Modelling Asthma Patients' Responsiveness to Treatment Using Feature Selection and Evolutionary Computation

Alejandro Lopez-Rincon[1(✉)], Daphne S. Roozendaal[1], Hilde M. Spierenburg[1], Asta L. Holm[1], Renee Metcalf[1], Paula Perez-Pardo[1], Aletta D. Kraneveld[1], and Alberto Tonda[2]

[1] Division of Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Faculty of Science, Utrecht University, Universiteitsweg 99, 3584 CG Utrecht, The Netherlands
a.lopezrincon@uu.nl

[2] UMR 518 MIA, INRAE, Université Paris-Saclay, Paris, France
alberto.tonda@inrae.fr

Abstract. For several medical treatments, it is possible to observe transcriptional variations in gene expressions between responders and non-responders. Modelling the correlation between such variations and the patient's response to drugs as a system of Ordinary Differential Equations could be invaluable to improve the efficacy of treatments and would represent an important step towards personalized medicine. Two main obstacles lie on this path: (i) the number of genes is too large to straightforwardly analyze their interactions; (ii) defining the correct parameters for the mathematical models of gene interaction is a complex optimization problem, even when a limited number of genes is involved. In this paper, we propose a novel approach to creating mathematical models able to explain patients' response to treatment from transcriptional variations. The approach is based on: (i) a feature selection algorithm, set to identify a minimal set of gene expressions that are highly correlated with treatment outcome, (ii) a state-of-the-art evolutionary optimizer, Covariance Matrix Adaptation Evolution Strategy, applied to finding the parameters of the mathematical model characterizing the relationship between gene expressions and patient responsiveness. The proposed methodology is tested on real-world data describing responsiveness of asthma patients to Omalizumab, a humanized monoclonal antibody that binds to immunoglobulin E. In this case study, the presented approach is shown able to identify 5 genes (out of 28,402) that are transcriptionally relevant to predict treatment outcomes, and to deliver a compact mathematical model that is able to explain the interaction between the different genes involved.

Keywords: Omalizumab · Machine learning · Evolutionary computation · Asthma · Mathematical model

P. A. Castillo and J. L. Jiménez Laredo (Eds.): EvoApplications 2021, LNCS 12694, pp. 359–372, 2021.
https://doi.org/10.1007/978-3-030-72699-7_23

1 Introduction

When patients are treated with medical drugs, it is possible to observe a variation in their gene expression. Such transcriptional variation can potentially be correlated with the responsiveness to treatment, and this relationship can be described through a system of Ordinary Differential Equations (ODE).

Such a model would be important not only to explain the differences in treatment outcome, but also to provide indications to medical personnel on how to improve the therapy. For example, if a gene is shown to be overexpressed in non-responsive patients, with respect to responsive patients, doctors might devise a new therapy, combining the current treatment with substances that lower the expression of that particular gene.

Nevertheless, obtaining such an ODE system is not a straightforward process. The genes potentially correlated to drug response number in the tens of thousands, with possible complex interconnections in expression levels. Not only that, but even when a limited number of genes is identified, finding satisfying values of the parameters for the ODE system describing their interactions is a complex optimization problem that cannot be tackled through gradient-based techniques.

In this paper, we propose a novel methodology to obtain compact mathematical models describing the correlation between gene expression levels and responsiveness to treatment. The methodology combines a technique for feature selection [19,20], able to identify a small set of genes highly correlated with treatment outcome, and a state-of-the-art evolutionary optimizer [15] in order to find good values for the ODE system characterizing their interaction.

The presented approach is tested on real-world data from N=40 patients affected by moderate-to-severe asthma, treated with the recent anti-IgE drug Omalizumab (30 responsive and 10 non-responsive). The results show that the methodology is effective in identifying 5 genes that are highly correlated with responsiveness to treatment, and it is able to deliver an ODE system that can reliably describe their interaction, explaining the responsiveness of patients to the Omalizumab treatment.

2 Background

In this section, we introduce the minimal notions that are necessary to introduce the scope of our work.

2.1 Feature Selection

In machine learning (ML), feature selection (FS) is defined as the process of identifying the features of a data set in order to obtain a minimal, informative subset. Features may not be part of this subset for two main reasons: they might be unrelated to the underlying nature of the problem, just adding noise; they

might be heavily correlated with other features, adding no relevant information for the task. Applications range from face recognition [31] to medicine [35], and approaches can be divided into two categories [12]: filters that score features according to a criterion (often a statistical test); and recursive procedures (forward or backwards) that attempt to reduce the features to a small set of non-redundant ones [8,18].

In the scope of this work, we focus on recursive FS algorithms, in particular Recursive Ensemble Feature Selection (REFS). The method is a variation of Recursive Feature Elimination (RFE) [13] that scores the features in a 10-fold cross-validation scheme, using 8 different classifiers: gradient boosting, passive aggressive classifier, logistic regression, Support Vector Machine classifier (SVC), random forest, Stochastic Gradient Descent (SGD), ridge classifier and bagging. The lowest scoring features are removed from the analysis and the process is repeated until the overall classification accuracy drops below a given threshold. The use of an ensemble of classifiers reduces the effects of the inherent bias in each ML algorithm, thus delivering a more objective feature ranking. This technique has been applied successfully for problems involving both mRNA [20] and miRNA [19], featuring number of variables ranging from 1,046 to 54,675.

2.2 Omalizumab Treatment for Asthma Patients

Omalizumab is the first humanized monoclonal antibody that binds to immunoglobuline E (anti-IgE) prescribed to patients with moderate-to-severe allergic asthma who do not respond to inhaled corticosteroids and long-acting β2-agonist bronchodilators. Omalizumab works by specifically binding free serum IgE [29], which characterizes allergic asthma. IgE binds to high affinity receptors (FcϵRI) expressed on effector cells such as basophils and mast cells, but also on other immune cells like eosinophils, thereby triggering an inflammatory cascade through the release of inflammatory mediators [14]. By binding to the Fc region of IgE and forming IgE-antibody complexes, Omalizumab prevents the binding of allergen specific IgE to FcϵRI and the subsequent inflammatory allergy reaction [29]. Consequently, FcϵRI expression is reduced, leading to less immune activation. Furthermore, Omalizumab decreases eosinophil numbers found in the airway of asthmatic patients, although the mechanism through which this effect is achieved is not entirely understood [29]. Eosinophils and their derived proinflammatory mediators are major contributors to airway inflammation and damage [26]. Omalizumab's ability to combat long-term airway remodeling is still under investigation [14].

It is important to understand the mechanism of action of Omalizumab treatment and to research the differences in responsiveness. Potential transcriptional variations between responders (R) and non-responders (NR) to Omalizumab can function as predictive biomarkers in the future. A recent study by Upchurch et al. 2020, with accession number GSE134544 at gene expression omnibus (GEO) [30] investigated whole blood transcriptomes of moderate-to-severe asthma patients (N = 40; 30 responders (R) and 10 non-responders (NR)), over the course of

Omalizumab treatment. Blood was collected at day 0, 7, 42, 98, 182 where the treatment started at day 7, and day 0 marks one week before the treatment.

Total RNA was isolated from whole blood, and all samples passing quality control were then amplified, (biotin-)labelled and hybridized to Illumina HT-12 V4 BeadChips (Illumina). Subsequent differential gene expression analysis was performed using Welch's T-test for comparisons between R and NR, and transcriptional changes within each group were assessed using a paired T-test. However, direct comparisons between R and NR did not provide sufficient gene lists after multiple testing corrections. Therefore, whole blood mRNA signature differences between groups were characterized using a gene cluster strategy, or modular-level type analysis adopted from studies by Chaussabel et al. [7] and Banchereau et al. [3]. Hierarchical clustering of genes revealed 8 similarly expressed transcript clusters in R and NR (i.e. protein synthesis (1); T cell/NK cell/ cytotoxicity (2); hematopoiesis (3); cell cycle control/proliferation (4); T cell regulation and activation (5); monocytes (6); glucose metabolism (7) and inflammation (8)). Of these, cluster 2 and 7 in R were reported to be higher, while clusters 3 and 8 were higher in NR, suggesting that clusters 2, 3, 7 and 8 can be used as predictors of response to Omalizumab treatment. These clusters, combined, contain a total of 1,776 genes.

Due to the unpredictability of an asthma patient's responsiveness to Omalizumab, there have been multiple studies into finding a reliable biomarker that can act as a predictor. A recent study [16] reported that interleukin (IL)-9, IL-13, IL-25, IL-33 and thymic stromal lymphopoietin (TSLP) levels were significantly higher in R compared to NR. Most R were of a high type-2 cytokine endotype whereas only one NR was [16]. Data from the INNOVATE trial of Omalizumab for severe persistent asthma [17] was analyzed by [5], and this study found that of the biomarkers recorded by INNOVATE, only baseline total IgE levels were a predictor of efficacy. However, pooled analysis showed that treatment was effective in some cases irrespective of IgE levels. It was concluded that the most meaningful measure of responsiveness was the physician's overall assessment [5]. An extensive look into gene expression relating to asthma patients vs controls in different tissue types, disease severity and response to allergens and corticosteroid treatment in several datasets reported multiple gene signatures and pathways [1]. Although this may prove useful in explaining Omalizumab response, it revealed no significant gene overlap.

3 Proposed Approach

We present a new approach to obtain compact, human-readable mathematical models to explain responsiveness to treatment in patients. The methodology first applies feature selection to identify a small set of relevant genes, and then uses state-of-the-art evolutionary optimization to find the parameters of an ODE system that describes the relationship between gene expression levels and patient's responsiveness to treatment.

3.1 Feature Selection

In a first step, our objective is to select the most meaningful genes to correctly predict and model patients' responsiveness to treatment. We apply the REFS algorithm, which uses the feedback of an ensemble of classifiers to rank each feature depending on its usefulness for the process of classification. Then, the lowest-scoring features are removed, and the classification/ranking is repeated, until the average classification accuracy falls below a user-defined threshold.

3.2 Mathematical Modeling

Once a small subset of the genes is identified, we create a mathematical model that interconnects the gene expression given the values at different time points. While other solutions to model the correlation between the gene expression values and responsiveness are possible, like black-box machine learning, white-box models are preferred by practitioners, as they are commonly considered more interpretable. We assume to have whole blood mRNA expression at different points in time available, and we will consider the average value for R and NR over all samples at each point in time, for each category of patients. Thus, the mathematical model we propose is a system of ordinary differential equations (ODEs) in the form of Eq. 1:

$$\begin{aligned} \frac{dg_0}{dt} &= -k_0 u(t) + \alpha_0 e^{-\beta_0 t}, \\ &..., \\ \frac{dg_n}{dt} &= -k_n u(t) + \alpha_n e^{-\beta_n t}, \\ u(t) &= K_{g_0} g_0 + ... + K_{g_n} g_n, \end{aligned} \tag{1}$$

where g_0 to g_n will be the most important genes, k_i, α_i, β_i and K_{g_i} are coefficients calculated by each gene and $u(t)$ is an unknown function that interconnects the gene expression of all the genes, as to consider a relationship between all the variables.

Then, to solve the model, we use Euler's numerical method which transforms system 1 into:

$$\begin{aligned} g_0^t &= g_0^{t-1} + \triangle t \frac{dg_0}{dt}, \\ &..., \\ g_n^t &= g_n^{t-1} + \triangle t \frac{dg_n}{dt}, \\ u(t) &= K_{g_0} g_0^t + ... + K_{g_n} g_n^t. \end{aligned} \tag{2}$$

$u(t)$ models the interconnection between different gene expressions, that we hypothesize exists to avoid trivial assumptions of independence. As the problem is not treatable resorting to classical gradient-based techniques, it is necessary to use state-of-the-art stochastic optimization, such as CMA-ES [15], to find satisfying values for k_i, α_i, β_i and K_{g_i}.

Given the measurement of gene expressions at different instants $t = \{t_0, ..., t_N\}$, from Eq. 2 we can define the cost function to be minimized by CMA-ES as:

$$error = \sum_{i=0}^{n} |g^t_{CRi} - g^t_{Ri}| + |g^t_{CNRi} - g^t_{NRi}| \tag{3}$$

where g^t_{CRi} is the average gene i expression calculated at time t for R, g^t_{Ri} is the average gene i expression measured t time t for responders, g^t_{CNRi} is the average gene i expression calculated at time t for non-responders and g^t_{NRi} is the average gene i expression measured t time t for non-responders.

4 Experimental Evaluation

All the necessary code for the experiments has been developed in Python, using the `scikit-learn` package [25] for machine learning, and the `cma` package for CMA-ES. The code is hosted on the open GitHub repository:[1]

4.1 Data

Moderate-to-severe asthma patients were prescribed Omalizumab, based on the manufacturer's dosing table. Patient blood was collected at day 0, 7, 42, 98, 182 where the treatment started at day 7 with Omalizumab in 40 patients: 30 R and 10 NR. For each patient, for each sample, for each instant of time, the dataset contains information about 28,402 gene expression levels. All data was used as provided, at GEO accession code GSE134544 [30].

4.2 Feature Selection

Running the REFS algorithm previously described 10 times, we identified a set of 5 features (out of 28,402). This compact set can predict the Omalizumab responsiveness in patients with a mean accuracy of 0.975 in a 10-fold cross-validation, considering the binary classification problem (R/NR) with all classifiers in the REFS ensemble. As the REFS process is stochastic, it was iterated 10 times and the feature set corresponding to the highest peak in accuracy was selected (see Fig. 1). From the figure, it is interesting to notice how using all 28,402 features actually provides a lower mean classification accuracy (0.703). Classification algorithms, usually exploiting optimization heuristics, often show a lower performance when asked to explore a larger feature search space.

The resulting most significant features uncovered by the presented algorithms are $ILMN\_3286286$; $ILMN\_1775520$; $ILMN\_1656849$; $ILMN\_1781198$ and $ILMN\_1665457$ (Fig. 2). Details of the Illumina probes are further specified in Table 1.

To further validate the selected features, we computed the area under the curve (AUC) and receiver operating characteristic (ROC) curve in a 10-fold

[1] https://github.com/steppenwolf0/modelingEvolutionaryComputation.

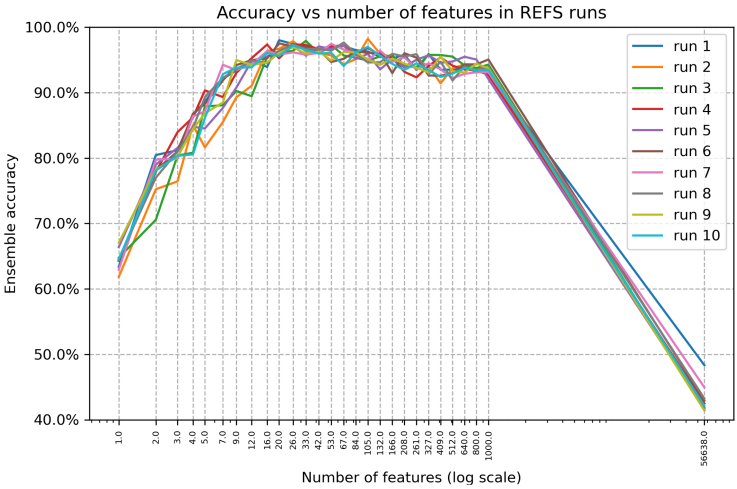

Fig. 1. The results of 10 runs of the REFS algorithm for the classification of Omalizumab responsiveness in allergic asthma patients. The x axis cuts at 5 variables, in correspondence with the highest peak.

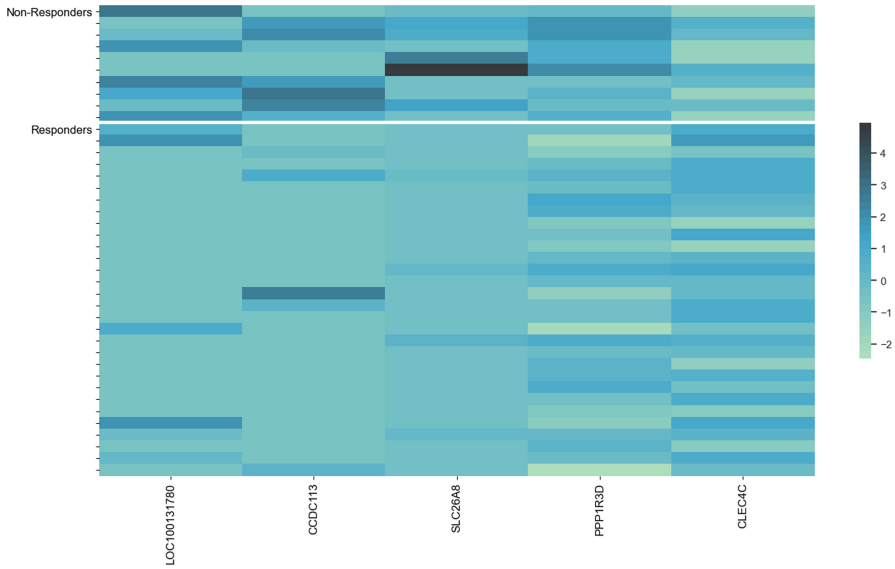

Fig. 2. Heatmap for the normalized gene expression for the 5 selected genes in all of the samples. Interestingly, samples of the two classes can be visually separated by just looking at the different normalized gene expression of the samples in the two groups.

Table 1. Information on the selected features to predict the responsiveness of the Omalizumab treatment.

Illumina probe	Corresponding gene	Sequence
ILMN_3286286	LOC100131780	GAGATTGCGAAACTGGACAAACTGCTGA ACCTGGACAGGGGCCAGGGCTG
ILMN_1775520	CCDC113	GGACATGAGAACATATTTCCAAGACAGA GGATTCTATGGGGACGGGTCAC
ILMN_1656849	SLC26A8	TGGGCGTATTGGGTTTGGGCTTCATTGC CACTTACCTTCCGGAGTCTGCA
ILMN_1781198	PPP1R3D	GGCCTTCACTGCTACGCCCCTGGCCCCA AAACAGAGAGCAAGACAGTTGT
ILMN_1665457	CLEC4C	GTGGTTCCAGTTGAAGGTCTGGTCCATG GCAGTCGTATCCATCTTGCTCC

cross-validation, using just the selected features, testing all classifiers in the REFS ensemble. The best AUC, 0.99, was obtained using Passive Aggressive classifier. This result is considered as an excellent diagnostic accuracy (AUC 0.9-1.0) by specialists of the field [21,27].

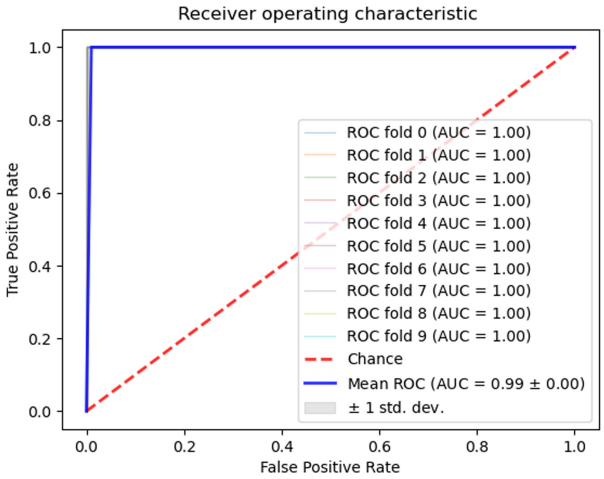

Fig. 3. ROC Curve in a 10-fold cross validation using Passive Aggressive classifier (the most effective in the REFS ensemble for this particular problem) for the 5 selected genes.

4.3 Mathematical Modeling

Considering we find 5 gene expressions to be the most meaningful, and using the template defined in Eq. 1, we can write the specific model linking gene expression levels to responsiveness to Omalizumab as:

$$
\begin{aligned}
\frac{dg_0}{dt} &= -k_0 u(t) + \alpha_0 e^{-\beta_0 t}, \\
\frac{dg_1}{dt} &= -k_1 u(t) + \alpha_1 e^{-\beta_1 t}, \\
\frac{dg_2}{dt} &= -k_2 u(t) + \alpha_2 e^{-\beta_2 t}, \\
\frac{dg_3}{dt} &= -k_3 u(t) + \alpha_3 e^{-\beta_3 t}, \\
\frac{dg_4}{dt} &= -k_4 u(t) + \alpha_4 e^{-\beta_4 t}, \\
u(t) &= K_{g_0} g_0 + K_{g_1} g_1 + K_{g_2} g_2 + K_{g_3} g_3 + K_{g_4} g_4, \qquad (4)
\end{aligned}
$$

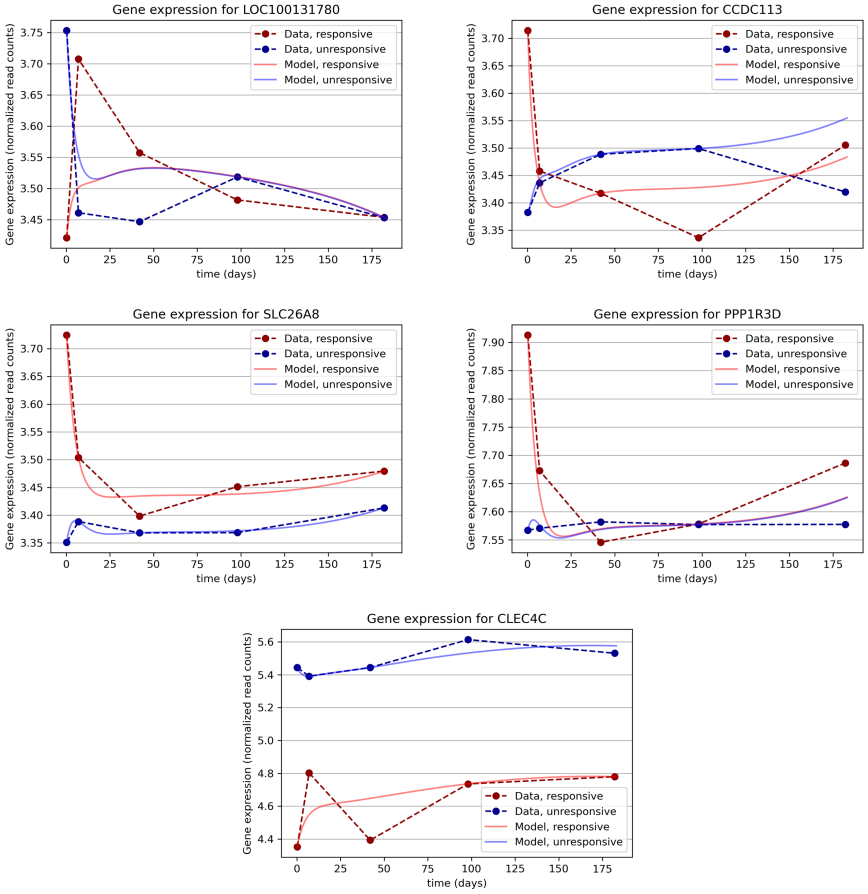

Fig. 4. Genes expression modeling and measured values for the 5 selected features at $t = 0, 7, 42, 98, 182$.

where we will need to find 20 parameters. We run CMA-ES with a $\sigma = 0.01$ and $\lambda = 1000$ with default stop conditions. Using CMA-ES with the cost function in Eq. 3, with $\triangle t = 0.25$, we find the following values; $k_0 = 0.5313, k_1 = 0.6440, k_2 = 0.4899, k_3 = 0.5504, k_4 = -0.4702, \alpha_0 = -0.0001, \beta_0 = -0.0168, \alpha_1 = -0.0386, \beta_1 = 0.9277, \alpha_2 = -0.0071, \beta_2 = 0.0733, \alpha_3 = -0.0176, \beta_3 = 0.1422, \alpha_4 = 0.0029, \beta_4 = 0.0079, K_{g_0} = 0.2598, K_{g_1} = 0.1078, K_{g_2} = 0.8069, K_{g_3} = -0.5708, K_{g_4} = 0.0565$ with an $error = 1.6036$ as the best of 20 runs.

As a baseline comparison, we also tested the `scipy` [32] implementation of the Nelder-Mead optimization algorithm [11], currently considered among the state-of-the-art for gradient-free optimization, for 20 runs. The best run of the Nelder-Mead algorithm yields a solution with $error = 4.3143$, of lower quality than that of CMA-ES.

From the results in Fig. 4, we can see that the model approximates considerably the behaviour of the genes, given the parameters and the initial values only; with a clear exception of $LOC100131780$. Although the increase of responders and decrease of non-responders functions are reflected, the amplitude does not match, therefore suggesting the necessity of increasing the degree of the answer or a more precise ODE solver.

Finally, from function $u(t)$ reported in Fig. 5, it is possible to notice that, just from its initial value $u(0)$, it is already possible to differentiate responders from non-responders, predicting the outcome of the treatment before its beginning.

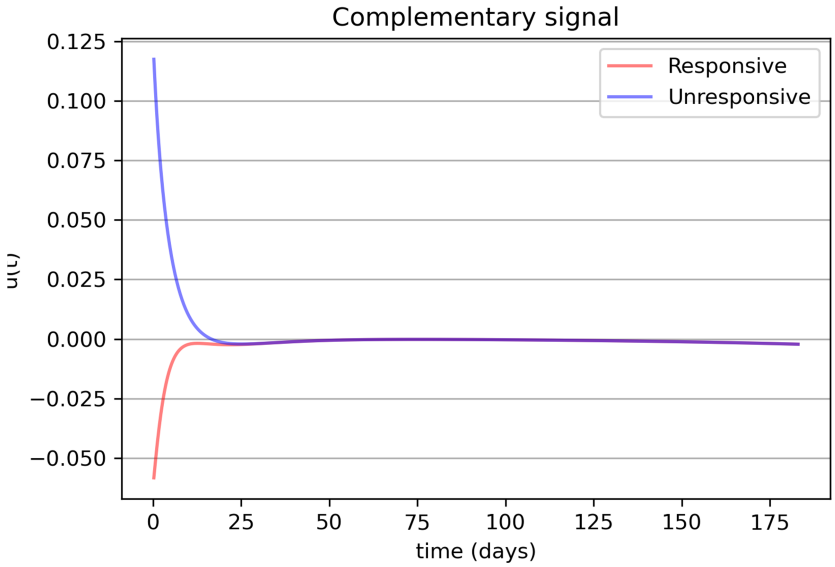

Fig. 5. Complementary signal $u(t) = K_{g_0} g_0 + K_{g_1} g_1 + K_{g_2} g_2 + K_{g_3} g_3 + K_{g_4} g_4$ that interconnects the expression values of all of the genes.

5 Discussion and Conclusions

Severe-to-moderate asthma patients appear to respond differently to the biological anti-IgE treatment Omalizumab. Total mRNA sequencing of whole blood from R and NR of Omalizumab treatment can help appreciate the differences at a transcriptional level. In this study, the GSE134544 dataset (that was recently published by Upchurch et al. [30]), was analyzed with the use of machine learning-based REFS. This novel study reveals 5 genes that are highly relevant in predicting Omalizumab responsiveness in asthma patients. In addition, we created a mathematical model to approximate the interdependence of the most significant genes to explain the effect of Omalizumab treatment using evolutionary computation.

Investigating gene function corresponding to the 5 mRNA (Table 1) predictors found in this study may illuminate new pathways involved in allergic asthma and the mechanism of Omalizumab resistance. For example, CCDC113 is vital for ciliogenesis and when knocked down, causes a reduction in cilium formation [10] which is previously related to severe asthma [28]. Also, this gene was found to be overexpressed in asthma patients as compared to controls [24], and in NR as compared to R.

CLEC4C is a marker for plasmacytoid dendritic cells (pDCs) subtypes [22, 23], which both have been implicated in driving acute asthma exacerbations [6, 33] and have shown to have a tolerogenic effect in asthma by inducing Treg cell differentiation [33]. CLEC4C is overexpressed in R when comparing against NR. PPP1R3D, a gene that codes for a subunit of PP1 (protein phosphatase 1) [2] regulates protein serine/threonine phosphatase activity and was found to be a causal key driver for acute peanut allergic response [34]. Genes coding for other subunits of PP1 have been shown to be upregulated in asthmatic patients (PPP1R16A) or, more specifically, in corticosteroid resistant patients (PPP1R15A) [1]. This gene is underexpressed in R, when compared against NR.

More unexpected in its relation to Omalizumab responsiveness in severe-to-moderate asthma is SLC26A8, which is most commonly associated with sperm motility and mutations that can cause male infertility [9]. It was also found to be upregulated in patients with severe asthma as compared to healthy controls [4]. Lastly, LOC100131780 is proprietary of Illumina and we, unfortunately, could not find a perfect gene match for this probe. The sequence did, however, overlap partially with DNAI1, which codes for dynein axonemal intermediate chain 1. The gene is strongly linked to primary ciliary dyskinesia (DNAI1 - Dynein Intermediate Chain 1, Axonemal - Homo Sapiens (Human) - DNAI1 Gene & Protein), which can cause respiratory infections and breathing problems (NHLBI).

An attempt to understand transcriptional variations between responders and non-responders on Omalizumab treatment was made by Upchurch et al. 2020, who recently analyzed changes in 8 gene clusters. In Omalizumab responders, the T cell/natural killer (NK) cell/cytotoxicity gene cluster and the glucose metabolism gene cluster were higher, whereas gene clusters involved in hematopoiesis and inflammation appeared to be higher in non-responders. 19 genes of the 8th (inflammatory) gene cluster were specifically annotated in their

paper [30]; however, none of these genes corresponded with the 5 most relevant genes we found with our methodology.

By setting a goal of predicting responsiveness instead of investigating transcriptional changes, the amount of data and complexity are reduced, since only the pre-testing time point needs to be considered. Using REFS provides the benefit of illuminating possible gene interactions without bias, as opposed to using a single algorithm or performing significance tests and clustering up- or down-regulated genes on the basis of known pathway function. The overall benefit of machine learning over basic statistics is that it is able to find predictive patterns which do not rely on assumptions about the data-generating system. Furthermore, it avoids the problem of eliminating significant mRNAs by not needing to take into account false discovery rate (FDR correction). Therefore, the technique used in this paper greatly reduced the number of meaningful mRNAs (5) compared to Upchurch et al. 2020 (1,776).

Further clinical testing and a replication dataset could reveal whether the 5 mRNAs proposed in this paper can reliably predict moderate-to-severe asthma patients' response to Omalizumab treatment on a larger scale. Implementing this sort of pre-treatment testing can both reduce the cost of asthma treatment, as Omalizumab is a relatively expensive drug, and prevent unnecessary and unproductive treatment time. Although currently, we limit our study to Omalizumab, it could be applied to similar studies.

This study not only proposes 5 specific genes that are transcriptionally relevant to predict Omalizumab responsiveness in moderate-to-severe asthma patients, but also puts forward a novel technique that aims to reduce the necessary information to the smallest set of whole blood mRNAs. The presented methodology elucidates the power of machine learning versus more general univariate/multivariate statistical analysis strategies.

References

1. Alrashoudi, R.H., Crane, I.J., Wilson, H.M., Al-Alwan, M., Alajez, N.M.: Gene expression data analysis identifies multiple deregulated pathways in patients with asthma. Biosci. Rep. **38**(6), (2018)
2. 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**(1–2), 210–214 (1997)
3. Banchereau, R., et al.: Personalized immunomonitoring uncovers molecular networks that stratify lupus patients. Cell **165**(3), 551–565 (2016)
4. Bigler, J., et al.: A severe asthma disease signature from gene expression profiling of peripheral blood from u-biopred cohorts. Am. J. Respir. Crit. Care Med. **195**(10), 1311–1320 (2017)
5. Bousquet, J., et al.: Predicting and evaluating response to omalizumab in patients with severe allergic asthma. Respir. Med. **101**(7), 1483–1492 (2007)
6. Chairakaki, A.D., et al.: Plasmacytoid dendritic cells drive acute asthma exacerbations. J. Allergy Clin. Immunol. **142**(2), 542–556 (2018)
7. Chaussabel, D., et al.: A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. Immunity **29**(1), 150–164 (2008)

8. Chien, Y., Fu, K.S.: On the generalized karhunen-loève expansion (corresp.). IEEE Trans. Inf. Theory **13**(3), 518–520 (1967)
9. Dirami, T., et al.: Missense mutations in slc26a8, encoding a sperm-specific activator of CFTR, are associated with human asthenozoospermia. Am. J. Hum. Genetics **92**(5), 760–766 (2013)
10. 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**(19), 4128–4133 (2014)
11. Gao, F., Han, L.: Implementing the nelder-mead simplex algorithm with adaptive parameters. Comput. Optim. Appl. **51**(1), 259–277 (2012)
12. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. J. Mach. Learn. Res. **3**, 1157–1182 (2003)
13. Guyon, I., Weston, J., Barnhill, S., Vapnik, V.: Gene selection for cancer classification using support vector machines. Machine learning **46**(1–3), 389–422 (2002)
14. Hamelmann, E.: The rationale for treating allergic asthma with anti-ige. Eur. Respir. Rev. **16**(104), 61–66 (2007)
15. Hansen, N., Ostermeier, A.: Completely derandomized self-adaptation in evolution strategies. Evol. Comput. **9**(2), 159–195 (2001). https://doi.org/10.1063/1.2713540
16. Huang, Y.C., Weng, C.M., Lee, M.J., Lin, S.M., Wang, C.H., Kuo, H.P.: Endotypes of severe allergic asthma patients who clinically benefit from anti-ige therapy. Clin. Exp. Allergy **49**(1), 44–53 (2019)
17. Humbert, M., et al.: Benefits of omalizumab as add-on therapy in patients with severe persistent asthma who are inadequately controlled despite best available therapy (gina 2002 step 4 treatment): Innovate. Allergy **60**(3), 309–316 (2005)
18. Lewis, P.: The characteristic selection problem in recognition systems. IRE Trans. Inform. Theory **8**(2), 171–178 (1962)
19. 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 Bioinform. **20**(1), 480 (2019)
20. Lopez-Rincon, A., et al.: Machine learning-based ensemble recursive feature selection of circulating mirnas for cancer tumor classification. Cancers **12**(7), 1785 (2020)
21. Mandrekar, J.N.: Receiver operating characteristic curve in diagnostic test assessment. J. Thorac. Oncol. **5**(9), 1315–1316 (2010)
22. Murray, L., Xi, Y., Upham, J.W.: Clec4c gene expression can be used to quantify circulating plasmacytoid dendritic cells. J. Immunol. Methods **464**, 126–130 (2019)
23. 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. (2020)
24. Pandey, G., et al.: A nasal brush-based classifier of asthma identified by machine learning analysis of nasal RNA sequence data. Sci. Rep. **8**(1), 1–15 (2018)
25. Pedregosa, F., et al.: Scikit-learn: machine learning in python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
26. Possa, S.S., Leick, E.A., Prado, C.M., Martins, M.A., Tibério, I.F.L.C.: Eosinophilic inflammation in allergic asthma. Front. Pharmacol. **4**, 46 (2013)
27. Šimundić, A.M.: Measures of diagnostic accuracy: basic definitions. Ejifcc **19**(4), 203 (2009)
28. Thomas, B., et al.: Ciliary dysfunction and ultrastructural abnormalities are features of severe asthma. J. Allergy Clin. Immunol. **126**(4), 722–729 (2010)

29. Thomson, N.C., Chaudhuri, R.: Omalizumab: clinical use for the management of asthma. Clinical Medicine Insights: Circulatory, Respiratory and Pulmonary Medicine 6, CCRPM-S7793 (2012)
30. Upchurch, K., et al.: Whole blood transcriptional variations between responders and non-responders in asthma patients receiving omalizumab. Clinical & Experimental Allergy (2020)
31. Vignolo, L.D., Milone, D.H., Scharcanski, J.: Feature selection for face recognition based on multi-objective evolutionary wrappers. Expert Syst. Appl. **40**(13), 5077–5084 (2013)
32. Virtanen, P., et al.: SciPy 1.0 Contributors: SciPy 1.0: fundamental algorithms for scientific computing in python. Nature Methods **17**, 261–272 (2020). https://doi.org/10.1038/s41592-019-0686-2
33. Vroman, H., Hendriks, R.W., Kool, M.: Dendritic cell subsets in asthma: impaired tolerance or exaggerated inflammation? Front. Immunol. **8**, 941 (2017)
34. Watson, C., et al.: Integrative transcriptomic analysis reveals key drivers of acute peanut allergic reactions. Nat. Commun. **8**(1), 1–13 (2017)
35. Zhou, Z., Li, S., Qin, G., Folkert, M., Jiang, S., Wang, J.: Multi-objective based radiomic feature selection for lesion malignancy classification. IEEE J. Biomed. Health Inform. (2019)