ISSN 1022-7954, Russian Journal of Genetics, 2023, Vol. 59, No. 2, pp. 169–179. © Pleiades Publishing, Inc., 2023.
Russian Text © The Author(s), 2023, published in Genetika, 2023, Vol. 59, No. 2, pp. 196–208.

HUMAN GENETICS

Polygenic Analysis of Cytokine and Inflammatory Gene Polymorphisms in Chronic Obstructive Pulmonary Disease

G. F. Korytina[a], *, Y. G. Aznabaeva[b], O. V. Kochetova[a], T. R. Nasibullin[a], L. Z. Akhmadishina[a], N. N. Khusnutdinova[a], N. Sh. Zagidullin[b], and T. V. Victorova[b]

[a] *Institute of Biochemistry and Genetics, Subdivision of the Ufa Federal Research Center, Russian Academy of Sciences (IBG UFRC RAS), Ufa, 450054 Russia*

[b] *Bashkortostan State Medical University, Ufa, 450000 Russia*

**e-mail: guly_kory@mail.ru*

Received March 22, 2022; revised May 16, 2022; accepted May 31, 2022

Abstract—Chronic obstructive pulmonary disease (COPD) is a complex lung disease characterized by progressive airflow limitation and abnormal inflammatory response of the lungs to inhaled noxious particles or gases. COPD pathogenesis has been linked to oxidative stress and systemic inflammation. We aimed to assess the association of cytokines and inflammatory genes polymorphisms and their combinations with COPD. SNPs of the inflammatory genes *FASLG* (rs763110), *IL19* (rs2243193), *IL20* (rs2981573), *IL24* (rs291107), *PPBP* (rs352010), *IL4* (rs2243250), *IL4* (rs2070874), *C5* (rs17611), *FAS* (rs1800682), *IL4RA* (rs1805010), and *TGFb1* (rs1800469) were genotyped by the real-time polymerase chain reaction (PCR) among 601 COPD patients and 617 controls. Significant associations with COPD in the study group under an additive genetic model were identified for *IL19* (rs2243193) ($P = 0.00001$, OR = 0.73), *IL4* (rs2243250) ($P = 0.024$, OR = 1.27), *IL4* (rs2070874) ($P = 0.00001$, OR = 0.62), and for *PPBP* (rs352010) under the recessive model ($P = 0.00001$, OR = 2.34). Using the APSampler algorithm, we obtained gene–gene combinations that remained significantly associated with COPD; the *A* allele of *IL19* (rs2243193) and *C* allele of *PPBP* (rs352010) were the core elements of the majority of protective patterns associated with COPD. The highest risk of COPD was conferred by combination of alleles: *G* of *IL12A* (rs2243115) with *A* of *IL13* (rs20541) and *C* of *IL4* (rs2070874) (OR = 2.72). The receiver operating characteristic (ROC) analysis resulted in an area under the curve (AUC) of 0.895 (95%CI 0.874–0.916) for model including SNPs: the *A* allele of *IL19* (rs2243193) and *AA* genotype of *IL20* (rs2981573) combination, *IL19* (rs2243193), *IL12A* (rs2243115), *PPBP* (rs352010), and *IL4* (rs2070874) together with age and smoking pack/years, indicating a high ability of the model to correctly classify individuals with and without COPD.

Keywords: chronic obstructive pulmonary disease, cytokine, inflammation, IL19, PPBP, gene–gene interactions

DOI: 10.1134/S1022795423020047

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a multifactorial lung disease characterized by progressive restriction of airflow and hyperinflammatory response of the lungs to inhalation of harmful particles [1]. COPD is one of the most common chronic respiratory diseases with a high incidence of morbidity and mortality. Tobacco smoking, environmental pollution, and genetic predisposition are major risk factors for COPD [2]. Exposure to tobacco smoke induces pathological processes in the lungs, leading to a systemic inflammatory response. Inflammation is a key factor in the development of COPD. When activated as a result of oxidative stress macrophages and neutrophils secrete pro-inflammatory cytokines, chemokines, which circulate in the bloodstream and stimulate the secretion of acute phase proteins [1]. The genetic mechanisms of COPD development are actively studied around the world [3]. There is evidence that genes encoding cytokines and inflammatory mediators play an important role in the development of COPD [4]. Previously, using polygenic analysis, we demonstrated the presence of associations with COPD of a number of polymorphic variants of chemokine genes [5]. COPD is a multifactorial polygenic disease that develops as a result of the complex interactions of multiple genes, and as a result, the contribution of individual polymorphic markers may be small or not manifested at all.

The results we obtained suggest that the multilocus approach is more effective in identifying genetic predictors of multifactorial disease. In this work, we continued the study of gene–gene interactions in the development of COPD. The purpose of this study was

to identify the association of polymorphic variants of the cytokine and the immune response genes *IL19*, *IL20*, *IL24*, *PPBP*, *IL4*, *IL4RA*, *C5*, *FAS*, *FASLG*, and *TGFb1* with COPD.

MATERIALS AND METHODS

The study design is a candidate case-control study. We used DNA samples of unrelated individuals who live in the territory of the Republic of Bashkortostan, Tatars by ethnicity. The study was approved by the Ethics Committee of Institute of Biochemistry and Genetics, Ufa Research Center of the Russian Academy of Sciences (Protocol no. 17 of December 7, 2010). All study participants received informed voluntary consent for the use of biological material in the planned studies. The group of patients included 601 individuals (of which 522 were men (86.85%) and 79 women (13.15%)); the average age was 63.38 ± 11.81 years. Among COPD patients were smokers and former smokers, 484 people (80.53%), and nonsmokers, 117 (19.47%). The smoking index for smokers and former smokers was 44.58 ± 25.92 pack/years. The control group included 617 individuals (548 men (88.88%) and 69 women (11.12%)); the average age was 58.44 ± 14.79; smokers and former smokers, 517 (83.79%); nonsmokers, 100 (16.21%); the smoking index in smokers was 38.54 ± 23.12 pack/years. In all patients, the function of external respiration was examined by spirometry, the vital capacity (VC), the forced vital capacity (FVC), and the forced expiratory volume in the first second (FEV1) were evaluated; the ratio of forced expiratory volume in 1 s to the vital capacity (FEV1/VC). In the patient group, the indicators (in percent of the norm) were: FEV1 = 41.68 ± 19.32, FVC = 44.22 ± 17.88, VC = 49.02 ± 15.54, and FEV1/FVC = 58.66 ± 13.66. A detailed description of the criteria for inclusion and exclusion from study groups was described earlier [6].

Genotyping

DNA was isolated from peripheral blood leukocytes using phenolic-chloroform extraction. The following polymorphic loci were selected for our study: *FASLG* (rs763110, c.-844C>T), *IL19* (rs2243193, c.*258A>G), *IL20* (rs2981573, c.379-152A>G), *IL24* (rs291107, c.108-172T>C), *PPBP* (rs352010, c.-1411T>C), *IL4* (rs2243250, c.-589C>T), *IL4* (rs2070874, c.-33C>T), *C5* (rs17611, c.2422G>A, p.Val802Ile), *FAS* (rs1800682, c.-671A>G), *IL4RA* (rs1805010, c.223A>G, p.Ile75Val), and *TGFb1* (rs1800469, c.-1347T>C). The functional significance of polymorphic loci was investigated by the RegulomeDB Version 1.1 (https://regulomedb.org), SNPinfo Web Server (https://snpinfo.niehs.nih.gov) and HaploReg v3 databases [7]. Polymorphic gene variants were analyzed by real-time polymerase chain reaction (PCR) with commercial fluorescent detection kits (https://www.oligos.ru, OOO DNK Sintez, Russia) on the BioRad CFX96™ instrument (Bio-Rad Laboratories, Inc., United States). The detailed methods of analysis were described by us in [5, 6].

Statistical processing of results. Statistical data processing was carried out using the software packages of Statistica v. 6.0 (StatSoft Inc., United States) and PLINK v. 1.07 [8]. A detailed description of the standard methods of statistical analysis was provided earlier [6]. The analysis of allele/genotype combination associations with COPD was performed using the APSampler 3.6.1 program (http://sourceforge.net/projects/apsampler/). The main algorithm was described in [9]. Correction for multiple testing was performed using the FDR false-positive rate (false discovery rate, B. Hochberg), using the program (http//www.sdmproject.com/utilinies/?show=FDR) and a new value of $P_{\text{cor-FDR}}$ was obtained. Haplotype frequencies and linkage disequilibrium LD (D') were calculated in the Haploview 4.2 program. In order to study the mutual influence of clinical and genetic factors and assess the predictive value of polymorphic loci on the risk of COPD development, statistical processing of the obtained results by multiple regression and ROC analysis methods was carried out, with the calculation of the area under the curve (AUC) to evaluate the effectiveness of the predictive model using IBM SPSS Statistica 22.0 software.

RESULTS

Polymorphic loci of cytokine and immune response genes were analyzed in the formed COPD and control groups. The genes and SNPs for analysis were selected according to the criteria of belonging to those of functional significance and/or for which previous associations with other multifactorial human diseases were shown and a frequency of rare alleles (MAF) ≥5% in populations of Caucasians according to the National Center for Biotechnology Information database (http://www.ncbi.nlm.nih.gov/projects/SNP/). Functional analysis showed that the *IL19* locus (rs2243193) is located in the 3'-untranslated region of the gene and alters the binding sites for several miRNAs (hsa-miR-1259, hsa-miR-135b, hsa-miR-27a, hsa-miR-27b, hsa-miR-450b-5p, and hsa-miR-641). The *PPBP* locus (rs352010) is located in the 2KB region of a gene that has multiple transcription factor binding sites (ARNT and HIF). The *IL4* gene loci rs2070874 and rs2243250 are located in a 5'-untranslated region of DNA and the 2KB region of the gene, respectively, having binding sites for several transcription factors. The *FAS* locus (rs1800682) is located in the region of the gene promoter and changes the binding site with several transcription factors (STAT, ZBRK1, PAX3, and SP-1), and the *FASLG* locus (rs763110), according to the GTEx portal (https://www.gtexportal.org), causes a change in the expression of the *FASLG* gene in various tissues, including lung tissues. *TGFb1*

(rs1800469) is located in the 2KB region of the gene and alters the binding site for the transcription factor NR2F1, indicating the unambiguous effect of this polymorphism on gene expression. Thus, most of the selected polymorphic loci of the cytokine genes and immune response influenced gene expression or were linked to functional loci of the gene.

Before we begin to analyze the association of allelic variants of candidate genes with the development of COPD, we calculated the frequencies of alleles and genotypes in groups of COPD patients and control and the correspondence of the frequency distribution of genotypes to the Hardy–Weinberg equilibrium (Table 1). The genotypic frequencies of all polymorphic loci tested in the control group were found to be consistent with the Hardy–Weinberg equilibrium: *FASLG* (rs763110) ($P_{X\text{-}B}$ = 0.24), *IL19* (rs2243193) ($P_{X\text{-}B}$ = 0.084), *IL20* (rs2981573) ($P_{X\text{-}B}$ = 0.7), *IL24* (rs291107) ($P_{X\text{-}B}$ = 0.31), *PPBP* (rs352010) ($P_{X\text{-}B}$ = 0.19), *IL4* (rs2243250) ($P_{X\text{-}B}$ = 0.54), *IL4* (rs2070874) ($P_{X\text{-}B}$ = 0.81), *C5* (rs17611) ($P_{X\text{-}B}$ = 0.23), *FAS* (rs1800682) ($P_{X\text{-}B}$ = 0.43), *IL4RA* (rs1805010) ($P_{X\text{-}B}$ = 0.15), *TGFb1* (rs1800469) ($P_{X\text{-}B}$ = 0.46).

Next, we evaluated the statistical significance of the differences between the groups in the distribution of allele and genotypes frequencies and calculated the odds ratios for the rare allele of each locus. In the next step, using the logistic regression method, we analyzed the association of individual polymorphic loci or haplotypes of linked loci in different models (additive, dominant, recessive) taking into account quantitative and binary traits (sex, age, and body mass index) introduced into the regression equation as independent variables; the individual regression coefficient exponent (beta) was interpreted as the odds ratio (OR) with the calculation of the 95% confidence interval (in Table 2 statistically significant results of regression analysis of individual loci are presented, and in Table 3, an analysis of haplotypes of linked loci.

Given that in multifactorial diseases the contribution of individual genes to the development of the disease may be small or not manifested at all, we searched for informative gene–gene combinations that lead to the development of COPD using the APSampler (Allelic Pattern Sampler) program. This program uses the Monte Carlo Markov Chain Method (MCMC), based on Baeisov approaches, which allows one to identify combinations of alleles and genotypes of multiple loci associated with a feature under study [9].

When it comes to the diagnostic and predictive value of the results of molecular genetic analysis, one of the preferred analytical tools for analysis is the search for complex clinical and genetic models of the risk of developing a pathology. Therefore, at the final stage, we conducted statistical processing of the results obtained by methods of multiple regression analysis with step-by-step inclusion of the most significant predictors, followed by ROC analysis to assess the effectiveness of the prognostic model.

Analysis of the Association of Individual Polymorphic Variants of Cytokines and the Immune Response Genes with COPD

Statistically significant differences in the frequency distribution of polymorphic variants between COPD groups and healthy individuals were identified by polymorphic loci of the *IL19* (rs2243193), *IL4* (rs2243250), *IL4* (rs2070874), and *PPBP* (rs352010) genes (Table 1). The association of the *IL19* locus (rs2243193) with COPD was established in the dominant (P_{adj} = 0.00001, OR = 0.50) and log-additive model (P_{adj} = 0.00001, OR = 0.73) (Table 2). Given that the *IL19* (rs2243193), *IL20* (rs2981573) and *IL24* (rs291107) genes are located on the same chromosome 1q32.1 and are in the same linkage group, we conducted an analysis of the haplotypes of these loci. A significant linkage disequilibrium between polymorphic loci *IL19* (rs2243193) and *IL20* (rs2981573) (D' = 0.828, r^2 = 0.458) was shown (Table 3). Statistically significant differences in the distribution of haplotype frequencies between COPD and control groups were established (P = 0.00001). The frequency of haplotype G-A on *IL19* (rs2243193) and *IL20* (rs2981573) loci was significantly higher in the COPD group (P_{adj} = 2.073 × 10^{-6}, OR = 1.55). The *IL4* locus (rs2243250) was associated with COPD in a recessive (P_{adj} = 0.037, OR = 1.74) and log-additive model (P_{adj} = 0.024, OR = 1.27) (Tables 1, 2). The association of the *IL4* locus (rs2070874) with COPD was established in recessive (P_{adj} = 0.00001, OR = 0.07) and log additive models (P_{adj} = 0.00001, OR = 0.62). Linkage disequilibrium between the rs2243250 and rs2070874 loci of the *IL4* gene (D' = 0.46, r^2 = 0.15) was detected, as well as statistically significant differences in the pattern of frequency distribution of haplotypes of the *IL4* gene between the COPD and control groups (P = 0.00001) (Table 3). The frequency of the T-C haplotype of the rs2243250 and rs2070874 loci was significantly higher in the COPD group (P_{adj} = 0.0037, OR = 1.40). The *PPBP* locus (rs352010) was associated with COPD in the dominant (P_{adj} = 0.0044, OR = 0.70) and recessive models (P_{adj} = 0.00001, OR = 2.34) (Table 2).

Analysis of Gene–Gene Combinations of Cytokine and Immune Response Genes Polymorphic Loci with COPD

Using the APSampler program, we searched for informative gene–gene combinations associated with COPD. In addition to the 11 polymorphic loci we studied, five more cytokine gene loci that we previously studied were included in the analysis: *IL12RB2* (rs3762317), *IL12B* (rs3212227), *IL12A* (rs568408),

Table 1. The polymorphic loci genotypes and alleles frequency distribution of studied genes in COPD and control groups

Gene, SNP	A rare allele	Genotype, alleles	COPD *n* (%) (*N* = 601)	Control *n* (%) (*N* = 617)	*P*	OR (95% CI)
FASLG rs763110 C>T	*T*	*CC/CT/TT*	279/227/95 (46.42/37.77/15.81)	264/268/85 (42.79/43.44/13.78)	0.1250	0.97 (0.82–1.15)
		C/T	785/417 (65.31/34.69)	796/438 (64.51/35.49)	0.7117	0.96 (0.82–1.14)
IL19 rs2243193 A>G	*A*	*GG/GA/AA*	288/188/125 (47.92/31.28/20.80)	195/283/139 (31.60/45.87/22.53)	0.00001	0.73 (0.62–0.86)
		G/A	764/438 (63.56/36.44)	673/561 (54.54/45.46)	2.866×10^{-5}	0.68 (0.58–0.81)
IL20 rs2981573 A>G	*G*	*AA/AG/GG*	295/245/61 (49.08/40.77/10.15)	281/267/69 (45.54/43.27/11.18)	0.457	0.90 (0.75–1.08)
		A/G	835/367 (69.47/30.53)	829/405 (67.18/32.82)	0.2437	0.89 (0.76–1.07)
IL24 rs291107 T>C	*C*	*TT/TC/CC*	198/272/131 (32.95/45.26/21.80)	176/294/147 (28.53/47.65/23.82)	0.239	0.88 (0.75–1.04)
		T/C	668/534 (55.57/44.43)	646/588 (52.35/47.65)	0.1258	0.87 (0.74–1.03)
PPBP rs352010 T>C	*T*	*CC/CT/TT*	418/106/77 (69.55/17.64/12.81)	378/202/37 (61.26/32.74/6.00)	0.00001	0.97 (0.81–1.16)
		C/T	942/260 (78.37/21.63)	958/276 (77.63/22.37)	0.7808	0.96 (0.79–1.16)
IL4 rs2243250 C>T	*T*	*CC/CT/TT*	319/234/48 (53.08/38.94/7.99)	362/225/30 (58.67/36.47/4.86)	0.033	1.27 (1.03–1.56)
		C/T	872/330 (72.55/27.45)	949/285 (76.90/23.10)	0.024	1.26 (1.05–1.51)
IL4 rs2070874 C>T	*T*	*CC/CT/TT*	412/186/3 (68.55/30.95/0.50)	367/219/31 (59.48/35.49/5.02)	0.00001	0.62 (0.50–0.78)
		C/T	1010/192 (84.03/15.97)	953/281 (77.23/22.77)	5.407×10^{-5}	0.64 (0.53–0.79)
C5 rs17611 G>A	*A*	*GG/GA/AA*	200/276/125 (33.28/45.92/20.80)	205/288/124 (33.23/46.68/20.10)	0.946	1.01 (0.86–1.20)
		G/A	676/526 (56.24/43.76)	698/536 (56.56/43.44)	0.8729	1.01 (0.86–1.19)

Table 1. (Contd.)

Gene, SNP	A rare allele	Genotype, alleles	COPD *n* (%) (*N* = 601)	Control *n* (%) (*N* = 617)	*P*	OR (95% CI)
FAS rs1800682 A>G	*G*	*AA/AG/GG*	210/269/122 (34.94/44.76/20.30)	193/294/130 (31.28/47.65/21.07)	0.392	0.92 (0.77–1.09)
		A/G	689/513 (57.32/42.68)	680/554 (55.11/44.89)	0.300	0.91 (0.78–1.07)
IL4RA rs1805010 A>G	*G*	*AA/AG/GG*	213/251/137 (35.44/41.76/22.80)	202/286/129 (32.74/46.35/20.91)	0.272	0.99 (0.84–1.16)
		A/G	677/525 (56.32/43.68)	690/544 (55.92/44.08)	0.8943	0.98 (0.84–1.15)
TGFb1 rs1800469 T>C	*T*	*CC/CT/TT*	290/240/71 (48.25/39.93/11.81)	261/273/83 (42.30/44.25/13.45)	0.112	0.85 (0.71–1.02)
		C/T	820/382 (68.22/31.78)	795/439 (64.42/35.58)	0.053	0.84 (0.71–0.99)

P is the significance of group differences in allele and genotype frequencies (sample χ^2 homogeneity test); OR is the odds ratio indicator for the rare allele (baseline allele test).

IL12A (rs2243115), and *IL13* (rs20541) [10]. There are 2587 patterns associated with COPD; Table 4 shows the results of the most significant combinations with P_{FDR} less than 0.05 and/or more than 1.5 (for risk combinations) or less than 0.4 for protective combinations. Most of the combinations identified included alleles or genotypes of *PPBP* (rs352010) genes (nine gene combinations), *IL19* (rs2243193) (six gene combinations), and *IL12A* (rs568408) genes (six gene combinations). The most significant combinations associated with a reduced risk of COPD included the *IL19* locus allele *A* (rs2243193) and the *C* allele or the *PPBP* locus *TC* genotype (rs352010), as well as *AA* genotype of the *IL20* locus (rs2981573). The *PPBP* locus *CC* genotype (rs352010) was a mandatory part of three patterns associated with COPD risk. However, the greatest risk of COPD was determined by a combination of alleles of three functionally interconnected cytokines: *G* of *IL12A* (rs2243115), *A* of *IL13* (rs20541) and *C* of *IL4 (*rs2070874) (OR = 2.72). The analysis of allele/genotype combinations of the polymorphic loci studied allowed us to determine the associations of polymorphic loci of the *C5* (rs17611), *FASLG* (rs763110) and *TGFb1* (rs1800469) genes, which showed their effects only in combination with the *PPBP* (rs352010), *IL12A* (rs568408), and *IL19* (rs2243193) genes.

Assessing the Prognostic Significance of the Studied Cytokine and Inflammatory Response Molecules Genes Polymorphic Loci

Based on the results of multiple regression and ROC analysis, the first predictive model of the risk of COPD formation included such signs as age and the smoking index and polymorphic gene variants: a combination of the *IL19* (rs2243193) allele *A* and *IL20* (rs2981573) genotype *AA*, *IL19* (rs2243193), *IL12A* (rs2243115), *PPBP* (rs352010), and *IL4 (* rs2070874). The model was characterized by a high predictive power of AUC = 0.895 (Table 5, Fig. 1). This indicates the good ability of the model to correctly classify individuals with and without COPD. The second significant prognostic model included only polymorphic variants of genes: the combination of the *IL19* (rs2243193) allele *A* and *IL20* (rs2981573) genotype *AA*, *IL19* (rs2243193), *IL12A* (rs2243115), *PPBP* (rs352010), and *IL4* (rs2070874) (Table 5, Fig. 1). ROC analysis of the obtained model showed its moderate predictive ability at AUC = 0.676. This model can effectively identify people suffering from COPD.

DISCUSSION

In order to study the combined effect of cytokine and the immune response genes on the risk of COPD, as well as to search for complex clinical genetic models

Table 2. Statistically significant results of the association analysis of studied SNPs with COPD (log-regression analysis)

Gene, SNP	A rare allele	*N*	Genotype, model	OR_{adj} (CI 95%)	P_{adj}	$P_{cor\text{-}FDR}$
IL19 rs2243193 A>G	*A*	1218	*GG* *GA* + *AA* Dominant	1.00 0.50 (0.39–0.64)	0.00001	1.8×10^{-5}
			Log-additive	0.73 (0.62–0.86)	0.00001	1.8×10^{-5}
IL4 rs2243250 C>T	*T*	1218	*CC* + *CT* *TT* Recessive	1.00 1.74 (1.02–2.95)	0.039	0.043
			Log-additive	1.27 (1.03–1.56)	0.024	0.0308
IL4 rs2070874 C>T	*T*	1218	*CC* + *CT* *TT* Recessive	1.00 0.07 (0.02–0.31)	0.00001	1.8×10^{-5}
			Log-additive	0.62 (0.50—0.78)	0.00001	1.8×10^{-5}
PPBP rs352010 T>C	*T*	1218	*CC* *CT* + *TT* Dominant	1.00 0.70 (0.54–0.89)	0.0044	0.0066
			CC + *CT* *TT* Recessive	1.00 2.34 (1.52–3.60)	0.00001	1.8×10^{-5}
			Log-additive	0.97 (0.81–1.16)	0.76	0.76

N is the number of individuals included in the regression analysis; P_{adj} is the significance for the likelihood ratio test of the log-regression model, taking into account age, status and smoking index, body mass index, sex. OR_{adj} is the odds ratio with all factors in mind, CI 95% is the 95% confidence interval for or; $P_{cor\text{-}FDR}$ is the significance of the test after FDR correction; log-additive model for rare allele dose is the rare allele dose increase in the series: homozygote in the frequent allele (0), heterozygote (1), homozygote in the rare allele (2).

of the disease risk, we conducted an association analysis of eleven polymorphic loci of the *IL19*, *IL20*, *IL24*, *PPBP*, *IL4*, *IL4RA*, *C5*, *FAS*, *FASLG*, and *TGFb1* genes in combination with previously studied polymorphic loci of the *IL12RB2*, *IL12B*, *IL12A*, and *IL13* cytokine genes with COPD [10]. As a result of the study, using a sequential analysis of the association of individual loci of cytokine and the immune response genes and gene–gene combinations of the polymorphic loci of the studied genes, we have established a significant association of polymorphic variants of the *IL19*, *IL20*, *PPBP*, and *IL4* genes with COPD.

We have shown an association of the *IL19* gene (rs2243193) loci with COPD for the first time; the risk of COPD was associated with the GG genotype, while the rare allele marked a reduced risk of developing the disease. IL19, IL20, and IL24 belong to the cytokine family IL10 [11], are located on chromosome 1q32.1, and have a similar structure. Representatives of this family of cytokines play an important role in the development of infectious and inflammatory diseases [11]. IL19 is produced by macrophages and monocytes when activated by extracellular pathogens, which may lead to additional activation of other cytokines (TNFA, IL6, and IL12) [12]. Previously, polymorphic variants of the *IL19* gene have been shown to be associated with the development of autoimmune diseases and cardiovascular complications in rheumatoid arthritis [13, 14]. Increases in serum IL19 were correlated with progression of COPD [15]. The contribution of polymorphic variants of the *IL19* gene to COPD has not previously been studied.

An association has been established with COPD and the polymorphic locus (rs352010) of the *PPBP* gene encoding the proplatelet basic protein (PPBP), which is located at the 4q13.3 site (https://www.ncbi.nlm.nih.gov/gene/5473). PPBP is a platelet-derived growth factor or platelet chemokine belonging to the CXC family of chemokines (CXCL7) that activates neutrophils via the CHCR2 receptor [16]. Information on the relationship between COPD and PPBP (CXCL7) is limited, but there is evidence of its involvement in the pathogenesis of acute lung injury [17]. The data from our study confirm the association of the *PPBP* (rs352010) gene with COPD; more functional studies are needed to interpret the results.

Significant associations with COPD were identified with *IL4* polymorphic loci rs2070874 and rs2243250. The *IL4* gene is located on chromosome

Table 3. The association of haplotypes of polymorphic loci genes *IL19*, *IL20,* and *IL4* with COPD

Haplotype	COPD frequency/control	OR (95%CI)	P_{adj}
IL19 (rs2243193) A>G–*IL20* (rs2981573) A>G ($N = 1218$)			
G-A	0.617/0.509	1.55 (1.29–1.86)	2.073×10^{-6}
A-G	0.274/0.292	0.90 (0.75–1.09)	0.3883
A-A	0.082/0.162	0.43 (0.32–0.58)	1.078×10^{-7}
G-G	0.027/0.037	0.58 (0.35–0.98)	0.043
P-value by haplotype frequency distribution between groups			0.00001
IL4 (rs2243250) C>T–*IL4* (rs2070874) C>T ($N = 1218$)			
C-C	0.627/0.653	0.896 (0.74–1.14)	0.3711
T-C	0.180/0.119	1.40 (1.05–1.86)	0.0037
T-T	0.139/0.123	0.82 (0.61–1.11)	0.4022
C-T	0.054/0.106	0.50 (0.35–0.72)	0.0016
P-value by haplotype frequency distribution between groups			0.00001

N is the number of individuals included in the regression analysis; P_{adj} is the significance for the likelihood ratio test of the log-regression model, taking into account age, status and smoking index, body mass index, sex. OR_{adj} is the ratio of odds taking into account all factors, CI 95% is the 95% confidence interval for OR; *D*' is the coefficient of imbalance in the linkage between the two loci (Levontin coefficient); r^2 is the correlation coefficient; *IL19* (rs2243193) and *IL20* (rs2981573) $D' = 0.828$, $r^2 = 0.458$; rs2243250C>T and rs2070874C>T of the *IL4* gene ($D' = 0.46$, $r^2 = 0.15$).

5q31.1 in the same cluster with the *IL3*, *IL5*, and *IL13* genes and *CSF2*. IL4 is an important cytokine involved in tissue repair, mediating and regulating various allergic and acute inflammatory reactions (https://www.ncbi.nlm.nih.gov/gene/3565). According to a number of studies, the role of polymorphic variants of the *IL4* gene in the pathogenesis of bronchial asthma [18], COPD [19] has been proved.

The multi-locus analysis showed that the *IL19* allele *A* (rs2243193) and *IL20 AA* genotype (rs2981573) are among the most significant combinations associated with a reduced risk of COPD. Along with polymorphic loci variants *IL12A* (rs2243115, rs568408), *PPBP* (rs352010), and *IL4* (rs2070874, rs2243250). The *PPBP* (rs352010) *CC* genotype was part of three informative combinations associated with a high risk of COPD. The allele *C* of the *IL4* (rs2070874) locus was one of the most informative combinations marking the risk of developing the disease, in combination with the alleles *G* of the *IL12A* (rs2243115) locus and *A* locus *IL13* (rs20541), which are functionally related cytokines. When conducting a polygenic association analysis, genetic patterns significantly associated with COPD were identified, which included polymorphic loci of the *FASLG*, *TGFb1,* and *C5* genes that provide their effect only in combination with other genes of cytokines and immune response. FASLG is a member of the tumor necrosis factor superfamily; its primary function is induction of apoptosis triggered by binding to FAS (https://www.ncbi.nlm.nih.gov/gene/356). The role of *FASLG* genes in the development of COPD has not been studied before, but there is evidence about the key role of FASL/FAS, as a system in the development of pulmonary inflammation, damage, and fibrosis [20]. The *TGFb1* gene, which is located on chromosome 19q13.2, encodes the secreted ligand of the superfamily of transforming growth factor beta proteins (https://www.ncbi.nlm.nih.gov/gene/7040), involved in the regulation of various cellular processes such as growth, development, differentiation, proliferation, cell mobility, adhesion, and apoptosis [21]. TGFβ1 plays a key role in tissue damage caused by smoking and airway remodeling [22]. The *C5* gene, which encodes component 5 of the complement system, is located on chromosome 9q33.2. C5 is an integral part of the innate immunity system, which plays an important role in the development of the inflammatory response, homeostasis, and protection against pathogens (https://www.ncbi.nlm.nih.gov/gene/727). The results we obtained show the promise of polygenic analysis of associations in the study of a complex heterogeneous disease as COPD.

The analysis of individual risk factors of the disease does not take their interactions into account, which is necessary to predict the risk of developing pathology. Based on multiple regression and ROC analysis, the predictive model of COPD risk includes factors such as age and the smoking index, as well as genetic predictors obtained from the analysis of gene–gene interactions and individual loci: the combination of the

Table 4. Combinations of alleles and/or genotypes of cytokine and immune response molecules genes polymorphic loci most significantly associated with COPD

Combinations	COPD (frequency)	Control (frequency)	*P*	P_{FDR}	OR	CI (95%)
IL20 (rs2981573) *AA* + *IL19* (rs2243193) *A* + *PPBP* (rs352010) *C*	0.04	0.17	1.68e-11	6.92e-08	0.20	0.12–0.34
IL19 (rs2243193) *A* + *IL4* (rs2243250) *C* + *PPBP* (rs352010) *TC*	0.07	0.23	2.6e-11	5.34e-08	0.25	0.16–0.39
IL20 (rs2981573) *AA* + *IL19* (rs2243193) *A* + *IL12A* (rs568408) *G*	0.033	0.16	6.34e-11	3.72e-08	0.18	0.09–0.32
IL20 (rs2981573) *AA* + *IL19* (rs2243193) *A*	0.046	0.172	1.22e-10	4.17e-08	0.23	0.14–0.38
IL12A (rs568408) *G* + *IL4* (rs2243250) *C* + *PPBP* (rs352010) *TC*	0.12	0.299	1.94e-10	3.98e-08	0.32	0.22–0.46
IL19 (rs2243193) *A* + *C5* (rs17611) *G* + *PPBP* (rs352010) *T* + *FASLG* (rs763110) *C*	0.075	0.212	8.61e-10	9.58e-08	0.30	0.20–0.45
IL4 (rs2243250) *C* + *C5* (rs17611) *G* + *TGFb1* (rs1800469) *C* + *PPBP* (rs352010) *T*	0.129	0.269	6.19e-08	1.25e-06	0.39	0.27–0.56
IL12A (rs2243115) *G* + *IL13* (rs20541) *A* + *IL4* (rs2070874) *C*	0.236	0.102	1e-07	1.72e-06	2.72	1.85–3.97
IL12A (rs568408) *GG* + *IL4* (rs2243250) *C* + *PPBP* (rs352010) *T*	0.068	0.189	1.47e-07	2.27e-06	0.31	0.19–0.50
IL12A (rs2243115) *G* + *IL4* (rs2070874) *C*	0.333	0.189	1.77e-07	2.59e-06	2.14	1.59–2.86
IL12RB2 (rs3762317) *A* + *PPBP* (rs352010) *CC*	0.696	0.565	1.34e-05	5.87e-05	1.76	1.35–2.29
TGFb1 (rs1800469) *C* + *PPBP* (rs352010) *CC*	0.67	0.542	2.04e-05	8.26e-05	1.72	1.32–2.21
IL12A (rs568408) *A* + *PPBP* (rs352010) *CC*	0.392	0.288	0.000517	0.0012	1.60	1.21–2.09
IL12A (rs568408) *A* + *FASLG* (rs763110) *CC*	0.268	0.184	0.00151	0.0032	1.62	1.19–2.21
IL19 (rs2243193) *G* + *IL12A* (rs568408) *A*	0.445	0.346	0.00155	0.0032	1.52	1.16–1.98

P—value is the Fisher's test significance level, P_{FDR} is the FDR value after FDR correction, OR odds ratio, 95% CI is the 95% confidence interval for the OR.

Table 5. Predictive regression models of COPD development

Variables in the equation	β	P_{Wald}	OR	95% CI_{OR}
Model 1				
IL19 (rs2243193) *A* + *IL20* (rs2981573) *AA*	−1.05	0.0028731	0.35	0.17–0.7
IL19 (rs2243193) *GG*	0.19	0.056	1.20	0.99–1.46
IL19 (rs2243193) *A*	−0.19	0.056	0.83	0.69–1.01
IL12A (rs2243115) *T*	−1.44	1.57×10^{-5}	0.24	0.12–0.45
IL12A (rs2243115) *GG*	1.44		4.23	2.2–8.14
PPBP (rs352010) *CC*	0.04	0.0167	1.04	0.73–1.5
PPBP (rs352010) *CT*	−0.53		0.59	0.4–0.88
PPBP (rs352010) *TT*	0.49		1.63	0.87–3.04
IL4 (rs2070874) *C*	0.79	0.054	2.19	0.99–4.88
IL4 (rs2070874) *TT*	−0.79		0.46	0.21–1.01
Age	0.10	1.3×10^{-30}	1.10	1.08–1.12
Index of smoking	0.05	2.83×10^{-20}	1.05	1.04–1.06
Intercept	−5.93	3.33×10^{-17}		
$\chi^2 = 499.415$ $df = 8$ $P = 1.99 \times 10^{-103}$				
Model 2				
IL19 (rs2243193) *A* + *IL20* (rs2981573) *AA*	−1.27	1.09×10^{-5}	0.28	0.16–0.49
IL19 (rs2243193) *A*	−0.22	0.003	0.80	0.69–0.93
IL19 (rs2243193) *GG*	0.22		1.25	1.08–1.44
IL12A (rs2243115) *GG*	1.18	2.256×10^{-5}	3.25	1.88–5.59
IL12A (rs2243115) *T*	−1.18		0.31	0.18–0.53
PPBP (rs352010) *CC*	0.23	0.0001	1.26	0.96–1.65
PPBP (rs352010) *CT*	−0.50		0.61	0.45–0.83
PPBP (rs352010) *TT*	0.27		1.31	0.82–2.1
IL4 (rs2070874) *TT*	−1.10	0.003	0.33	0.16–0.69
IL4 (rs2070874) *C*	1.10		3.01	1.44–6.25
Intercept	−0.02	0.974	0.98	
$\chi^2 = 114.817$ $df = 6$ $P = 1.99 \times 10^{-22}$				

β is the beta coefficient for variable, P_{wald} is the significance for Wald statistics, the OR is (exp (β)), χ^2 is the likelihood ratio test (LR), *df* is the number of degrees of freedom, *P* is the value for the likelihood ratio test.

allele *A IL19* (rs2243193) and the *AA* genotype *IL20* (rs2981573) locus, as well as polymorphic variants of the genes of functionally interconnected cytokines: *IL19* (rs2243193), *IL12A* (rs2243115), *PPBP* (rs352010), and *IL4* (rs2070874).

In conclusion, we have identified significant associations with COPD of both individual polymorphic loci and complex polygenic combinations of the studied cytokine and immune response genes. A significant association of the *IL19* (rs2243193) and *PPBP* (rs352010) genes with COPD is shown for the first time in our study. Using a multi-locus association analysis, we demonstrated the presence of COPD associations of the *IL20*, *FASLG*, *TGFb1*, and *C5*

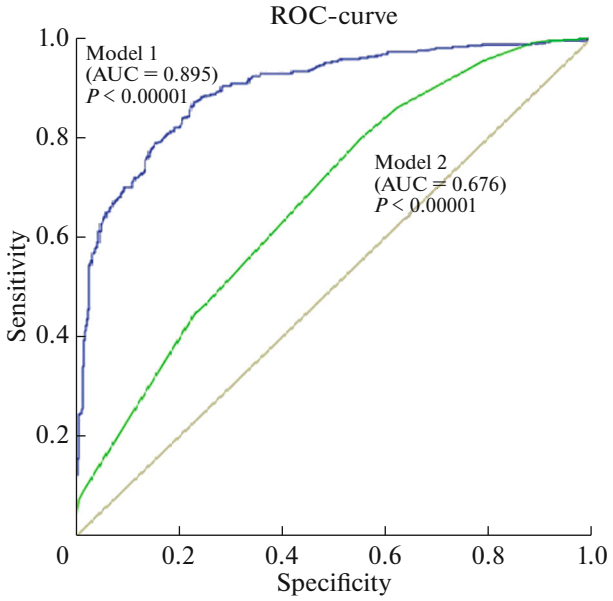

Fig. 1. The ROC curve to estimate the predictive power of regression models for COPD. AUC is the area under curve. Model 1 (AUC = 0.895) includes the following genes: combination of the allele *A* of *IL19* (rs2243193) and the *AA* genotype of *IL20* (rs2981573) locus, *IL19* (rs2243193), *IL12A* (rs2243115), *PPBP* (rs352010) and *IL4* (rs2070874) together with demographic indicators: the age and smoking index ($P = 1.99 \times 10^{-103}$; sensitivity, 85.5%; specificity; 77.5%). Model 2 (AUC = 0.679) includes only genes: combination of *A* allele of *IL19* (rs2243193) and genotype *AA* of *IL20* (rs2981573), *IL19* (rs2243193), *IL12A* (rs2243115), *PPBP* (rs352010) and *IL4* (rs2070874) ($P = 11.99 \times 10^{-22}$; sensitivity; 76.4%; specificity; 45.1%).

genes polymorphic variants only in informative combinations with *PPBP, IL12A* and *IL19* genes alleles, which may indicate the synergy of the genes.

ACKNOWLEDGMENTS

The biological material (DNA) for the study was taken from the Collection of Human Biological Materials of the Institute of Biochemistry and Genetics, Ufa Research Center, Russian Academy of Sciences supported by the program of bioresource collections of FANO Russia; the work was performed using the equipment of the Central Enterprise Biomika and UNU KODINK (Institute of Biochemistry and Genetics, Ufa Research Center, Russian Academy of Sciences).

FUNDING

The study was conducted within the framework of research no. 122041400169-2; the study was partially supported by the megagrant of the Government of the Russian Federation (2020-220-08-2197, agreement no. 075-15-2021-595.

COMPLIANCE WITH ETHICAL STANDARDS

Conflict of interest. The authors declare that they have no conflicts of interest.

Statement of compliance with standards of research involving humans as subjects. All procedures performed in the human study comply with the ethical standards of the institutional committee on research ethics and the 1964 Helsinki Declaration and its subsequent changes or comparable standards of ethics.

Informed voluntary consent was obtained from each of the participants enrolled in the study.

REFERENCES

1. Barnes, P.J., Inflammatory mechanisms in patients with chronic obstructive pulmonary disease, *J. Allergy Clin. Immunol.*, 2016, vol. 138, no. 1, pp. 16—27. https://doi.org/10.1016/j.jaci.2016.05.011
2. Hurst, J.R., Siddiqui, M.K., Singh, B., et al., A systematic literature review of the humanistic burden of COPD, *Int. J. Chron. Obstruct. Pulmon. Dis.*, 2021, vol. 16, pp. 1303—1314. https://doi.org/10.2147/COPD.S296696
3. Silverman, E.K., Genetics of COPD, *Annu. Rev. Physiol.*, 2020, vol. 82, pp. 413—431. https://doi.org/10.1146/annurev-physiol-021317-121224
4. Ahmadi, A., Ghaedi, H., Salimian, J., et al., Association between chronic obstructive pulmonary disease and interleukins gene variants: a systematic review and meta-analysis, *Cytokine*, 2019, vol. 117, pp. 65—71. https://doi.org/10.1016/j.cyto.2019.02.003
5. Korytina, G.F., Aznabaeva, Y.G., Akhmadishina, L.Z., et al., The relationship between chemokine and chemokine receptor genes polymorphisms and chronic obstructive pulmonary disease susceptibility in Tatar population from Russia: a case control study, *Biochem. Genet.*, 2022, vol. 60, no. 1, pp. 54—79. https://doi.org/10.1007/s10528-021-10087-2
6. Korytina, G.F., Akhmadishina, L.Z., Aznabaeva, Y.G., et al., Associations of the NRF2/KEAP1 pathway and antioxidant defense gene polymorphisms with chronic obstructive pulmonary disease, *Gene*, 2019, vol. 692, pp. 102—112. https://doi.org/10.1016/j.gene.2018.12.061
7. Ward, L.D. and Kellis, M., HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease, *Nucleic Acids Res.*, 2016, vol. 44, no. D1, pp. D877—D881. https://doi.org/10.1093/nar/gkv1340
8. Purcell, S., Neale, B., Todd-Brown, K., et al., PLINK: a toolset for whole-genome association and population-based linkage analysis, *Am. J. Hum. Genet.*, 2007, vol. 81, no. 3, pp. 559—575. https://doi.org/10.1086/519795
9. Favorov, A.V., Andreewski, T.V., Sudomoina, M.A., et al., A Markov chain Monte Carlo technique for identification of combinations of allelic variants underlying complex diseases in humans, *Genetics*, 2005, vol. 171, no. 4, pp. 2113—2121. https://doi.org/10.1534/genetics.105.048090

10. Korytina, G.F., Zulkarneev, Sh.R., Aznabaeva, Yu.G., et al., The contribution of *IL12A, IL12B, IL13* and *IL12RB2* gene polymorphisms to the development of chronic obstructive pulmonary disease in the ethnic group of Tatars, *Yakutsk. Med. Zh.*, 2021, vol. 75, no. 3, pp. 21—25. https://doi.org/10.25789/YMJ.2021.75.05
11. Ouyang, W., Rutz, S., Crellin, N.K., et al., Regulation and functions of the IL-10 family of cytokines in inflammation and disease, *Annu. Rev. Immunol.*, 2011, vol. 29, pp. 71—109. https://doi.org/10.1146/annurev-immunol-031210-101312
12. Azuma, Y.T., Fujita, T., Izawa, T., et al., IL-19 contributes to the development of nonalcoholic steatohepatitis by altering lipid metabolism, *Cells*, 2021, vol. 10, no. 12, p. 3513. https://doi.org/10.3390/cells10123513
13. Wang, W., Wang, X., Yang, K., et al., Association of *BCL2* polymorphisms and the *IL19* single nucleotide polymorphism rs2243188 with systemic lupus erythematosus, *J. Int. Med. Res.*, 2021, vol. 49, no. 5. https://doi.org/10.1177/03000605211019187
14. Leonard, D., Svenungsson, E., Dahlqvist, J., et al., Novel gene variants associated with cardiovascular disease in systemic lupus erythematosus and rheumatoid arthritis, *Ann. Rheum. Dis.*, 2018, vol. 77, no. 7, pp. 1063—1069. https://doi.org/10.1136/annrheumdis-2017-212614
15. Rong, B., Liu, Y., Li, M., et al., Correlation of serum levels of HIF-1α and IL-19 with the disease progression of COPD: a retrospective study, *Int. J. Chron. Obstruct. Pulmon. Dis.*, 2018, vol. 13, pp. 3791—3803. https://doi.org/10.2147/COPD.S177034
16. Brown, A.J., Joseph, P.R., Sawant, K.V., et al., Chemokine CXCL7 heterodimers: structural insights, CXCR2 receptor function, and glycosaminoglycan interactions, *Int. J. Mol. Sci.*, 2017, vol. 18, no. 4, p. 748. https://doi.org/10.3390/ijms18040748
17. Bdeir, K., Gollomp, K., Stasiak, M., et al., Platelet-specific chemokines contribute to the pathogenesis of acute lung injury, *Am. J. Respir. Cell Mol. Biol.*, 2017, vol. 56, no. 2, pp. 261—270. https://doi.org/10.1165/rcmb.2015-0245OC
18. Kousha, A., Mahdavi Gorabi, A., Forouzesh, M., et al., Interleukin 4 gene polymorphism (−589C/T) and the risk of asthma: a meta-analysis and met-regression based on 55 studies, *BMC Immunol.*, 2020, vol. 21, no. 1, p. 55. https://doi.org/10.1186/s12865-020-00384-7
19. Choudhury, P., Biswas, S., Singh, G., et al., Immunological profiling and development of a sensing device for detection of IL-13 in COPD and asthma, *Bioelectrochemistry*, 2022, vol. 143, p. 107971. https://doi.org/10.1016/j.bioelechem.2021.107971
20. Dosreis, G.A., Borges, V.M., and Zin, W.A., The central role of Fas-ligand cell signaling in inflammatory lung diseases, *J. Cell Mol. Med.*, 2004, vol. 8, no. 3, pp. 285—293. https://doi.org/10.1111/j.1582-4934.2004.tb00318.x
21. Caramori, G., Adcock, I.M., Di Stefano, A., et al., Cytokine inhibition in the treatment of COPD, *Int. J. Chron. Obstruct. Pulmon. Dis.*, 2014, vol. 9, pp. 397—412. https://doi.org/10.2147/COPD.S42544
22. Tam, A., Leclair, P., Li, L., et al., FAM13A as potential therapeutic target in modulating TGF-β-induced airway tissue remodeling in COPD, *Am. J. Physiol.: Lung Cell Mol. Physiol.*, 2021, vol. 321, no. 2, pp. L377—L391. https://doi.org/10.1152/ajplung.00477.2020