous week, use of short-acting β-agonists (>= 2x per week) and night-time sleep disruption (>= 2 per week), unchanged asthma control through medication, indications of little/no improvements in asthma by physicians. 6

**2.3 Feature Selection**

To analyze the mRNA expression data, genes from the microarray platform Illumina were converted to gene IDs using the web probe conversion tool in the Ensembl Genome Browser.67 This step was done only with the final short list of genes. Next, the data was normalized with z-score normalization. 68

Non-responders were labelled as 0 (n=10), and responders were labelled as 1 (n=30).

*Algorithm 1: REFS*

REFS 69 was used to identify a gene signature in moderate-to-severe asthma patients to separate omalizumab responders from non-responders. Briefly, the REFS algorithm runs a machine learning algorithm 10 times and it ranks features (in this study genes), repetitively eliminating the lowest scoring features. After every iteration, the most relevant gene-signature for therapy response is selected. The REFS algorithm is an ensemble which means it uses 8 different classifiers to rank the features. From these classifiers (metrics for algorithm performance) the most important features are extracted and ranked accordingly. The REFS algorithm was implemented on 28,402 genes and validated with a 10-fold cross validation.

To visualize the diagnostic ability of classifiers and evaluate the performance of learning models, Receiver Operating Characteristic (ROC) and area under the curve (AUC) were used.

**2.4 Logic Explained Networks (LEN)**

While REFS is designed to directly perform feature selection for classification, the idea behind LENs is to automatically extract a set of rules that use a minimal amount of features. In other words, REFS sees feature selection and classification as separate, while LENs attempts to perform both classification and feature selection in a single step. LEN was used to create a set of rules for predicting responsiveness to omalizumab on the same GSE134544 dataset.

*Algorithm 2: LEN*

The LENs algorithm 63 was implemented on 28,402 genes. LENs provides rules on sets of mRNA expression compared to healthy controls. It uses healthy controls as baseline.

In our research we compare responders and non-responders and data from healthy controls used from LENs. LENs algorithm predicts a connected network of genes and their shared up/down regulated gene expression in omalizumab treatment comparing responders and non-responders to mean mRNA expression of healthy controls. In this work we used an Entropy-based LEN 64 implemented in the python package pytorch\_explain.70

**2.5 Biological Context Of Gene Signatures**

To determine the biological context of the identified predictive genes for omalizumab therapy response and explain the (un)responsiveness to omalizumab, literature research was conducted. PubMed and Google Scholar were used for the assessment of gene function of each gene identified from REFS (algorithm 1) and LENs (algorithm 2). Protein function was investigated in the context of asthma pathology and omalizumab response.

# Results

**3.1 Feature selection**

The GSE134544 dataset containing 40 samples was used to run the REFS algorithm ten times (see Fig. 3). With the machine-learning REFS approach (algorithm 1) the total number of genes was reduced from 28,402 to 5 genes associated with therapeutic responsiveness. The optimal selection of genes depends on classification accuracy of the binary problem (responders/non-responders). The 5-gene signature corresponds with the highest peak in accuracy over all 8 classifiers in the REFS ensemble (See, Fig. S1).

**3.2 Genes of interest**

REFS (algorithm 1) identified 5 individual genes as predictors of omalizumab response: CCDC113, SLC26A, PPP1R3D, CLEC4C and LOC100131780. The mRNA expression levels of the identified 5 genes in non-responder patients’ samples compared to responders are shown in Fig. 1. Fig. 1 shows that moderate-to-severe asthma patients that did not respond well to omalizumab have a higher mRNA expression of 4 out of 5 identified genes when compared to responders: LOC100131780, CCDC113, SLC26A and PPP1R3D. The mRNA expression of one gene, CLEC4C, is downregulated in non-responders compared to responders.

**3.3 ROC analysis of gene signature**

The AUC of the ROC curve was used to verify the efficacy/performance of the identified 5-gene signature over all 8 classifiers in a 10-fold cross validation. As shown in Fig. 2, the AUC of 0.99 was obtained using the Passive Aggressive classifier, a performance metric, which represents the best discriminatory accuracy for our model.27, 28

**3.4 Relationship with treatment outcome: LEN**

In parallel to the REFS algorithm (algorithm 1), a LEN on the same GSE134544 dataset was used to compare results (See, Fig. 3)

As can be seen from Fig.3, the LEN algorithm returns three simple rules to explain why identified genes from the algorithm predict responsiveness to omalizumab.

The LEN rules 1,2 or 3 predict the best responsiveness of omalizumab in moderate-to-severe asthma patients.

Rule 1 [LOC100131780, CCDC113 and SLC26A8] indicates a combination of response-predictive gene set 1 . The first rule implies that the mRNA expression of all genes in this trio, is downregulated in the responder group as compared to the average mRNA expression of non-asthmatic healthy controls.

Rule 2 [LOC100131780, SLC26A8 and PPP1R3D] indicates combination of response-predictive gene set 2. The second rule implies that the mRNA expression of all genes in this trio are downregulated in the responder group as compared to the average gene expression of non-asthmatic healthy controls.

Rule 3 [CCDC113 , SLC26A8 and PPP1R3D] indicates combination of response-predictive gene set 3. The third rule implies that the mRNA expression of all genes in this trio, are downregulated in the responder group as compared to the average gene expression of non-asthmatic healthy controls.

**3.5 Combined results of REFS and LEN**

The REFS (algorithm 1) and the rule-based LEN (algorithm 2) both allow identification of compact gene signatures and provide a mean accuracy of 97.5%. An overlap in the results from both algorithms was found for 4 of the 5 genes identified by REFS. The following genes are identified as omalizumab response-predictive genes: LOC100131780, CCDC113, SLC26A and PPP1R3D, which are all upregulated in non-responders and downregulated in responders . This is summarized in Fig 3.

**3.6 Biological interpretation of the gene signature predicting responsiveness to omalizumab**

To determine the biological context of our findings, a literature research on the proteins encoded by the five identified genes was conducted. Literature research showed that the identified response-predictive genes can be associated with: Mucosal immunity, Cellular energy, Airway hypersensitivity and remodeling. Details of the genes are enlisted in Table 1. and described below.

*CLEC4C*, also denoted as the *BDCA2* or *CD303* gene, encodes a member of the C-type Lectin superfamily, C-type lectin domain family 4 member C. This Lectin-type cell surface receptor is linked to various functions including antigen-uptake by dendritic cells 29 for internalization/presentation to T cells 30 and attachment to serum IgG.31 *CLEC4C* triggers the src-family protein- tyrosine kinases signaling pathways to inhibit induction of IFN-α/β expression in plasmacytoid dendritic cells (pDC).29,30This inhibition, in turn, leads to the production of pro- or anti-inflammatory cytokines and consequently, fine-tunes innate and adaptive immune responses to viral infections. High mRNA expression of *CLEC4C* in whole blood has been associated with low mRNA expression of Toll-like receptor 7 (TLR 7) expression and increased risks for common colds in asthmatic subjects. TLR-7 mediated induction of IFN- α/β, and inflammatory cytokine production is critical in antiviral immune responses.32 Hence, impaired anti-viral response in asthmatic patients may be coordinated by *CLEC4C*.33 (See, supporting information S1) Taken together, the present set of findings indicate that severe asthma patients suffering from frequent (viral) exacerbations may have enhanced pDC (*CLEC4C*+DC) expression. This conclusion seems sensible as pDCs are critical mediators for anti-viral responses. Our results obtained from the REFS algorithm show that *CLEC4C* expression in whole blood of non-responders to omalizumab was downregulated when compared to responders. Thus, for proper response to omalizumab, pDC may be central in improving asthma symptoms.

*CCDC113* gene encodes Coiled-coil domain-containing protein 113 *(CCDC113)* which is a centrosome-associated protein critical for cilia formation.34 This gene coordinates ciliary beating frequency (CBF), beat direction and cilia stroke.35 *CCDC113* has been shown to be overexpressed in nasal brushes of asthmatic subjects when compared to healthy controls.36 Upregulation of *CCDC113* could possibly be a rebound effect for ciliary defects or damage. The upregulation of CCDC113 shown in our results from LENs and REF can be expected based on the pathology of asthma. In the airways, cilia work along with airway mucus to facilitate mucociliary clearance acting as a pulmonary defense.37, 38 It therefore follows that ciliary dysfunction could lead to impaired mucociliary clearance making asthmatic patients more susceptible to airway infection or inflammation. Overexpression of *CCDC113* mRNA in whole blood of non-responders compared to omalizumab responders might indicate ciliary defects in airways of non-responders. Thus as CCDC113 is involved in cilium assembly it may simply increase the severity of asthma.

*SLC26A8* gene encodes Testis anion transporter 1 (Tat1), which is a sperm-specific member of the SLC26 family of multifunctional anion exchangers. *SLC26A8* functional relevance is linked to sperm motility and mutations appear to cause male infertility.39 In the whole blood mRNA of severe asthmatics, upregulated *SLC26A8* compared to healthy controls has been reported.40 The presented LENs and REFS approach showed similar results for non-responders. Increased levels of *SLC26A8* are associated with asthma and lower sperm count/sperm motility.42 Given the limited information, it is hypothesized that *SLC26A8*, is merely an additional biomarker in non-responders as no logical context related to omalizumab therapy response can be formulated

*PPP1R3D*, encodes for Protein Phosphatase 1 (PP1) regulatory Subunit 3D.41 This gene is associated with the cellular energy sensor 5' AMP-activated protein kinase (AMPK). 41 The interaction between AMPK and PP1 is regulated by intracellular glycogen content.42 Upregulated mRNA expression of *PPP1R3D* in blood cells is associated with IgE-mediated peanut allergies in children.43 Food allergies are also associated with increased asthma severity in later life.44–46 Interestingly, *PPP1R3D* has been reported as a novel obesity candidate gene.47 Obese adults tend to have more severe asthma as compared to lean adults.48 Indeed, *PPP1R3D* was found overexpressed by LENs and REFS in the non-responder group. Based on the limited information on this gene, it can be suggested that non-responders to omalizumab might exhibit a severe non-IgE-mediated form of asthma or might be more obese than responders. Data on body weight of asthmatic subjects was not available to include in the present study.

*LOC100131780* is an illumina gene ID, which could not be matched to a specific gene in literature. As such, *LOC100131780* association with asthma remains obscure. Limited information could be recovered computationally from its sequence overlap with *DNAI1* that codes for dynein axonemal intermediate chain 1. *LOC100131780* as *CCDC113* also points to the important role of primary cilia *Mutations of the LOC100131780* gene 49 have been linked to primary ciliary dyskinesia (PCD), a rare inherited disease with dysfunctional mucocilIary clearance eventually leading to airway remodeling in thickened airway walls and obstruction.50, 51 Elevated levels of *LOC100131780* were found by LENs and REFS in omalizumab non-responders. Speculatively, PCD is a confounding condition that mimics asthma symptoms. As PCD and asthma rarely co-exist 50 patients with overexpressed LOC100131780 classified as non-responders might be misdiagnosed with severe asthmatic instead of PCD.

# Discussion

Pre-therapeutic screening helps in unwarranted drug exposure in patients suffering from severe forms of asthma. Treatment with a biological is the most logical step for this group of patients. However, different biologicals are available and biomarkers to determine which biological would be appropriate for the individual patient are lacking. Omaluzimab (anti-IgE) is one of the commonly used biologicals in asthma patients, however patients do not respond equally well to therapy with this drug.21, 55, 56 Studies examining transcriptional expression profiles in whole blood and sputum of asthmatic patients have identified gene signatures that may be associated with asthmatic phenotypes57–59 or therapeutic responsiveness to ICS 60,61 and omalizumab.6,18 However, no definitive set of biomarkers for omalizumab therapy response was found till date, while omalizumab response prediction could be very beneficial for proper disease management. To select the best treatment strategy for severe asthma patients, better predictive biomarkers for omalizumab treatment response are therefore needed.

ML prediction can help to find a small and easy to measure set of biomarkers that predicts treatment responses. Yet, finding accurate and robust response-predictive genes using ML has proven to be a challenge due to underpowered studies, poor explanatory models and the use of single genes or gene sets with no or few overlapping genes.22,23 REFS algorithm offers a solution for this issue 25,62 in terms of better accuracy and robustness. LEN can explain why the algorithm arrives at a certain decision, such that the results of LEN algorithm can be interpreted by clinicians. Explainability matters especially in clinical context because predictions are useful when they can be understood for acceptance of the AI decisions by physicians. LENs 26,63,64 is an interesting approach to extract biologically meaningful gene associations and thereby increase prediction model interpretability.

In the present study, a novel feature selection algorithm combining 8 classifiers, REFS, was used on open accessible mRNA expression data (GSE13544) from moderate-to-severe asthma patients to predict treatment responsiveness of omalizumab. In addition to REFS, the entropy-based LEN model 26 was used, as it provides the highest flexibility, generalization accuracy, and less complex logic formulas.The machine-learning based REFS algorithm returned a 5-gene signature: CLEC4C, CCDC113, SLC26A8, PPP1R3D and LOC100131780. The LEN algorithm returned a short 4-gene signature: CCDC113, SLC26A8, PPP1R3D and LOC100131780. Groups of responsive predictive genes: ([LOC100131780, SLC26A8 and PPP1R3D]; [LOC100131780, CCDC113 and SLC26A8]; [CCDC113, SLC26A8 and PPP1R3D]) were found through the LEN algorithm. Four overlapping genes (CCDC113, SLC26A8, PPP1R3D and LOC100131780) were found to be upregulated in non-responders with both methodologies. Our compact set of genes was able to differentiate between responders and non-responders with a mean accuracy 0.975.

Zhang et al (2021), 18 used the same GEO dataset GSE134544 employed in this study. The study reported one useful genetic biomarker, the T-cell surface glycoprotein CD3 epsilon chain (CD3E) using weighted gene co-expression network analysis (WGCNA). This biomarker was downregulated in non-responders. However, a single biomarker might be insufficient to capture interactions between other key genes relevant to accurately predict patient responsiveness to omalizumab treatment. For example, some patients exhibiting high CD3E expression might still not respond to omalizumab, while conversely some patients with low expression of CD3E could respond to treatment. Identifying a short list of responsiveness predictive genes as in our results could be useful in clinical setting to show coordinated changes in gene expression that may impact therapeutic response, which cannot be identified by a single biomarker. With REFS/LENs method a mean accuracy of 0.975 was computed and the proposed methodology outperformed Zhang et al. (2021) 18 that reported an accuracy of 0.763.

Upchurch et al. (2020) 6 identified biomarkers through gene clustering but predicted a large set of response-predictive genes on GSE134544. With the REFS and LENs approach a more compact set of genes (4) in contrast to Upchurch et al. (2020) 6 methodology (1776) was identified. Overall, the results from REFS and LENs show robustness and higher accuracy results.

To validate the relevance of overlapping genes from both computational methodologies and understand the mechanisms involved in omalizumab responsiveness, biological functions of the genes were investigated. From a clinical perspective, every single gene in the predicted gene set by REFS is a potential biomarker for omalizumab therapy response. How large the contribution of genes in the predicted gene trios identified by LENs is, is a matter of debate. As mRNA of all genes are overexpressed in the non-responder group, it can mean that these combinations are of interest for further research due to their supposed increased biological significance in activating or overloading specific biological pathways related to asthma pathogenesis and omalizumab response.

Co-expression of response-predictive genes does not simply mean interaction between their associated proteins, but the results may propose similarities in their regulation by transcription factors such as NFκb which controls various aspects of innate and adaptive immune functions, serving as critical mediator of inflammatory responses.65 In the current study, genes that were differentially co-expressed between responders and non-responders were associated in pathways regulating diverse pathological processes in chronic inflammatory conditions (the AMPK pathway *PPP1R3D* or tyrosine kinase signaling pathway (*CLEC4C*) associated with disturbed mucosal clearance (*CCDC113* & *LOC100131780*) and tissue remodeling (*LOC100131780*) (See, Table 1). These genes appear to be somewhat connected in terms of function. It is possible that the upregulated genes in non-responders are associated with a more severe asthmatic phenotype. That might be a reason why patients with severe asthma might respond badly to omalizumab.

The severity of asthma could result from an impaired immune activation or dysregulation of immune response. Zhang et al (2021), 18 hypothesized that omalizumab inhibits airway inflammation by reducing Th2 inflammation cascade. By calculating the immune enrichment score they found that CD4+ T cells and dendritic cells numbers were lower in blood of non-responders. A low T cell signature and high inflammatory gene cluster in non-responders to omalizumab was also reported by Upchurch et al. (2020). 6 In the patient cohort, it is possible that type 2 inflammation indicative for IgE-mediated asthma and non-type 2 inflammation might not be entirely mutually exclusive in responders and non-responders as found by Zhang et al (2021)18 or Upchurch et al. (2020).6 Although we did not investigate this aspect in our study, mixed phenotypes (type 2 and non-type 2) might be associated in the same individual, resulting in less likelihood of response to omalizumab.

One of the most intriguing results was the identification *SLC26A8* linked to male infertility as an mRNA predictor in all LEN proposed gene combinations. Although not much is known about this gene biological function in asthma, it might play a pivotal role in regulating gene expression changes observed in the airways of male non-responders to omalizumab therapy (See, Table 1). Determining interplay between asthma severity, gender and reproductive health was beyond the scope of this study.

Our findings thus reflect that predicting response to omalizumab therapy is powerful when identified genes are given a biological context. However, our approach does have some limitations. Not all pathways involved in our gene set have been fully characterized in literature. Consequently, our findings depend on the current state of knowledge and pathway annotations of genes may be sparse and skewed in comparison to well-defined gene sets. To complement our literature search, *in vitro* experiments with whole blood sample might provide us with broader insight regarding gene function and mechanisms associated with response to omalizumab. For clinical applications, a simple blood test can be useful to assess mRNA expression of the 4 overlapping genes for future treatment with omalizumab. If mRNA expressions of all 4 overlapping genes are found downregulated in responder group compared to healthy control mRNA expression, anti-IgE treatment might be beneficial for the patient.

In summary, we suggest that our small set of mRNA predictors returned from two different algorithms is highly accurate and holds potential value as clinical biomarkers for predicting omalizumab treatment response in moderate-to-severe asthma. We conclude that our work represents a first step towards a more tailored prediction of omalizumab response. Future studies in other cohorts should validate our computational approach. Furthermore, a prospective study is necessary to test the clinical utility of the biomarkers. As a next step, add-on studies comparing biological medications could be valuable to identify patients that should be treated with omalizumab versus another biologic. To support a physicians’ treatment decision for a biologic, we recommend confirming the diagnosis of severe asthma, exclusion of conditions mimicking asthma symptoms such as PCD and assessment of comorbidities associated with severe forms of asthma. Patients who might attain incomplete benefit from omalizumab could benefit from switching to a different biologic which targets an alternative mechanism if GINA criteria are met. Alternatively, non-responders might benefit from add-on therapy with a second biologic.

Word count: 4354

# References

1. Masoli, M., Fabian, D., Holt, S., Beasley, R. & for Asthma (GINA) Program, G. I. The global burden of asthma: executive summary of the gina dissemination committee report. *Allergy* 59, 469–478 (2004).
2. Larsson, K. *et al.* Prevalence and management of severe asthma in primary care: an observational cohort study in sweden (pacehr). *Respir. research* 19, 1–10 (2018).
3. FitzGerald, J. M. *et al.* The burden of exacerbations in mild asthma: a systematic review. *ERJ open research* 6 (2020).
4. Chung, K. F. *et al.* International ERS/ATS guidelines on definition, evaluation and treatment of severe asthma TASK FORCE REPORT ERS/ATS GUIDELINES ON SEVERE ASTHMA Executive Summary. *J* 43, 343–373, DOI: 10.1183/ 09031936.00202013 (2014).
5. Mincheva, R., Ekerljung, L., Bossios, A., Lundbäck, B. & Lötvall, J. High prevalence of severe asthma in a large random population study. *J. Allergy Clin. Immunol.* 141, 2256–2264, DOI: 10.1016/j.jaci.2017.07.047 (2018).
6. Upchurch, K. *et al.* Whole blood transcriptional variations between responders and non-responders in asthma patients receiving omalizumab. *Clin. Exp. Allergy* 50, 1017–1034, DOI: 10.1111/cea.13671 (2020).
7. Incorvaia, C., Mauro, M., Makri, E., Leo, G. & Ridolo, E. Two decades with omalizumab: What we still have to learn. *Biol. Targets Ther.* 12, 135–142, DOI: 10.2147/BTT.S180846 (2018).
8. Belliveau, P. P. Omalizumab: A Monoclonal Anti-IgE Antibody. *Medscape Gen. Medicine* 7, 27 (2005).
9. Busse, W. *et al.* Omalizumab, anti-IgE recombinant humanized monoclonal antibody, for the treatment of severe allergic asthma. *J. Allergy Clin. Immunol.* 108, 184–190, DOI: 10.1067/mai.2001.117880 (2001).
10. Chanez, P. *et al.* Omalizumab-induced decrease of FcεRI expression in patients with severe allergic asthma. *Respir. Medicine* 104, 1608–1617, DOI: 10.1016/j.rmed.2010.07.011 (2010).
11. Plewako, H. *et al.* The effect of omalizumab on nasal allergic inflammation. *J. Allergy Clin. Immunol.* 110, 68–71, DOI: 10.1067/mai.2002.125488 (2002).
12. Eger, K., Kroes, J. A., Ten Brinke, A. & Bel, E. H. Long-term therapy response to anti–il-5 biologics in severe asthma—a real-life evaluation. *The J. Allergy Clin. Immunol. In Pract.* 9, 1194–1200 (2021).
13. Bleecker, E. R. *et al.* Efficacy and safety of benralizumab for patients with severe asthma uncontrolled with high-dosage inhaled corticosteroids and long-acting β 2-agonists (sirocco): a randomised, multicentre, placebo-controlled phase 3 trial. *The Lancet* 388, 2115–2127 (2016).
14. Castro, M. *et al.* Dupilumab efficacy and safety in moderate-to-severe uncontrolled asthma. *New Engl. J. Medicine* 378, 2486–2496 (2018).
15. Bousquet, J. *et al.* The effect of treatment with omalizumab, an anti-IgE antibody, on asthma exacerbations and emergency medical visits in patients with severe persistent asthma. *Allergy: Eur. J. Allergy Clin. Immunol.* 60, 302–308, DOI: 10.1111/j.1398-9995.2004.00770.x (2005).
16. Busse, W. W., Massanari, M., Kianifard, F. & Geba, G. P. Effect of omalizumab on the need for rescue systemic corticosteroid treatment in patients with moderate-to-severe persistent IgE-mediated allergic asthma: A pooled analysis. *Curr. Med. Res. Opin.* 23, 2379–2386, DOI: 10.1185/030079907X226258 (2007).
17. Walker, S., Monteil, M., Phelan, K., Lasserson, T. J. & Walters, E. H. Anti-IgE for chronic asthma in adults and children. *Cochrane Database Syst. Rev.* DOI: 10.1002/14651858.cd003559.pub3 (2006).
18. Zhang, Q. *et al.* Cd3e as a new predictive biomarker of response to omalizumab treatment in asthma patients: Evidence from bioinformatic analysis. *Int. Immunopharmacol.* 93, 107423 (2021).
19. Li, B. *et al.* Prediction of clinical response to omalizumab in moderate-to-severe asthma patients using the change in total serum ige level. *J. Thorac. Dis.* 12, 7097 (2020).
20. Riccio, A. M. *et al.* Plasma galectin-3 and urine proteomics predict fev1 improvement in omalizumab-treated patients with severe allergic asthma: Results from the proxima sub-study. *World Allergy Organ. J.* 13, 100095 (2020).
21. Bousquet, J. *et al.* Predicting and evaluating response to omalizumab in patients with severe allergic asthma. *Respir. Medicine* 101, 1483–1492, DOI: 10.1016/j.rmed.2007.01.011 (2007).
22. Ruberg, S. J. & Shen, L. Personalized Medicine: Four Perspectives of Tailored Medicine. *http://dx.doi.org.proxy.library.uu.nl/10.1080/19466315.2015.1059354* 7, 214–229, DOI: 10.1080/19466315.2015.1059354 (2015).
23. Sechidis, K. *et al.* Distinguishing prognostic and predictive biomarkers: an information theoretic approach. *Bioinformatics* 34, 3365–3376, DOI: 10.1093/BIOINFORMATICS/BTY357 (2018).
24. Lopez-Rincon, A., Martinez-Archundia, M., Martinez-Ruiz, G. U., Schoenhuth, A. & Tonda, A. Automatic discovery of 100-mirna signature for cancer classification using ensemble feature selection. *BMC bioinformatics* 20, 1–17 (2019).
25. Lopez-Rincon, A. *et al.* Machine learning-based ensemble recursive feature selection of circulating mirnas for cancer tumor classification. *Cancers* 12, 1785 (2020).
26. Barbiero, P. *et al.* Entropy-based Logic Explanations of Neural Networks. (2021).
27. Šimundic ́, A.-M. Measures of diagnostic accuracy: basic definitions. *Ejifcc* 19, 203 (2009).
28. Mandrekar, J. N. Receiver operating characteristic curve in diagnostic test assessment. *J. Thorac. Oncol.* 5, 1315–1316 (2010).
29. Riboldi, E. *et al.* Human c-type lectin domain family 4, member c (clec4c/bdca-2/cd303) is a receptor for asialo-galactosyl- oligosaccharides. *J. Biol. Chem.* 286, 35329–35333 (2011).
30. Dzionek, A. *et al.* Bdca-2, a novel plasmacytoid dendritic cell–specific type ii c-type lectin, mediates antigen capture and is a potent inhibitor of interferon α/β induction. *The J. experimental medicine* 194, 1823–1834 (2001).
31. Jégouzo,S.A.*etal.*Anovelmechanismforbindingofgalactose-terminatedglycansbythec-typecarbohydraterecognition domain in blood dendritic cell antigen 2. *J. Biol. Chem.* 290, 16759–16771 (2015).
32. Petes, C., Odoardi, N. & Gee, K. The toll for trafficking: toll-like receptor 7 delivery to the endosome. *Front. immunology* 8, 1075 (2017).
33. Murray, L. M., Yerkovich, S. T., Ferreira, M. A. & Upham, J. W. Risks for cold frequency vary by sex: role of asthma, age, tlr7 and leukocyte subsets. *Eur. Respir. J.* 56 (2020).
34. Firat-Karalar, E. N., Sante, J., Elliott, S. & Stearns, T. Proteomic analysis of mammalian sperm cells identifies new components of the centrosome. *J. Cell Sci.* 127, 4128–4133, DOI: 10.1242/jcs.157008 (2014).
35. Bazan, R. *et al.* Ccdc113/ccdc96 complex, a novel regulator of ciliary beating that connects radial spoke 3 to dynein g and the nexin link. *PLoS genetics* 17, e1009388 (2021).
36. Pandey, G. *et al.* A nasal brush-based classifier of asthma identified by machine learning analysis of nasal rna sequence data. *Sci. reports* 8, 1–15 (2018).
37. Wanner, A., Salathé, M. & O’Riordan, T. G. Mucociliary clearance in the airways. *Am. journal respiratory critical care medicine* 154, 1868–1902 (1996).
38. Knowles, M. R., Boucher, R. C. *et al.* Mucus clearance as a primary innate defense mechanism for mammalian airways. *The J. clinical investigation* 109, 571–577 (2002).
39. Dirami, T. *et al.* Missense mutations in SLC26A8, encoding a sperm-specific activator of CFTR, are associated with human asthenozoospermia. *Am. J. Hum. Genet.* 92, 760–766, DOI: 10.1016/j.ajhg.2013.03.016 (2013).
40. Bigler, J. *et al.* A severe asthma disease signature from gene expression profiling of peripheral blood from u-biopred cohorts. *Am. journal respiratory critical care medicine* 195, 1311–1320 (2017).
41. Armstrong, C. G., Browne, G. J., Cohen, P. & Cohen, P. T. PPP1R6, a novel member: Of the family of glycogen-targetting subunits of protein phosphatase 1. *FEBS Lett.* 418, 210–214, DOI: 10.1016/S0014-5793(97)01385-9 (1997).
42. Garcia-Haro, L. *et al.* The PP1-R6 protein phosphatase holoenzyme is involved in the glucose-induced dephosphorylation and inactivation of AMP-activated protein kinase, a key regulator of insulin secretion, in MIN6 beta cells. *The FASEB journal : official publication Fed. Am. Soc. for Exp. Biol.* 24, 5080–5091, DOI: 10.1096/fj.10-166306 (2010).
43. Watson, C. T. *et al.* Integrative transcriptomic analysis reveals key drivers of acute peanut allergic reactions. *Nat. Commun.* 8, DOI: 10.1038/s41467-017-02188-7 (2017).
44. Illi, S. *et al.* The pattern of atopic sensitization is associated with the development of asthma in childhood. *J. Allergy Clin. Immunol.* 108, 709–714, DOI: 10.1067/mai.2001.118786 (2001).
45. Liu, A. H. *et al.* National prevalence and risk factors for food allergy and relationship to asthma: Results from the National Health and Nutrition Examination Survey 2005-2006. *J. Allergy Clin. Immunol.* 126, DOI: 10.1016/j.jaci.2010.07.026 (2010).
46. Wang, J., Visness, C. M. & Sampson, H. A. Food allergen sensitization in inner-city children with asthma. *J. Allergy Clin. Immunol.* 115, 1076–1080, DOI: 10.1016/j.jaci.2005.02.014 (2005).
47. Morton, N. M. *et al.* A stratified transcriptomics analysis of polygenic fat and lean mouse adipose tissues identifies novel candidate obesity genes. *PloS one* 6, e23944 (2011).
48. Peters, U., Dixon, A. E. & Forno, E. Obesity and asthma. *J. Allergy Clin. immunology* 141, 1169–1179 (2018).
49. Zariwala, M. A. *et al.* Mutations of dnai1 in primary ciliary dyskinesia: evidence of founder effect in a common mutation. *Am. journal respiratory critical care medicine* 174, 858–866 (2006).
50. Lucas, J. S. *et al.* Diagnosis and management of primary ciliary dyskinesia. *Arch. Dis. Child.* 99, 850–856, DOI: 10.1136/archdischild-2013-304831 (2014).
51. Lucas, J. S., Walker, W., Kuehni, C. & Lazor, R. Primary ciliary dyskinesia. *Orphan Lung Dis. Eur Respir Monogr* 54, 201–217 (2011).
52. Parsons, S. J. & Parsons, J. T. Src family kinases, key regulators of signal transduction. *Oncogene* 23, 7906–7909 (2004).
53. Roskoski Jr, R. Src protein-tyrosine kinase structure, mechanism, and small molecule inhibitors. *Pharmacol. research* 94, 9–25 (2015).
54. Jeon, S.-M. Regulation and function of ampk in physiology and diseases. *Exp. & molecular medicine* 48, e245–e245 (2016).
55. Chang, T. W., Wu, P. C., Hsu, C. L. & Hung, A. F. Anti-ige antibodies for the treatment of ige-mediated allergic diseases. *Adv. immunology* 93, 63–119 (2007).
56. Harris, J. M., Wong, D. A. & Kapp, A. V. Development of the asthma control composite outcome measure to predict omalizumab response. *Annals Allergy, Asthma & Immunol.* 107, 273–280 (2011).
57. Kuo, C.-H. S. *et al.* T-helper cell type 2 (Th2) and non-Th2 molecular phenotypes of asthma using sputum transcriptomics in U-BIOPRED. *Eur. Respir. J.* 49, DOI: 10.1183/13993003.02135-2016 (2017).
58. Modena, B. D. *et al.* Gene expression in relation to exhaled nitric oxide identifies novel asthma phenotypes with unique biomolecular pathways. *Am. J. Respir. Critical Care Medicine* 190, 1363–1372, DOI: 10.1164/rccm.201406-1099OC (2014).
59. Rossios, C. *et al.* Sputum transcriptomics reveal upregulation of IL-1 receptor family members in patients with severe asthma. *J. Allergy Clin. Immunol.* 141, 560–570, DOI: 10.1016/j.jaci.2017.02.045 (2018).
60. Baines, K. J., Simpson, J. L., Wood, L. G., Scott, R. J. & Gibson, P. G. Transcriptional phenotypes of asthma defined by gene expression profiling of induced sputum samples. *J. Allergy Clin. Immunol.* 127, DOI: 10.1016/j.jaci.2010.10.024 (2011).
61. Berthon, B. S., Gibson, P. G., Wood, L. G., MacDonald-Wicks, L. K. & Baines, K. J. A sputum gene expression signature predicts oral corticosteroid response in asthma. *The Eur. respiratory journal* 49, DOI: 10.1183/13993003.00180-2017 (2017).
62. Metselaar, P. I. *et al.* Recursive ensemble feature selection provides a robust mrna expression signature for myalgic encephalomyelitis/chronic fatigue syndrome. *Sci. reports* 11, 1–11 (2021).
63. Ciravegna, G. *et al.* Logic explained networks. *arXiv preprint arXiv:2108.05149* (2021).
64. Barbiero, P. *et al.* Entropy-based logic explanations of neural networks. *arXiv preprint arXiv:2106.06804* (2021).
65. Hayden, M. S. & Ghosh, S. Nf-κb, the first quarter-century: remarkable progress and outstanding questions. *Genes & development* 26, 203–234 (2012).
66. Global initiative for asthma. Global strategy for asthma management and prevention. https://ginasthma.org/gina-reports/ (2022). Online; accessed 05 May 2022.
67. Hubbard, T. *et al.* The Ensembl genome database project. *Nucleic acids research* 30, 38–41, DOI: 10.1093/NAR/30.1.38 (2002).
68. Pedregosa FABIANPEDREGOSA, F. *et al.* Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot. *J. Mach. Learn. Res.* 12, 2825–2830 (2011).
69. Lopez-Rincon, A., Martinez-Archundia, M., Martinez-Ruiz, G. U., Schoenhuth, A. & Tonda, A. Automatic discovery of 100-miRNA signature for cancer classification using ensemble feature selection. *BMC Bioinforma. 2019 20:1* 20, 1–17, DOI: 10.1186/S12859-019-3050-8 (2019).
70. Barbiero, P., Ciravegna, G., Georgiev, D. & Giannini, F. Pytorch, explain! a python library for logic explained networks. *arXiv preprint arXiv:2105.11697* (2021).

# Tables

**Table 1. Restricted set of features to predict responsiveness of Omalizumab (anti-IgE therapy)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Publication | Gene | Protein | Cell | Functional context | Key findings |
| Chairakaki et al., 2018  Vroman et al., 2017 | *CLEC4C* | C-type lectin domain family 4 member C | pDC | Mucosal and anti-viral immunity (src pathway) | -implicated in driving acute asthma exacerbations  -Tolerogenic effect in asthma by inducing Treg cell differentiation |
| Firat-Karalar et al., 2014  Thomas, B et al., 2010 | *CCDC113* | Coiled-coil domain-containing protein 113 | Epithelial cells | Mucociliary clearance | -Vital for ciliogenesis  -Reduction in cilium formation in knockdown models |
| Dirami, T et al., 2013 | *SLC26A8* | Solute Carrier Family 26 Member 8 | Sperm cells | Male fertility | -Associated with sperm motility  -Mutations may cause male infertility |
| Watson, C et al ., 2017 | PPP1R3D | Protein Phosphatase 1 Regulatory Subunit 3D | Blood cells | Cell metabolism (AMPK pathway) | -Regulates protein serine/threonine phosphatase activity  -Driver for acute peanut allergic responses |
| Zariwala, M. A et al., 2006  Lucas, J.S et al., 2014 | LOC100131780 | Non-annotated | Epithelial cells | Possible: Mucociliary clearance and airway remodelling | -Overlaps partially with gene DNAI1  -May be linked to primary ciliary dyskinesia |

# *Src family kinases (SFKs) are non-receptor tyrosine kinases signaling coordinated cell proliferation, differentiation, apoptosis, migration, and metabolism.52, 53 5-adenosine monophosphate (AMP)-activated protein kinase (AMPK) signalling pathway coordinates cell growth, autophagy and metabolism.54*

# Figure legends

**Figure 1. Heatmap of normalized gene expressions for the 5 selected genes in all samples identified by the REFS algorithm (algorithm 1).** The heatmap legend displays a color gradient range where -2 denotes the lowest gene expression and 4 the highest gene expression. Merely from visual inspection, samples can be differentiated into two groups: responders and non-responders to omalizumab treatment.

**Figure 2. ROC curve over all 8 classifiers from REFS** **to validate the identified 5-gene signature.** The ROC curve shows the binarization threshold from 0 (all moderate-to-severe asthma patients as omalizumab responders and both the TPR and FPR =1; upper right corner of ROC) to 1 (all moderate-severe asthma patients classified as non-responders to omalizumab and both TPR and FPR=0; lower left corner of ROC). The AUC is the area in the plot which stays under the ROC curve. The Passive aggressive classifier which produced the blue ROC curve shows the best predictive accuracy as it covers a larger area compared to the straight ROC curve with the random classifier (red dashed line).

**Figure 3.** **Summarized results of the two algorithms predicting treatment responsiveness of omalizumab in moderate-to-severe asthma.** Whole blood mRNA expression profiles in samples collected day 0 (1 week before the start of the treatment) were used for the REFS analysis (algorithm 1). In total 40 moderate-to-severe asthmatic patients, n= 30 responders, and n=10 non-responders were included. For LEN analysis (algorithm 2), n=17 healthy controls were also included. With REFS, five independent responsiveness-predictive genes are identified, whereas rule-based LEN identified three gene groups that predicted responsiveness. Comparing both approaches an overlap of four genes was found. The relationship between responder status (R/NR) is shown in the heatmap. The mRNA expression of responders compared to healthy controls is shown in the table.