ISSN 0026-8933, Molecular Biology, 2013, Vol. 47, No. 6, pp. 852–862. © Pleiades Publishing, Inc., 2013.
Original Russian Text © V.A. Stepanov, E.A. Trifonova, 2013, published in Molekulyarnaya Biologiya, 2013, Vol. 47, No. 6, pp. 976–986.

GENOMICS.
TRANSCRIPTOMICS

UDC 575.174:577.21

Multiplex SNP Genotyping by MALDI-TOF Mass Spectrometry: Frequencies of 56 Immune Response Gene SNPs in Human Populations

V. A. Stepanov and E. A. Trifonova

Research Institute for Medical Genetics, Siberian Branch of the Russian Academy of Medical Sciences, Tomsk, 634050 Russia;

e-mail: vadim.stepanov@medgenetics.ru

Received May 13, 2013; in final form, June 27, 2013

Abstract—Single-nucleotide polymorphisms (SNPs) are the most common type of genetic polymorphisms. Despite the progress in sequencing and postgenomic technologies, targeted SNP genotyping continues to be in highest demand in the approach to human and medical genetics. In this work, we describe the application of multiple SNP genotyping by MALDI-TOF mass spectrometry for analysis of genetic diversity of immune response genes in human populations. It was shown that MALDI-TOF mass spectrometry is a rapid, accurate, and efficient method of medium-scale SNP genotyping. Allele frequencies of 56 SNPs in 41 genes implicated in the regulation of immune response were similar in four populations studied (Russians, Komi, Khanty, and Buryats). These populations had similar levels of genetic diversity and were clustered according to their geographic location. The cost efficiency of MALDI-TOF mass spectrometry was evaluated compared to real-time PCR technology.

DOI: 10.1134/S0026893313060149

Keywords: single nucleotide polymorphisms (SNPs), MALDI-TOF mass spectrometry, multiplex genotyping, real-time PCR, genetic diversity, human populations, immune response genes

INTRODUCTION

Single-nucleotide polymorphisms (SNPs) are the most common type of genetic polymorphism. The current version of the dbSNP database (build 137), which accumulates information on single-nucleotide genome variation, describes approximately 38 million validated SNPs in the human genome [1]. Thus, in the human genome, on average, 1 nucleotide in 100 is variable. The advent of next-generation sequencing (NGS) techniques and the work of the international HapMap and 1000 Genomes Projects have much contributed to the progress in collecting information on the human genome variation [2, 3]. In each individual genome, NGS resequencing identifies 3–5 million SNPs [4–7]. Presumably, on average, two unrelated individual genomes differ by 1 nucleotide per 1000.

Although the progress in sequencing and postgenomic techniques has been considerable, targeted SNP genotyping is still the most widely used approach in human genetics and medical genetics. An analysis of SNP variations is used in the genetic mapping of human diseases, in association studies on common diseases (including genome-wide studies), in human evolution and population genetics, in DNA diagnostics of monogenic diseases, in forensic DNA identification, and in many other basic research areas and practical applications [8].

In the past 30 years, SNP genotyping techniques have developed from the manual genotyping of individual markers by nuclear acid hybridization or by PCR with the identification of amplicon differences (PCR-RFLP, allele-specific PCR, etc.) to automated multiplex analysis of hundreds of thousands to millions of SNPs on high-density DNA biochips [8–11].

In current research and laboratory practice, SNPs are most commonly genotyped by real-time PCR using TaqMan probes with a fluorescent label [12]. Large-scale studies employ high-density biochip-based genotyping [9, 13, 14]. While the first approach allows the genotyping of individual markers in hundreds or thousands of specimens in a single experiment, the other one delivers information on dozens or hundreds of thousands of SNPs in dozens of specimens. At the same time, there are still not enough modern technological solutions for intermediate-scale genotyping of dozens or hundreds of markers in hundreds of specimens, which is most commonly required for the majority of purposes. Approaches involving low-density DNA microarrays [15–19] or SNaPShot analysis by capillary electrophoresis [20, 21] remain unpopular because of the restrictions on multiplexing and limited capacity of amplicon resolution.

One of the techniques that address the latter group of research problems is the mass spectrometry of biological macromolecules, including DNA [22]. Mass spectrometry-based approaches are used in genetics and molecular biology for genotyping, the analysis of gene expression, and DNA methylation studies [23–25]. In this work, we describe an example of applying multiplex SNP genotyping by mass spectrometry for an analysis of 56 markers in immune response genes in four human populations, and evaluate the economic efficiency of the suggested approach compared to real-time PCR.

EXPERIMENTAL

Principle of MALDI-TOF mass spectrometry. The method of mass spectrometry is based on measuring the ratio between the masses of charged particles and their charges. There are different methods of ionizing organic molecules depending on the nature and phase of the analyte. The method of matrix assisted laser desorption/ionisation (MALDI) enables ionization of large biomolecules without fragmenting them; i.e., a specimen in the form of a solid-phase solution of the analyte in an organic matrix is irradiated by short laser pulses. Molecules of the matrix heavily absorb photons emitted by an ultraviolet or an infrared laser, and dense high-temperature plasma is produced over the specimen surface. In addition to ions and molecules of the matrix material, the plume contains molecules of the compound under study. These molecules are ionized by photon absorption or in ion–molecule reactions, and the resulting positive or negative ions move down the potential gradient from the ionization zone into the analyzer. In the analyzer, the charged particles are separated by their time of flight (TOF) for a particular distance; this time is proportional to the ratio between the mass of the particle and its charge. The MALDI-TOF technique is currently widely used to analyze peptides, proteins, nucleotides, polysaccharides, synthetic polymers, humic acids, organic complex compounds, and other biomolecules with molecular masses of up to several hundreds of kilodaltons [22].

Principal steps of the experiment. Multiplex genotyping was performed on a MassARRAY Analyzer 4 mass spectrometer (Sequenom, United States). The procedure of specimen preparation and analysis included the following consecutive stages:

(1) Multiplex PCR;

(2) SAP reaction;

(3) iPLEX reaction;

(4) ionization and spectral analysis.

Following iPLEX reaction, specimens were desalted, loaded on SpectroCHIPs, and analyzed by MALDI-TOF spectrometry. The principal stages of genotyping using a MassARRAY Analyzer 4 are shown in Fig. 1.

DNA specimens used in the study were isolated from peripheral blood lymphocytes by conventional methods. Altogether, 192 blood specimens were obtained from individuals representing four different populations: Russians from Tomsk ($N = 48$), Komi from Syktyvkar, the Republic of Komi ($N = 48$), Khanty from the settlement of Kazym, Khanty–Mansi autonomous okrug of Tyumen oblast ($N = 48$), and Buryats from the settlement of Kurumkan, the republic of Buryatia ($N = 48$).

Genes and SNPs. The study involved 58 SNPs of genes, the products of which participate in the regulation of immune response. Data from the genome-wide association studies (GWAS) database of the United States National Human Genome Research Institute (NHGRI) [26] were used to select markers associated with autoimmune diseases (e.g., systemic lupus erythematosus, multiple sclerosis, rheumatoid arthritis) or with cytokine or immunoglobulin E production. Markers were selected that take into account the following parameters: highly significant phenotype associations according to GWASs ($p \le (5 \times 10^{-8})$), phenotype association with a gene or marker reproduced in several GWASs, marker localization in the coding or regulatory gene regions, and functional significance of SNPs confirmed with published data.

Primer design. The selected SNPs were combined in two multiplexes, i.e., one of 34 polymorphisms and another of 24 polymorphisms (Table 1). Primers were designed using the Sequenom Assay Design software (Sequenom) available online at www.sequenom.com. For each SNP, two PCR primers (forward and reverse) and one iPLEX (extension) primer were designed. For successful marker and allele identification by mass spectrometry of multiplex PCR products, the masses of extension primers should differ by at least 30 Da among different SNPs and by 5 Da between alternative alleles of the same marker. Primers were synthesized by Metabion (Germany). Primer sequences are available from the authors on request.

Multiplex PCR was used to accumulate amplicons containing SNPs to be analyzed. It was performed in 96-well plates in a Thermo Scientific (Germany) and an Applied Biosystems (United States) thermal cyclers; the reaction mixture (5 μL) contained 10 ng of genomic DNA template and the appropriate multiplex primer mixture. The reaction protocol included initial denaturation at 94°C (5 min) and 42 amplification cycles consisting of 20 s denaturation at 94°C, 40 s annealing at 56°C, and 60 s elongation at 72°C; finally, specimens were incubated at 72°C for additional 5 min.

SAP reaction was used to dephosphorylate unincorporated dNTPs with alkaline phosphatase (SAP). The procedure is required to clear amplification products from excessive dNTPs. It was performed in 96-well plates after PCR amplification. SAP mixture (Sequenom) comprising SAP buffer and 0.5 units of shrimp alkaline phosphatase per specimen was added

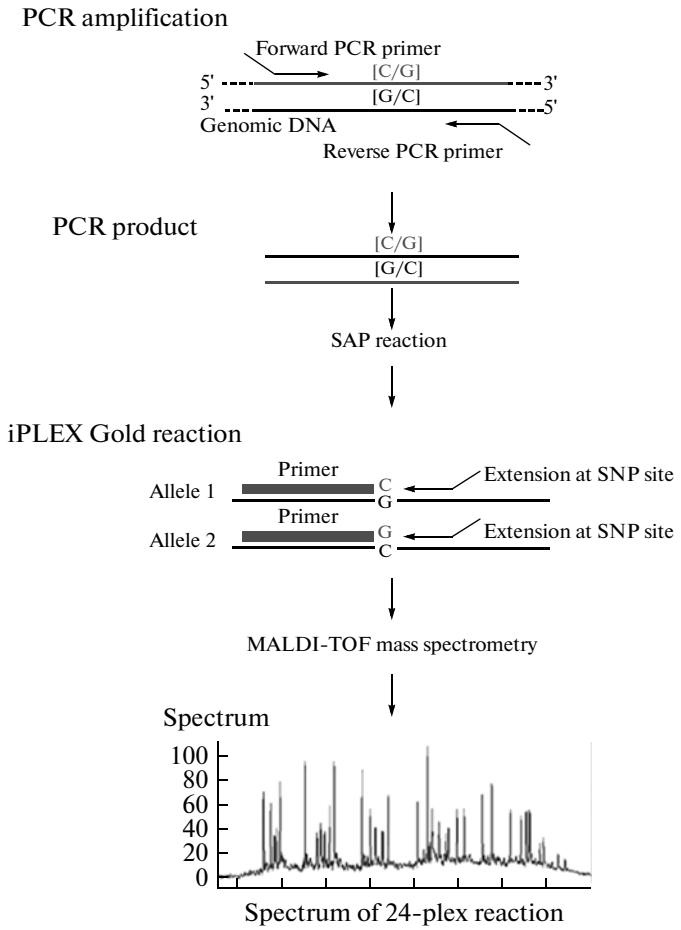

Fig. 1. Steps of specimen preparation and analysis in genotyping by MALDI-TOF mass spectrometry.

to reaction products obtained at the first step. Next, PCR plates were centrifuged, incubated at 37°C for 40–50 min, and heated to 85°C for 5–20 min to inactivate SAP.

iPLEX reaction. This step involved multiplex extension of iPLEX primers by incorporating modified dideoxynucleotide triphosphates (ddNTP) with altered molecular masses complementary to the nucleotides found in respective polymorphic positions. As a result, a mixture of short PCR products representing alleles of polymorphic markers included in the multiplex is obtained. Following SAP reaction, specimens were supplemented with a solution containing tenfold iPLEX buffer (Sequenom), DNA polymerase (iPLEX enzyme, Sequenom), four ddNTPs with altered molecular masses, and the extension primer mixture. Importantly, since the MALDI-TOF signal intensity depends on the mass of the analyte, concentrations of different extension primers in the reaction mixture must be appropriately adjusted, i.e., concentrations of low-weight primers must be lower than those of higher-weight primers. The amplification protocol included initial denaturation at 94°C for 30 s, next 40 cycles of 5 s denaturation at 94°C, 5 s annealing at 52°C, and 5 s elongation at 80°C; finally, specimens were incubated at 72°C for 5 min. The reaction was performed in the same plate as the previous steps of the experiment.

PCR product purification (desalting). To reduce background noise during mass spectrometry, the reaction mixture must be purified from salts that accumulated in multiplex PCR and iPLEX reaction. To remove Na^+, K^+, and Mg^{2+} ions, reaction products were treated with SpectroCLEAN resin (Sequenom). After adding 15 mg of resin to each well of the 96-well plate, the plate was covered with film, rotated for 15 min, and then centrifuged at 3200 g for 5 min.

Specimen transfer onto a spectroCHIP. Prior to ionization and mass spectrometry, specimens must be transferred onto a spectroCHIP. A Sequenom spectroCHIP is a 96- or 384-well carrier with a MALDI ionization matrix. Specimens were automatically transferred onto the chip from PCR plates on a Nanodispenser RS1000 workstation (Sequenom). Each spectroCHIP well was loaded with 8–12 nL of specimen.

Specimen ionization and analysis of mass spectra. The spectroCHIP was placed into the working chamber of the mass spectrometer, where specimen ioniza-

Table 1. Multiplexes used in PCR and mass spectrometry

Multiplex 1			Multiplex 2		
No.	SNP (rs)	gene	No.	SNP (rs)	gene
1	rs1042713	*ADRB2*	1	rs1024161	*CTLA4*
2	rs11150610	*ITGAM*	2	rs11865121	*CLEC16A*
3	rs1128334	*ETS1*	3	rs1800875	*CMA1*
4	rs12722489	*IL2RA*	4	rs20541	*IL13*
5	rs1295685	*IL13*	5	rs2227306	*IL8*
6	rs1327474	*IFNgR1*	6	rs2230926	*TNFAIP3*
7	rs13277113	*BLK*	7	rs2234711	*IFNgR1*
8	rs1335532	*CD58*	8	rs2240335	*PADI4*
9	rs144651842	*IL4RA*	9	rs2300747	*CD58*
10	rs1800693	*TNFRSF1A*	10	rs2305480	*GSDMB*
11	rs1800896	*IL10*	11	rs231735	*CTLA4*
12	rs1801275	*IL4RA*	12	rs2381416	*IL33*
13	rs1805015	*IL4RA*	13	rs2476601	*PTPN22*
14	rs1837253	*TSLP*	14	rs2569190	*CD14*
15	rs2056626	*CD247*	15	rs3087243	*CTLA4*
16	rs2070874	*IL4*	16	rs324015	*STAT6*
17	rs2104286	*IL2RA*	17	rs3890745	*STAT4*
18	rs2227284	*IL4*	18	rs485499	*IL12A*
19	rs2430561	*IFNg*	19	rs5744455	*CD14*
20	rs2546890	*LOC285626*	20	rs6441286	*IL12A*
21	rs2618476	*BLK*	21	rs7097397	*WDFY4*
22	rs2736340	*BLK*	22	rs907092	*IKZF3*
23	rs3761847	*TRAF1*	23	rs9303277	*IKZF3*
24	rs3790567	*IL12RB2*	24	rs9888739	*ITGAM*
25	rs3821236	*STAT4*			
26	rs4810485	*CD40*			
27	rs4986790	*TLR4*			
28	rs5743708	*TLR2*			
29	rs6604026	*RPL5*			
30	rs6679677	*PTPN22*			
31	rs6897932	*IL7R*			
32	rs6920220	*TNFAIP3*			
33	rs743777	*IL2RB*			
34	rs7574865	*STAT4*			

tion and the mass spectrometry of ionized particles were performed consequently for each well of the chip. The whole procedure of specimen ionization and obtaining the mass spectra required approximately 20 min for a 96-well chip. The real-time analysis of mass spectra was performed with MassARRAY TYPER 4.0 software (Sequenom). The same program package was used for the subsequent analysis and recording of the results.

Statistical analysis of the data involved conventional techniques of population genetics and mathematical statistics. The agreement of the genotype distribution with the Hardy–Weinberg equilibrium was verified using the exact Guo and Thompson's test, and

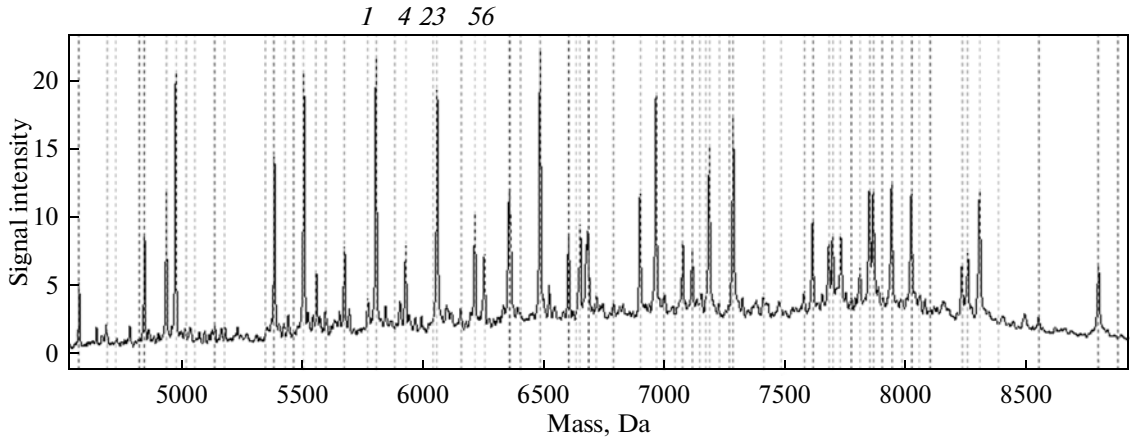

Fig. 2. Mass spectrum obtained for a specimen in the 24-plex reaction (multiplex 2). A homozygous genotype (rs20541), peak positions: (*1*) Native extension primer, no ddNTPs incorporated; (*2*) T allele oligonucleotide (no peak); (*3*) C allele oligonucleotide (peak present). Heterozygous genotype (rs6441286), peak positions: (*4*) Native extension primer, no ddNTPs incorporated; (*5*) G allele oligonucleotide (peak present); (*6*) T allele oligonucleotide (peak present).

interpopulation differences were evaluated by the coefficient of genetic differentiation (Gst). Multidimensional allele frequency analysis was performed using the principle component approach with Statistica software (StatSoft, United States).

RESULTS AND DISCUSSION

Analysis of Mass Spectra and Genotyping

Figure 2 represents an example of the mass spectrum obtained for one of specimens using multiplex 2 (for 24 SNPs). The masses of molecules in the analysis ranged from 5000 to 9000 Da. In this mass range, 72 different products can be detected, including three products that correspond to each SNP, i.e., the initial extension primer, and two iPLEX products that consist of the extension primer with an attached nucleotide that corresponds to either alternative allele. Vertical dashed lines indicate the positions of the expected mass spectrum peaks for 72 products. In the given specimen, all 24 markers were successfully amplified and extended; accordingly, 24 iPLEX primer peaks and 29 peaks of specific iPLEX products (17 homozygous and 6 heterozygous markers) can be seen. It should be noted that the intensities of native primer peaks did not exceed 50 units and were always lower than the peaks of iPLEX products for the corresponding SNP, which indicates that the iPLEX reaction was highly efficient and that most primer molecules were incorporated in specific iPLEX products.

Figure 3 represents a fragment of the mass spectrum obtained in the 24-plex reaction for three specimens with different rs20541 genotypes. The masses of allele-specific iPLEX products differ by 16 Da, so alleles and genotypes can be easily identified even by manual analysis of the mass spectrum.

In the MassARRAY TYPER 4.0 program package, automated genotype reading from the multiplex mass spectrum is based on computing the ratio between the peak heights of iPLEX products that represent two alternative alleles. Figure 4 shows an example of specimen clustering by rs20541 genotypes in the 24-plex reaction. Although the peak heights differed considerably among specimens, all 95 specimens were successfully genotyped. The only specimen that both produced no peaks (alleles not detected) and was located at the origin of coordinates was the control specimen without DNA.

Genotyping Efficiency

Our work included four experiments that analyzed the method efficiency and reproducibility by genotyping the two suggested multiplexes (34-plex and 24-plex) in two plates. One plate carried DNA specimens obtained from Komi and Russians, the other one carried those from Buryats and Khanty. Each of the 96-well plates contained 95 DNA specimens from respective populations and a control specimen without DNA. Some specimens were replicated to verify the genotyping reproducibility.

In the 24-plex reaction, the overall genotyping efficiency was very high (Table 2). For all SNPs, genotypes were determined in nearly all specimens. The total call rate (the proportion of determined genotypes among all possible genotypes) was 98.64%.

The efficiency of the 34-plex was slightly lower. Two of the 34 SNPs included in this multiplex could not be genotyped. That is, rs4810485 genotypes could not be determined in any specimen; furthermore, for rs3761847, the call rate was 42.1%, so these two markers were excluded from the further analysis. Without taking into account these two SNPs, the 32-plex genotyping efficiency was 97.87%. For most SNPs of both

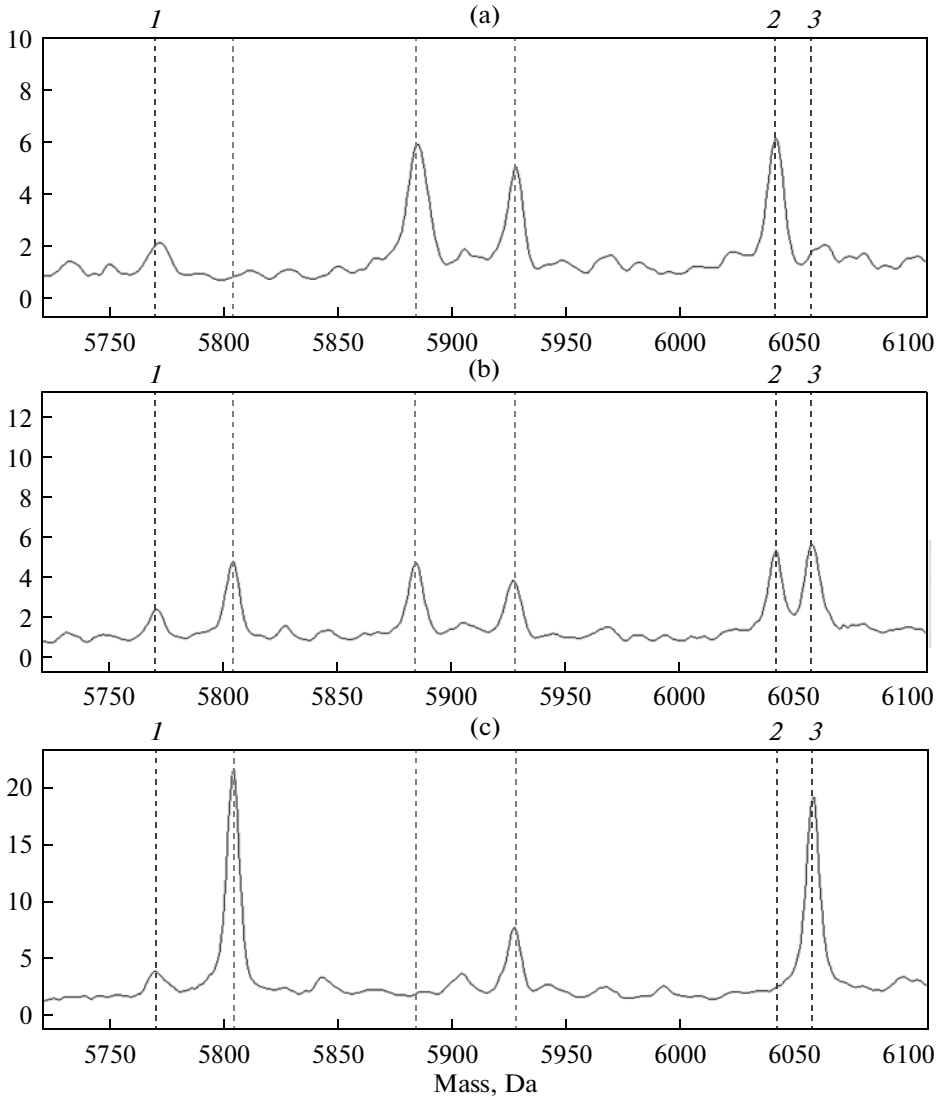

Fig. 3. Fragment of mass spectra of three specimens with different rs20541 genotypes: TT (a), TC (b), CC (c). Peak positions: (*1*) Native extension primer, no ddNTPs incorporated (5771 Da); (*2*) T allele oligonucleotide (6042 Da); (*3*) C allele oligonucleotide (6058 Da).

multiplexes (52 of 56), call rates ranged from 98 to 100%. The lowest genotyping efficiency was observed for rs6441286, rs22403350 (96.84% each), rs6604026 (93.15%), and rs6920220 (87.89%). The overall call rate for 56 SNPs was 98.2% (10449 genotypes of 10640 were determined, Table 2).

Genetic Analysis of Allele Frequencies in Population

The data on the allele frequencies of the 56 SNPs studied, genetic diversity, and genetic differentiation among populations are given in Table 3. (Individual genotype distributions for each marker are available from the authors on request.) Among the 224 genotype distributions obtained for the total data array (56 markers in four populations), 12 distributions did not agree with the Hardy–Weinberg equilibrium, which did not exceed the expected number of random deviations. Individual markers or populations did not deviate from the Hardy–Weinberg equilibrium.

In all four populations studied, the allele frequency spectra were similar. The only exceptions were rs6604026 in *RPL5* and rs1128334 in *ETS1*; for these SNPs, the allele rare in Russians and Komi was frequently found in Buryats and Khanty. The populations also had similar general levels of genetic diversity by the markers studied. The mean expected heterozygosity ranged from 0.3522 in Buryats to 0.3698 in Komi.

For most individual loci (30 of 56 markers), the coefficients of genetic differentiation were lower than 2% (Table 3). Interpopulation differentiation exceeded 10% for only three SNPs, including the above-mentioned rs6604026 and rs1128334, which were charac-

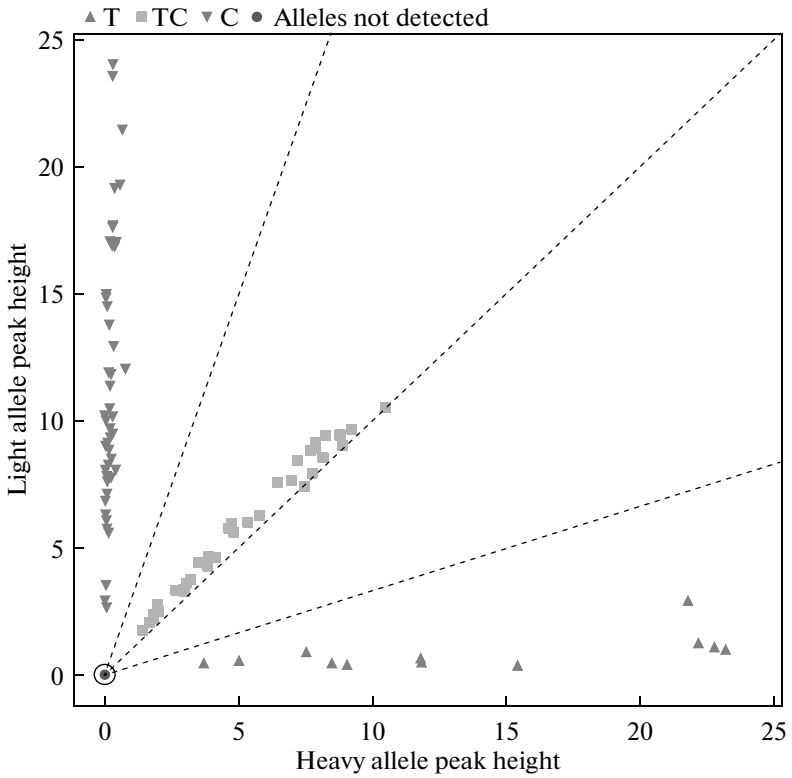

Fig. 4. Specimen clusters by rs20541 genotypes determined with MassARRAY TYPER 4.0.

terized with exceptionally high differentiation coefficients (Gst = 34.9%). The mean level of genetic differentiation by 56 markers was 4.2%, which is in agreement with the levels of interpopulation genetic differentiation determined using different autosomal marker systems, including genome-scale SNP sets [7, 14].

A multidimensional analysis of genetic diversity based on 56 markers across the four populations studied was performed using the principal component (PC) approach. Taken together, the first two PCs accounted for 87.5% of allele frequency variation in the populations. The positions of the four populations in the PC1–PC2 space reflected their geographical location (Fig. 5), as expected from the data obtained in previous studies [27, 28]. Populations from the European part of Russia were clustered in the right-hand part of the chart, while populations from Siberia were clustered on the left. Thus, PC1 probably represents the wedge-type variation of allele frequencies with longitude, which was previously described for autosomal markers, Y chromosome lineages, and mtDNA [7, 14, 27, 28]. PC2 apparently corresponds to the North–South axis.

Evaluation of Time- and Cost-Efficiency of SNP Genotyping by MALDI-TOF Mass Spectrometry in Comparison to Real-Time PCR

Real-time PCR is currently the most widely used technique of small-scale SNP genotyping. Mass spec-

Table 2. Proportion of detected genotypes (call rate) in multiplex reactions

Experiment	Multiplex 1 (34-plex)*			Multiplex 2 (24-plex)			Overall call rate (%)
	total number of genotypes	genotype not determined	call rate (%)	total number of genotypes	genotype not determined	call rate (%)	
Plate 1	3040	113	96.29	2280	26	98.85	97.39
Plate 2	3040	16	99.47	2280	36	98.42	99.02
Overall call rate (%)	6080	129	97.87	4560	62	98.64	98.20

* rs4810485 and rs3761847 not accounted for (32-plex).

Table 3. Allele frequencies, genetic diversity, and genetic differentiation in the populations studied

SNP	Allele	Allele frequencies				He*	Gst**
		Russians	Komi	Buryats	Khanty		
rs1024161	T	0.50	0.58	0.70	0.72	0.4536	0.0337
rs1042713	G	0.66	0.73	0.52	0.62	0.4532	0.0244
rs11150610	C	0.61	0.63	0.41	0.49	0.4818	0.0320
rs1128334	G	0.86	0.81	0.28	0.20	0.3183	0.3598
rs11865121	A	0.32	0.30	0.22	0.24	0.3894	0.0085
rs12722489	G	0.84	0.89	0.89	0.94	0.1939	0.0148
rs1295685	C	0.67	0.68	0.69	0.77	0.4148	0.0066
rs1327474	A	0.50	0.63	0.84	0.84	0.3751	0.1029
rs13277113	G	0.79	0.73	0.46	0.52	0.4323	0.0787
rs1335532	T	0.78	0.73	0.42	0.48	0.4328	0.0983
rs144651842	G	0.97	0.92	0.81	0.77	0.2181	0.0573
rs1800693	A	0.86	1.00	0.97	0.97	0.0865	0.0592
rs1800875	G	0.48	0.49	0.50	0.53	0.4991	0.0018
rs1800896	A	0.51	0.53	0.80	0.73	0.4271	0.0680
rs1801275	G	0.17	0.14	0.16	0.16	0.2635	0.0013
rs1805015	T	0.87	0.89	0.91	0.97	0.1650	0.0166
rs1837253	T	0.24	0.27	0.46	0.45	0.4397	0.0433
rs20541	C	0.67	0.69	0.68	0.75	0.4197	0.0044
rs2056626	G	0.37	0.39	0.15	0.27	0.3963	0.0427
rs2070874	T	0.32	0.22	0.50	0.29	0.4205	0.0495
rs2104286	A	0.79	0.81	0.87	0.92	0.2545	0.0222
rs2227284	A	0.38	0.28	0.54	0.38	0.4609	0.0355
rs2227306	C	0.58	0.51	0.61	0.61	0.4841	0.0070
rs2230926	T	0.97	0.94	0.98	0.99	0.0565	0.0096
rs2234711	C	0.38	0.49	0.46	0.47	0.4917	0.0064
rs2240335	G	0.64	0.66	0.52	0.49	0.4785	0.0211
rs2300747	A	0.81	0.73	0.42	0.47	0.4221	0.1157
rs2305480	C	0.56	0.50	0.52	0.53	0.4973	0.0020
rs231735	G	0.35	0.24	0.23	0.18	0.3686	0.0193
rs2381416	A	0.86	0.81	0.94	0.91	0.2074	0.0229
rs2430561	T	0.62	0.57	0.73	0.76	0.4311	0.0266
rs2476601	G	0.93	0.89	0.96	0.98	0.1129	0.0197
rs2546890	A	0.53	0.53	0.46	0.50	0.4981	0.0036
rs2569190	G	0.55	0.49	0.43	0.38	0.4884	0.0173
rs2618476	T	0.79	0.73	0.42	0.52	0.4279	0.0968
rs2736340	C	0.78	0.73	0.40	0.50	0.4312	0.1011
rs3087243	G	0.69	0.76	0.77	0.81	0.3635	0.0091
rs324015	A	0.27	0.40	0.40	0.43	0.4610	0.0164
rs3790567	A	0.28	0.18	0.24	0.23	0.3561	0.0064
rs3821236	G	0.78	0.77	0.64	0.72	0.3888	0.0160
rs3890745	G	0.42	0.38	0.46	0.40	0.4830	0.0041
rs485499	T	0.56	0.56	0.88	0.74	0.3959	0.0844
rs4986790	A	0.94	0.89	0.97	0.99	0.0991	0.0295
rs5743708	G	0.95	0.97	0.96	0.98	0.0711	0.0031
rs5744455	C	0.78	0.74	0.76	0.69	0.3799	0.0054
rs6441286	T	0.62	0.65	0.53	0.56	0.4789	0.0095
rs6604026	T	0.30	0.34	0.92	0.91	0.2948	0.3754
rs6679677	C	0.94	0.88	0.96	0.98	0.1135	0.0259
rs6897932	C	0.74	0.67	0.77	0.76	0.3883	0.0078
rs6920220	G	0.87	0.89	0.91	0.90	0.1894	0.0027
rs7097397	G	0.66	0.64	0.53	0.50	0.4765	0.0198
rs743777	G	0.34	0.20	0.06	0.11	0.2709	0.0749
rs7574865	T	0.22	0.28	0.33	0.27	0.3944	0.0079
rs907092	G	0.49	0.56	0.48	0.44	0.4965	0.0067
rs9303277	T	0.53	0.45	0.58	0.57	0.4930	0.0097
rs9888739	T	0.08	0.08	0.07	0.09	0.1473	0.0007
Mean He		0.3690	0.3698	0.3522	0.3471	0.3595	0.0419

* He, expected heterozygosity. ** Gst, genetic differentiation coefficient.

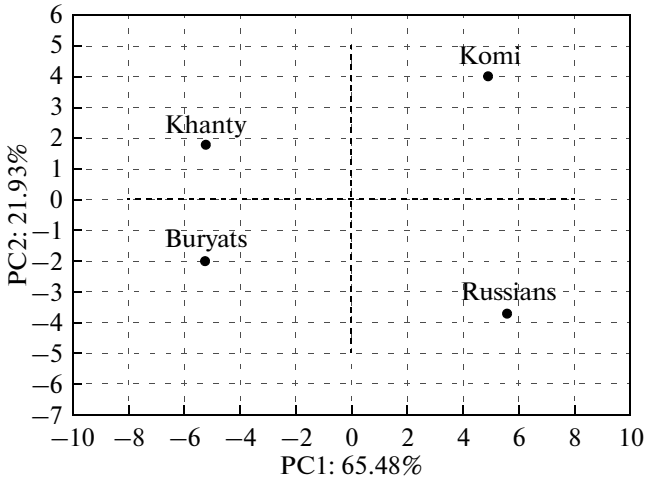

Fig. 5. Space of the two first principal components that describe the allele frequency variation by 56 SNPs in studied populations.

trometry-based genotyping can be used for the purposes of genotyping on both small scales (individual markers in hundreds or thousands of specimens) and medium scales (dozens and hundreds of markers in hundreds or thousands of specimen). Let us consider a hypothetical problem of genotyping 30 markers in 960 specimens by MALDI-TOF mass spectrometry and by real-time PCR using the 96-well format in both approaches. In PCR with fluorescent dye-labeled TaqMan probes, each marker must be genotyped individually. For MALDI-TOF mass spectrometry, 30 markers can be combined in a single multiplex. Costs of expendable materials and reagents for both techniques were calculated based on the official prices of Russian distributors as of April 2013 (Table 4). We also accounted for the possibility of reducing the volume of the real-time PCR mixture by three to four times compared to the original protocols of TaqMan manufacturers. Our calculations showed that the cost of genotyping one marker in one specimen by MALDI-TOF mass spectrometry is 3.5 times lower than by real-time PCR. It should also be noted that a 30-marker multiplex is by far not the largest possible; the resolution of the mass spectrometer used in our study actually allows the use of multiplexes comprising up to 40 SNPs. The higher throughput of mass spectrometry also seems a considerable advantage, i.e., the number of genotypes that can be obtained per working day on one workstation is 15 times higher than that determined by real-time PCR in the same conditions (Table 4). However, the advantages of mass spectrometry in genotyping costs and throughput are to a certain extent counterbalanced by the fact that by no means all markers can be included in multiplexes, as well as by the high cost of the MALDI-TOF equipment. The difference in the costs of equipment for PCR and mass spectrometry will only pay off for relatively large amounts of genotyping: with the calculated net costs of a single marker, it will be overcome by genotyping approximately 400000 genotypes. Apparently, mass spectrometry can be considered as the optimal genotyping technique for large laboratories or collective use centers, whereas real-time PCR and other techniques of small-scale genotyping will remain the methods of choice for smaller research units; alternatively, genotyping can be outsourced to service centers.

Table 4. Time- and cost efficiency of SNP genotyping using the MALDI-TOF technique on a Sequenom MassARRAY 4 mass spectrometer in comparison to real-time PCR

Parameter	MALDI-TOF	Real-time PCR
Reagents and expendable materials for genotyping 30 markers in 960 specimens (RUB)*	347908	1228027
Which includes (RUB):		
—Reagent set	263 093	608304
—Oligonucleotide primers	82758	–
—TaqMan probes	–	558010
—Plasticware (plates and tips)	2057	61713
The cost of genotyping one marker in one specimen (RUB)**	12.08	42.64
Number of genotypes obtained by one operator per working day***	5760	384

* For the 96-well format; calculated based on the prices of official Russian distributors as of April 2013.

** Only reagents and expendable materials; the cost and amortization of the equipment, the operator's labor, overhead, and other costs not taken into account.

*** In the 96-well format: MALDI-TOF, two plates for 96 specimens each, 30 markers per specimen; real-time PCR, four plates for 96 specimens each, one marker per specimen.

In this work, we described an example of applying multiplex SNP genotyping by tandem mass spectrometry for an analysis of genetic diversity of immune response genes in four human populations of Russia. It was shown that MALDI-TOF mass spectrometry is a rapid, accurate, and efficient technique appropriate for medium-scale SNP genotyping. The allele frequencies of 56 SNPs in 41 genes implicated in immune response regulation were largely similar among the four populations studied. Populations of Russians, Komi, Khanty, and Buryats had identical genetic diversity levels and were clustered according to their geographical location. The cost-efficiency of genotyping using MALDI-TOF mass spectrometry was evaluated compared to real-time PCR.

ACKNOWLEDGMENTS

This work was partially supported by the Federal Target Programs "Research and Development in Priority Areas of Science and Technology" (state contract no. 11.519.11.2036), and "Human Resources in Science and Teaching in Innovative Russia" (agreement no. 8042), and the Russian Foundation for Basic Research (project no. 12-04-00595a).

REFERENCES

1. http://www.ncbi.nlm.nih.gov/projects/SNP/snp_summary.cgi
2. The 1000 Genomes Project Consortium. 2012. An integrated map of genetic variation from 1092 human genomes. *Nature*. **491**, 56–65.
3. The International HapMap 3 Consortium. 2010. Integrating common and rare genetic variation in diverse human populations. *Nature*. **467**, 52–58.
4. Levy S., Sutton G., Ng P.C., et al. 2007. The diploid genome sequence of an individual human. *PLoS Biol*. **5**, 2113–2144.
5. Schuster S.C., Miller W., Ratan A., et al. 2007. Complete Khoisan and Bantu genomes from southern Africa. *Nature*. **463**, 963–947.
6. Skryabin G.K., Prokhortchouk E.B., Mazur A.M., et al. 2009. Combining two technologies for full genome sequencing of human. *Acta Naturae*. **1**, 102–107.
7. Stepanov V.A. 2010. Genomes, populations, and diseases: Ethnic genomics and personalized medicine. *Acta Naturae*. **2**, 16–31.
8. Kim S., Misra A. 2007. SNP genotyping: Technologies and biomedical applications. *Annu. Rev. Biomed. Eng*. **9**, 289–320.
9. Steemers F.J., Gunderson K.L. 2007. Whole genome genotyping technologies on the BeadArray platform. *Biotechnol. J*. **2**, 41–49.
10. Kwok P.Y. 2001. Methods for genotyping single nucleotide polymorphisms. *Annu. Rev. Genomics Hum. Genet*. **2**, 235–258.
11. Kumar S., Banks T.W., Cloutier S. 2012. SNP discovery through next-generation sequencing and its applications. *Int. J. Plant Genomics*. doi10.1155/2012/831460
12. Shen G.Q., Abdullah K.G., Wang Q.K. 2009. The TaqMan method for SNP genotyping. In: *Single Nucleotide Polymorphisms: Methods and Protocols*. Methods in Molecular Biology, vol. 578. Komar A.A. Ed. Totowa, NJ: Humana Press, pp. 293–306.
13. Karunas A.S., Yunusbaev B.B., Fedorova Yu.Yu., Gimalova G.F., Ramazanova N.N., Gur'eva L.L., Mukhtarova L.N., Zagidullin Sh.Z., Etkina E.I., Khusnutdinova E.K. 2011. Genome-wide association study of bronchial asthma in the Volga–Urals region of Russia. *Mol. Biol*. (Moscow). **45**, 911–920.
14. Stepanov V.A., Prokhorchuk E.B., Khusnutdinova E.K., Mazur A.M., Tesluk A., Chekanov N.N., Khrameeva E., Kharkov V.N., Gubina M.A., Kutuev I.A., Khusainova R.I., Marusin A.V., Spiridonova M.G., Akhmetova V.L., Khidiyatova I.M., Litvinov S.S., Bulygina E., Puzyrev V.P., Skryabin K.G. 2010. Eurasian genetic landscape: Gene geography in the mirror of genome-wide variability. *9th Asian Pacific Conference on Human Genetics, Hong Kong, 2010: Program and Abstracts*, pp. 118–119.
15. Tebbutt S.J., He J.Q., Burkett K.M. et al. 2004. Microarray genotyping resource to determine population stratification in genetic association studies of complex disease. *Biotechniques*. **37**, 977–985.
16. Fedorova Yu.Yu., Gra O.A., Karunas A.S., Khuzina A.Kh., Ramazanova N.N., Yuldasheva A.A., Biktasheva A.R., Etkina E.I., Nasedkina T.V., Goldenkova–Pavlova I.V., Khusnutdinova E.K.. 2009. Association of polymorphisms of xenobiotic metabolism genes with childhood atopic diseases in Russian patients from Bashkortostan. *Mol. Biol*. (Moscow). **43**, 961–967.
17. Gryadunov D.A., Zimenkov D.V., Mikhailovich V.M., et al. 2009. Hydrogel biochiptechnology and its applications in medical laboratory diagnosis, *Med. Alfavit*. **3**, 10–14.
18. Gra O.A., Kozhekbaeva Zh.M., Gra D.V., et al. 2009. Biochip technologies in the diagnosis of socially significant multifactorial diseases. *Ovchinnikov Bull. Biotechnol. Phys. Chem. Biol*. (Moscow). **5**, 5–25.
19. Pavlov A.E., Apalko S.V., Vorobyev E.V. 2012. Molecular genetic diagnosis of cystic fibrosis in a microchip format. *Laboratoriya*. **4**, 16–19.
20. Pati N., Schowinsky V., Kokanovic A., et al. 2004. A comparison between SNaPshot, pyrosequencing, and biplex invader SNP genotyping methods: Accuracy, cost, and throughput. *Biomed. Biophys. Methods*. **60**, 1–12.
21. Esteves L.M., Bulhoes S.M., Brilhante M.J., Mota-Vieira L. 2013. The multiplex snapshot assays for SNP genotyping in candidate innate immune genes. *BMC Res. Notes*. **6**, 54. doi 10.1186/1756-0500-6-54
22. Cole R.B. 2012. *Electrospray and MALDI Mass Spectrometry: Fundamentals, Instrumentation, Practicalities,*

and Biological Applications, 2nd ed. Hoboken, NJ: Wiley.

23. Ratnayake M., Reynard L.N., Raine E., et al. 2012. Allelic expression analysis of the osteoarthritis susceptibility locus maps to MICAL3. *BMC Med. Genet.* **13**, 12. doi 10.1186/1471-2350-13-12
24. Bradic M., Costa J, Chelo I.V. 2011. Genotyping with Sequenom. In: *Molecular Methods for Evolutionary Genetics*. Methods in Molecular Biology, vol. 772, pp. 193–210.
25. Gabriel S., Ziaugra L., Tabbaa D. 2009. SNP genotyping using the Sequenom MassARRAY iPLEX platform. *Curr. Protoc. Hum. Genet.* **60**, 2.12.1–2.12.18.
26. Hindorff L.A., Sethupathy P., Junkins H.A., et al. 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 9362–9367.
27. Stepanov V.A. 2002. *Etnogenomika naseleniya Sibiri i Srednei Azii* (Ethnogenomics of the Population of Siberia and Middle Asia). Tomsk: Pechatnaya Manufaktura.
28. Stepanov V.A., Khar'kov V.N., Puzyrev V.P. 2006. The evolution and phylogeography of human Y-chromosome lineages. *Inf. Vestn. Vavilovskogo O-va. Genet. Sel.* **10** (1), 57–73.

Translated by D. Timchenko