


# A robust mRNA signature obtained via recursive ensemble feature selection predicts the responsiveness of omalizumab in moderate-to-severe asthma

Sarah Kidwai<sup>1</sup>  | Pietro Barbiero<sup>2</sup> | Irma Meijerman<sup>1</sup> | Alberto Tonda<sup>3</sup> | Paula Perez-Pardo<sup>1</sup> | Pietro Lio<sup>2</sup> | Anke H. van der Maitland-Zee<sup>4</sup> | Daniel L. Oberski<sup>5</sup> | Aletta D. Kraneveld<sup>1</sup> | Alejandro Lopez-Rincon<sup>1,5</sup>

<sup>1</sup>Division of Pharmacology, Utrecht Institute for Pharmaceutical Science, Faculty of Science, Utrecht University, Utrecht, The Netherlands

<sup>2</sup>Department of Computer Science and Technology, University of Cambridge, Cambridge, UK

<sup>3</sup>UMR 518 MIA, INRAE, Université Paris-Saclay, Paris, France

<sup>4</sup>Department of Pulmonary Medicine, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands

<sup>5</sup>Department of Data Science, University Medical Center Utrecht, Utrecht, The Netherlands

## Correspondence

Alejandro Lopez-Rincon.  
Email: [a.lopezrincon@uu.nl](mailto:a.lopezrincon@uu.nl)

## Abstract

**Background:** Not being well controlled by therapy with inhaled corticosteroids and long-acting  $\beta_2$  agonist bronchodilators is a major concern for severe-asthma patients. The current treatment option for these patients is the use of biologicals such as anti-IgE treatment, omalizumab, as an 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 was reduced from 28,402 genes to 5 genes while obtaining a cross-validated accuracy of 0.975. The 5 responsiveness predictive genes encode 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 in clinical settings. Future studies in relevant cohort data should validate our computational approach.

## KEYWORDS

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

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. Clinical and Translational Allergy published by John Wiley & Sons Ltd on behalf of European Academy of Allergy and Clinical Immunology.

## 1 | INTRODUCTION

Asthma is a common chronic respiratory disease, affecting more than 300 million individuals worldwide.<sup>1</sup> Around 50%–70% of the total asthmatic population represents individuals with moderate-to-severe asthma.<sup>2,3</sup> This form of asthma is difficult to treat, resulting in frequent asthma exacerbations, hospitalizations, and even death.<sup>4,5</sup> Patients diagnosed with moderate-to-severe asthma cannot always control their disease with long acting inhaled  $\beta$  2-agonists and corticosteroids.<sup>6</sup> As an add-on therapy for these patients, a biological such as omalizumab<sup>7</sup> can be prescribed. This anti-IgE agent specifically binds free IgE and prevents IgE binding to the high affinity IgE receptor Fc $\epsilon$ RI expressed on immune cells such as mast cells, basophils and antigen-presenting cells.<sup>8</sup> It reduces allergic airway inflammation by decreasing IgE levels in blood, and by consequent downregulation of the expression of Fc $\epsilon$ RI on airway inflammatory cells and lowers eosinophil counts.<sup>9–11</sup>

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),<sup>12</sup> anti-IL-5R (benralizumab)<sup>13</sup> and anti-IL-4R (dupilumab)<sup>14</sup> 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 costs that arise due to the use of improper or suboptimal medications. However, predictors of response to omalizumab are still limited.<sup>15–17</sup> 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<sup>18</sup> or clinical measurements (e.g., total serum IgE, FEV1).<sup>19–21</sup> 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 relatively small number of samples, making it difficult to analyse using standard univariate techniques.<sup>22,23</sup>

Using ML gene cluster analysis, Upchurch et al. (2020)<sup>6</sup> 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),<sup>22</sup> 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.<sup>24,25</sup> REFS allows the identification of compact gene signatures. Logic explained networks (LEN), an advanced computational algorithm, gives in contrast to other algorithms<sup>26</sup> an interpretable result by providing an explanation for the possible relationships between the features. It simplifies how to read the mRNA expression as a set of logical rules which may serve to take clinical action/decision. Measuring mRNA expression in patients' whole blood could provide 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.<sup>24,25</sup>

We implemented the two-state-of-art computational algorithm in ML that is, 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 those used by Upchurch et al. (2020)<sup>6</sup> to find predictors of omalizumab response. Single Response-predictive genes were identified through ML-based REFS and validated with 10-fold cross-validation.<sup>24,25</sup> In addition, we used LEN<sup>26</sup> which identifies gene groups and their interrelationship as predictors of omalizumab response. The LEN algorithm provides rules on sets of mRNA expression compared with 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.

## 2 | 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).<sup>6</sup>

The GSE134544 database has 239 blood mRNA samples from 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 the 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 using an 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)<sup>6</sup> selected asthmatic patients who had uncontrolled asthma despite treatment with inhaled corticosteroids (ICS) and/or long-acting  $\beta$ -agonists (LABA). Patients and non-asthmatic healthy controls<sup>6</sup> were recruited under protocols approved by the Institutional Review Board.

According to the Global Initiative for Asthma guideline,<sup>27</sup> 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.<sup>6</sup>

Omalizumab responsiveness was defined by Upchurch et al (2020)<sup>6</sup> 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  $\beta$ -agonists ( $\geq 2 \times$  per week) and night-time sleep disruption ( $\geq 2$  per week), unchanged asthma control through medication, and indications of little/no improvements in asthma by physicians.<sup>6</sup>

## 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.<sup>28</sup> This step was done only with the final short list of genes. Next, the data were normalized with z-score normalization.<sup>29</sup>

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

### 2.3.1 | Algorithm 1: REFS

REFS<sup>30</sup> is an algorithm to discover features (genes) in our study. REFS was used to identify a gene signature in moderate-to-severe asthma patients to separate omalizumab responders from non-responders. The REFS algorithm decreases the number of genes in the dataset to the most important genes associated with treatment outcome. REFS uses 8 different classifier algorithms: Gradient boosting, Passive Aggressive, Logistic Regression, Support Vector Machine classifier

(SVC), Random Forest, Stochastic Gradient Descent, Ridge and Bagging. REFS merges the results of these individual classifiers. From these classifiers (metrics for algorithm performance), important genes are extracted, ranked and the lowest-scoring genes are removed.

The REFS algorithm gives each gene associated with treatment outcome a 'score' based on how each different classifier algorithm uses it. For instance, in a tree-based algorithm, the score is based on how many times a feature/gene was found in the tree. In the case of a coefficient-based algorithm, the scoring depends on the coefficient value, with the highest weight being the most relevant.

As REFS is a stochastic algorithm, it was iterated 10 times. Running the algorithm several times with the same settings produces slightly different results and is thus a method to obtain more reliable results. After every iteration, the most relevant gene-signature for therapy response is selected until the algorithm reaches 1 feature with a 20% step reduction and a first cut-off of 1000 features.

The REFS algorithm was implemented on 28,402 genes and validated through 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

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 number of features. In other words, REFS sees feature selection and classification as separate, while LEN 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.

### 2.4.1 | Algorithm 2: LEN

The LENs algorithm<sup>31</sup> was implemented on 28,402 genes. LENs provide rules on sets of mRNA expression compared with healthy controls. It uses healthy controls as the baseline.

The LENs algorithm uses logic to transform gene expression values into qualitative descriptors that can be evaluated using a set of rules. LENs are a special family of concept-based neural networks providing first-order logic explanations for their decisions.<sup>31</sup> Concept-based models provide reasoning as in human-interpretable symbols (the *concepts*). As an example, a concept-based explanation to describe the category (*human*) can be through its characteristics (*head AND hand*), which leads to the explanation 'a human has *hands* and a *head*'.

Logic-based explanations may be merged to describe groups of observations. For instance, in an image dataset of humans, an image showing the anterior side of the face, a possible explanation to describe the category (*human*) can be through the attributes (*nose AND lips*), which reads 'being *human* implies having a *nose* AND *lips*'.

Similarly, another image showing the posterior side of the *human* can be described through the attributes (*feet AND hair AND ears*). An explanation for the class *human* is then ‘being *human* implies having *feet AND hair AND ears*. Both explanations can then be merged using the attributes ‘(*nose AND lips*) OR (*feet AND hair AND ears*)’. The merged explanation reads then as ‘being *human* implies having a ‘(*nose AND lips*) OR (*feet AND hair AND ears*)’.

The quality of logic-based explanations can be quantitatively measured to check their correctness and completeness (measurability). For instance, once the explanation “(*nose AND lips*) OR (*feet AND hair AND ears*)” is extracted for the class *human*, this logic formula can be applied to a test set to check its generality in terms of quantitative metrics like accuracy, and consistency.

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<sup>32</sup> implemented in the python package `pytorch_explain`.<sup>33</sup>

## 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 (Section 2.3.1) and LENs (Section 2.4.1). Protein function was investigated in the context of asthma pathology and omalizumab response.

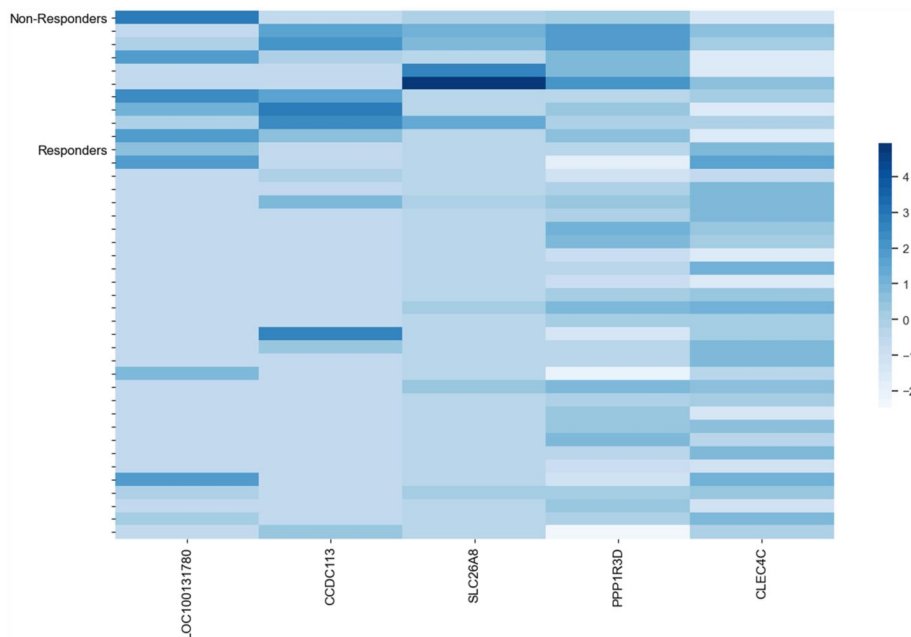
## 3 | RESULTS

### 3.1 | Feature selection

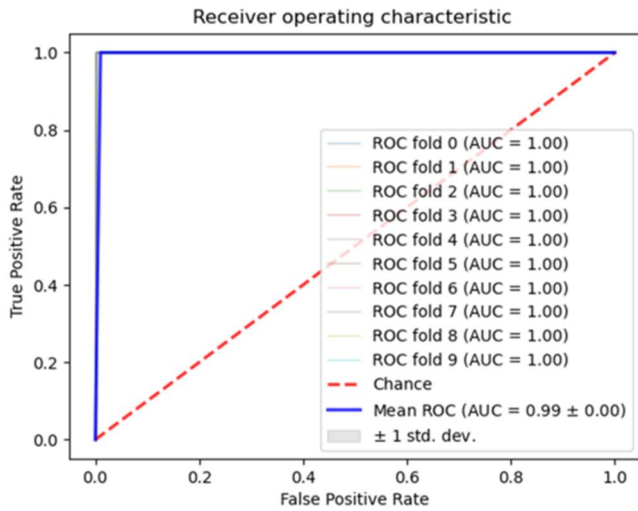
The GSE134544 dataset containing 40 samples was used to run the REFS algorithm ten times (see Figure 3). With the machine-learning REFS approach (Section 2.3.1) the total number of genes was reduced from 28,402 to 5 genes associated with therapeutic responsiveness. The optimal selection of genes depends on the classification accuracy of the binary problem (responders/non-responders). The 5-gene signature corresponds to the highest peak in accuracy over all 8 classifiers in the REFS ensemble (See, Figure S1). The gene expression values of responders, non-responders and non-asthmatic controls are shown in Figure 2.

### 3.2 | Genes of interest

REFS (Section 2.3.1) identified 5 individual genes as predictors of omalizumab response: *CCDC113*, *SLC26A8*, *PPP1R3D*, *CLEC4C* and *LOC100131780*. The mRNA expression levels of the identified 5 genes in non-responder patients' samples compared to responders are shown in Figure 1. Figure 1 shows that moderate-to-severe



**FIGURE 1** Heatmap of normalized gene expressions for the 5 selected genes in all samples identified by the Recursive Ensemble Feature Selection (REFS) algorithm (Section 2.3.1). 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** Receiver Operating Characteristic (ROC) curve over all 8 classifiers from Recursive Ensemble Feature Selection (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 area under the curve (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).

asthma patients who did not respond well to omalizumab have a higher mRNA expression of 4 out of 5 identified genes when compared to responders: *LOC100131780*, *CCDC113*, *SLC26A8* and *PPP1R3D*. The mRNA expression of one gene, *CLEC4C*, is down-regulated in non-responders compared to responders.

### 3.3 | ROC analysis of gene signature

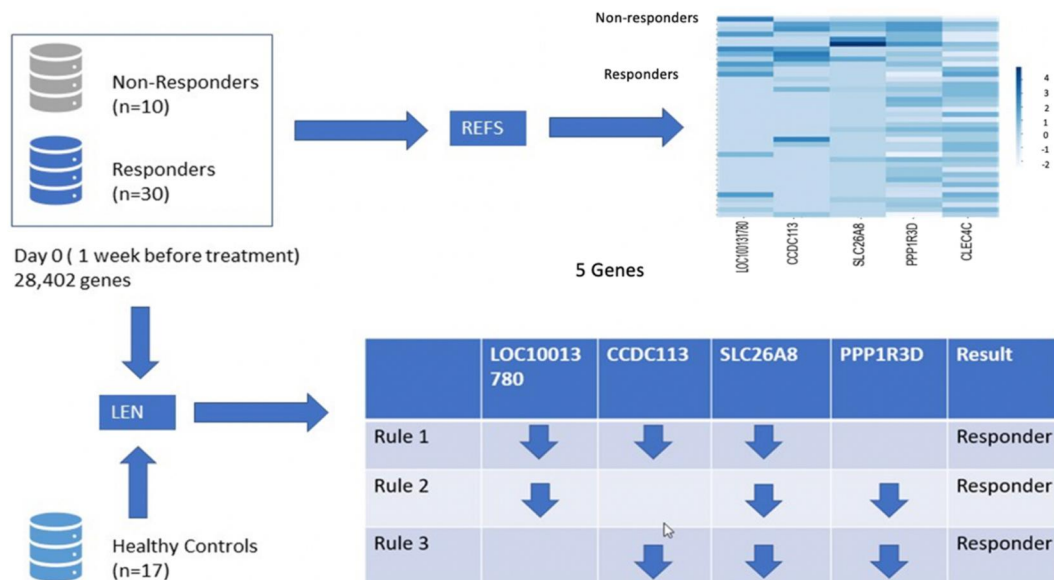
The AUC of the ROC curve was used to verify the efficacy/performance of the identified 5-gene signature over all 8 classifiers in 10-fold cross validation. As shown in Figure 2, the AUC of 0.99 was obtained using the Passive Aggressive classifier, a performance metric, which represents the best discriminatory accuracy for our model.<sup>34,35</sup>

### 3.4 | Relationship with treatment outcome: LEN

In parallel to the REFS algorithm (Section 2.3.1), a LEN on the same GSE134544 dataset was used to compare results (See, Figure 3).

As can be seen from Figure 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.



**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 Recursive Ensemble Feature Selection (REFS) analysis (Section 2.3.1). In total, 40 moderate-to-severe asthmatic patients,  $n = 30$  responders, and  $n = 10$  non-responders were included. For LEN analysis (Section 2.4.1),  $n = 17$  healthy controls were also included. With REFS, five independent responsiveness-predictive genes were 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 with healthy controls is shown in the table.



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 a combination of response-predictive gene set 2. The second rule implies that the mRNA expression of all genes in this trio is downregulated in the responder group as compared to the average gene expression of non-asthmatic healthy controls.

Rule 3 [CCDC113, SLC26A8 and PPP1R3D] indicates a combination of response-predictive gene set 3. The third rule implies that the mRNA expression of all genes in this trio is 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 (Section 2.3.1) and the rule-based LEN (Section 2.4.1) 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*, *SLC26A8* and *PPP1R3D*, which are all upregulated in non-responders and downregulated in responders. This is summarized in Figure 3.

### 3.6 | Biological interpretation of the gene signature predicting responsiveness to omalizumab

To determine the biological context of our findings, 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<sup>39</sup> for internalization/presentation to T cells<sup>40</sup> and attachment to serum IgG.<sup>41</sup> *CLEC4C* triggers the src-family protein- tyrosine kinases signaling pathways to inhibit the induction of IFN- $\alpha/\beta$  expression in plasmacytoid dendritic cells (pDC).<sup>30,39</sup> This 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-  $\alpha/\beta$ , and inflammatory cytokine production is critical in antiviral immune responses.<sup>42</sup> Hence, impaired anti-viral response in asthmatic patients may be coordinated by *CLEC4C*.<sup>43</sup> (See, Supporting Information S1) Taken together, the present set of findings indicate that severe asthma

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

Note: Src family kinases (SFKs) are non-receptor tyrosine kinases signaling coordinated cell proliferation, differentiation, apoptosis, migration, and metabolism.<sup>36,37</sup> 5-adenosine monophosphate (AMP)-activated protein kinase (AMPK) signalling pathway coordinates cell growth, autophagy and metabolism.<sup>38</sup>

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 a 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.<sup>44</sup> This gene coordinates ciliary beating frequency, beat direction and cilia stroke.<sup>45</sup> *CCDC113* has been shown to be overexpressed in the nasal brushes of asthmatic subjects when compared to healthy controls.<sup>46</sup> 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.<sup>47,48</sup> Therefore, it 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 the whole blood of non-responders compared with omalizumab responders might indicate ciliary defects in the airways of non-responders. Thus, as *CCDC113* is involved in the cilium assembly, it may simply increase the severity of asthma.

The *SLC26A8* gene encodes Testis anion transporter 1, 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.<sup>49</sup> In the whole blood mRNA of severe asthmatics, upregulated *SLC26A8* compared to healthy controls has been reported.<sup>50</sup> 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.<sup>51</sup> 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.<sup>52</sup> This gene is associated with the cellular energy sensor 5' AMP-activated protein kinase (AMPK).<sup>52</sup> The interaction between AMPK and PP1 is regulated by intracellular glycogen content.<sup>51</sup> Upregulated mRNA expression of *PPP1R3D* in blood cells is associated with IgE-mediated peanut allergies in children.<sup>53</sup> Food allergies are also associated with increased asthma severity in later life.<sup>54–56</sup> Interestingly, *PPP1R3D* has been reported as a novel obesity candidate gene.<sup>57</sup> Obese adults tend to have more severe asthma compared to lean adults.<sup>58</sup> Indeed, *PPP1R3D* was found to be 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 were not available for inclusion in the present study.

*LOC100131780* is an illumina gene ID, which could not be matched to a specific gene in the literature. As such, the association of *LOC100131780* 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<sup>59</sup> 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.<sup>60,61</sup> 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<sup>60</sup> patients with overexpressed *LOC100131780* classified as non-responders might be misdiagnosed with severe asthma instead of PCD.

## 4 | DISCUSSION

Pre-therapeutic screening helps in unwarranted drug exposure in patients suffering from severe forms of asthma. Treatment with a biological agent 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.<sup>21,62,63</sup> Studies examining transcriptional expression profiles in whole blood and sputum of asthmatic patients have identified gene signatures that may be associated with asthmatic phenotypes<sup>64–66</sup> or therapeutic responsiveness to ICS<sup>67,68</sup> and omalizumab.<sup>6,18</sup> However, no definitive set of biomarkers for omalizumab therapy response has been found till date, while omalizumab response prediction could be very beneficial for proper disease management. Therefore, better predictive biomarkers for omalizumab treatment response are needed to select the best treatment strategy for severe asthma patients.

ML prediction can help to find a small and easy to measure a set of biomarkers that predicts treatment responses. However, 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.<sup>22,23</sup> REFS algorithm offers a solution for this issue<sup>25,69</sup> in terms of better accuracy and robustness. LEN can explain why the algorithm arrives at a certain decision, such that the results of the LEN algorithm can be interpreted by clinicians. Explainability matters especially in a clinical context because predictions are useful when they can be understood for the acceptance of the AI decisions by physicians. LENs<sup>26,31,32</sup> 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<sup>26</sup> 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. The gene expression values of 5 selected genes across healthy controls, non-responders, and responders are shown in Figure S2. Our compact set of genes was able to differentiate between responders and non-responders with a mean accuracy 0.975. As shown in Figure S2, after treatment with omalizumab, the gene expression levels of the 5 identified genes of responders were closer to the healthy controls.

Zhang et al (2021),<sup>18</sup> 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. 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, whereas conversely, some patients with low *CD3E* expression could respond to treatment. Identifying a short list of responsiveness predictive genes as in our results could be useful in clinical settings to show coordinated changes in gene expression that may impact the 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)<sup>18</sup> that reported an accuracy of 0.763. The utility and accuracy of REFS in comparison to other methodologies has been previously demonstrated in miRNA and mRNA datasets.<sup>24,25</sup>

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

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 the 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 NF $\kappa$ B which controls various aspects of innate and adaptive immune functions and serves as a critical mediator of inflammatory responses.<sup>70</sup> 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 impaired immune activation or dysregulation of the immune response. Zhang et al (2021),<sup>18</sup> hypothesized that omalizumab inhibits airway inflammation by reducing the Th2 inflammation cascade. By calculating the immune enrichment score, they found that CD4+ T and dendritic cell numbers were lower in the 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).<sup>6</sup> 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)<sup>18</sup> or Upchurch et al. (2020).<sup>6</sup> 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 biological function of this gene 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 the 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. Unfortunately, we could not find an independent cohort dataset to validate our gene signature. Not all pathways involved in our gene set have been fully characterized in the 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 samples 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 the mRNA expression of the 4 overlapping genes for future treatment with omalizumab. If mRNA expressions of all 4 overlapping genes are found to be downregulated in the responder group compared with 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 who should be treated with omalizumab versus another biological. To support physicians' treatment decision for a biological, 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.

#### AUTHOR CONTRIBUTIONS

**Sarah Kidwai:** Conceptualization (equal); Investigation (equal); Writing – original draft (lead). **Pietro Barbiero:** Methodology (lead); Validation (lead); Writing – review & editing (equal). **Irma Meijerman:** Supervision (lead); Writing – original draft (equal); Writing – review & editing (equal). **Alberto Tonda:** Formal analysis (lead); Methodology (equal); Writing – review & editing (equal). **Paula Perez-Pardo:** Resources (equal); Writing – original draft (equal); Writing – review & editing (equal). **Pietro Lio:** Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal). **Anke H. Maitland-van der Zee:** Conceptualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Daniel L. Oberski:** Supervision (lead); Writing – original draft (equal); Writing – review & editing (equal). **Aletta D. Kraneveld:** Conceptualization (lead); Supervision (lead); Writing – original draft (equal); Writing – review & editing (equal). **Alejandro Lopez-Rincon:** Conceptualization (equal); Methodology (lead); Visualization (lead); Writing – review & editing (equal).

#### ACKNOWLEDGEMENTS

We would like to thank 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.

#### CONFLICT OF INTEREST STATEMENT

All authors of the manuscript declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

#### DATA AVAILABILITY STATEMENT

All the necessary scripts to reproduce the experiments are stored on the public GitHub repository: <https://github.com/stephenwolf0/REFS>. We used clinical and array data from the Gene Expression Omnibus (GEO) with accession code GSE134544.

#### ORCID

Sarah Kidwai  <https://orcid.org/0000-0002-6457-2011>

#### REFERENCES

- Masoli M, Fabian D, Holt S, Beasley R, Global Initiative for Asthma (GINA) Program The global burden of asthma: executive summary of the gina dissemination committee report. *Allergy*. 2004;59(5):469-478. <https://doi.org/10.1111/j.1398-9995.2004.00526.x>
- Larsson K, Ställberg B, Lisspers K, et al. Prevalence and management of severe asthma in primary care: an observational cohort study in Sweden (pacehr). *Respir. Res.* 2018;19:1-10. <https://doi.org/10.1186/s12931-018-0719-x>
- FitzGerald JM, Barnes PJ, Chipps BE, et al. The burden of exacerbations in mild asthma: a systematic review. *ERJ open Res.* 2020;6(3):00359--02019. <https://doi.org/10.1183/23120541.00359-2019>
- Chung KF, Wenzel SE, Brozek JL, 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 Eur Respir.* 2014;43(2):343-373. <https://doi.org/10.1183/09031936.00202013>
- 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.* 2018;141(6):2256-2264. <https://doi.org/10.1016/j.jaci.2017.07.047>
- Upchurch K, Wiest M, Cardenas J, et al. Whole blood transcriptional variations between responders and non-responders in asthma patients receiving omalizumab. *Clin Exp Allergy.* 2020;50(9):1017-1034. <https://doi.org/10.1111/cea.13671>
- Incorvaia C, Mauro M, Makri E, Leo G, Ridolo E. Two decades with omalizumab: what we still have to learn. *Biol Targets & Ther.* 2018;12:135-142. <https://doi.org/10.2147/BTT.S180846>
- Belliveau PP. Omalizumab: a monoclonal anti-IgE antibody. *Med-scape Gen Med.* 2005;7:27.
- Busse W, Corren J, Lanier BQ, et al. Omalizumab, anti-IgE recombinant humanized monoclonal antibody, for the treatment of severe allergic asthma. *J Allergy Clin Immunol.* 2001;108(2):184-190. <https://doi.org/10.1067/mai.2001.117880>
- Chanez P, Contin-Bordes C, Garcia G, et al. Omalizumab-induced decrease of FcεRI expression in patients with severe allergic asthma. *Respir Med.* 2010;104(11):1608-1617. <https://doi.org/10.1016/j.rmed.2010.07.011>
- Plewako H, Arvidsson M, Petruson K, et al. The effect of omalizumab on nasal allergic inflammation. *J Allergy Clin Immunol.* 2002;110(1):68-71. <https://doi.org/10.1067/mai.2002.125488>
- Eger K, Kroes JA, Ten Brinke A, Bel EH. Long-term therapy response to anti-il-5 biologics in severe asthma—a real-life evaluation. *J Allergy Clin Immunol Pract.* 2021;9(3):1194-1200. <https://doi.org/10.1016/j.jaip.2020.10.010>
- Bleecker ER, FitzGerald JM, Chanez P, 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. *Lancet.* 2016;388(10056):2115-2127. [https://doi.org/10.1016/s0140-6736\(16\)31324-1](https://doi.org/10.1016/s0140-6736(16)31324-1)
- Castro M, Corren J, Pavord ID, et al. Dupilumab efficacy and safety in moderate-to-severe uncontrolled asthma. *N Engl J Med.* 2018;378(26):2486-2496. <https://doi.org/10.1056/nejmoa1804092>

15. Bousquet J, Cabrera P, Berkman N, 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.* 2005;60(3):302-308. <https://doi.org/10.1111/j.1398-9995.2004.00770.x>
16. Busse WW, Massanari M, Kianifard F, Geba GP. 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.* 2007;23(10):2379-2386. <https://doi.org/10.1185/030079907X226258>
17. Walker S, Monteil M, Phelan K, Lasserson TJ, Walters EH. Anti-IgE for chronic asthma in adults and children. *Cochrane Database Syst Rev.* 2006. <https://doi.org/10.1002/14651858.cd003559.pub3>
18. Zhang Q, Li H, Gao S, et al. Cd3e as a new predictive biomarker of response to omalizumab treatment in asthma patients: evidence from bioinformatic analysis. *Int Immunopharm.* 2021;93:107423. <https://doi.org/10.1016/j.intimp.2021.107423>
19. Li B, Huang M, Huang S, 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.* 2020;12:7097-7105. <https://doi.org/10.21037/jtd-20-2073>
20. Riccio AM, Mauri P, De Ferrari L, 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.* 2020;13(1):100095. <https://doi.org/10.1016/j.waojou.2019.100095>
21. Bousquet J, Rabe K, Humbert M, et al. Predicting and evaluating response to omalizumab in patients with severe allergic asthma. *Respir Med.* 2007;101(7):1483-1492. <https://doi.org/10.1016/j.rmed.2007.01.011>
22. Ruberg SJ, Shen L. Personalized medicine: four perspectives of tailored medicine. *Stat Biopharm Res.* 2015;7(3):214-229. <https://doi.org/10.1080/19466315.2015.1059354>
23. Sechidis K, Papangelou K, Metcalfe PD, Svensson D, Weatherall J, Brown G. Distinguishing prognostic and predictive biomarkers: an information theoretic approach. *Bioinformatics.* 2018;34(19):3365-3376. <https://doi.org/10.1093/BIOINFORMATICS/BTY357>
24. Lopez-Rincon A, Martinez-Archundia M, Martinez-Ruiz GU, Schoenhuth A, Tonda A. Automatic discovery of 100-mirna signature for cancer classification using ensemble feature selection. *BMC Bioinf.* 2019;20:1-17. <https://doi.org/10.1186/s12859-019-3050-8>
25. Lopez-Rincon A, Mendoza-Maldonado L, Martinez-Archundia M, et al. Machine learning-based ensemble recursive feature selection of circulating mirnas for cancer tumor classification. *Cancers.* 2020;12(7):1785. <https://doi.org/10.3390/cancers12071785>
26. Barbiero P, et al. Entropy-based Logic Explanations of Neural Networks; 2021.
27. Global initiative for asthma. Global Strategy for Asthma Management and Prevention; 2022. accessed 05 May 2022. <https://ginasthma.org/gina-reports/>
28. Hubbard T, et al. The Ensembl genome database project. *Nucleic Acids Res.* 2002;30(1):38-41. <https://doi.org/10.1093/NAR/30.1.38>
29. Pedregosa F, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res.* 2011;12:2825-2830.
30. Lopez-Rincon A, Martinez-Archundia M, Martinez-Ruiz GU, Schoenhuth A, Tonda A. Automatic discovery of 100-miRNA signature for cancer classification using ensemble feature selection. *BMC Bioinf.* 2019;20:1-17. <https://doi.org/10.1186/S12859-019-3050-8>
31. Ciravegna G, et al. Logic Explained Networks; 2021. arXiv preprint arXiv:2108.05149.
32. Barbiero P, et al. Entropy-based Logic Explanations of Neural Networks; 2021. arXiv preprint arXiv:2106.06804.
33. Barbiero P, Ciravegna G, Georgiev D, Giannini F. Pytorch, Explain! a python Library for Logic Explained Networks; 2021. arXiv preprint arXiv:2105.11697.
34. Šimundić A-M. Measures of diagnostic accuracy: basic definitions. *EJIFCC.* 2009;19:203.
35. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol.* 2010;5(9):1315-1316. <https://doi.org/10.1097/jto.0b013e3181ec173d>
36. Parsons SJ, Parsons JT. Src family kinases, key regulators of signal transduction. *Oncogene.* 2004;23(48):7906-7909. <https://doi.org/10.1038/sj.onc.1208160>
37. Roskoski R, Jr. Src protein-tyrosine kinase structure, mechanism, and small molecule inhibitors. *Pharmacol Res.* 2015;94:9-25. <https://doi.org/10.1016/j.phrs.2015.01.003>
38. Jeon S-M. Regulation and function of ampk in physiology and diseases. *Exp & Mol Med.* 2016;48(7):e245. <https://doi.org/10.1038/emm.2016.81>
39. Riboldi E, Daniele R, Parola C, 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.* 2011;286(41):35329--35333. <https://doi.org/10.1074/jbc.c111.290494>
40. Dzionek A, Sohma Y, Nagafune J, 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  $\alpha/\beta$  induction. *J. Exp Med.* 2001;194(12):1823-1834. <https://doi.org/10.1084/jem.194.12.1823>
41. Jégouzo SA, Feinberg H, Dungarwalla T, Drickamer K, Weis WI, Taylor ME. Anovelmechanismforbindingofgalactose-terminatedglycansbythec-typecarbohydraterecognition domain in blood dendritic cell antigen 2. *J Biol Chem.* 2015;290(27):16759-16771. <https://doi.org/10.1074/jbc.m115.660613>
42. Petes C, Odoardi N, Gee K. The toll for trafficking: toll-like receptor 7 delivery to the endosome. *Front Immunol.* 2017;8:1075. <https://doi.org/10.3389/fimmu.2017.01075>
43. Murray LM, Yerkovich ST, Ferreira MA, Upham JW. Risks for cold frequency vary by sex: role of asthma, age, tlr7 and leukocyte subsets. *Eur Respir J.* 2020;56(4):1902453. <https://doi.org/10.1183/13993003.02453-2019>
44. Firat-Karalar EN, Sante J, Elliott S, Stearns T. Proteomic analysis of mammalian sperm cells identifies new components of the centrosome. *J Cell Sci.* 2014;127:4128-4133. <https://doi.org/10.1242/jcs.157008>
45. Bazan R, Schröfel A, Joachimiak E, Poprzeczko M, Pigino G, Wloga D. Ccdc113/ccdc96 complex, a novel regulator of ciliary beating that connects radial spoke 3 to dynein g and the nexin link. *PLoS Genet.* 2021;17:e1009388. <https://doi.org/10.1371/journal.pgen.1009388>
46. Pandey G, Pandey OP, Rogers AJ, et al. A nasal brush-based classifier of asthma identified by machine learning analysis of nasal rna sequence data. *Sci Rep.* 2018;8:1-15. <https://doi.org/10.1038/s41598-018-27189-4>
47. Wanner A, Salathé M, O'Riordan TG. Mucociliary clearance in the airways. *Am J Respir Crit Care Med.* 1996;154(6):1868-1902. <https://doi.org/10.1164/ajrccm.154.6.8970383>
48. Knowles MR, Boucher RC, et al. Mucus clearance as a primary innate defense mechanism for mammalian airways. *J. Clin Invest.* 2002;109(5):571-577. <https://doi.org/10.1172/jci0215217>
49. Dirami T, Rode B, Jollivet M, et al. Missense mutations in SLC26A8, encoding a sperm-specific activator of CFTR, are associated with human asthenozoospermia. *Am J Hum Genet.* 2013;92(5):760-766. <https://doi.org/10.1016/j.ajhg.2013.03.016>
50. Bigler J, Boedigheimer M, Schofield JPR, et al. A severe asthma disease signature from gene expression profiling of peripheral blood from u-biopred cohorts. *Am J Respir Crit Care Med.* 2017;195(10):1311-1320. <https://doi.org/10.1164/rccm.201604-0866oc>
51. Garcia-Haro L, Garcia-Gimeno MA, Neumann D, Beullens M, Bollen M, Sanz P. 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. *FASEB J Off Publ Fed Am Soc Exp Biol.* 2010;24(12):5080-5091. <https://doi.org/10.1096/fj.10-166306>
52. Armstrong CG, Browne GJ, Cohen P, Cohen PT. PPP1R6, a novel member: of the family of glycogen-targetting subunits of protein phosphatase 1. *FEBS Lett.* 1997;418(1-2):210-214. [https://doi.org/10.1016/S0014-5793\(97\)01385-9](https://doi.org/10.1016/S0014-5793(97)01385-9)
  53. Watson CT, Cohain AT, Griffin RS, et al. Integrative transcriptomic analysis reveals key drivers of acute peanut allergic reactions. *Nat Commun.* 2017;8(1):1943. <https://doi.org/10.1038/s41467-017-02188-7>
  54. Illi S, von Mutius E, Lau S, et al. The pattern of atopic sensitization is associated with the development of asthma in childhood. *J Allergy Clin Immunol.* 2001;108(5):709-714. <https://doi.org/10.1067/mai.2001.118786>
  55. Liu AH, Jaramillo R, Sicherer SH, 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.* 2010;126(4):798-806.e14. <https://doi.org/10.1016/j.jaci.2010.07.026>
  56. Wang J, Visness CM, Sampson HA. Food allergen sensitization in inner-city children with asthma. *J Allergy Clin Immunol.* 2005;115(5):1076-1080. <https://doi.org/10.1016/j.jaci.2005.02.014>
  57. Morton NM, Nelson YB, Michailidou Z, et al. A stratified transcriptomics analysis of polygenic fat and lean mouse adipose tissues identifies novel candidate obesity genes. *PLoS One.* 2011;6(9):e23944. <https://doi.org/10.1371/journal.pone.0023944>
  58. Peters U, Dixon AE, Forno E. Obesity and asthma. *J Allergy Clin Immunol.* 2018;141(4):1169-1179. <https://doi.org/10.1016/j.jaci.2018.02.004>
  59. Zariwala MA, Leigh MW, Ceppa F, et al. Mutations of *dnai1* in primary ciliary dyskinesia: evidence of founder effect in a common mutation. *Am J Respir Crit Care Med.* 2006;174(8):858-866. <https://doi.org/10.1164/rccm.200603-370oc>
  60. Lucas JS, Burgess A, Mitchison HM, Moya E, Williamson M, Hogg C. Diagnosis and management of primary ciliary dyskinesia. *Arch Dis Child.* 2014;99(9):850-856. <https://doi.org/10.1136/archdischild-2013-304831>
  61. Lucas JS, Walker W, Kuehni C, Lazor R. Primary ciliary dyskinesia. *Orphan Lung Dis Eur Respir Monogr.* 2011;54:201-217.
  62. Chang TW, Wu PC, Hsu CL, Hung AF. Anti-ige antibodies for the treatment of ige-mediated allergic diseases. *Adv Immunol.* 2007;93:63-119.
  63. Harris JM, Wong DA, Kapp AV. Development of the asthma control composite outcome measure to predict omalizumab response. *Ann Allergy Asthma Immunol.* 2011;107(3):273-280. <https://doi.org/10.1016/j.anai.2011.06.005>
  64. Kuo C-HS, Pavlidis S, Loza M, et al. T-helper cell type 2 (Th2) and non-Th2 molecular phenotypes of asthma using sputum transcriptomics in U-BIOPRED. *Eur Respir J.* 2017;49:1602135. <https://doi.org/10.1183/13993003.02135-2016>
  65. Modena BD, Tedrow JR, Milosevic J, et al. Gene expression in relation to exhaled nitric oxide identifies novel asthma phenotypes with unique biomolecular pathways. *Am J Respir Crit Care Med.* 2014;190(12):1363-1372. <https://doi.org/10.1164/rccm.201406-1099OC>
  66. Rossios C, Pavlidis S, Hoda U, et al. Sputum transcriptomics reveal upregulation of IL-1 receptor family members in patients with severe asthma. *J Allergy Clin Immunol.* 2018;141(2):560-570. <https://doi.org/10.1016/j.jaci.2017.02.045>
  67. Baines KJ, Simpson JL, Wood LG, Scott RJ, Gibson PG. Transcriptional phenotypes of asthma defined by gene expression profiling of induced sputum samples. *J Allergy Clin Immunol.* 2011;127(1):153-160. <https://doi.org/10.1016/j.jaci.2010.10.024> e9
  68. Berthon BS, Gibson PG, Wood LG, MacDonald-Wicks LK, Baines KJ. A sputum gene expression signature predicts oral corticosteroid response in asthma. *Eur. Respir J.* 2017;49(6):1700180. <https://doi.org/10.1183/13993003.00180-2017>
  69. Metselaar PI, Mendoza-Maldonado L, Li Yim AYP, et al. Recursive ensemble feature selection provides a robust mrna expression signature for myalgic encephalomyelitis/chronic fatigue syndrome. *Sci Rep.* 2021;11:1-11. <https://doi.org/10.1038/s41598-021-83660-9>
  70. Hayden MS, Ghosh S. Nf-kb, the first quarter-century: remarkable progress and outstanding questions. *Gene Dev.* 2012;26(3):203-234. <https://doi.org/10.1101/gad.183434.111>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Kidwai S, Barbiero P, Meijerman I, et al. A robust mRNA signature obtained via recursive ensemble feature selection predicts the responsiveness of omalizumab in moderate-to-severe asthma. *Clin Transl Allergy.* 2023;e12306. <https://doi.org/10.1002/ctt2.12306>