GEOGRAPHIC DISTRIBUTION OF THE *LZTFL1* SNP MARKERS ASSOCIATED WITH SEVERE COVID-19 IN RUSSIA AND WORLDWIDE

Balanovska EV[1,2] ✉, Gorin IO[1], Petrushenko VS[1], Chernevsky DK[1], Koshel SM[1,3], Temirbulatov II[1,4], Pylev VYu[1,2], Agdzhoyan AT[1]

[1] Research Centre for Medical Genetics, Moscow, Russia
[2] Autonomous non-profit organization "Biobank of North Eurasia", Moscow, Russia
[3] Lomonosov Moscow State University, Moscow, Russia
[4] Russian Medical Academy of Continuous Professional Education, Ministry of Healthcare of the Russian Federation, Moscow, Russia

The correlation between the risk of death from COVID-19 and the patient's ethnogeographic origin has been previously detected. *LZTFL1* gene was identified as a potential marker of a two times higher risk of severe COVID-19. The study was aimed to assess spatial variation in the *LZTFL1* SNP markers in indigenous populations of Russia and the world. Spatial variation in the *LZTFL1* polymorphic markers was analyzed in 28 metapopulations (97 ethnic groups) of North Eurasia ($n = 1980$) and 34 world's metapopulations ($n = 3637$) by bioinformatics, statistical and cartographic methods. In North Eurasia, the major geographic variation vectors, North–South and West–East, are generally in line with the Caucasoid–Mongoloid anthropological vector. Global variation also corresponds to anthropological features: each cluster of indigenous populations includes only those from the place where it originates: Africa, Asia, or America. Indo-European cluster integrates Caucasoid populations of Europe and Asia. All four clusters of the world's indigenous population are separated from each other. The huge genetic diversity of Russia peoples and neighboring countries forms a bridge between three continents: Europe, Asia and America. Cartographic atlas for spatial variation in 11 *LZTFL1* markers in the populations has been created. The following major patterns have been revealed: a) the world's extrema fall on the indigenous populations of Africa and America; 2) Eurasia constitutes a transition zone between these two extrema, but has its own patterns and shows enormous scale of variation shows enormous variation on a global scale; 3) the genetic landscape of Russia tends to be seamlessly integrated into the Eurasian landscape.

Keywords: COVID-19, LZTFL1, SNP, gene geography, populations, indigenous people, Russia, world

ГЕНОГЕОГРАФИЯ В РОССИИ И МИРЕ SNP-МАРКЕРОВ ГЕНА *LZTFL1*, АССОЦИИРОВАННЫХ С ТЯЖЕЛЫМ ТЕЧЕНИЕМ COVID-19

Е. В. Балановская[1,2] ✉, И. О. Горин[1], В. С. Петрушенко[1], Д. К. Черневский[1], С. М. Кошель[1,3], И. И. Темирбулатов[1,4], В. Ю. Пылёв[1,2], А. Т. Агджоян[1]

[1] Медико-генетический научный центр имени Н. П. Бочкова, Москва, Россия
[2] Автономная некоммерческая организация «Биобанк Северной Евразии», Москва, Россия
[3] Московский государственный университет имени М. В. Ломоносова, Москва, Россия
[4] Российская медицинская академия непрерывного профессионального образования Минздрава России, Москва, Россия

Ранее была обнаружена корреляция между риском смерти от COVID-19 и этногеографическим происхождением пациента. Ген *LZTFL1* отмечается как потенциальный маркер, ассоциированный с двухкратным увеличением риска тяжелого течения COVID-19. Целью исследования было изучить пространственную изменчивость SNP-маркеров гена *LZTFL1* в коренном населении России и мира. Биоинформатическими, статистическими и картографическими методами был проведен анализ пространственной изменчивости полиморфных маркеров гена *LZTFL1* в 28 метапопуляциях (97 этносов) Северной Евразии ($n = 1980$) и 34 метапопуляциях мира ($n = 3637$). В Северной Евразии основные географические векторы изменчивости «север–юг» и «запад–восток» в целом согласуются с антропологическим вектором «европеоидность–монголоидность». Глобальная изменчивость тоже соответствует антропологии: каждый кластер коренного населения включает популяции только «своей» части света — Африки, Азии или Америки. «Индоевропейский» кластер объединяет европеоидные популяции Европы и Азии. Все четыре кластера коренного населения мира отдалены друг от друга, и только огромное генетическое разнообразие народов России и сопредельных стран является мостом, связующим три части света: Европу, Азию и Америку. Создан картографический атлас пространственной изменчивости 11 SNP-маркеров *LZTFL1* в популяциях. Выявлены основные закономерности: а) мировые экстремумы приходятся на коренное население Африки и Америки; 2) Северная Евразия является переходной зоной между мировыми экстремумами, но обладает собственными закономерностями и огромным размахом изменчивости в мировом масштабе; 3) генетический ландшафт России, как правило, органично вписан в Евразийский ландшафт.

Ключевые слова: COVID-19, LZTFL1, SNP, геногеография, популяции, коренное население, Россия, мир

The COVID-19 pandemic, with its high mortality and severe complications, forced the world's scientific community to engage in the search for DNA markers associated with the SARS-COV-2 infection. COVID-19 severity varies among representatives of various world populations, that is why the COVID-19 Host Genetics Initiative international project has started gathering information about the frequency of genome variants associated with severe COVID-19 [1]. Among these gene *LZTFL1* [2] is specified as a potential marker of a two times higher risk of severe COVID-19 [3].

LZTFL1 is expressed in human lungs and encodes a protein involved in transport of other proteins to the primary cilia of the ciliated epithelial cells [4]. The *LZTFL1* gene clinical significance has been discvered earlier: the gene is associated with Bardet-Biedl syndrome-17 (OMIM 615994) [5], the autosomal recessive ciliopathy [6, 7]. Seven BBS and BBIP10 proteins form a stable complex referred to as BBSome. This complex ensures protein transport to the ciliary membrane [8, 9], while the reduced *LZTLF1* function can compensate for a lack of BBS proteins and restore ciliary motility [10].

Alterations related to severe COVID-19 and associated with *LZTFL1* were found in the patients' lung epithelial cells [3, 11]: alterations in the chromosome 3p21.31 region carrying gene *LZTFL1* resulted in the twofold increased risk of respiratory failure [12, 13] and more than twofold increased risk of mortality in people under the age of 60 [11]. The study of polymorphism of one of the *LZTFL1* gene variants in the UK population revealed association between the risk of death from COVID-19 and the patient's origin: ones from South Asia had a four times higher risk than patients of European descent. These differences partly explain higher mortality rates among the representatives of South Asian peoples living in the UK [3].

A significant association between *LZTFL1* and severe COVID-19, as well as therapeutic potential and ethnogeographic differences, calls for examining interpopulation variations among the world's population. The research team has an information base that includes both own and literature data on the world's peoples' genomes. The information base, that has already enabled the analysis of variation in SNP markers (*rs11385942*, *rs657152*) associated with severe COVID-19 among the world's population (more specifically in Russia) [14], makes it possible to perform similar study of 11 *LZTFL1* SNP markers.

The study was aimed to assess spatial variation in SNP markers of the LZTFL1 gene associated with severe COVID-19 [15] in the human population: 1) to perform the search for polymorphic *LZTFL1* SNP markers provided information about the SNP marker abundance in indigenous peoples of the world and North Eurasia (in more detail); 2) to provide two representative pools of population data on these SNP markers; 3) to perform multivariate statistical analysis of these data; 4) to create a cartographic atlas of the *LZTFL1* polymorphic SNP marker distribution among indigenous populations of North Eurasia and the world.

METHODS

Two original pools of DNA markers

The pool of data on the indigenous population of North Eurasia is represented by the populations of 97 ethnic groups, mostly of Russian origin, but also from the majority of post-Soviet states and Mongolia. DNA samples were provided by Biobank of North Eurasia. The sampling method was described earlier [16]: the samples comprised specimens obtained exclusively from unrelated individuals, whose grandparents belonged to the studied ethnic group. Specimens obtained from geographically and historically proximate populations, but from small samples,

were merged into metapopulations [17]. The resulting pool of data on the indigenous population of North Eurasia included 28 metapopulations (*n* = 1980) with the average sample size of 140 chromosomes.

The pool of data on the indigenous population of other world's regions (*n* = 1657) comprises data on the genome-wide panels by Illumina reported in scientific literature and accumulated in the GG-Base [18]. The geographically and historically proximate groups were merged into metapopulations in order to provide a representative sample. The resulting pool of data on the world's indigenous population (*n* = 3637) included 34 metapopulations (64 populations of the maps).

Selection of polymorphic *LZTFL1* markers

Bioinformatics analysis of both data pools revealed 10 *LZTFL1* SNPs characteristically represented in the indigenous population of North Eurasia and other regions of the world. Of these two SNP markers are specified as strongly associated with COVID-19: *rs1058961* (3′-untranslated region) and *rs12493471* (intron 2) [1]. The other eight SNP markers were studied for the first time: *rs11130077* (intron 3), *rs17078408*, *rs1860264*, *rs2191031*, *rs2236938*, *rs6441929*, *rs7614952* and *rs9842595* (intron 2). There was no information about the *rs17713054* marker associated with severe COVID-19 [3] (*LZTFL1* enchancer) in the GG-Base, so that analysis was provided only for North Eurasia.

Evaluation of linkage disequilibrium (LD R2) for the studied 11 *LZTFL1* SNP markers is based on the North Eurasian data pool (appendix, Table 1). Tight linkage to *rs12493471* associated with severe COVID-19 [1], as well as to *rs2191031* and *rs11130077* was revealed for the *rs17713054* marker [3].

Statistical and cartographic analysis

Multivariate statistical analysis was performed using the frequencies of 10 *LZTFL1* SNP markers, for which information about the populations in North Eurasia and the world was available (appendix, Tables 2, 3): multidimensional scaling (MDS) was used for North Eurasia to perform analysis based on the frequencies of all 10 SNPs and 5 SNPs showing the least tight linkage with each other (appendix, Table 1); principal component analysis (PCA) and MDS were used for the world to perform analysis based on the frequencies of all 10 markers. The MDS algorithm involved calculations based on pairwise Nei's genetic distances (d).

The *LZTFL1* SNP marker frequencies were used to create maps of marker distribution in indigenous populations of North Eurasia (11 SNP markers) and the world (10 SNP markers) in the original GeneGeo software package. In maps and tables (Appendix, Tables 2, 3; Fig. C) each population was assigned a number making it easy to identify. The maps were created based on the digital grid model representing the matrix of interpolated marker frequency values at the regular grid nodes calculated by the weighted average interpolation with the weights decreasing with the cube of the distance based on all values at all the reference points which fell into the circle of radius R (R = 3000 km for North Eurasia, R = 4200 km for the world). Uniform color and numerical scales used all maps ensured the atlas unity.

RESULTS

Heterogeneity of North Eurasian populations based on the *LZTFL1* SNP markers

The data on certain SNP marker frequencies in the North Eurasian populations are provided in Table 2 of Appendix.
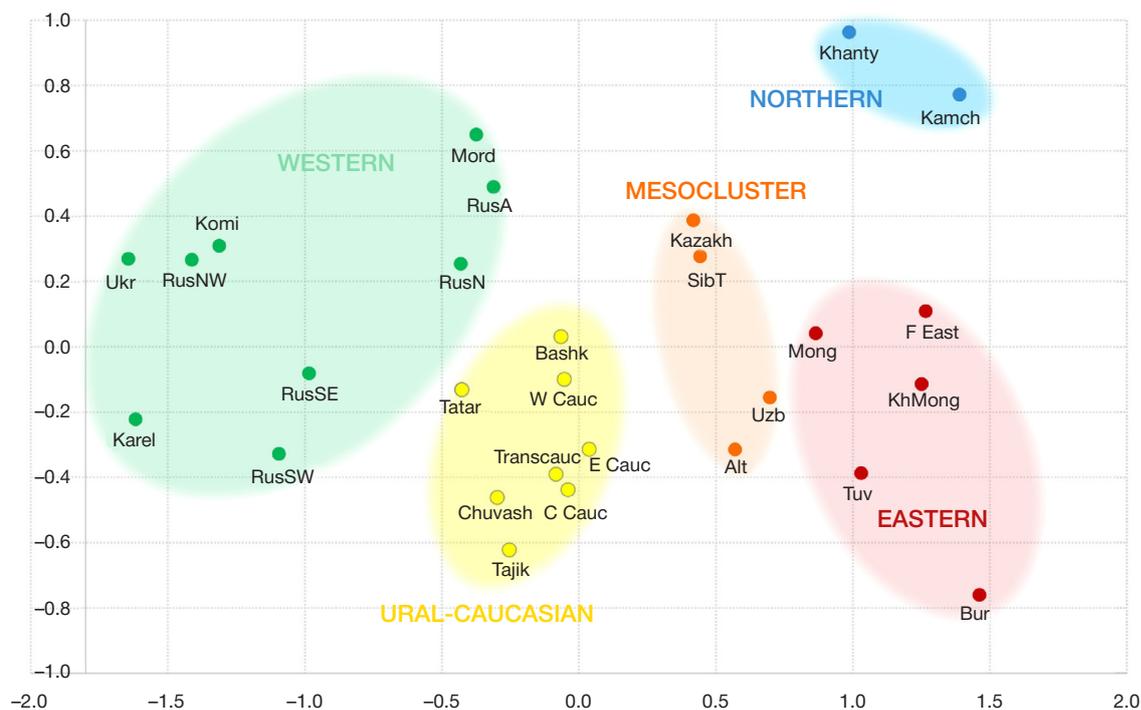
Fig. 1. Position of the populations of Russia and neighboring countries in the genetic space mapped by multidimensional scaling based on the data on 10 *LZTFL1* SNP markers. Note: multidimensional scaling plot, coefficient of alienation 0.1, stress SO = 0.09. Northern cluster is highlighted in blue, Western cluster in *green*, Ural-Caucasian cluster in *yellow*, Mesocluster in *orange*, Eastern cluster in *red*

However, it is necessary to recognize the patterns of variation in the entire marker set prior to analysis of the genetic landscape for each of the *LZTFL1* SNP markers (Fig. 1).

Their compliance with geographic variation along both axes, North–South and West–East, has turned out to be the most important feature of the relative position of populations from Russia and neighboring countries on the the MDS plot (Fig. 1).

The Northern cluster, which brings together the northernmost population of Western Siberia (Nenets, Mansi, Khanty) and the Far East (Itelmes, Koryaks, Chukchi), is opposed to all other clusters located strictly along the West–East axis. However, in each of the "southern" clusters the populations do not comply to their geographical position.

The Western cluster includes all Eastern Slavs (Belarusians, Russians, Ukranians) and Finnic-speaking peoples of Russia (Besermyans, Karelians, Komi, Mordvins, Udmurts).

The Ural-Caucasian cluster includes all peoples of the Caucasus, Transcaucasia and Tajikistan, Turks of the Urals (Bashkirs, Volga Tatars, Chuvash), and the Finnish-speaking Mari people included in the same metapopulation with the Chuvash.

Mesocluster visualizes gene pool transition from the Western to the Eastern cluster, bringing together those peoples of South Siberia and Eurasian Steppe, whose gene pools comprise both ancient Caucasoid stratum and potent later strata of Mongoloid populations (Altaians, Kazakhs, Karakalpaks, Kyrgyz, Nogais, Siberian Tatars, Turkmens, Uzbeks, Uyghurs, Khakas).

The Eastern cluster includes all the Mongolic-speaking peoples of North Eurasia (Buryats, Kalmyks, peoples of Mongolia), Tuvans (who were a part of Mongolia until the mid XX century) and small ethnic groups of the Far East (Nanais, Nivkhs, Ulchis, Evens).

Similar analysis performed based on the frequencies of 5 SNP markers showing the least tight linkage (LD R^2 < 0.2) reveals the same structure (Appendix, Fig. A), except the Volga Tatars' transition to the Western cluster.

In general, the overall trend of the *LZTFL1* SNP marker variation over North Eurasia is also in line with the West–East geographical vector and the Caucasoid–Mongoloid anthropological vector (Fig. 1).

Heterogeneity of the world's population based on the *LZTFL1* SNP markers

Global variation of distinct SNP markers is presented in Table 3 of Appendix, while the patterns for the set of SNP markers are provided in the PCA (Fig. 2) and MDS (Appendix, Fig. B) plots. Since the results and cluster structures obtained by both methods are similar, let us consider the PCA plot (Fig. 2).

What is striking is how high the variation of the populations of Russia and neighboring countries is on a global scale. These populations did not fit in any of five clusters within the space of principal components 1 and 2 of the world's gene pool (Fig. 2). The Western cluster of Russia is located in the upper part of the world's Indo-European cluster between the populations of North and Central Europe. The Ural-Caucasian cluster of Russia is in the opposite part of the Indo-European cluster, it is surrounded by the populations of Western Asia and South Europe. Three "Asian" clusters of North Eurasia, arranged in accordance with their origin, form their own North Asian cluster in the global genetic space: Russian Mesocluster gravitates to the world's Indo-European cluster, Eastern cluster of Russia to the world's South Asian Cluster, and Northern cluster of Russia (includes peoples of Chukotka and Kamchatka) is close to American cluster.

In general, the world's indigenous population is distributed over four clusters based on the parts of the world, however, it is adjusted to the anthropological features of the population (Fig. 2). Three clusters of indigenous populations includes only those from the place where it originates: Africa, Asia, or America.

However, Indo-European cluster juxtaposes geography and history of the populations. It includes Caucasoid populations of both Europe and Asia (India, Pakistan, Afghanistan, Middle East).

In other words, the main trend in variation of the whole *LZTFL1* SNP marker set across the world fits well with the
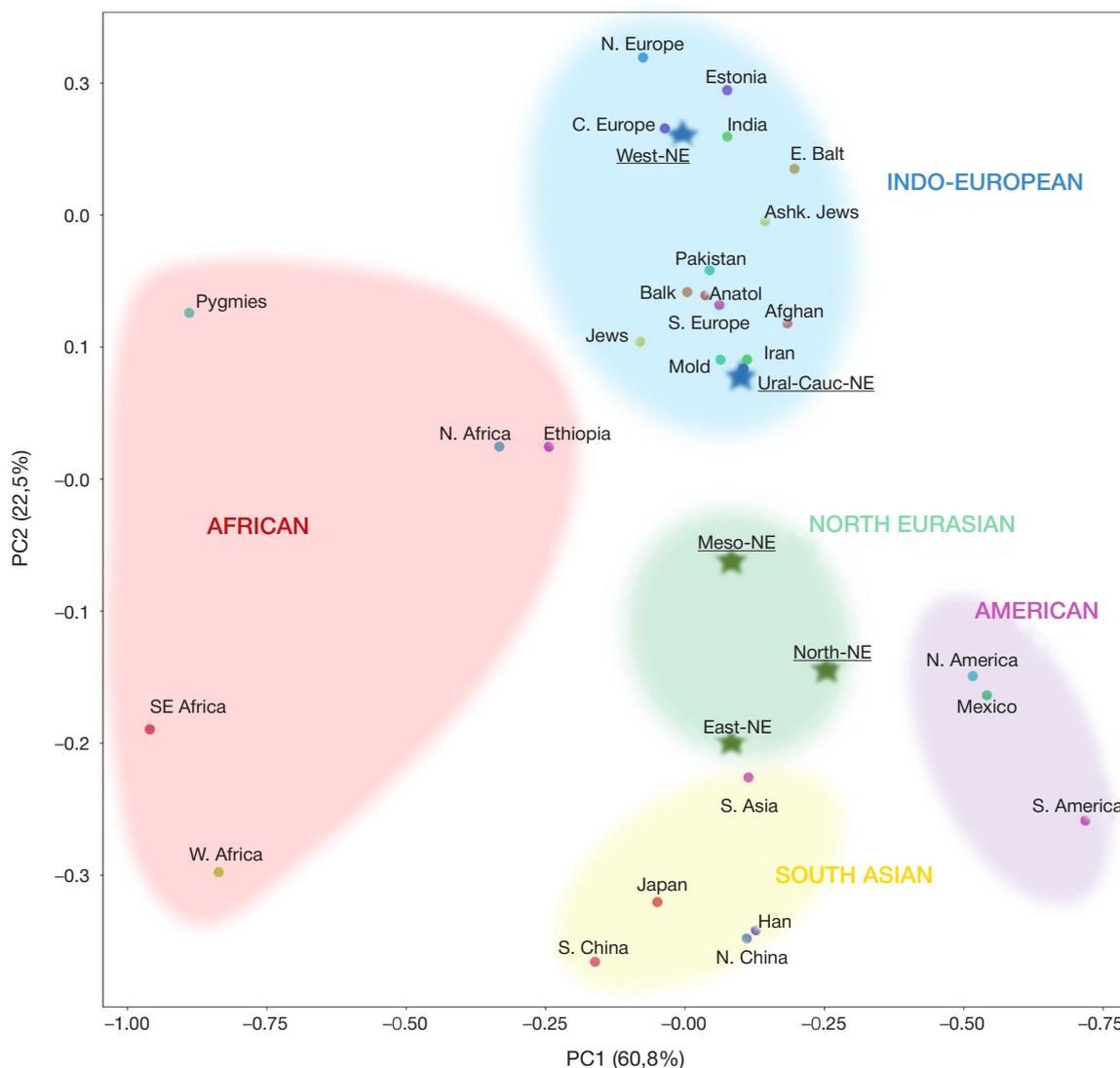
Fig. 2. Position of the world's indigenous populations in the space formed by principal components (PC) 1 and 2 based on variation in 10 *LZTFL1* SNP markers. Five clusters identified by multidimensional scaling of the populations of Russia and neighboring countries are marked with *asterisks*. Note: share of the described variance: PC 1 (61%), PC 2 (22%). Indo-European cluster is highlighted in *blue*, African cluster in *red*, North Asian cluster in *green*, South Asian cluster in *yellow*, American cluster in *violet*

world population anthropological division. Moreover, all four world's clusters are separated from each other. It's just huge genetic diversity of the peoples of Russia and neighboring countries that forms a bridge connecting three parts of the world: Europe, Asia and America.

Gene geography of 11 *LZTFL1* SNP markers in the populations of Russia and the world

The maps are not an illustration. They add two more dimensions of the geographic space to the tables to become an effective and powerful analysis tool. The ability to quickly capture a huge amount of information due to non-verbal representation is a specific advantage of this tool. We have constructed two variation maps per *LZTFL1* SNP marker (except *rs17713054*, since no information about the marker is available from global databases): for indigenous populations of North Eurasia and the world. Map comparison makes it possible both to reveal the global patterns and not to lose sight of the Russian genetic landscape. Each of 28 North Eurasian populations is marked with the number, allowing one to update both metapopulation name and the frequency of SNP marker in the population, in all maps (Appendix, Table 2). Information helps to navigate the

world's metapopulations (Appendix, Table 3, Fig. C). The maps are arranged in the same order as in Table 1 of Appendix.

rs17713054. Spatial variation in the *rs17713054*(*A*) frequency across North Eurasia (Fig. 3A) is low, however, it fits the West–East trend: with its minimum in the Far East and its maximum in the European part, where high frequency values are found in the west, (Ukraine, 16%), northwest (Karelia, 14%), Urals region (16%), and Caucasus (14%). That is why the European part of North Eurasia can be considered the region showing the highest frequency of this SNP marker. The other region of increased frequency is emerging in Tajikistan ($p = 0.15$), which could indirectly confirm the earlier conclusions [3] about the high *rs17713054* frequency in the southern regions of Asia.

rs1058961. The *rs1058961*(*A*) genetic landscape in North Eurasia (Fig. 3B) showing higher average frequency (30%) reflects similar, but more smoothed clinal variation in the form of frequency decline from the west (43% in Karelians and Veps) to the northeast (20% in the Far East). The local minima are found in the north of Western Siberia (8%), while the local maxima are observed in Central Asia (37%).

Comparison with global variability (Fig. 3C) shows that the North Eurasian genetic landscape is almost fully integrated into the overall pattern of world population. The frequency decline
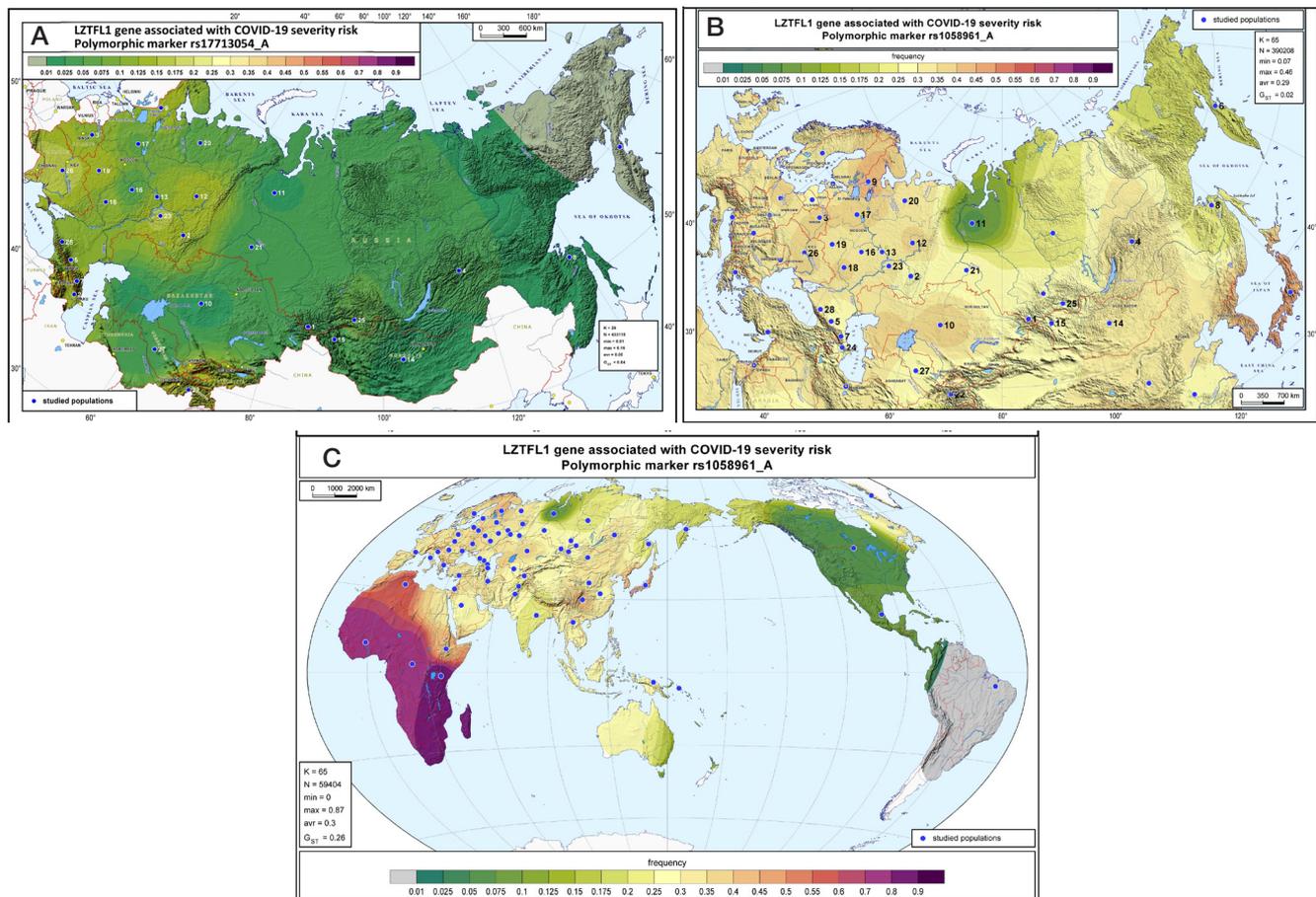
Fig. 3. Geographic variation in the frequencies of *LZTFL1* SNP markers in indigenous population: 3A — *rs17713054*(*A*) variation in the population of North Eurasia; 3B — *rs1058961*(*A*) variation in the population of North Eurasia; 3C — *rs1058961*(*A*) variation in the world's population. Note: The numbers of the populations on the map of North Eurasia correspond to that presented in Table 1 of Appendix, the numbers on the map of the world's population correspond to that provided in Table 2 and Fig. 1 of Appendix. No literature data on the indigenous population of Australia have been found, that is why the data on Oceania are interpolated on this region

found in the Far East smoothly turns into frequency decline in Alaska, decline in the indigenous population of North America, and decline to zero in South America. High frequencies found in Europe gradually turn into maximum frequencies (87%) found in Africa. Even the local maximum found in Japan (46%) (Fig. 3B) is reflected in the increasing frequency found in South-East Asia (42%).

rs12493471. The same West–East clinal variation across North Eurasia has been found in *rs12493471*(*A*) (Fig. 4A) linked with *rs17713054*(*A*) (Fig. 3A; Appendix, Table 1). However, the frequency drop gradient between west and east is much clearer: the maxima (≈ 50%) covering the European part of Russia barely move beyond the Urals. The frequency decline found in the east covers all regions and drops to zero in the Far East, Japan and China. The peak frequency found in Eastern Europe and Fennoscandia decreases in Western and Southern Europe.

The global variation map (Fig. 4B) shows that the peak frequency found in Eastern and Northern Europe is a world's maximum, from which the frequency decrease goes in all directions showing local maxima (40%) in Hindustan (the impact of which reaches Pamir) and Oceania.

rs11130077. In North Eurasia, the West–East clinal variation is also typical for *rs11130077*(*G*) (Fig. 4C). Minor differences are associated with the maximum shift to Fennoscandia (54%), however, the maxima do not move beyond the Urals and do not enter the Caucasus. Yet, the pattern observed in the Asian part is less clear than the patterns found in the previous maps. When the frequency is reduced to 16–17% in Siberia, there is

a slight increase to 20% in the Far East and to 27% in South Siberia.

The *rs11130077*(*G*) variation in the world's population (Fig. 4D) is generally in line with the previous map. However, there is one exception: the world's maximum fall not on Europe (53%), but on the African population showing significant differences based on this marker (40% in North Africa to 82% in Pygmies of Africa).

rs7614952. Unlike the previous maps, the *rs7614952*(*A*) genetic landscape (Fig. 4E) in North Eurasia shows no clear pattern. Despite the maximum values are still found in the western part of the region and decrease toward the Far East, the frequency minimum falls on the northern part of West Siberia, and the local maxima are found in both Europe and Transbaikal.

In the context of global variation (Fig. 4F) we can see that the Baikal maximum is a high frequency part of the entire East Asian region. This is the most marked difference with the previous map: minimum *rs11130077*(*G*) frequencies are found in East Asia (Fig. 4G), but the frequencies of *rs7614952*(*A*) are high (Fig. 4F). In contrast, were see a rapid *rs7614952*(*A*) frequency drop instead of high frequencies in Hindustan and West Asia. However, the world's maximum is still in Africa, while the world's minimum is in America.

rs2191031. The *rs2191031*(*A*) variation in North Eurasia is unimpressive (Fig. 5A) due to alternation of local maxima and minima. Minimum is once again found in the Far East (10%), as well as in the western part of the region (18%) and in West Siberia (22%). High frequencies that stretch from Transbaikal
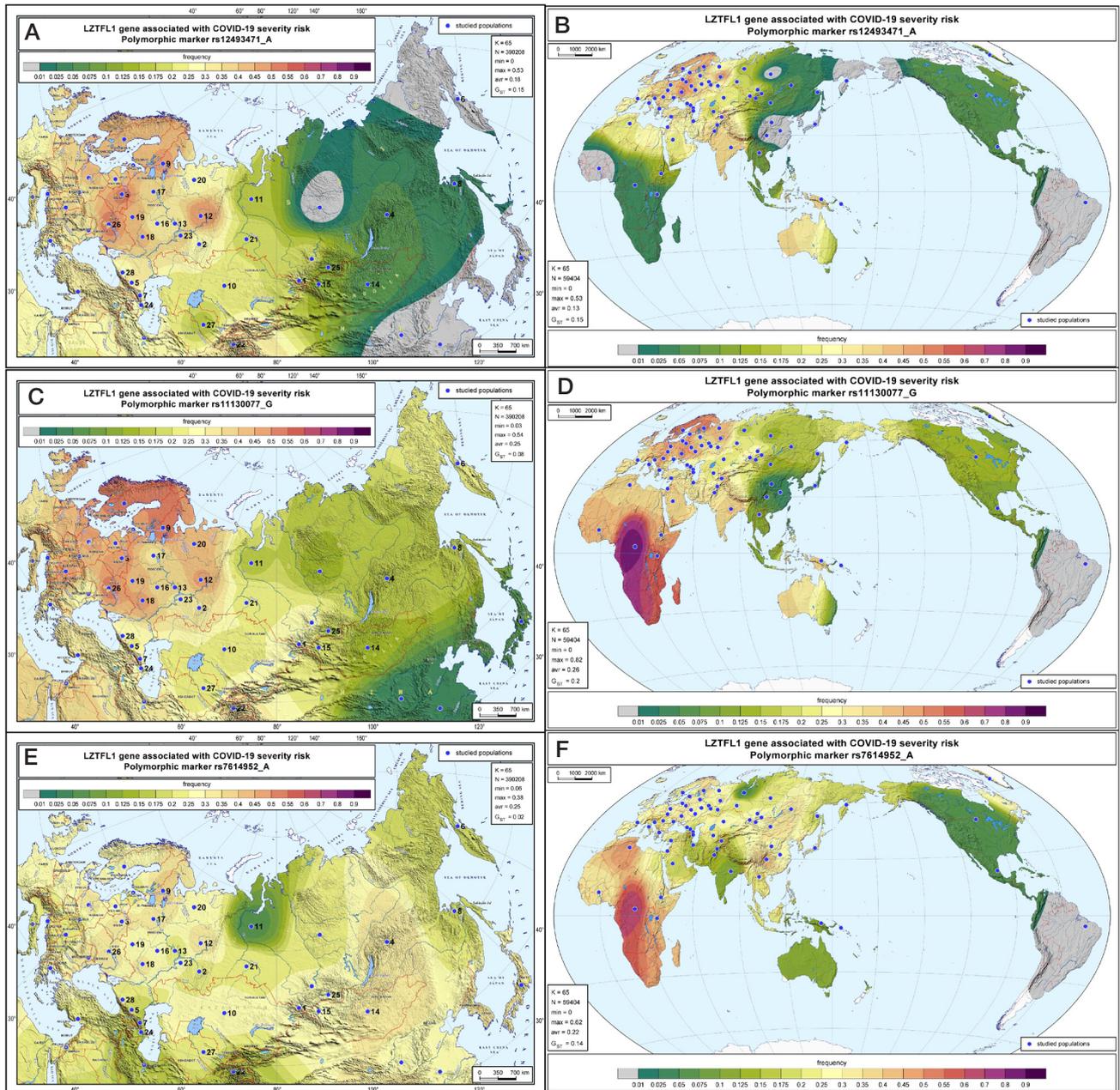
Fig. 4. Geographic variation in the frequencies of *LZTFL1* SNP markers in indigenous population: 4A — *s12493471*(*A*) variation in the population of North Eurasia; 4B — *rs12493471*(*A*) variation in the world's population; 4C — *rs11130077*(*G*) variation in the population of North Eurasia; 4D — *rs11130077*(*G*) variation in the world's population; 4E — *rs7614952*(*A*) variation in the population of North Eurasia; 4F — *rs7614952*(*A*) variation in the world's population. Note: The numbers of the populations on the map of North Eurasia correspond to that presented in Table 1 of Appendix, the numbers on the map of the world's population correspond to that provided in Table 2 and Fig. 1 of Appendix. No literature data on the indigenous population of Australia have been found, that is why the data on Oceania are interpolated on this region

(36%) through South Siberia (34%) to Central Asia (36%) are also typical for Povolzhye (37%) and the Caucasus (34%).

However, this pattern is fully integrated into the global variation landscape (Fig. 5B): the frequency decline observed in Western Europe (16%) turns into minimum in the West and East Africa (3%), while the frequency decline found in the Far East transforms into minima observed in the indigenous population of America (0–4%). The maxima found in the southern Siberia and Central Asia are a part of the high frequency region of Southeast and South Asia (35–40%), and the world's maximum is reached in Oceania (53%).

rs9842595. The *rs9842595*(*A*) genetic landscape is even less impressive (Fig. 5C): high frequencies are distributed through North Eurasia: Far East (19%), northern (17%) and southern (15%) parts of Europe, Ural region (12%) and the Caucasus (11%).

Similar is the global genetic landscape (Fig. 5D), covered almost entirely by the low frequency region, except sub-Saharan Africa, where the marker frequency rises to 33%.

rs1860264. In contrast, the *rs1860264*(*C*) marker frequency does not drop below 22% in North Eurasia (Fig. 5E) and generally fits the common West–East trend, although the local minima and maxima are scattered over various regions. Thus, the minimum frequency band stretches across the entire West Siberia towards Kazakhstan, but also shows up in the Ural region and Baltic States. High frequencies have been revealed not only in the western part of the region (50%), but also in Central Asia, South Siberia and the Baikal region (40–42%).

The global genetic landscape map (Fig. 5F) shows that Eurasia represents a gradual transition from the African maximum (97%) to American minimum (0%).
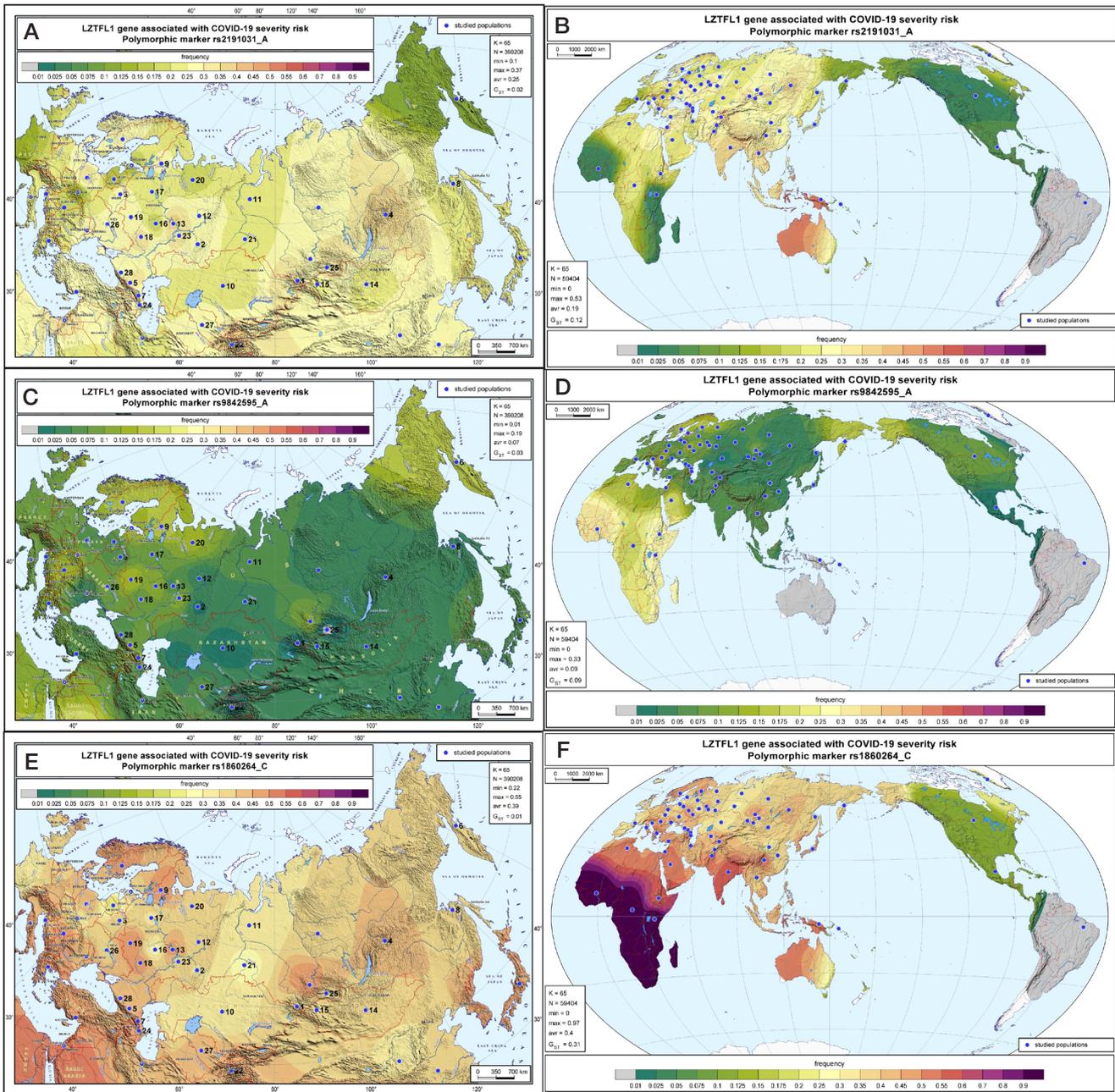
Fig. 5. Geographic variation in the frequencies of *LZTFL1* SNP markers in indigenous population: 5A — *rs2191031*(*A*) variation in the population of North Eurasia; 5B — *rs2191031*(*A*) variation in the world's population; 5C — *rs9842595*(*A*) variation in the population of North Eurasia; 5D — *rs9842595*(*A*) variation in the world's population; 5E — *rs1860264*(*C*) variation in the population of North Eurasia; 5F — *rs1860264*(*C*) variation in the world's population. Note: The numbers of the populations on the map of North Eurasia correspond to that presented in Table 1 of Appendix, the numbers on the map of the world's population correspond to that provided in Table 2 and Fig. 1 of Appendix. No literature data on the indigenous population of Australia have been found, that is why the data on Oceania are interpolated on this region

rs6441929. The *rs6441929*(*G*) variation pattern also fits the West–East trend (Fig. 6A), however, the highest values are found in the east with their maximum in Transbaikal and minimum frequencies in the Urals, Caucasus and Baltic States.

This landscape is fully integrated into the global one (Fig. 6B): the division of Eurasia into western, showing low values, and eastern, showing higher values, continues to the south down to the border between Hindustan and Southeast Asia. However, Eurasia globally contains no extrema. These are once again found in Africa (maxima) and America (minima).

rs2236938. The *rs2236938*(*A*) genetic landscape is similar to the previous one (Fig. 6C). The western zone of minima is the only clear one, while the zone showing a rapid frequency drop and gravitating towards the American world's minimum has emerged in the northeastern part of the eastern high frequency zone.

The *rs2236938*(*A*) global landscape, which is also similar to the previous one, appears to be more contrasting (Fig. 6D). In Africa the frequency rises to 53%, while Arabia, northern and northeastern Africa accede to the western Eurasian low frequency zone.

rs17078408. Finally, the marker is considered that is virtually abscent all over the world (Fig. 6E, F), except Africa, where the *rs17078408*(*C*) frequency reaches 48%.

DISCUSSION

Summarizing genetic landscapes of all the discussed *LZTFL1* markers associated with severe COVID-19, we shall refer to the main patterns: 1) the world's extrema are most typical for indigenous populations of Africa and America and are
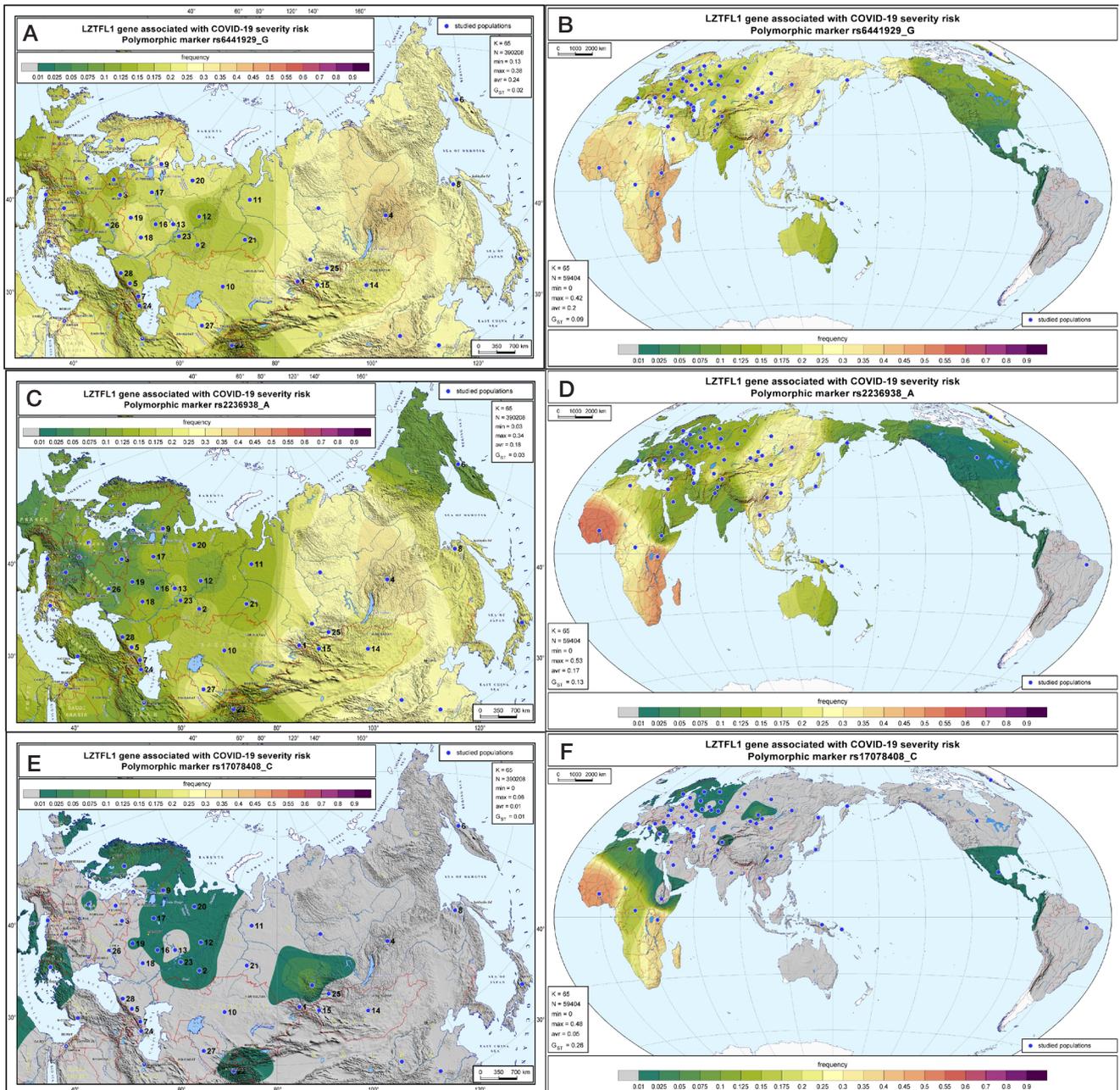
Fig. 6. Geographic variation in the frequencies of *LZTFL1* SNP markers in indigenous population: 6A — *rs6441929*(*G*) variation in the population of North Eurasia; 6B — *rs6441929*(*G*) variation in the world's population; 6C — *rs2236938*(*A*) variation in the population of North Eurasia; 6D — *rs2236938*(*A*) variation in the world's population; 6E — *rs17078408*(*C*) variation in the population of North Eurasia; 6F — *rs17078408*(*C*) variation in the world's population. Note: The numbers of the populations on the map of North Eurasia correspond to that presented in Table 1 of Appendix, the numbers on the map of the world's population correspond to that provided in Table 2 and Fig. 1 of Appendix. No literature data on the indigenous population of Australia have been found, that is why the data on Oceania are interpolated on this region

usually alternative; 2) Eurasia usually constitutes a transition zone between these two extrema, but shows its own patterns and enormous variation on a global scale; 3) the genetic landscape of Russia is seamlessly integrated into the Eurasian landscape.

These main patterns cannot always be relied on.

There are two exceptions to the first pattern. The minimum *rs12493471*(*A*) frequencies are clustered in Africa, America and East Asia, while the maximum values are clustered in Europe and Hindustan. Likewise, the *rs2191031*(*A*) minima are found in Africa and America; high frequencies are found across the entire Eurasia, however, the maximum is centered in Oceania. It should be noted that indigenous population of America always shows the common pattern: low frequencies of all the discussed markers drop to zero in South America.

There are also exceptions to the second pattern. For example, the world's maximum of *rs11130077*(*G*) is typical not only for Africa, but also for North Europe. Eurasia, like North America, represents a confinement of low *rs9842595*(*A*) frequencies. The *rs6441929*(*G*) Eurasian landscape is in sharp contrast to the term "transition zone", although it is an intermediate between two extrema: the maximum frequencies of the African continent share borders with the maximum frequencies found in Europe and Hindustan, while the world's lowest frequency is found in America which shares borders with high frequencies found in East Asia. The same Eurasian "patchwork" is observed for *rs2236938*(*A*). The division into west and east along the 80th meridian, separating Hindustan from Southeast Asia in the south and gradually blurring on its way to the Arctic Ocean, is most typical for Eurasia. The

following markers do not fit to this pattern: *rs1058961*(*A*) (high frequencies are observed in almost all Eurasia); *rs9842595*(*A*) and *rs17078408*(*C*) (the entire Eurasia is homogenous based on low frequencies of this marker); *rs2191031*(*A*) (frequency increases towards the south and turns into the Oceanic world's maximum); the pattern of *rs9842595*(*A*) is unclear. However, this pattern is clear in a half of the *LZTFL1* markers.

There are also exceptions to the third pattern. These are usually related to the local extrema found in the northern part of West Siberia, southern part of Middle Siberia and in the Ural region. In general, these patterns do not negate, but cast light on the overall integration of the Russian genetic landscape into the Eurasian landscape.

CONCUSIONS

1. The patterns typical for indigenous populations of Russia and the world were revealed in the spatial variation of the studied *LZTFL1* SNP markers associated with severe COVID-19. 2. The main pattern revealed in the North Eurasian genetic space is the compliance with geographic variation along two axes, North–South and West–East. This trend fits well with the Caucasoid–Mongoloid anthropological vector. 3. The main vector of global variation is fully in line with the world's population anthropological division. The clusters of indigenous populations of Africa, Asia and America include only the populations of their own parts of the world. The Indo-European cluster juxtaposes the population's geography and history, it includes Caucasoid populations of both Europe and Asia. 4. All four clusters of the world's indigenous population are separated from each other. It's just huge genetic diversity of the peoples of Russia and neighboring countries that forms a bridge connecting gene pools in three parts of the world: Europe, Asia and America. 5. A cartographic atlas for spatial variation of 11 *LZTFL1* markers in the populations of North Eurasia and the world showing the main patterns of the genetic landscapes has been created: a) the world's extrema fall on the indigenous populations of Africa and America; b) Eurasia constitutes a transition zone between these two extrema, but has its own patterns; c) the genetic landscape of Russia is seamlessly integrated into the Eurasian landscape.

References

1. The COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19. Nature. 2021; 600: 472–7. DOI: 10.1038/s41586-021-03767-x.
2. COVID19-hg GWAS meta-analyses round 6. The COVID-19 Host Genetics Initiative. [cited 2022 Sep 13]. Available from: https://www.covid19hg.org/results/r6/.
3. Downes DJ, Cross AR, Hua P, Roberts N, Schwessinger R, Cutler AJ, et al. Identification of LZTFL1 as a candidate effector gene at a COVID-19 risk locus. Nat Genet. 2021; 53: 1606–15. DOI: 10.1038/s41588-021-00955-3.
4. Vologzhanin DA, Golota AS, Kamilova TA, Shneider OV, Sherbak SG. Genetics of COVID-19. Journal of Clinical Practice. 2021; 12 (1): 41–52. DOI: 10.17816/clinpract64972. Russian.
5. Bardet-Biedl syndrome-17; BBS17. Online Mendelian Inheritance in Man — OMIM. [cited 2021 Nov 25]. Available from: https://omim.org/entry/615994.
6. Waters AM, Beales PL. Ciliopathies: an expanding disease spectrum. Pediatr Nephrol. 2011; 26: 1039–56. DOI: 10.1007/s00467-010-1731-7.
7. Potrokhova EA, Babayan ML, Baleva LS, Safonova MP, Sipyagina AE. Bardet–Biedl Syndrome. Rossiyskiy Vestnik Perinatologii i Pediatrii (Russian Bulletin of Perinatology and Pediatrics). 2020; 65 (6): 76–83. DOI: 10.21508/1027-4065-2020-65-6-76-83. Russian.
8. Seo S, Zhang Q, Bugge K, Breslow DK, Searby CC, Nachury MV, et al. A novel protein LZTFL1 regulates ciliary trafficking of the BBSome and Smoothened. PLoS Genet. 2011; 7 (11): e1002358. DOI: 10.1371/journal.pgen.1002358.
9. GeneCards: The Human Gene Database [Internet]. Rehovot, Israel: Weizmann Institute of Science. c1996-2022 — LZTFL1 Gene — Leucine Zipper Transcription Factor Like 1; [cited 2022 Sep 12]. Available from: https://www.genecards.org/cgi-bin/carddisp.pl?gene=LZTFL1.
10. Klink BU, Gatsogiannis C, Hofnagel O, Wittinghofer A, Raunser S. Structure of the human BBSome core complex. eLife. 2020; 9: e53910. DOI: 10.7554/eLife.53910.
11. Nakanishi T, Pigazzini S, Degenhardt F, Cordioli M, Butler-Laporte G, Maya-Miles D, et al. Age-dependent impact of the major common genetic risk factor for COVID-19 on severity and mortality. J Clin Invest. 2021; 131 (23): e152386. DOI: 10.1172/JCI152386.
12. Ellinghaus D, Degenhardt F, Bujanda L, Buti M, Albillos A, Invernizzi P, et al. Genomewide Association Study of Severe Covid-19 with Respiratory Failure. Engl J Med. 2020; 383 (16): 1522–34. DOI: 10.1056/NEJMoa2020283.
13. Pairo-Castineira E, Clohisey S, Klaric L, Bretherick AD, Rawlik K, Pasko D, et al. Genetic mechanisms of critical illness in COVID-19. Nature. 2021; 591 (7848): 92–8. DOI: 10.1038/s41586-020-03065-y.
14. Balanovsky O, Petrushenko V, Mirzaev K, Abdullaev S, Gorin I, Chernevskiy D, et al. The variation of genome sites associated with severe COVID-19 across populations the worldwide and national patterns. Pharmgenomics Pers Med. 2021; 14: 1391–402. DOI: 10.2147/PGPM.S320609.
15. Secolin R, de Araujo TK, Gonsales MC, Rocha CS, Naslavsky M, Marco L, et al. Genetic variability in COVID-19-related genes in the Brazilian population. Hum Genome Var. 2021; 8 (15). DOI: 10.1038/s41439-021-00146-w.
16. Balanovska EV, Zhabagin MK, Agdjoyan AT, Chuhryaeva MI, Markina NV, Balaganskaya OA, et al. Population biobanks: Organizational models and prospects of application in gene geography and personalized medicine . Russ J Genet. 2016; 52 (12): 1227–43. DOI: 10.1134/S1022795416120024.
17. Gorin IO, Petrushenko VS, Zapisetskaya YS, Koshel SM, Balanovsky OP. Application of the population biobank for analysis of the distribution of the clinically significant DNA markers in the Russian populations: bioinformatic aspects. Cardiovascular Therapy and Prevention. 2020; 19 (6): 2732. DOI: 10.15829/1728-8800-2020-2732. Russian.
18. GG-base [cited 2022 Sep 10]. Available from: https://gg-base.org/.

Литература

1. The COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19. Nature. 2021; 600: 472–7. DOI: 10.1038/s41586-021-03767-x.
2. COVID19-hg GWAS meta-analyses round 6. The COVID-19 Host Genetics Initiative. [cited 2022 Sep 13]. Available from: https://www.covid19hg.org/results/r6/.
3. Downes DJ, Cross AR, Hua P, Roberts N, Schwessinger R, Cutler AJ, et al. Identification of LZTFL1 as a candidate effector gene at a COVID-19 risk locus. Nat Genet. 2021; 53: 1606–15. DOI: 10.1038/s41588-021-00955-3.
4. Вологжанин Д. А., Голота А. С., Камилова Т. А., Шнейдер О. В., Щербак С. Г. Генетика COVID-19. Клиническая практика.

2021; 12 (1): 41–52. DOI: 10.17816/clinpract64972.

5. Bardet-Biedl syndrome-17; BBS17. Online Mendelian Inheritance in Man — OMIM. [cited 2021 Nov 25]. Available from: https://omim.org/entry/615994.

6. Waters AM, Beales PL. Ciliopathies: an expanding disease spectrum. Pediatr Nephrol. 2011; 26: 1039–56. DOI: 10.1007/s00467-010-1731-7.

7. Потрохова Е. А., Бабаян М. Л., Балева Л. С., Сафонова М. П., Сипягина А. Е. Синдром Барде–Бидля. Российский вестник перинатологии и педиатрии. 2020; 65 (6): 76–83. DOI: 10.21508/1027-4065-2020-65-6-76-83.

8. Seo S, Zhang Q, Bugge K, Breslow DK, Searby CC, Nachury MV, et al. A novel protein LZTFL1 regulates ciliary trafficking of the BBSome and Smoothened. PLoS Genet. 2011; 7 (11): e1002358. DOI: 10.1371/journal.pgen.1002358.

9. GeneCards: The Human Gene Database [Internet]. Rehovot, Israel: Weizmann Institute of Science. c1996-2022 — LZTFL1 Gene — Leucine Zipper Transcription Factor Like 1; [cited 2022 Sep 12]. Available from: https://www.genecards.org/cgi-bin/carddisp.pl?gene=LZTFL1.

10. Klink BU, Gatsogiannis C, Hofnagel O, Wittinghofer A, Raunser S. Structure of the human BBSome core complex. eLife. 2020; 9: e53910. DOI: 10.7554/eLife.53910.

11. Nakanishi T, Pigazzini S, Degenhardt F, Cordioli M, Butler-Laporte G, Maya-Miles D, et al. Age-dependent impact of the major common genetic risk factor for COVID-19 on severity and mortality. J Clin Invest. 2021; 131 (23): e152386. DOI: 10.1172/JCI152386.

12. Ellinghaus D, Degenhardt F, Bujanda L, Buti M, Albillos A, Invernizzi P, et al. Genomewide Association Study of Severe Covid-19 with Respiratory Failure. Engl J Med. 2020; 383 (16): 1522–34. DOI: 10.1056/NEJMoa2020283.

13. Pairo-Castineira E, Clohisey S, Klaric L, Bretherick AD, Rawlik K, Pasko D, et al. Genetic mechanisms of critical illness in COVID-19. Nature. 2021; 591 (7848): 92–8. DOI: 10.1038/s41586-020-03065-y.

14. Balanovsky O, Petrushenko V, Mirzaev K, Abdullaev S, Gorin I, Chernevskiy D, et al. The variation of genome sites associated with severe COVID-19 across populations the worldwide and national patterns. Pharmgenomics Pers Med. 2021; 14: 1391–402. DOI: 10.2147/PGPM.S320609.

15. Secolin R, de Araujo TK, Gonsales MC, Rocha CS, Naslavsky M, Marco L, et al. Genetic variability in COVID-19-related genes in the Brazilian population. Hum Genome Var. 2021; 8 (15). DOI: 10.1038/s41439-021-00146-w.

16. Балановская Е. В., Жабагин М. К., Агджоян А. Т., Чухряева М. И., Маркина Н. В., Балаганская О. А. и др. Популяционные биобанки: принципы организации и перспективы применения в геногеографии и персонализированной медицине. Генетика. 2016; 52 (12): 1371–87. DOI: 10.7868/S001667581612002X.

17. Горин И. О., Петрушенко В. С., Записецкая Ю. С., Кошель С. М., Балановский О. П. Применение популяционного биобанка для анализа распространенности клинически значимых ДНК-маркеров в населении России: биоинформатические аспекты. Кардиоваскулярная терапия и профилактика. 2020; 19 (6): 2732. DOI: 10.15829/1728-8800-2020-2732

18. GG-base [cited 2022 Sep 10]. Available from: https://gg-base.org/.