

RESEARCH

Open Access



TGF- β gene polymorphisms as risk factors for asthma control among clinic patients

Panek Michał^{1*}, Stawiski Konrad² and Kuna Piotr¹

Abstract

Background: TGF- β and its receptors play a crucial role in asthma pathogenesis, bronchial hyperreactivity, and bronchial remodeling. Expression of isoforms 1–3 of TGF β cytokine is influenced by tagging polymorphisms in the TGF β 1, TGF β 2 and TGF β 3 gene, and these SNPs may be associated with the risk of asthma development and severity as well as with other diseases. Polymorphic forms of TGF- β 1, TGF- β 2 and TGF- β 3 genes regulate the degree of bronchial inflammation, deterioration of lung functional parameters in spirometry and elevated level of total IgE. All this results in intensification of disease symptoms. According to current GINA 2020 guidelines, the Asthma Control Test (ACT™) should be applied to assess asthma symptoms.

Methods: An analysis of polymorphisms localized in TGF- β 1, TGF- β 2 and TGF- β 3 genes was conducted on 652 DNA samples with an application of the MassARRAY® system using the mass spectrometry technique MALDI TOF MS. The degree of asthma control was evaluated with ACT™.

Results: The occurrence of the T / C genotype in rs8109627 ($p = 0.0171$) in the TGF- β 1 gene is significantly associated with a higher ACT result (controlled asthma) in a multivariate linear regression analysis model after using backward stepwise selection of variables. In addition, in the linear model for prediction of ACT score we showed SNP rs8109627 ($p = 0.0497$) in the TGF- β 1 gene (improvement of the disease control - controlled asthma) and rs2796822 ($p = 0.0454$) in the TGF- β 2 gene (deterioration of the disease control - uncontrolled asthma) significantly modify the degree of asthma control.

Discussion: We described clinical significance of two SNPs in two genes TGF- β 1 and TGF- β 2, as yet unknown. We proved that the use of both genotypes and MAC allows to create a moderately correct prognostic model which is about 70% efficient on the entire set of analyzed SNPs in TGF- β 1, TGF- β 2, and TGF- β 3 genes.

Keywords: Asthma, TGF β 1, TGF β 2 and TGF β 3 gene, SNPs, Inflammation, Asthma control, The asthma control test (ACT™)

Background

Asthma is a chronic, inflammatory and heterogeneous respiratory disease. Many genes, inflammatory cells, and substances secreted by them are involved in the etiopathogenesis of this disease. The clinical picture depends on complex gene-gene and gene-environment interactions. These components make asthma a dynamic

disease, and its classifications can change both during treatment and irrespective of it. The etiology of bronchial asthma is multifactorial. The disease develops under the influence of variables, which, according to the report of the *Global Strategy for Asthma Management and Prevention* (GINA) can be divided into those which induce asthma and those which trigger its clinical symptoms. According to current GINA 2019 guidelines, the Asthma Control Questionnaire (ACQ) or the Asthma Control Test (ACT) can be used to assess its symptoms. The severity of the disease is assessed not based on the

* Correspondence: michalmp@poczta.onet.pl

¹Department of Internal Medicine, Asthma and Allergy, Medical University of Łódź, ul. Kopcińskiego 22, 90-153 Łódź, Poland

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

intensity of symptoms before the treatment is initiated, but only after several months of the treatment, when the level of intensity necessary to achieve and maintain asthma control has been established [1–4]. TGF- β (transforming growth factor β) is the most important gene responsible for bronchial hyperreactivity and the development of structural changes in bronchial walls is [5–8]. The key role of TGF- β as a growth factor responsible for prolonged reparatory processes and activation of myofibroblasts has been demonstrated in many studies on airway inflammation and its bronchial remodeling in asthma [5–12]. Increased TGF β expression has been observed in asthma, chronic obstructive pulmonary disease, idiopathic pulmonary fibrosis, cystic fibrosis, systemic scleroderma, non-alcoholic steatohepatitis, renal failure, myocardial infarction, or heart failure. It should be mentioned here that deregulation of TGF β cell signaling homeostasis is the main mechanism initiating and regulating fibrosis progression in various tissues. TGF β is a pleiotropic cytokine that regulates cell proliferation, contributes to epithelial-to-mesenchymal transition (EMT), suppresses the function of immune cells, compromising immune response, contributes to the conversion of fibroblasts to myofibroblasts, and causes overproduction of extracellular matrix (ECM) in tissues undergoing fibrosis. TGF β upregulates the expression and synthesis of many matrix proteins, primarily through the recruitment of myofibroblasts. Proteins upregulated by TGF β include collagens I - V, basement membrane proteins (laminin, entactin, perlecan) and ECM proteins (fibronectin, osteopontin, thrombospondin, tenascin, osteonectin/SPARC, elastin, biglycan, decorin, and hyaluronan). Additionally, at the early stages of fibrosis, TGF β stimulates myofibroblasts and other stromal cells to enhance the synthesis of collagen crosslinking enzymes, particularly lysyl oxidase, which increases the rigidity of the collagen network. Moreover, TGF β downregulates the synthesis of matrix-depleting proteins, such as matrix metalloproteinases (MMP-1, -8, -13). As a result, the increase in matrix protein synthesis and decrease in matrix proteinase activity, owing to the TGF β activity, contribute to the remodeling of the bronchial ECM and can result in a fibrotic response [11–16]. Polymorphic forms of the TGF β gene can significantly contribute to the development of asthma, induce disease progression and its complications, and have a significant clinical impact on symptom control. Single Nucleotide Polymorphism (SNP) rs8179181 in the TGF β 1 gene induces childhood asthma and atopy development. It increases the risk of a severe course of the disease and correlates with high levels of TGF β 1 mRNA [17]. rs4803455 has been reported to increase the risk of asthma development. This SNP down-regulates lung function and induces airway remodeling in asthma [18].

rs1800469 in the TGF- β 1 gene promoter increases the plasma level of TGF- β 1 and the level of total IgE. It may also reduce the risk of progression of bronchial remodeling and inhibit the development of asthma [12, 19–21]. rs11083616 correlates with bronchial obstruction and with airway wall phenotypes - airway wall thickness. The presented SNP is an important marker of the development of all obstructive diseases [22]. The role of rs8109627 in the TGF β 1 gene has not been explicitly and precisely described in the etiopathogenesis of asthma. SNPs tagging in the TGF β 2 gene (rs10495098, rs17047703, rs17558745, rs2799085, rs2009112, rs10482751, rs2027567, rs10779329, rs2796821, rs2796822, rs4846479, rs2798631, rs10863399) and the TGF β 3 gene (rs4903359, rs3917187, rs2284792, rs2268626) have not been analyzed as risk factors for development of obstructive diseases [23]. Table 1 presents detailed characteristics of TGF β 1, TGF β 2, and TGF β 3 genes and their SNPs.

Aims

Our purpose was to determine whether genotypes of MAC/SNP TGF β 1, TGF β 2, and TGF β 3 are related to the level of asthma control, measured with the application of the Asthma Control Test (ACT™) in asthmatics and healthy controls.

Methods

Consent of the bioethics committee

The study was approved by the local ethics committee (Consent of Research Review Board at the Medical University of Lodz, Poland, No RNN/133/09/KE). At the commencement of the study, the participants were invited to get involved voluntarily. Before the enrolment, written informed consent was obtained from each patient.

Subjects

Asthma diagnosis was established according to GINA (The Global Initiative For Asthma) recommendations, based on clinical asthma symptoms and a lung function test. The level of asthma severity and control was determined on the basis of the GINA Report Guidelines. All the participants underwent subjective examinations (structuralized anamnesis including, besides the element of subjective examination, also an analysis of factors such as: gender, obesity, tobacco smoking, duration of bronchial asthma, allergy to house dust mites, animal fur, mold spores, cockroaches allergens, hypersensitivity to non-steroid anti-inflammatory drugs (NSAIDs), etc. Results of pulmonary function tests and allergological tests were obtained from individual medical records of the patients. If results of spirometry or allergological tests were not available, such examinations were additionally performed during the recruitment visit. Subjects

Table 1 Presentation of SNPs TGF β genes on the base of the NCBI gene database, available on <https://www.ncbi.nlm.nih.gov>

Name/ Gene ID	Description	Location	Aliases	MIM	SNPs	Role in asthma
TGFB1 ID: 7040	transforming growth factor beta 1 [<i>Homo sapiens</i> (human)]	Chromosome 19, NC_000019.10 (41,330,323..41353922, complement)	CED, DPD1, IBDIMDE, LAP, TGF-beta1, TGFB, TGFbeta	190,180	rs8109627, rs8179181, rs4803455, rs1800469, rs11083616	dimerization and phosphorylation of receptor-activated Smad proteins
TGFB2 ID: 7042	transforming growth factor beta 2 [<i>Homo sapiens</i> (human)]	Chromosome 1, NC_000001.11 (218,345,284..218444619)	G-TSF, LDS4, TGF-beta2	190,220	rs10495098, rs17047703, rs17558745, rs2799085, rs2009112, rs10482751, rs2027567, rs10779329, rs2796821, rs2796822, rs4846479, rs2798631, rs10863399	expressed by eosinophils; predominant in severe allergic asthma
TGFB3 ID: 7043	transforming growth factor beta 3 [<i>Homo sapiens</i> (human)]	Chromosome 14, NC_000014.9 (75,958,061..75983011, complement)	ARVD, ARVD1, LDS5, RNHF, TGF-beta3	190,230	rs4903359, rs3917187, rs2284792, rs2268626	lung cells become TGF- β 3 positive under inflammatory conditions

<https://www.ncbi.nlm.nih.gov/gene/?term=TGFB1>

<https://www.ncbi.nlm.nih.gov/gene/?term=TGFB2>

<https://www.ncbi.nlm.nih.gov/gene/?term=TGFB3>

are suffering from clinically significant exacerbations and using drugs that might induce resistance to glucocorticoids (such as rifampicin, phenobarbital, phenytoin, ephedrine), subjects with signs of viral infections, either generalized or affecting the respiratory tract, as well as subjects failing to comply with the doctor's recommendations, were excluded from the patient group. The control arm included a group of healthy adults who met the following criteria: no history or symptoms of either bronchial asthma or other pulmonary diseases, no history or symptoms of allergy, no history or symptoms of atopic dermatitis, no history or signs of hypersensitivity to aspirin, negative results of skin tests for 12 common allergens, no first-degree relatives with bronchial asthma or atopic disorders. Spirometry tests were conducted in the Outpatient Clinic according to ERS (European Respiratory Society)/ATS (American Thoracic Society) standards, and allergological tests according to EAACI (European Academy of Allergy and Clinical Immunology) guidelines [12, 19–21, 24–26].

Note: The authors of the article for this scientific project used their own patient database containing detailed

statistical and genetic characteristics used in previous publications. However, this manuscript does not present results published earlier in any other study (the article contains original results not published elsewhere).

Six hundred fifty-two subjects were included in the study. Of this number, there were 345 asthmatic patients at the average age of 48.6 years and 307 healthy subjects at 46.0 years. Detailed patients' characteristics were presented in Table 2.

The whole group of patients participating in the study was stratified and subjected to statistical analysis. The genotyping was performed by two investigators who were unaware of the phenotypes.

Genomic DNA extraction and SNPs analysis (MassARRAY® system)

Venous blood samples were collected from the participants onto EDTAK3, and DNA was obtained from peripheral blood leukocyte fraction. The genetic material was isolated using QIAamp DNA Blood Mini Kit (QIAGEN Inc.) according to guidelines provided by the manufacturer (12,13,17,18,19).

Table 2 Clinical characteristics of the recruited cohort and spirometric characteristics of the studied groups

Characteristics of the studied groups	Asthmatic Patients (n = 345)	Healthy Controls (n = 307)	p-value
Age [years]	48.6 ± 15.4	46.0 ± 16.3	0.04
Sex	Females: 222 Males: 123	Females: 197 Males: 110	0.96
FEV1(%)	75.81 ± 21.37	95.83 ± 19.95	< 0.001
FVC(%)	93.52 ± 18.78	101.63 ± 17.69	< 0.001
FEV1/FVC(%)	83.25 ± 14.71	95.88 ± 10.05	< 0.001
Smoking pack years	5.82 ± 10.98	5.68 ± 11.71	0.88

FEV1 (forced expiratory volume in 1 s) expressed in %, FEV1% (A/N% - percentage ratio of the measured to expected value) expressed as a percent of the expected value; FVC (forced vital capacity) expressed in %, FVC% (A/N% - percentage ratio of the measured to expected value) expressed as a percent of the predicted value; FEV1% FVC index (FEV1 to FVC ratio - forced vital capacity) expressed in %

An analysis of polymorphisms localized in TGF- β 1, TGF- β 2 and TGF- β 3 genes was conducted on 652 DNA samples with the application of the MassARRAY® system (Bionanopark, Lodz, Poland), using the mass spectrometry technique (MALDI TOF MS – matrix-assisted laser desorption ionization–time-of-flight mass spectrometry), dedicated to nucleic acids.

The authors measured the output level of genomic human DNA and diluted DNA to obtain final cDNA levels = 15 [ng/ μ l]. In samples with a DNA level lower than 15 [ng/ μ l], a sufficient undiluted amount of sample was used for analytical purposes. The absorbance ratio A260/A280, identified as a degree of DNA impurities, ranged from 1.7 to 2.0. The DNA levels were measured in all the studied samples.

The first stage of laboratory tests involved performing multiplex-PCR, specific for particular loci, in which 20 fragments of DNA, around 100 bp (base pairs) long, were amplified. Each DNA fragment had a polymorphic locus in its sequences.

After performing PCR, the remaining unbound dNTPs were dephosphorylated using Shrimp Alkaline Phosphatase (SAP).

The next stage involved performing iPLEX reaction. On the matrix which was formed during the multiplex-PCR reaction, terminal dideoxynucleotides (ddNTPs), corresponding to the allele variant present in the gene, were bonded to MassEXTEND primers with modified mass and complementary to the sequence, located above the polymorphic locus.

After performing the iPLEX reaction, 15 mg resin was placed on each well of a 96-well plate to bind metal ions, such as Na+, K+, Mg2+, which might cause high background interferences in mass spectra. After centrifugation of the plate (5 min., 4000 rpm), the analyzed material was transferred to SpectroCHIP using Nanodispenser RS1000. Its volume ranged from 3 to 25 nl and bonded with the agent filling each of the 96 wells on the chip.

The first stage of the mass spectrometer analysis included matrix laser desorption, needed for DNA analysis. SpectroCHIP is placed inside the spectrometer, and each place with a DNA sample is ionized with a laser in the vacuum. An energy beam used in laser ionization must be well adjusted so that the particles do not get fragmented but bounced only. Next, the particles proceed to a time-of-flight analyzer. The particles going to the analyzer are accelerated with an electric impulse and start drifting through the analyzer chamber. At the end of the analyzer, an ion detector is connected with a device that registers the time between the acceleration impulse and the moment the particle hits the sensor. A result of the mass-to-charge ratio value corresponds to a value we obtain if we follow a law of physics saying that

for an impulse of particular amperage, the time-of-flight of ions gets longer if their molecular weight increases. An analysis of the sample using the MassARRAY mass spectrometer enabled us to obtain raw results presented in the form of mass spectra. The mass spectra, in turn, allowed the identification of polymorphisms in the following studied genes: TGF- β 1, TGF- β 2, and TGF- β 3. The graphs could enable us to determine the reaction yield and probability of the obtained results.

Statistical analysis

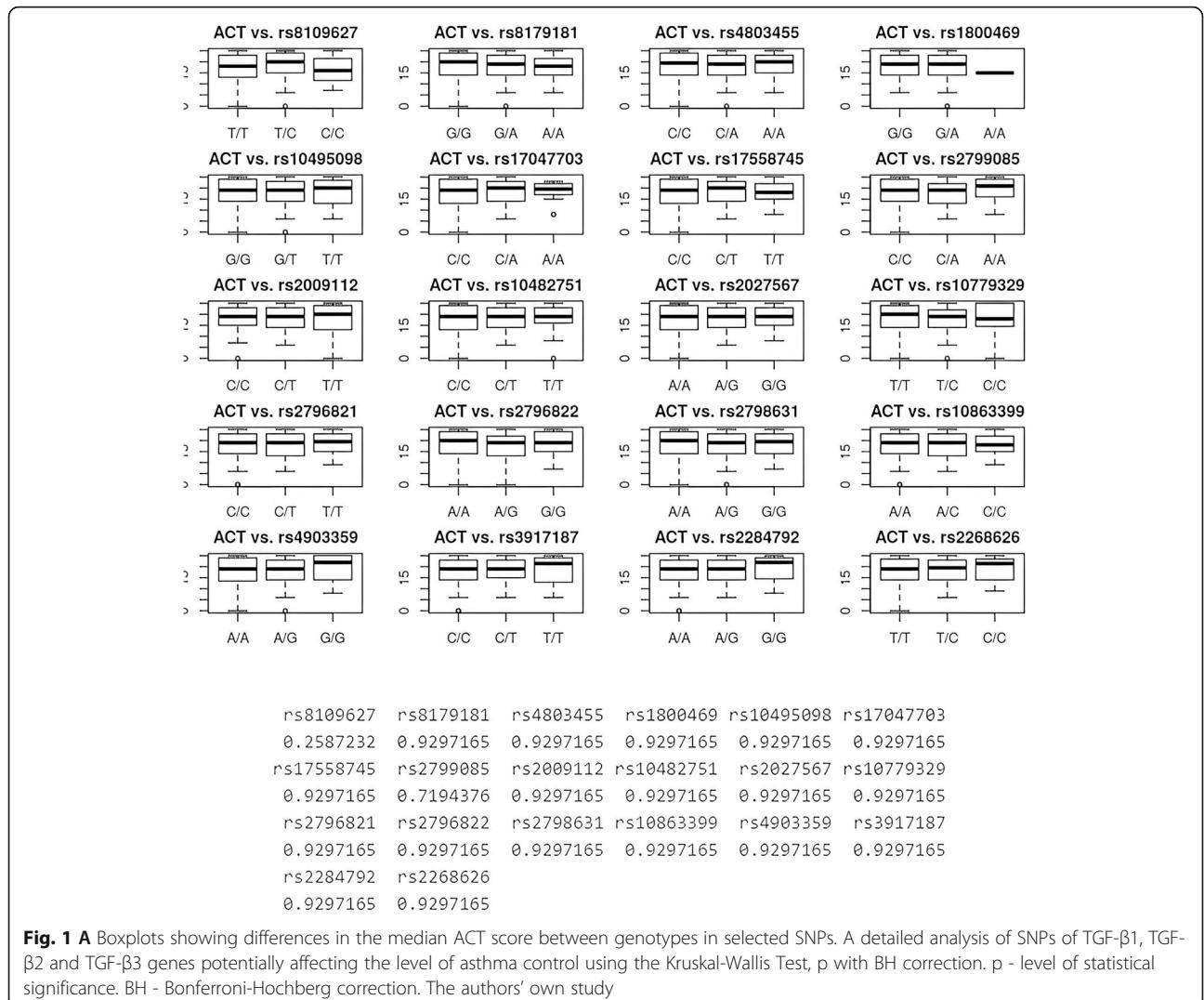
The Kruskal-Wallis rank-sum test was applied to compare asthma control test scores in genotypes, whereas the two-sample Wilcoxon test was used to compare the scores in minor allele carriers (MAC). Due to a large number of comparisons, the p -values were adjusted using the Benjamini & Hochberg procedure. Treating the ACT score as continuous, linear regression models were developed for both genotypes and MAC statuses to assess the influence of variables on the ACT score. The backward stepwise selection was based on the Akaike Information Criterion for a fitted parametric model. The overall performance of the models was assessed utilizing multiple R2 and RMSE. At the same time, the influence of certain factors required a calculation of standardized beta coefficients after estimating them using standard deviation. Although measured on a continuous scale, the response variable (ACT score) can be treated as an ordinal variable. Below I check whether the application of ordinal regression (i.e., proportional odds logistic regression) would bring about any changes. The logistic regression was designed in the same manner to test whether the genotypes or MAC statuses will allow predicting reasonable asthma control (ACT \geq 25). The analysis was performed in STATISTICA 13.1 (Dell, USA) and R programming language (packages: rms, cutpointr, caret).

Results

Whole genotypes and MACs were tested in TGF- β 1, TGF- β 2, and TGF- β 3 genes. The authors assessed their correlation with the degree of asthma control. It was found that the median of the Asthma Control Test (ACT, score in points) did not differ on genotypes of selected SNPs ($p > 0.05$ for all comparisons, BH [Bonferroni-Hochberg] correction). Details are shown in Fig. 1A.

After using correction for multiple comparisons, the median ACT did not differ depending on the minority allele carrier (MAC) of selected SNPs ($p > 0.05$ for all comparisons, BH correction). Details are shown in Fig. 2A.

For prediction of ACT based on genotype, the full model without feature selection showed low R2, equal to



0.084 and RMSE equal to 5.47 points. After applying the backward stepwise procedure, rs8109627, rs2799085, rs3917187 and rs2284792 were the only remaining SNPs in the model. Despite the feature selection, the R² did not improve and was equal to 0.05 with RMSE equal to 5.57 points. The model is shown in Fig. 2A.

The greatest standardized beta coefficient, thus the most significant influence on the predicted ACT, was shown for genotype G/G rs2284792; however, this was not significant in the t-test ($p = 0.06$). T/C genotype in rs8109627 was the only statistically significant variable in the t-test associated with linear modeling. It is further shown in the calibration plot. See Fig. 2B.

Calibration plot showing a relationship between predicted and actual results in the ACT score. Although the correlation is significant, the rho coefficient does not prove a strong relationship

In a similar analysis involving MAC, the final model after feature selection revealed that rs8109627,

rs2796822, and rs2798631 are most important. The multiple R² was, however, low and equal to 0.02, with RMSE equivalent to 5.66 points. The estimates are shown in Table 3.

Predictive analysis of the ACT result, depending on the genotypes of the studied SNPs with the application of backward selection, showed that the T / C genotype in rs8109627 is significantly associated with a higher ACT result (standardized beta = 0.11). In general, despite the variable selection, the model was characterized by low efficiency (adjusted R² 0.03). A GVIF analysis also showed that the model could have been described with collinearity of variables. Details are presented in Fig. 3.

Considering only the MAC analysis, in multivariate linear regression analysis, after applying stepwise backward selection of variables, the authors managed to reveal that MAC in rs8109627 is significantly associated with a higher ACT result (better asthma control), and in rs2796822 with a lower ACT result (worse asthma

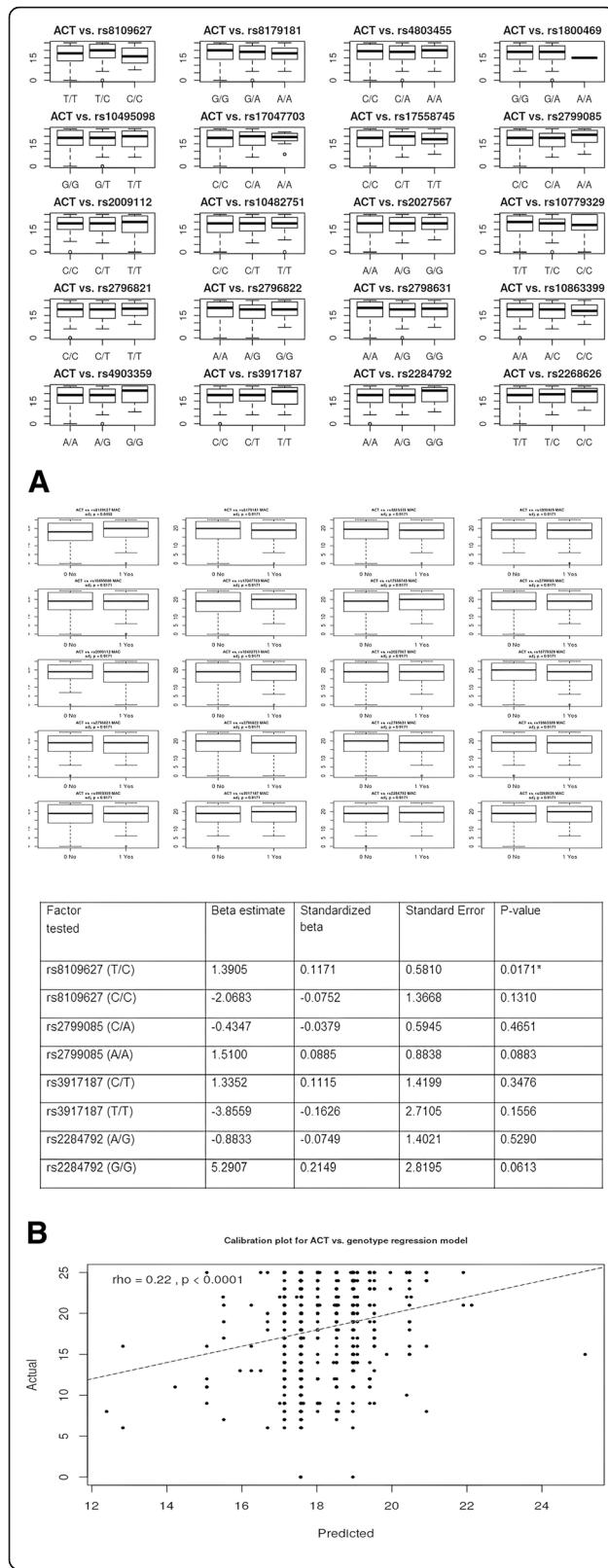


Fig. 2 A and B Coefficients of the linear model for prediction of ACT score. A detailed analysis of MAC in SNPs of TGF- β 1, TGF- β 2 and TGF- β 3 genes potentially affecting the level of asthma control using the U Mann-Whitney Test, with BH correction. p - level of statistical significance. BH - Bonferroni-Hochberg correction. The authors' own study

control). The selection of variables showed that MAC is most important in rs8109627, rs2796822, and rs2798631. However, the model was characterized with low efficiency (adjusted R² 0.01). Details of the analysis are presented in Fig. 4.

The authors made an attempt to create a decision tree (ctree), allowing to prognosticate the degree of disease control in ACT, based on particular SNPs in the studied genes. An attempt to model trees appeared to be unsuccessful in this case. The authors did not manage to obtain results which would be more resistant to over-adjusting after application of other techniques (e.g. ctree - decision trees), as illustrated in Fig. 5.

An analysis of ROC curves showed that statistical modeling might be helpful in clinical practice. Reasonable asthma control could be predicted by analyzing all twenty SNPs and not by analyzing MAC results in SNPs. These observations are illustrated in Fig. 6.

Due to the class imbalance, better results were obtained after performing the SMOTE-type modeling. Predicting good asthma control (ACT $>= 25$) with the use of this method enables to achieve 86.11% specificity and 49.17% sensitivity. Results are shown in Fig. 7.

Predicting good asthma control is possible with the SMOTE method. It should be noted that both the use of genotypes and MAC allows creating a moderately correct predictive model, which is about 70% efficient on the entire (training) data set. The description is shown in Fig. 8.

Discussion

The role of polymorphic genes in asthma, especially single ones, has been widely discussed for years in international professional literature [12, 13, 19–21]. Our work, unlike many others, presents a holistic and unique analysis along with a detailed statistical methodology that allows us to assess the effect of twenty SNPs, simultaneously in TGF- β 1, TGF- β 2 and TGF- β 3 genes on the degree of asthma control. Our results are also an attempt to present various statistical approaches to the problem analysis and interpretation of the said problem. The work is unique in this respect [27, 28].

Firstly, we did not note significantly different medians in ACT in genotypes and MACs (adjusted p -value > 0.05 for all comparisons). See Fig. 1A. For ACT prediction, based on the genotypes, the full model without feature selection showed low R² equal to 0.084 and RMSE equal

Table 3 Coefficients of the linear model for prediction of the ACT score. The MAC status was considered interesting. Although rs8109627 and rs2796822 show significant association with the ACT score, the effect seems too slight

SNP tested	Beta estimate	Standardized beta	Standard Error	P-value
rs8109627	1.1115	0.0957	0.5646	0.0497*
rs2796822	-1.7930	-0.1534	0.8932	0.0454*
rs2798631	1.3661	0.1123	0.9293	0.1423

to 5.47 points. After applying the backward stepwise procedure, rs8109627, rs2799085, rs3917187, and rs2284792 were the only remaining SNPs in the model. Despite feature selection, the R² did not improve and was equal to 0.05, with RMSE equal to 5.57 points. See Fig. 2A. The greatest standardized beta coefficient, thus the greatest influence on the predicted ACT, was shown for genotype G/G rs2284792; however, this was not significant in the t-test ($p = 0.06$). See Fig. 2A. The only statistically significant variable in the t-test and associated with linear modeling was T/C genotype in rs8109627. It is further shown in the calibration plot. See Fig. 2B. In a similar analysis involving MAC, the final model after feature selection revealed that rs8109627, rs2796822, and rs2798631 are most important. The multiple R² was,

however, low and equal to 0.02, with RMSE equal to 5.66 points. The estimates are shown in Table 3.

Next, in order to check whether the genotypes or MAC statuses will allow predicting reasonable asthma control (ACT \geq 25), the logistic regression was designed in the same manner. The analysis allowed us to conclude that after performing the feature selection, the final model included only rs8109627 and rs17047703 [27, 28]. See Supplementary File No 1.

It is difficult to discuss the issues described above because currently available databases (e.g., <https://www.ncbi.nlm.nih.gov/pubmed/?term=SNP+TGF+asthma+control>) do not have analyses of polymorphic forms of TGF- β 1, TGF- β 2, and TGF- β 3 genes, which influence the degree of asthma control. It also shows the unique nature of our

```

## lm(formula = ACT ~ rs8109627 + rs2799085 + rs3917187 + rs2284792,
##     data = temp)
##
## Residuals:
##   Min     1Q   Median     3Q    Max
## -18.961 -4.135   1.000  4.865  9.933
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 17.57017  0.00000 32.004 <2e-16 ***
## rs8109627/T/C 1.39051  0.11711  2.393  0.0171 *  
## rs8109627C/C -2.06833 -0.07523  1.36680 -1.513  0.1310  
## rs2799085C/A -0.43472 -0.03791  0.59455 -0.731  0.4651  
## rs2799085A/A  1.51001  0.08846  0.88382  1.709  0.0883 .  
## rs3917187C/T  1.33525  0.11146  1.41992  0.940  0.3476  
## rs3917187T/T -3.85590 -0.16262  2.71047 -1.423  0.1556  
## rs2284792A/G -0.88332 -0.07489  1.40209 -0.630  0.5290  
## rs2284792G/G  5.29066  0.21492  2.81947  1.876  0.0613 .  
## ---      
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.628 on 411 degrees of freedom
## Multiple R-squared:  0.05058,    Adjusted R-squared:  0.0321 
## F-statistic: 2.737 on 8 and 411 DF,  p-value: 0.005956

```

Fig. 3 Multivariate linear regression analysis after stepwise backward selection of variables, describing the effect of the studied SNPs genotypes on the level of asthma control (ACT). The authors' own study

```

## 
## Call:
## lm(formula = ACT ~ rs8109627_MAC + rs2796822_MAC + rs2798631_MAC,
##      data = temp)
##
## Residuals:
##    Min     1Q   Median     3Q    Max
## -18.5727 -4.4612  0.9049  4.8094  8.9049
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)           17.88811   0.00000 32.834 <2e-16 ***
## rs8109627_MAC1 Yes  1.11149   0.09565  0.56465  1.968  0.0497 *
## rs2796822_MAC1 Yes -1.79302  -0.15341  0.89322 -2.007  0.0454 *
## rs2798631_MAC1 Yes  1.36612   0.11230  0.92932  1.470  0.1423
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.686 on 416 degrees of freedom
## Multiple R-squared:  0.01931,   Adjusted R-squared:  0.01224
## F-statistic: 2.731 on 3 and 416 DF, p-value: 0.04354

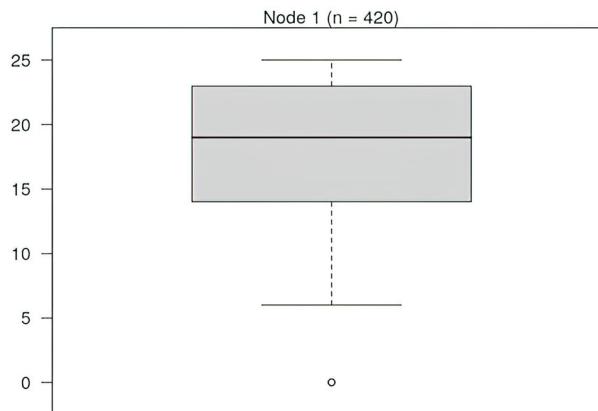
```

Fig. 4 Analysis of ACT result prediction, depending on MAC SNPs (full analysis model including all SNPs). The authors' own study

```

library(caret)
train_control <- trainControl(method="LOOCV")
model = train(ACT ~ ., data = temp, trControl=train_control, method="ctree")
plot(model$finalModel)

```



```
postResample(pred = predict(model, newdata = temp), obs = temp$ACT)
```

```
##      RMSE Rsquared      MAE
## 5.714010     NA 4.850726
```

Fig. 5 Modeling the c-tree, allowing to prognosticate the degree of asthma control, on the basis of SNPs analysis of TGF- β 1, TGF- β 2, and TGF- β 3 genes. The authors' own study

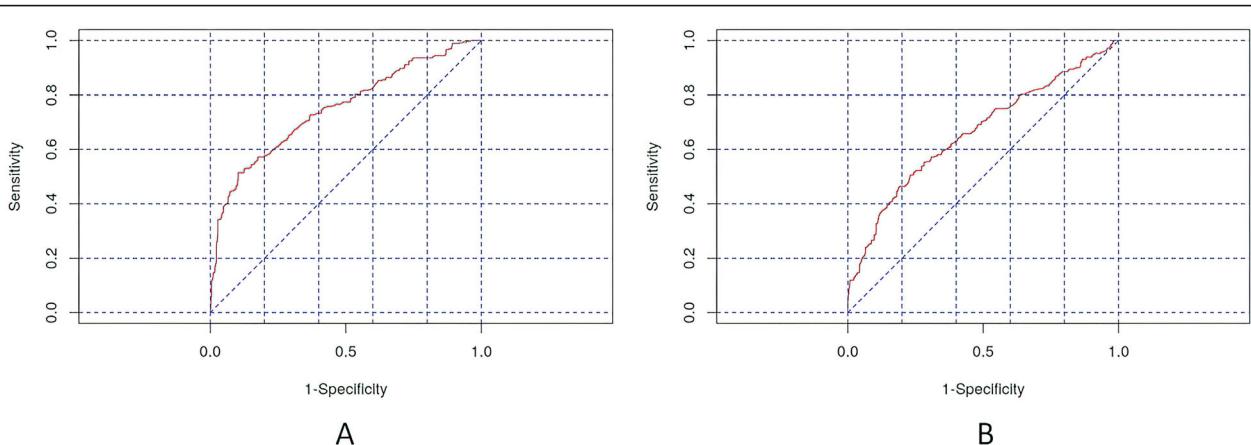


Fig. 6 An analysis of prognosis of effective asthma control ($ACT \geq 25$) on the basis of ROC analysis. **A** Logistic regression with the application of all genotypes in SNPs contributed to $AUC = 0.749$. **B** Logistic regression with the application of all MACs in SNPs contributed to $AUC = 0.661$. The authors' own study

work. Although literature reports an impact of single SNPs on the clinical course of asthma, there are no studies on the association between the TGF- β gene and the control of asthma symptoms [12, 13, 19, 20].

Secondly, we examined the possibilities of assessing correlations with polymorphic forms of TGF- β genes using various statistical methods. We showed that in the multivariate linear regression analysis model, after using backward stepwise selection of variables, the occurrence of the T/C genotype in rs8109627 is significantly associated with a higher ACT score; however, this effect is not spectacular (beta standardized = 0.11). It should be pointed out here that it has not been well documented that this rs8109627 in the $TGF\beta 1$ gene plays a role in increasing the risk of asthma. We obtained slightly

different results after analyzing not SNP but MAC in TGF- β genes. After backward stepwise selection of variables, the multivariate linear regression analysis of MAC showed that MAC in rs8109627 is significantly associated with a higher ACT™ result, and in rs2796822 - with a lower one. Interestingly, this rs2796822 in the TGF- $\beta 2$ gene occurs with different frequencies [1.71 (1.07,2.71) $p = 0.022$] in patients with non-severe and severe asthma. The risk of severe asthma in MAC carriers for the SNP mentioned above was 71% higher (the author's data in the publication procedure). A further selection of variables confirmed that MAC is most important in rs8109627, rs2796822, and rs2798631, but this model was characterized with low efficiency (adjusted R² = 0.01). SNPs rs8109627 and rs2798631 correlated with a

Reference		OR	lower95ci	upper95ci	Pr(> Z)
Prediction 0_Not 1_Well	`rs8109627T/C`	2.319171e+00	1.71127236	3.1430141	5.828667e-08
0_Not 496 183	`rs4803455C/A`	7.628586e-01	0.54015444	1.0773830	1.243439e-01
1_Well 80 177	`rs4803455A/A`	5.897361e-01	0.38325404	0.9074625	1.632525e-02
Accuracy : 0.719	`rs10495098G/T`	5.461033e-01	0.38260887	0.7794612	8.607962e-04
95% CI : (0.689, 0.7476)	`rs10495098T/T`	5.727335e-01	0.36477837	0.8992409	1.546191e-02
No Information Rate : 0.6154	`rs17047703C/A`	1.673850e+00	1.12579559	2.4887059	1.091283e-02
P-Value [Acc > NIR] : 1.782e-11	`rs17047703A/A`	3.500510e-07	0.0000000	Inf	9.734126e-01
Kappa : 0.3728	`rs17558745C/T`	7.359208e-01	0.51143956	1.0589313	9.862532e-02
	`rs17558745T/T`	1.689126e-01	0.05184589	0.5503131	3.166557e-03
Mcnemar's Test P-Value : 3.183e-10	`rs2009112C/T`	5.925143e-01	0.43000826	0.8164336	1.374565e-03
Sensitivity : 0.4917	`rs10482751T/T`	2.111074e+00	1.31466988	3.3899253	1.987096e-03
Specificity : 0.8611	`rs2027567A/G`	7.842569e-01	0.57150460	1.0762100	1.323005e-01
Pos Pred Value : 0.6887	`rs10779329T/C`	5.294641e-01	0.38373863	0.7305292	1.080701e-04
Neg Pred Value : 0.7305	`rs2796822G/G`	2.149451e+00	1.47062126	3.1416237	7.758931e-05
Prevalence : 0.3846	`rs4903359A/G`	5.703417e-01	0.39466149	0.8242243	2.799242e-03
Detection Rate : 0.1891	`rs4903359G/G`	3.333997e+00	1.53843109	7.2252402	2.276258e-03
Detection Prevalence : 0.2746	`rs2284792A/G`	1.346804e+00	0.89951437	2.0165107	1.482516e-01
Balanced Accuracy : 0.6764	`rs2284792G/G`	3.104222e-01	0.14215347	0.6778724	3.328629e-03
'Positive' Class : 1_Well	`rs2268626T/C`	1.469691e+00	1.00475395	2.1497723	4.721061e-02

Fig. 7 A SMOTE analysis and analysis with the application of backward stepwise selection, enabling to predict good asthma control ($ACT >= 25$), on the basis of all twenty SNPs. The authors' own study

Reference		OR	lower95ci	upper95ci	Pr(> Z)
Prediction 0_Not 1_Well					
0_Not 511 244	rs8109627_MAC1 Yes`	1.4757251	1.1090888	1.9635620	7.574183e-03
1_Well 65 116	rs8179181_MAC1 Yes`	0.7304810	0.5449112	0.9792466	3.571021e-02
Accuracy : 0.6699	rs4803455_MAC1 Yes`	0.7394319	0.5431559	1.0066348	5.511744e-02
95% CI : (0.6387, 0.7)	rs10495098_MAC1 Yes`	0.6149851	0.4511447	0.8383267	2.100809e-03
No Information Rate : 0.6154	rs17558745_MAC1 Yes`	0.6460032	0.4793593	0.8705790	4.098985e-03
P-Value [Acc > NIR] : 0.0003085	rs2009112_MAC1 Yes`	0.7900859	0.5764990	1.0828046	1.428578e-01
Kappa : 0.2309	rs2027567_MAC1 Yes`	0.7403198	0.5525748	0.9918539	4.392834e-02
Mcnemar's Test P-Value : < 2.2e-16	rs10779329_MAC1 Yes`	0.6715985	0.4972768	0.9070292	9.428945e-03
Sensitivity : 0.3222	rs2796821_MAC1 Yes`	1.9664779	1.3716801	2.8191962	2.336161e-04
Specificity : 0.8872	rs10863399_MAC1 Yes`	0.7181097	0.4904748	1.0513925	8.869448e-02
Pos Pred Value : 0.6409	rs4903359_MAC1 Yes`	0.7114986	0.5151967	0.9825959	3.877567e-02
Neg Pred Value : 0.6768	rs2284792_MAC1 Yes`	2.1825396	1.5208863	3.1320418	2.283663e-05
Prevalence : 0.3846	rs2268626_MAC1 Yes`	0.7190727	0.5005167	1.0330635	7.442341e-02
Detection Prevalence : 0.1239					
Balanced Accuracy : 0.6047					
'Positive' Class : 1_Well					

Fig. 8 An analysis of the application of a prognostic model on the basis of SNP and MAC genotypes on the entire (training) data set to predict good asthma control. The authors' own study

clinically significant higher level of asthma control (controlled asthma) in the ACT™ Test, whereas rs2796822 was associated with a lower result (loss of disease control). See Fig. 2B. Nevertheless, this is the first report about the role of these three tagging SNPs in asthma in TGF- β 1, TGF- β 2, and TGF- β 3 genes. An attempt to optimize the prediction of good asthma control was also made ($ACT > 25$). Due to the class imbalance, better results were obtained after performing the SMOTE-type modeling than after applying multivariate linear regression analysis. For details, see Figs. 6, 7, and 8. Conducting analysis using various statistical models enables both genotypes and MAC, which creates a moderately correct prognostic model that is about 70% efficient on the entire (training) data set (Fig. 8).

In Supplementary File No 1, we checked whether the genotypes or MAC statuses would allow predicting good asthma control ($ACT \geq 25$); the logistic regression was designed in the same manner. The conducted procedure concluded that after feature selection, the final model included only rs8109627 and rs17047703. See Supplementary File No 1. The role of tagging SNP rs8109627 in the TGF- β 1 gene and rs17047703 in TGF- β 2 had not yet been described. We were the first in the above analysis to test it with advanced statistical methods [27, 28].

In our work, we analyzed twenty tagging SNPs in TGF- β 1, TGF- β 2, and TGF- β 3 genes. We critically interpreted the results with statistically different methods. It should be noted that only three SNPs, such as rs8109627, rs2796822, and rs2284792, play a significant clinical role in the regulation of asthma control. This is an important discovery since the role of these SNPs in the TGF- β gene in asthma had not been previously described. It should be critically stated that the prognostic efficiency of different statistical models was up to 70%. Nevertheless, the work, not carried out before, is a

unique and holistic analysis of twenty polymorphic forms in TGF- β 1, TGF- β 2, and TGF- β 3 genes using state-of-the-art techniques of molecular biology (MassARRAY® system).

Limitations

Both the use of genotypes and MAC allow the creation of only a moderately efficient prognostic model. In the multivariate linear regression analysis, we managed to show that only single genotypes significantly correlated with higher ACT results after using backward stepwise selection of variables. However, this effect was not spectacular. In general, despite the variable selection, the efficiency of the model was still low. A GVIF analysis also showed that the model could have been characterized with collinearity of variables.

Moreover, the selection of variables showed that MAC was most important in three SNPs, but this model was not also highly efficient. The authors did not manage to obtain results that would be more resistant to over-adjusting with the application of other techniques (e.g. ctree - decision trees). Due to the class imbalance, better results were obtained after performing the SMOTE-type modeling. After completing the SMOTE-type modeling, the logistic regression model was trained on the whole dataset. The application of SMOTE vastly inflates those results. By creating new cases that are similar to the original ones, the performance was only minimally inflated. According to the authors of the above statistical analyses, linear regression models are the only ones that can be used in this type of scientific study. The small size of the surveyed populations is another limitation of the study. It would be advisable to verify our theses on subpopulations consisting of about 1000 respondents. Our results should be treated as preliminary.

Conclusion

This study is the first one ever conducted in the Polish population. It attempts to analyze the functional impact of twenty SNPs of TGF- β 1, TGF- β 2, and TGF- β 3 genes on the degree of asthma control, assessed using the international ACT™ Questionnaire. We showed new relationships between the occurrence of SNP rs8109627 in the TGF- β 1 gene (improved disease control - controlled asthma) and rs2796822 in the TGF- β 2 gene (worsened disease control - uncontrolled asthma) and a significantly changed level of asthma control. This is a particularly important observation because the TGF- β gene affects eosinophil levels, bronchial hyperreactivity and obstruction, and clinical signs of asthma. In contrast, the TGF- β 1–3 gene complex is an essential regulator of the immune response in asthma. We also proposed a new statistical analysis model (multivariate linear regression analysis with backward stepwise selection of variables and the SMOTE model) to assess the effect of many SNPs, simultaneously in three TGF- β genes, on disease control. The analyzed results and methods were a subject of our internal discussion, during which we identified those that may be clinically useful. The presented results constitute an innovative approach to understanding the importance of SNP forms in TGF- β 1, TGF- β 2, and TGF- β 3 genes in the regulation of asthma control.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12950-021-00294-4>.

Additional file 1.

Additional file 2.

Acknowledgements

Not applicable.

Authors' contributions

M.P. conceived and planned the experiments. M.P. carried out the experiments. K.S. and M.P. planned and carried out the simulations. M.P. and K.S. contributed to sample preparation. M.P. and K.S. contributed to the interpretation of the results. M.P. took the lead in writing the manuscript. All authors provided critical feedback and helped shape the research, analysis and manuscript. The author(s) read and approved the final manuscript.

Funding

The study is financed from a grant from Polpharm Scientific Foundation, No 16/XIV/2015, Poland.

Availability of data and materials

All data and materials are presented in the section of the article. The authors will always provide other indicated data on request. Raw data were attached as the [supplementary dataset](#).

Declarations

Ethics approval and consent to participate

The local ethics committee approved the study (Consent of Research Review Board at the Medical University of Lodz, Poland, No RNN/133/09/KE). At the commencement of the study, the participants were invited to get involved

voluntarily. Before the enrolment, written informed consent was obtained from each patient.

Consent for publication

Prof. Piotr Bolesław Kuna, M.D., Ph.D. – Head of Department of Internal Medicine, Asthma and Allergy, Medical University of Lodz, Lodz, PL.

Competing interests

The authors declare that there is no conflict of interest regarding the publication of this article.

Author details

¹Department of Internal Medicine, Asthma and Allergy, Medical University of Lodz, ul. Kościuszki 22, 90-153 Łódź, Poland. ²Department of Biostatistics and Translational Medicine, Medical University of Lodz, Lodz, Poland.

Received: 11 December 2020 Accepted: 3 September 2021

Published online: 07 October 2021

References

1. GINA Main Report dostępny na 2019 <https://ginasthma.org/>
2. Droszcz W, Grzanka A. Asthma in adolescents and adults. Warszawa: Wydawnictwo Lekarskie PZWL; 2010.
3. Genuneit J, Cantelmo JL, Weirnmayr G, Wong GW, Cooper PJ, Riikjärv MA, et al. A multi-centre study of candidate genes for wheeze and allergy: the International Study of Asthma and Allergies in Childhood Phase 2. *Clin Exp Allergy*. 2009;39:1875–88.
4. Anderson GP. Endotyping asthma: new insights into key pathogenic mechanisms in a complex, heterogeneous disease. *Lancet*. 2008;372(9643):1107–19. [https://doi.org/10.1016/S0140-6736\(08\)61452-X](https://doi.org/10.1016/S0140-6736(08)61452-X).
5. Djukanović R, Roche WR, Wilson JW, Beasley CR, Twentyman OP, Howarth RH, et al. Mucosal inflammation in asthma. *Am Rev Respir Dis*. 1990;142(2):434–57. <https://doi.org/10.1164/ajrccm/142.2.434>.
6. Sumi Y, Hamid Q. Airway remodeling in asthma. *Allergol Int*. 2007;56(4):341–8. <https://doi.org/10.2332/allergolintR-07-153>.
7. Romagnani S. The increased prevalence of allergy and the hygiene hypothesis: missing immune deviation, reduced immune suppression, or both? *Immunology*. 2004;112(3):352–63. <https://doi.org/10.1111/j.1365-2567.2004.01925.x>.
8. Wang J, Young IG. Eosinophilic inflammation: mechanisms regulating IL-5 transcription in human T lymphocytes. *Allergy*. 2007;62(10):1131–8. <https://doi.org/10.1111/j.1398-9995.2007.01510.x>.
9. Yang YC, Zhang N, Van Crombrugge K, Hu GH, Hong SL, Bachert C. Transforming growth factor-beta1 in inflammatory airway disease: a key for understanding inflammation and remodeling. *Allergy*. 2012;67(10):1193–202. <https://doi.org/10.1111/j.1398-9995.2012.02880.x>.
10. Duvernelle C, Freund V, Frossard N. Transforming growth factor-beta and its role in asthma. *Pulm Pharmacol Ther*. 2003;16(4):181–96. [https://doi.org/10.1016/S1094-5539\(03\)00051-8](https://doi.org/10.1016/S1094-5539(03)00051-8).
11. Park HK, Park HW, Jeon SG, Shin ES, Gho YS, Cho SH, et al. Distinct association of genetic variations of vascular endothelial growth factor, transforming growth factor- β , and fibroblast growth factor receptors with atopy and airway hyperresponsiveness. *Allergy*. 2008;63(4):447–53. <https://doi.org/10.1111/j.1398-9995.2007.01593.x>.
12. Panek M, Pietras T, Fabijan A, Ziolo J, Wieteska L, Małachowska B, et al. Identification and association of the single nucleotide polymorphisms, C-509T, C+466T and T+869C, of the TGF- β 1 gene in patients with asthma and their influence on the mRNA expression level of TGF- β 1. *Int J Mol Med*. 2014;34(4):975–86. <https://doi.org/10.3892/ijmm.2014.1894>.
13. Koćwin M, Jonakowski M, Przemęcka M, Ziolo J, Panek M, Kuna P. The role of the TGF-SMAD signalling pathway in the etiopathogenesis of severe asthma. *Adv Respir Med*. 2016;84(5):290–301. <https://doi.org/10.5603/PiAP.2016.0037>.
14. Heldin CH, Moustakas A. Signaling Receptors for TGF- β Family Members. *Cold Spring Harb Perspect Biol*. 2016;8(8):a022053. <https://doi.org/10.1101/cshperspect.a022053>.
15. Huang T, Hinck A. Production, isolation, and structural analysis of ligands and receptors of the TGF- β superfamily. *Methods Mol Biol*. 2016;1344:63–92. https://doi.org/10.1007/978-1-4939-2966-5_4.

16. Bonnans C, Chou J, Werb Z. Remodelling the extracellular matrix in development and disease. *Nat Rev Mol Cell Biol.* 2014;15(12):786–801. <https://doi.org/10.1038/nrm3904>.
17. Li H, Romieu I, Wu H, Sienra-Monge JJ, Ramírez-Aguilar M, del Río-Navarro BE, et al. Genetic polymorphisms in transforming growth factor beta-1 (TGFB1) and childhood asthma and atopy. *Hum Genet.* 2007;121(5):529–38. <https://doi.org/10.1007/s00439-007-0337-z>.
18. Hur GY, Broide DH. Genes and pathways regulating decline in lung function and airway remodeling in asthma. *Allergy Asthma Immunol Res.* 2019;11(5):604–21. <https://doi.org/10.4168/aair.2019.11.5.604> Review.
19. Panek M, Jonakowski M, Ziolo J, Wieteska Ł, Małachowska B, Pietras T, et al. A novel approach to understanding the role of polymorphic forms of the NR3C1 and TGF-β1 genes in the modulation of the expression of IL-5 and IL-15 mRNA in asthmatic inflammation. *Mol Med Rep.* 2016;13(6):4879–87. <https://doi.org/10.3892/mmr.2016.5104>.
20. Panek M, Pietras T, Fabijan A, Ziolo J, Wieteska Ł, Małachowska B, et al. The NR3C1 glucocorticoid receptor gene polymorphisms may modulate the TGF-beta mRNA expression in asthma patients. *Inflammation.* 2015;38(4):1479–92. <https://doi.org/10.1007/s10753-015-0123-3>.
21. Panek M, Pietras T, Fabijan A, Milanowski M, Wieteska L, Górska P, et al. Effect of glucocorticoid receptor gene polymorphisms on asthma phenotypes. *Exp Ther Med.* 2013;5(2):572–80. <https://doi.org/10.3892/etm.2012.809>.
22. Kim WJ, Hoffman E, Reilly J, Hersh C, Demeo D, Washko G, et al. Association of COPD candidate genes with computed tomography emphysema and airway phenotypes in severe COPD. *Eur Respir J.* 2011;37(1):39–43. <https://doi.org/10.1183/09031936.00173009>.
23. National Center for Biotechnology Information (NCBI) gene database dostępna na <https://www.ncbi.nlm.nih.gov/gene>
24. Pietras T, Panek M, Tworek D, Oszajca K, Wujcik R, Górska P, et al. The Bcl I single nucleotide polymorphism of the human glucocorticoid receptor gene h-GR/NR3C1 promoter in patients with bronchial asthma: pilot study. *Mol Biol Rep.* 2011;38(6):3953–8. <https://doi.org/10.1007/s11033-010-0512-5>.
25. Pietras T, Panek M, Kuprys-Lipińska I, Oszajca K, Wujcik R, Kuna P, et al. Frequencies of Bcl I, E22E, and N363S of h-GR/NR3C1 restriction fragment length polymorphisms of glucocorticoid receptor gene in polish adult population. *Med Sci Monit.* 2010;16(10):CR475–9.
26. Panek M, Pietras T, Antczak A, Fabijan A, Przemęcka M, Górska P, et al. The N363S and I559N single nucleotide polymorphisms of the h-GR/NR3C1 gene in patients with bronchial asthma. *Int J Mol Med.* 2012;30(1):142–50. <https://doi.org/10.3892/jimm.2012.956>.
27. Ding C, Peng H. Minimum redundancy feature selection from microarray gene expression data. *J Bioinform Comput Biol.* 2005;3(2):185–205. <https://doi.org/10.1142/S0219720005001004>.
28. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* 1988;44(3):837–45. <https://doi.org/10.2307/2531595>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

