

© КОЛЛЕКТИВ АВТОРОВ, 2026

UDC 616.353.1-006:575.174.4(574)

DOI 10.59598/ME-2305-6053-2026-120-3-100-115

I. V. Korshukov^{1*}, I. A. Kadyrova¹, Ye. A. Kolesnikova², S. I. Kolesnichenko³, N. A. Kabildina¹, D. Zh. Taizhanova¹, A. A. Turmukhambetova⁴

ALLELE FREQUENCIES OF SINGLE-NUCLEOTIDE POLYMORPHISMS ASSOCIATED WITH COLORECTAL CANCER IN THE KAZAKH POPULATION OF CENTRAL KAZAKHSTAN

¹NC JSC «Karaganda Medical University» (100008, Republic of Kazakhstan, Karaganda c., Gogolya str., 40; e-mail: info@qmu.kz)

²Federal State Autonomous Educational Institution of Higher Education «N. I. Pirogov Russian National Research Medical University» (117513, Russian Federation, Moscow c., Ostrovityanova str., 1; e-mail: kolesnikova_ea6@rsmu.ru)

³University Hospital Center Sainte-Justine, Research Centre, University of Montreal (QC H3S 2G4, Canada, Montreal c., Côte-Sainte-Catherine Road, 3175; e-mail: svetlana.kolesnichenko@umontreal.ca)

⁴NC JSC «Astana Medical University» (010000, Republic of Kazakhstan, Astana city, Beibitshilik str., 49a; e-mail: mail@amu.kz)

*Ilya Vasilyevich Korshukov – NC JSC «Karaganda Medical University»; 100008, Republic of Kazakhstan, Karaganda, Gogolya str., 40; e-mail: Korshukov@qmu.kz

Aim. To estimate the allele frequencies of single nucleotide polymorphisms (SNPs) associated with the risk of colorectal cancer (CRC) in the Kazakh population of Central Kazakhstan and compare the obtained results with international reference data.

Materials and methods. The study included 97 apparently healthy individuals with no diagnosed CRC or family history of malignant neoplasms, who self-identified themselves, their biological parents, and their maternal and paternal grandparents as Kazakh. Genotyping was performed using the TaqMan OpenArray platform. Following missingness control and assessment of genotype distributions for Hardy-Weinberg equilibrium, the data were harmonized with the 1000 Genomes Project Phase 3 reference dataset. The population position of the local group was assessed using principal component analysis and hierarchical clustering. For each SNP, absolute differences in allele frequencies between the local group and the Eurasian reference superpopulations were calculated, and the closest reference superpopulation was identified. Statistical outliers were excluded using Tukey's fences. A total of 55 SNPs were included in the further analysis.

Results and discussion. Principal component analysis showed that the local group occupied an intermediate position within the Eurasian genetic space, predominantly between the European and East Asian superpopulations. In hierarchical clustering, the local group was represented in all four clusters, with most individuals distributed between two clusters of mixed population composition. The allele frequency profile of the analyzed SNPs showed pronounced locus-dependent variability. For 29 of the 55 SNPs (52,7%), the absolute allele frequency difference from the closest reference superpopulation did not exceed 0,025, whereas for 12 variants (21,8%) it was greater than 0,05. The closest reference superpopulation was EAS for 19 SNPs, EUR for 18 SNPs, and SAS for 18 SNPs.

Conclusions. The findings are consistent with the established understanding of the admixed genetic structure of the Kazakh population and demonstrate locus-dependent mosaicism in the allele frequency profile for the analyzed SNP panel. The nearly equal distribution of similarity across the Eurasian reference superpopulations indicates that no single reference superpopulation is the closest to the local group across the entire panel of variants. That should be taken into account while transferring findings from genome-wide association studies, applying polygenic risk scores, and planning medical genetic studies.

Key words: colorectal cancer; single-nucleotide polymorphisms; allele frequencies; Kazakh population

INTRODUCTION

According to worldwide GLOBOCAN estimates for 2024, colorectal cancer (CRC) ranks third among all malignancies by the number of newly diagnosed cases. In the Republic of Kazakhstan, CRC is also one of the most common oncological diseases, ranking second in terms of morbidity among all malignant

neoplasms (12,1% of all new cases) and third in terms of mortality from cancer (10,5% of all deaths) [1].

About 75% of CRC cases are sporadic, occurring in patients without a hereditary predisposition or family history of the disease [2]. The development of sporadic CRC is caused by the sequential accumulation of genetic and epigenetic abnormalities lead-

ing to the transformation of normal colon epithelium into adenocarcinoma, which in most cases takes 10–15 years [3]. Early stages of tumor transformation are usually associated with inactivation of the APC gene and activation of the WNT/ β -catenin signaling pathway, while the accumulation of alterations in the TP53, KRAS, PIK3CA genes, components of the TGF- β signaling pathway and other molecular regulators accompanies further progression [2, 3]. CRC is considered a molecularly heterogeneous disease, the formation of which is determined by the interaction of genetic changes, epigenetic alterations, the tumor microenvironment and immune response [3]. In addition to somatic molecular disorders, a significant contribution to the development of sporadic CRC is made by low penetrance hereditary predisposition and numerous environmental factors [4, 5]. The existing two-stage CRC screening program in the Republic [6] is based on fecal occult blood test results, followed by colonoscopy when indicated. This facilitates the detection of CRC cases in early stages of the existing tumor process, but does not take into account the molecular-genetic mechanisms underlying the sporadic forms' development.

Active search for common genetic variants associated with disease risk is conducted primarily using genome-wide association study (GWAS) methodology to identify genetic susceptibility loci [3]. Over the almost last 20 years, large GWAS and international meta-analyses have consistently identified new single nucleotide polymorphisms (SNPs) associated with CRC. As sample sizes increased from thousands to tens of thousands of cases and controls, the number of known risk loci expanded significantly, and pooling results from independent studies allowed the identification of variants with small effects that were undetectable earlier [7, 8, 9, 10, 11, 12, 13, 14, 15, 16]. It is also necessary to take into account the genetic architecture features of complex traits and the limited ability of GWAS to detect rare variants. All this complicates the generalization of accumulated information and its use for the creation of clinical diagnostic panels and the formation of polygenic risk scores. The development of such tools is necessary for the implementation of personalized preventive measures and the stratification of CRC development risk.

Most of the signals detected in the GWAS are not represented by direct causal variants, but by index (lead) or marker (tag) SNPs that are only statistically related to the corresponding locus. As a result, following studies usually involve fine mapping and functional validation aimed at locating the true causal variants, elucidating the molecular mechanisms of their action, and independently replicating them in different populations [7, 8, 10, 12, 13, 15, 17, 18, 19, 20].

Approaches aimed at solving the above-mentioned problems were outlined at the early stages of the genome-wide studies development. Possi-

ble measures include, for example, increasing the sample size, expanding the coverage of genotyped variants, using direct sequencing of loci identified by the GWAS, and taking into account the evolutionary characteristics of genetic variants, including their frequency, origin, and age [21].

Subsequent studies have shown that many risk loci are replicated in different populations, but population-specific variants that are characteristic for residents of East Asia, Latin America, Northern Europe and other ethnic groups have also been identified. These results suggest that the genetic architecture of CRC susceptibility includes both loci common to different populations and variants whose prevalence and contribution depend on population structure [17, 22, 23, 24, 25, 26]. Despite the high reproducibility of many associations, differences in the genetic structure of populations can alter the observed strength and, in some cases, the direction of associations. One of the reasons for this is the interpopulation differences in allele frequencies and the overall genetic structure that determine the genetic background on which the locus effect is realized. As a result, the impact of the same causal variant may vary depending on the genetic environment, including the distribution of other genetic variants and features of the genome architecture [7, 10, 12, 13, 15, 18, 25, 26, 27, 28, 29].

In addition, the transferability of the results is determined by the features of the structure of linkage disequilibrium (LD). Interpopulation differences in LD change the degree of statistical relationship between the index SNP and the causal variant, so the same genetic marker may reflect the influence of the causal variant unevenly in different populations. This can lead to a change in the observed strength of the association, decrease the reproducibility of the results, and complicate the localization of causal variants [7, 10, 12, 13, 15, 25, 29, 30].

When combining GWAS results from different ethnic populations, heterogeneity between studies becomes an additional problem [31]. A possible solution [30] is the use of transethnic GWAS analysis, which allows for the simultaneous associations reproducibility assessment in different populations and takes into account differences in their genetic structure for more accurate mapping of causal variants. Transethnic analysis of admixed populations is further complicated by the need for appropriate adjustments, the lack of which can lead to spurious associations and reduce the transferability of results between populations. The authors [32] note that a significant proportion of GWAS have been performed in populations of single ancestry, while admixed populations remain underrepresented in genetic studies. To address this issue, broader inclusion of mixed populations in GWAS, the use of methods that take into account the origin of study participants, including population and principal components, and the devel-

opment of specialized approaches for the analysis of genetically mixed groups are proposed.

Further development of transethnic approaches has shown that heterogeneity between populations is associated not only with differences in allele frequencies, but also with differences in the size of allele effects [33]. The authors note that the allelic effects of individual SNPs can vary between populations, and such heterogeneity often correlates with the origin of the samples studied. The reasons may be differences in the structure of LD, the interaction of causal variants with environmental factors, and the features of the imputation quality in different populations.

For many diseases, transethnic genetic correlation is significantly less than one, which limits the transferability of GWAS and polygenic risk assessment results between populations. However, existing methods typically estimate a single transethnic genetic correlation for the entire genome and make it virtually impossible to determine why the effects of causal variants differ between populations [34]. Regarding this, it is of interest to study transethnic genetic correlation in various functional categories of the genome and to identify areas where causal effects are more population-specific, which contributes to a better understanding of the biological mechanisms of these differences and the improvement of transethnic genetic research.

The mentioned problems of transferability of GWAS results are of great importance for the main ethnic group of Kazakhstan, given its admixed genetic background [35, 36]. Data obtained using whole genome sequencing show a high level of genetic admixture and an intermediate position of the Kazakh population between European and East Asian populations, but the number of participants in these studies is rather small.

The research results obtained in Kazakhstan agreed with published data on the Kazakh population in Xinjiang (China). Studies using panels of ancestry informative markers of 165 SNPs [37] and 56 InDels [38] independently showed an admixed genetic structure with a predominance of the East Asian component over the European one. Similar results are presented in a genome-wide study [39], which concludes that there is no division into two genetically distinct populations when compared with previously published genotypic data on a limited number of samples from Kazakhstan [40].

There are limitations in the study of genetic predisposition to CRC related to the transferability of identified associations between populations of different origins, which is especially relevant for the multi-ethnic population of Kazakhstan, characterized by a complex population structure. Due to the heterogeneity of genetic effects and the underrepresentation of genetically mixed populations in existing studies, the frequency characteristics of known SNPs associated

with CRC in the Kazakh population remain insufficiently studied.

Aim of the study was to estimate the allele frequencies of single nucleotide polymorphisms associated with the risk of colorectal cancer in the Kazakh population of Central Kazakhstan and compare the results with published international reference data.

MATERIALS AND METHODS

This study used materials obtained within the framework of the scientific and technical program O.0821 «Personalized approach in the management of a number of significant diseases», carried out at the Karaganda Medical University, the Regional Oncology Dispensary, and polyclinic organizations in Karaganda. The study was approved by the local bioethics committee of the Karaganda Medical University (decision No. 305 dated May 19, 2017). All study participants signed a written informed voluntary consent to participate in the study and to the use of biological material for scientific purposes before inclusion.

For further analysis, the local group included apparently healthy individuals without established CRC and a family history of malignant neoplasms among the relatives of the first, second, and third degree of kinship. All participants were examined as part of the national CRC screening program. The analysis included only individuals of Kazakh nationality who identified themselves, their biological parents, and their maternal and paternal grandparents as Kazakhs. In total, the local group included 97 participants: 37 women and 60 men. The median age was 48 years (Q25-Q75: 40-57 years).

For the molecular-genetic study, genomic DNA isolated from peripheral venous blood samples collected in EDTA tubes was used. DNA extraction was performed using the GeneJET Genomic DNA Purification Kit (Thermo Scientific, USA) according to the manufacturer's protocol. The quality and concentration of DNA were controlled spectrophotometrically using an Implen P330 nanospectrophotometer (Implen, Germany), after which the samples were stored at -20°C until genotyping.

For genotyping, a panel of single nucleotide polymorphisms associated with the risk of developing CRC according to published studies was used. Genotyping was performed using TaqMan® OpenArray® technology on the QuantStudio™ 12K Flex Real-Time PCR System platform (Applied Biosystems, USA) according to the manufacturer's protocol. Primary processing of the results was carried out using Thermo Fisher Cloud software, after which automatic determination of genotypes and visual control of allelic signal clustering were performed.

Information on published GWAS [7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 19, 20, 22, 23, 24, 25, 26, 41] for the SNPs included in this study was obtained

Table 1 – Sex distribution in reference superpopulations

Superpopulation	Females	Males	Total
AFR	342	319	661
AMR	177	170	347
EAS	260	244	504
EUR	263	240	503
SAS	229	260	489
Total	1271	1233	2504

from the NHGRI-EBI GWAS Catalog [42]. Obtaining information in batch mode was carried out through the web interface of the OpenCRAVAT platform [43]. In the case of a multiallelic variant, only the allelic pair corresponding to local samples was used. Functional annotation of SNPs was performed using the VEP ENSEMBLE web interface [44].

Missingness control in genotypic data was performed using PLINK software version 1.9 [45], available at <https://www.cog-genomics.org/plink/>. All samples and SNPs included in further analysis passed the missingness thresholds of 10% (--mind 0.10, --geno 0.10). The conformity of the genotype distribution with the Hardy-Weinberg equilibrium was assessed in the local group using the test implemented in PLINK 1.9 (--hardy). SNPs with a significance level of $p < 0.05$ were excluded from subsequent statistical analyses (rs16869961, rs647161, rs12682374, rs4948317, rs174537, rs4572213 and rs12953717).

For comparative population analysis, data from the 1000 Genomes Project (Phase 3) [46] were used, which included representatives of five superpopulations: African (AFR), Admixed American (AMR), East Asian (EAS), European (EUR) and South Asian (SAS). Data on the sex distribution and size of superpopulations are presented in Table 1. The 1000 Genomes Project does not provide information on the age of participants; it only uses the inclusion criterion of “age over 18 years”. After quality control, the local sample was combined with reference data to form a single set of genotypes. Before comparative analysis of the population structure, the orientation of alleles in the local sample was harmonized with the reference data by converting the A1/A2 designations to a single system, which ensured comparability of allele frequencies between populations.

Allele frequencies (AF) were calculated using PLINK v1.9 (--freqx) for the local group and the reference superpopulations EUR, EAS, and SAS. The similarity measure of allele frequencies from each reference Eurasian superpopulation for each SNP was estimated by the absolute difference between the allele frequency in the local group and the corresponding frequencies for the reference superpopulation ($ABS \Delta AF$). For each SNP, the closest reference superpopulation was determined as having the minimum allele frequency difference. The threshold

for detecting statistical outliers of minimal differences in allele frequencies was determined using Tukey's fences as the sum of the distribution's third quartile value and 1.5 times the interquartile range, where the interquartile range was the difference between the third and first quartiles ($Q3 - Q1$), calculated from SNPs that passed the Hardy-Weinberg criterion. Based on the results of applying this criterion, four variants (rs9929218, rs11196172, rs2189261 and rs10774214) were identified and excluded, as their values exceeded the established threshold for statistical outliers. A total of 55 SNPs were included in the subsequent analysis.

To assess the population structure, principal component analysis (PCA) was performed in PLINK v1.9 (--pca) for two data sets obtained by combining the local group with all reference populations and with the Eurasian reference superpopulations (EUR, EAS, SAS). For the Eurasian subsample, hierarchical clustering was performed using a Ward minimum variance method based on the values of the first ten principal components using Euclidean distance. The resulting clusters were used to assess the position of the local group relative to the reference populations. All statistical calculations were performed in the RStudio environment (version 2025.5.1.513; Posit Team, 2025).

RESULTS AND DISCUSSION

To assess the position of the local group relative to the major global superpopulations, a principal component analysis was performed on the pooled sample. Figure 1 shows the first two components (PC1 and PC2), which explain 13,6% and 9% of the total genetic variance, respectively.

Along the plane of the first two principal components, the African superpopulation was separated from the other groups. The East Asian and European superpopulations were located in opposite directions along the second principal component, while the South Asian superpopulation occupied an intermediate position between them. The American superpopulation overlapped significantly with the European and South Asian superpopulations. The local group occupied an intermediate position in the Eurasian genetic space, mainly between the European and East Asian superpopulations, with substantial overlap between the distribution areas of EUR, EAS, and SAS.

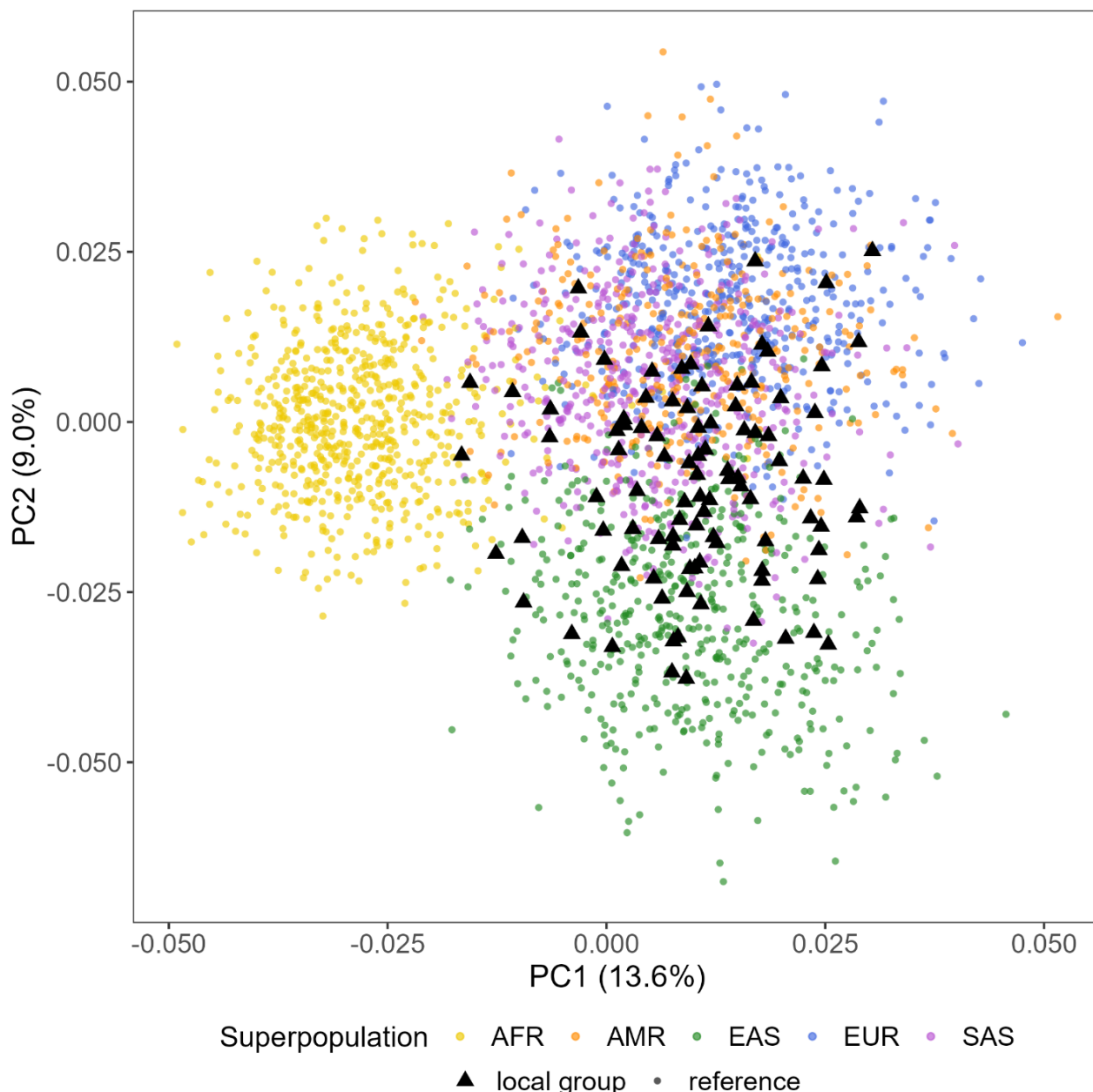


Figure 1 – Results of principal component analysis for all reference superpopulations and the local group

After assessing the position of the local group relative to the global reference superpopulations, the analysis was restricted to the Eurasian genetic space, including the EAS, EUR, and SAS superpopulations, as well as the local group. To assess the internal structure of this pooled sample, hierarchical clustering was performed using the first ten principal components of the PCA. The number of clusters was set to four, based on the assumption of the possible formation of separate groups corresponding to three reference superpopulations and a local group, the results of the cluster analysis are shown in Figure 2.

The local group was located predominantly in the overlapping area of the reference Eurasian superpopulations, without forming a visually distinctive cluster.

The composition of the obtained clusters is presented in Table 2; all of them were characterized by different population composition. In the first cluster, representatives of EUR prevailed, while cluster 2 mainly included representatives of EAS and SAS (289 and 333, respectively). At the same time, clusters 3 and 4 had a distinctly mixed composition. The local group was represented in all four clusters, with the largest number belonging to clusters 2 and 4. Thus, the local sample did not form a separate cluster and was distributed among several components of the Eurasian population structure.

The results of the cluster analysis are consistent with the position of the local group on the principal component plot and with published data on the admixed genetic structure of the Kazakh population [35, 36].

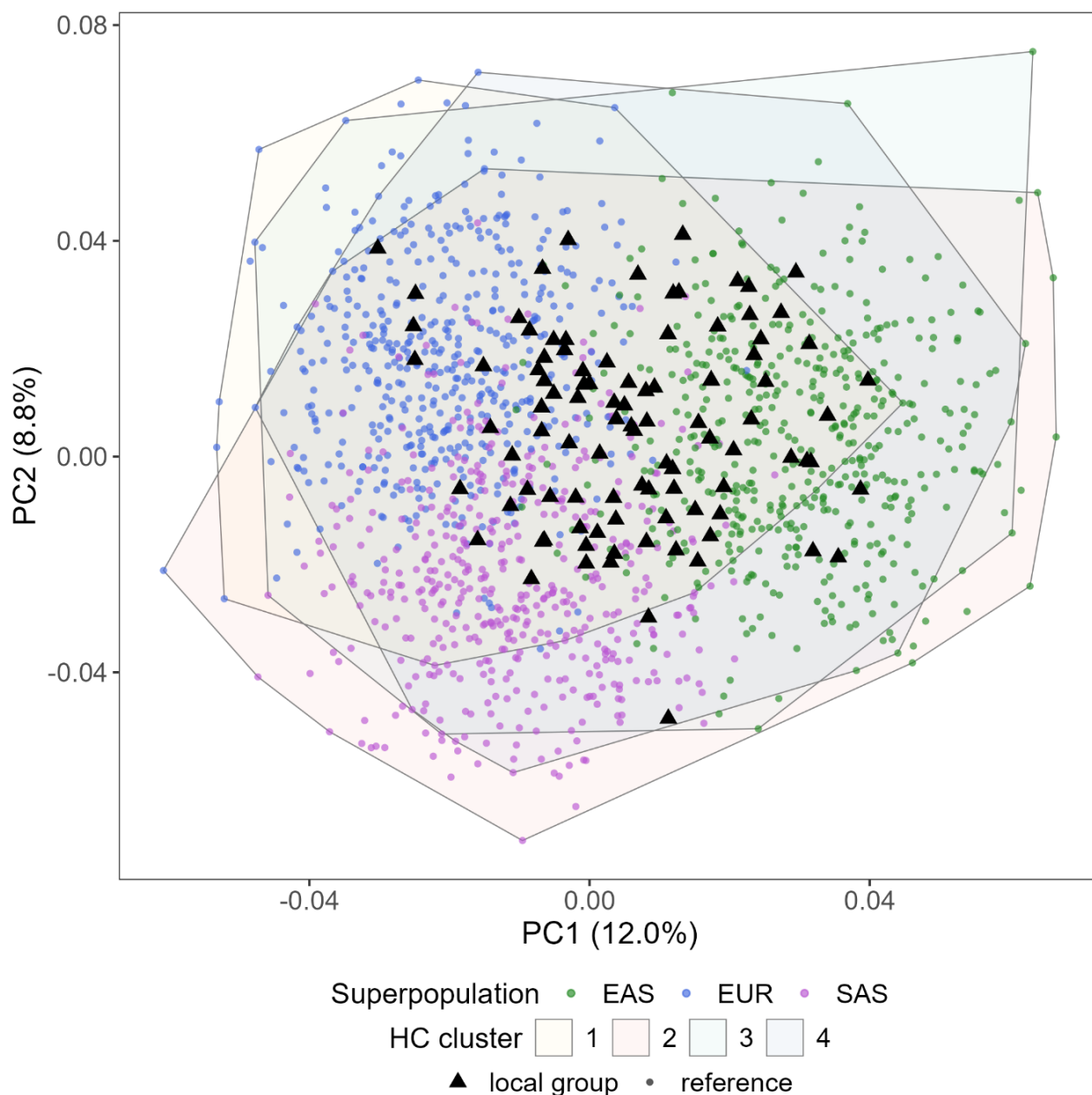


Figure 2 – Results of hierarchical cluster analysis for the Eurasian reference superpopulations and the local group, superimposed on the plane of 1 and 2 principal components

Table 2 – The results of a hierarchical cluster analysis of Eurasian reference superpopulations and the local group

Cluster	EAS	Local group	SAS	EUR
1	18	19	33	199
2	289	38	333	81
3	82	9	56	108
4	115	31	67	115

The population gradients of the A1 allele frequency distribution are shown in Figure 3. The rs36053993 variant was characterized by an almost

monomorphic distribution in all the populations studied. For the SNPs studied, pronounced interlocus variability in frequency differences between the local

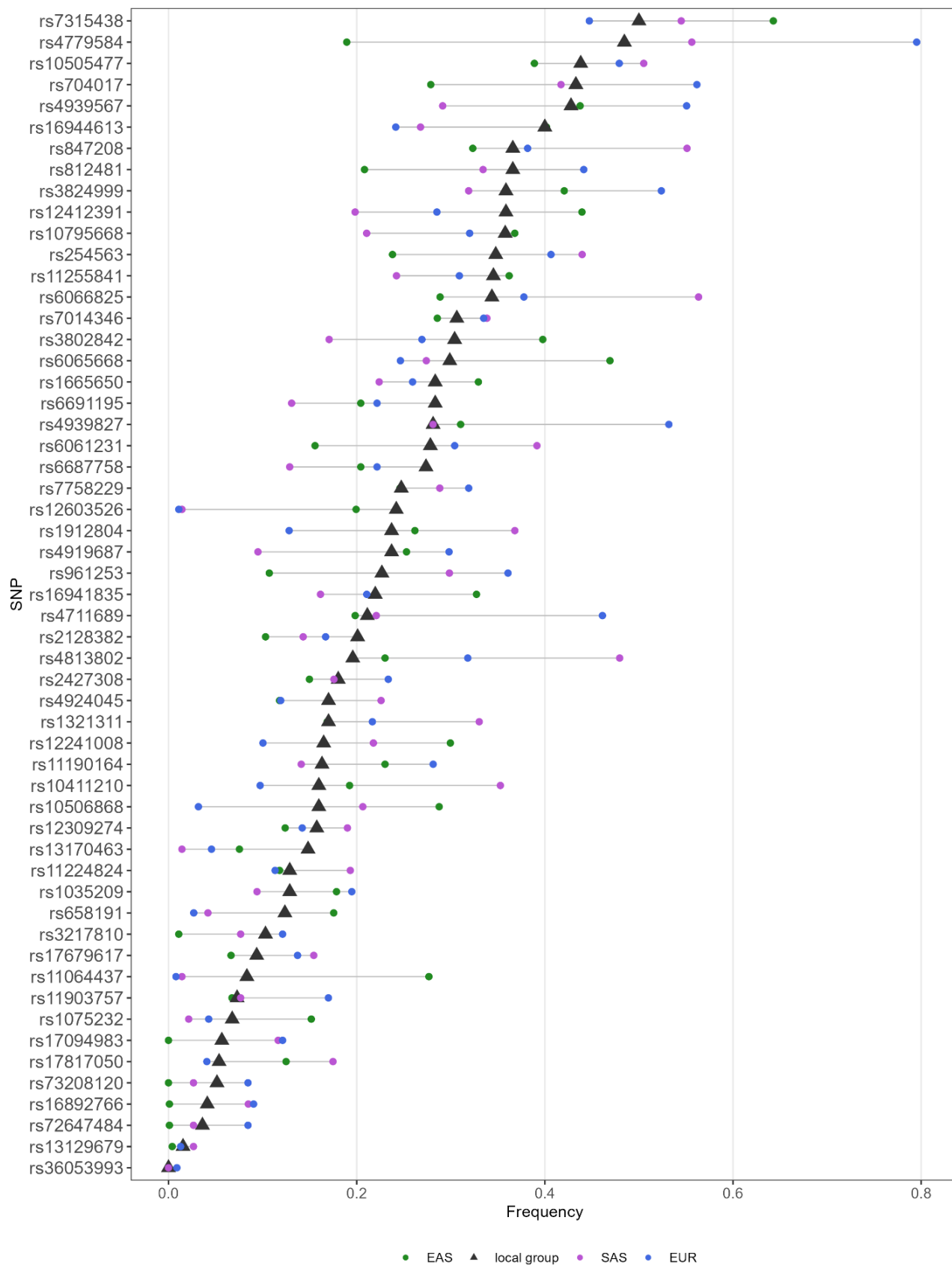


Figure 3 – Population gradients of A1 allele frequencies for 55 SNPs in the local group and the EAS, SAS, and EUR reference superpopulations. SNPs are sorted by A1 allele frequency in the local group

group and reference superpopulations was observed. At the same time, the A1 allele frequency was close to that of various Eurasian superpopulations.

Given the identified intermediate position of the local sample within the Eurasian genetic space, a comparison of allele frequencies was then performed

between the local group, represented by apparently healthy individuals, and the reference superpopulations EAS, SAS, and EUR. For each SNP, allele frequencies (AF) and the absolute allele frequency difference (ΔAF) relative to the reference populations were calculated and presented in Table 3.

Table 3 – Allele frequencies (AF A1) for the local group and reference superpopulations, absolute differences in allele frequencies between the local group and each of the reference superpopulations (ABS ΔAF), name of the closest superpopulation (Min ABS ΔAF) and results of the functional annotation of SNPs: functional consequence (FC) and gene symbol (GS)

Nr	SNP	A1	A2	AF A1 local group	AF A1 EAS	AF A1 EUR	AF A1 SAS	ABS ΔAF EUR	ABS ΔAF EAS	ABS ΔAF SAS	Min ABS ΔAF	FC	GS
1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	rs1035209	T	C	0,129	0,179	0,195	0,094	0,066	0,050	0,035	SAS	IgV	-
2	rs10411210	T	C	0,160	0,192	0,097	0,353	0,062	0,033	0,193	EAS	InV	RHPN2
3	rs10505477	A	G	0,438	0,389	0,479	0,505	0,041	0,049	0,067	EUR	InV, nCTV	CASC8
4	rs10506868	T	C	0,160	0,288	0,032	0,207	0,128	0,128	0,047	SAS	InV	VTI1A
5	rs1075232	A	C	0,068	0,152	0,043	0,021	0,025	0,084	0,046	EUR	IgV	-
6	rs10795668	A	G	0,358	0,368	0,320	0,211	0,038	0,010	0,147	EAS	InV, nCTV	LOC105376400
7	rs11064437	T	C	0,083	0,277	0,008	0,014	0,075	0,193	0,069	SAS	SpAV	SPSB2
8	rs11190164	G	A	0,163	0,230	0,281	0,141	0,118	0,067	0,022	SAS	IgV	-
9	rs11224824	A	G	0,129	0,118	0,113	0,193	0,016	0,011	0,064	EAS	InV	TRPC6
10	rs11255841	A	T	0,345	0,362	0,309	0,242	0,036	0,017	0,103	EAS	IgV	-
11	rs11903757	C	T	0,073	0,067	0,170	0,077	0,097	0,005	0,004	SAS	InV, nCTV	LOC124908062
12	rs12241008	C	T	0,165	0,300	0,100	0,218	0,065	0,135	0,053	SAS	InV	VTI1A
13	rs12309274	G	T	0,158	0,124	0,142	0,190	0,015	0,034	0,033	EUR	InV	WNK1
14	rs12412391	G	A	0,359	0,439	0,285	0,198	0,073	0,081	0,160	EUR	InV, nCTV	LINC01475
15	rs12603526	C	T	0,242	0,199	0,011	0,014	0,231	0,043	0,228	EAS	InV	NXN
16	rs13129679	A	G	0,015	0,004	0,013	0,027	0,003	0,011	0,011	EUR	InV	FAM193A
17	rs13170463	A	G	0,148	0,075	0,046	0,014	0,103	0,073	0,134	EAS	IgV	-
18	rs1321311	A	C	0,170	0,169	0,217	0,330	0,047	0,001	0,160	EAS	IgV	-
19	rs1665650	T	C	0,284	0,329	0,259	0,224	0,024	0,046	0,060	EUR	InV	HSPA12A
20	rs16892766	C	A	0,041	0,001	0,090	0,085	0,049	0,040	0,044	EAS	DsgV	LOC105375712
21	rs16941835	C	G	0,220	0,327	0,211	0,162	0,009	0,108	0,058	EUR	RRV	-
22	rs16944613	A	G	0,400	0,402	0,242	0,268	0,158	0,002	0,132	EAS	InV	CRTC3
23	rs17094983	A	G	0,057	0	0,121	0,117	0,065	0,057	0,060	EAS	IgV	-
24	rs17679617	T	C	0,094	0,066	0,137	0,154	0,043	0,027	0,061	EAS	InV	CDH13
25	rs17817050	G	A	0,054	0,125	0,041	0,175	0,013	0,071	0,121	EUR	UsgV	TMEM238L
26	rs1912804	C	G	0,237	0,262	0,128	0,368	0,109	0,025	0,131	EAS	InV	WWOX
27	rs2128382	T	C	0,201	0,103	0,167	0,143	0,034	0,098	0,058	EUR	IgV	-

Table 3 – Allele frequencies (AF A1) for the local group and reference superpopulations, absolute differences in allele frequencies between the local group and each of the reference superpopulations (ABS Δ AF), name of the closest superpopulation (Min ABS Δ AF) and results of the functional annotation of SNPs: functional consequence (FC) and gene symbol (GS) (continuation)

Nr	SNP	A1	A2	AF A1 local group	AF A1 EAS	AF A1 EUR	AF A1 SAS	ABS Δ AF EUR	ABS Δ AF EAS	ABS Δ AF SAS	Min ABS Δ AF	FC	GS
1	2	3	4	5	6	7	8	9	10	11	12	13	14
28	rs2427308	T	C	0,180	0,150	0,234	0,176	0,053	0,031	0,005	SAS	InV	CABLES2
29	rs254563	A	G	0,348	0,238	0,407	0,440	0,059	0,110	0,092	EUR	InV, nCTV	PITX1-AS1
30	rs3217810	T	C	0,103	0,011	0,121	0,077	0,018	0,092	0,026	EUR	InV	CCND2
31	rs36053993	T	C	0	0	0,009	0	0,009	0	0	EAS	MV, SRV	MUTYH
32	rs3802842	C	A	0,304	0,398	0,269	0,171	0,035	0,094	0,133	EUR	InV, nCTV	COLCA1
33	rs3824999	G	T	0,359	0,421	0,524	0,319	0,165	0,062	0,040	SAS	InV, nCTV	POLD3
34	rs4711689	G	A	0,211	0,198	0,461	0,221	0,250	0,013	0,010	SAS	RRV	-
35	rs4779584	C	T	0,485	0,189	0,795	0,556	0,311	0,295	0,072	SAS	IgV	-
36	rs4813802	G	T	0,196	0,230	0,318	0,480	0,122	0,034	0,284	EAS	RRV	-
37	rs4919687	A	G	0,237	0,253	0,298	0,095	0,061	0,016	0,142	EAS	InV	CYP17A1
38	rs4924045	T	C	0,170	0,118	0,119	0,226	0,051	0,052	0,056	EUR	UsgV	GOLGA8B
39	rs4939567	G	A	0,428	0,438	0,551	0,291	0,123	0,010	0,136	EAS	InV	SMAD7
40	rs4939827	T	C	0,281	0,311	0,532	0,281	0,251	0,029	0,000	SAS	InV	SMAD7
41	rs6061231	A	C	0,278	0,156	0,304	0,392	0,026	0,123	0,113	EUR	RRV	-
42	rs6065668	T	C	0,299	0,469	0,247	0,274	0,052	0,170	0,025	SAS	IgV	-
43	rs6066825	G	A	0,344	0,289	0,378	0,563	0,034	0,055	0,220	EUR	InV	PREX1
44	rs658191	G	C	0,124	0,176	0,027	0,042	0,097	0,052	0,082	EAS	InV	MACF1
45	rs6687758	G	A	0,274	0,204	0,222	0,129	0,052	0,069	0,145	EUR	IgV	-
46	rs6691195	A	C	0,284	0,204	0,222	0,131	0,062	0,079	0,153	EUR	DsgV	LINC02474
47	rs7014346	A	G	0,306	0,286	0,335	0,338	0,029	0,021	0,032	EAS	InV, nCTV	CASC8
48	rs704017	G	A	0,433	0,279	0,562	0,417	0,129	0,154	0,016	SAS	RRV	-
49	rs72647484	C	T	0,036	0,001	0,084	0,027	0,048	0,035	0,009	SAS	DsgV	LOC107985377
50	rs7315438	C	T	0,500	0,643	0,447	0,545	0,053	0,143	0,045	SAS	InV, nCTV	LOC105370003
51	rs73208120	G	T	0,052	0	0,084	0,027	0,033	0,052	0,025	SAS	InV	NOS1
52	rs7758229	T	G	0,247	0,246	0,319	0,288	0,072	0,001	0,041	EAS	InV	SLC22A3
53	rs812481	C	G	0,366	0,208	0,441	0,334	0,075	0,158	0,032	SAS	InV	LRIG1
54	rs847208	C	A	0,366	0,323	0,382	0,551	0,016	0,043	0,185	EUR	RRV	-
55	rs961253	A	C	0,227	0,107	0,361	0,299	0,134	0,120	0,072	SAS	UsgV	CASC20

Note: IgV – intergenic variant; InV – intron variant; nCTV – non coding transcript variant; SpAV – splice acceptor variant; DsgV – downstream gene variant; RRV – regulatory region variant; UsgV – upstream gene variant; MV – missense variant; SRV – splice region variant

The similarity of the allele frequencies of the local group with the reference superpopulations showed pronounced locus-dependent variability. For 29 of 55 SNPs (52,7%), the absolute frequency difference with the closest reference superpopulation did not exceed 0,025, while for 12 variants (21,8%) it was greater than 0,05.

At the same time, the closest reference superpopulation varied depending on the locus: for 19 SNPs, it was EAS, while for EUR and SAS, it was 18 each. The virtually equal distribution of variants according to their closest reference superpopulations indicates the absence of a single reference superpopulation closest to the local group based on the frequency profile of the entire investigated panel.

Limitations of the study. The relatively small sample size and the inclusion of participants primarily from Central Kazakhstan limit the generalizability of the obtained frequency characteristics to the Kazakh population as a whole.

The analysis covered a limited set of previously described SNPs associated with CRC; as a result, the identified features of the frequency distribution relate directly to the studied panel of variants.

CONCLUSION

The obtained results are consistent with modern concepts about the admixed and complex multicomponent genetic structure of the Kazakh population.

For the studied panel of CRC-associated SNPs, a locus-dependent mosaicism of the allele frequency profile is observed, characterized by an almost equal distribution of similarity in allele frequencies with reference superpopulations, which must be taken into account while transferring the existing GWAS results across populations, applying polygenic risk scores, and planning medical genetic studies.

Authors' contributions:

A. A. Turmukhambetova, D. Zh. Taizhanova, N. A. Kabildina, I. A. Kadyrova, I. V. Korshukov – conceptualization, methodology.

D. Zh. Taizhanova, N. A. Kabildina, S. I. Kolesnichenko, Y. A. Kolesnikova, I. A. Kadyrova, I. V. Korshukov – data curation, investigation.

I. A. Kadyrova, I. V. Korshukov – formal analysis, software, visualization.

I. A. Kadyrova, I. V. Korshukov – original draft writing.

A. A. Turmukhambetova, D. Zh. Taizhanova, N. A. Kabildina, S. I. Kolesnichenko, Y. A. Kolesnikova, I. A. Kadyrova, I. V. Korshukov – review & editing.

Conflict of interest:

The authors declare no conflict of interest.

REFERENCES

1. Sung H., Filho A.M., Laversanne M. Global cancer statistics 2024: GLOBOCAN estimates of inci-

dence and mortality worldwide for 34 cancers in 186 countries. *CA A Cancer J. Clinicians*. 2026; 76 (4): e70090. <https://doi.org/10.3322/caac.70090>

2. Yamagishi H., Kuroda H., Imai Y., Hiraishi H. Molecular pathogenesis of sporadic colorectal cancers. *Chin. J. Cancer*. 2016; 35 (1): 4. <https://doi.org/10.1186/s40880-015-0066-y>

3. Li J., Ma X., Chakravarti D., Shalapour S., DePinho R.A. Genetic and biological hallmarks of colorectal cancer. *Genes Dev*. 2021; 35 (11-12): 787-820. <https://doi.org/10.1101/gad.348226.120>

4. Matsuda T., Fujimoto A., Igarashi Y. Colorectal cancer: epidemiology, risk factors, and public health strategies. *Digestion*. 2025; 106 (2): 91-99. <https://doi.org/10.1159/000543921>

5. Sawicki T., Ruszkowska M., Danielewicz A., Niedźwiedzka E., Arłukowicz T., Przybyłowicz K.E. A review of colorectal cancer in terms of epidemiology, risk factors, development, symptoms and diagnosis. *Cancers*. 2021; 13 (9): 2025. <https://doi.org/10.3390/cancers13092025>

6. Приказ Министра здравоохранения Республики Казахстан от 16 марта 2011 года №145. «О внесении изменений в приказ и. о. Министра здравоохранения Республики Казахстан от 10 ноября 2009 года №685 «Об утверждении Правил проведения профилактических медицинских осмотров целевых групп населения». <https://adilet.zan.kz/rus/docs/V1100006902>

7. Cui R., Okada Y., Jang S.G. Common variant in 6q26-q27 is associated with distal colon cancer in an Asian population. *Gut*. 2011; 60 (6): 799-805. <https://doi.org/10.1136/gut.2010.215947>

8. Peters U., Jiao S., Schumacher F.R. Identification of genetic susceptibility loci for colorectal tumors in a genome-wide meta-analysis. *Gastroenterology*. 2013; 144 (4): 799-807. <https://doi.org/10.1053/j.gastro.2012.12.020>

9. Jia W.H., Zhang B., Matsuo K., Shin A., Xiang Y.B., Jee S.H., Kim D.H., Ren Z., Cai Q., Long J., et al. Genome-wide association analyses in East Asians identify new susceptibility loci for colorectal cancer. *Nat. Genet*. 2013; 45 (2): 191-196. <https://doi.org/10.1038/ng.2505>

10. Zhang B., Jia W.H., Matsuda K., Kweon S.S., Matsuo K., Xiang Y.B., Shin A., Jee S.H., Kim D.H., Cai Q., et al. Large-scale genetic study in East Asians identifies six new loci associated with colorectal cancer risk. *Nat. Genet*. 2014; 46 (6): 533-542. <https://doi.org/10.1038/ng.2985>

11. Al-Tassan N.A., Whiffin N., Hosking F.J. A new GWAS and meta-analysis with 1000Genomes imputation identifies novel risk variants for colorectal cancer. *Sci. Rep*. 2015; 5 (1): 10442. <https://doi.org/10.1038/srep10442>

12. Schumacher F.R., Schmit S.L., Jiao S. Genome-wide association study of colorectal cancer

- identifies six new susceptibility loci. *Nat. Commun.* 2015; 6 (1): 7138. <https://doi.org/10.1038/ncomms8138>
13. Zeng C., Matsuda K., Jia W.H. Identification of susceptibility loci and genes for colorectal cancer risk. *Gastroenterology*. 2016; 150 (7): 1633-1645. <https://doi.org/10.1053/j.gastro.2016.02.076>
14. Schmit S.L., Edlund C.K., Schumacher F.R. Novel common genetic susceptibility loci for colorectal cancer. *Journal of the National Cancer Institute*. 2019; 111 (2): 146-157. <https://doi.org/10.1093/jnci/djy099>
15. Law P.J., Timofeeva M., Fernandez-Rozadilla C., et al. Association analyses identify 31 new risk loci for colorectal cancer susceptibility. *Nat. Commun.* 2019; 10 (1): 2154. <https://doi.org/10.1038/s41467-019-09775-w>
16. Hofer P., Hagmann M., Brezina S., et al. Bayesian and frequentist analysis of an Austrian genome-wide association study of colorectal cancer and advanced adenomas. *Oncotarget*. 2017; 8 (58): 98623-98634. <https://doi.org/10.18632/oncotarget.21697>
17. Tanikawa C., Kamatani Y., Takahashi A., et al. GWAS identifies two novel colorectal cancer loci at 16q24.1 and 20q13.12. *Carcinogenesis*. 2018;39(5):652-660. <https://doi.org/10.1093/carcin/bgy026>
18. Fernandez-Rozadilla C., Timofeeva M., Chen Z., et al. Deciphering colorectal cancer genetics through multi-omic analysis of 100,204 cases and 154,587 controls of European and East Asian ancestries. *Nat. Genet.* 2023; 55 (1): 89-99. <https://doi.org/10.1038/s41588-022-01222-9>
19. Lu X., Cao M., Han S., Yang Y., Zhou J. Colorectal cancer risk genes are functionally enriched in regulatory pathways. *Sci. Rep.* 2016; 6 (1): 25347. <https://doi.org/10.1038/srep25347>
20. Montazeri Z., Li X., Nyiraneza C., et al. Systematic meta-analyses, field synopsis and global assessment of the evidence of genetic association studies in colorectal cancer. *Gut*. 2020; 69 (8): 1460-1471. <https://doi.org/10.1136/gutjnl-2019-319313>
21. Lachance J. Disease-associated alleles in genome-wide association studies are enriched for derived low frequency alleles relative to HapMap and neutral expectations. *BMC Med. Genomics*. 2010; 3 (1): 57. <https://doi.org/10.1186/1755-8794-3-57>
22. Wang M., Gu D., Du M., et al. Common genetic variation in ETV6 is associated with colorectal cancer susceptibility. *Nat. Commun.* 2016; 7 (1): 11478. <https://doi.org/10.1038/ncomms11478>
23. Schmit S.L., Schumacher F.R., Edlund C.K., et al. Genome-wide association study of colorectal cancer in Hispanics. *CARCIN*. 2016; 37 (6): 547-556. <https://doi.org/10.1093/carcin/bgw046>
24. Tanskanen T., Van Den Berg L., Välimäki N., et al. Genome-wide association study and meta-analysis in Northern European populations replicate multiple colorectal cancer risk loci. *Intl. Journal of Cancer*. 2018; 142 (3): 540-546. <https://doi.org/10.1002/ijc.31076>
25. Lu Y., Kweon S.S., Tanikawa C., et al. Large-scale genome-wide association study of East Asians identifies loci associated with risk for colorectal cancer. *Gastroenterology*. 2019; 156 (5): 1455-1466. <https://doi.org/10.1053/j.gastro.2018.11.066>
26. Lu Y., Kweon S.S., Cai Q., et al. Identification of novel loci and new risk variant in known loci for colorectal cancer risk in East Asians. *Cancer Epidemiology, Biomarkers & Prevention*. 2020; 29 (2): 477-486. <https://doi.org/10.1158/1055-9965.EPI-19-0755>
27. Wong M.C., Ding H., Wang J., Chan P.S., Huang J. Prevalence and risk factors of colorectal cancer in Asia. *Intest. Res.* 2019; 17 (3): 317-329. <https://doi.org/10.5217/ir.2019.00021>
28. Park S.L., Cheng I., Haiman C.A. Genome-wide association studies of cancer in diverse populations. *Cancer Epidemiology, Biomarkers & Prevention*. 2018;27(4):405-417. doi:10.1158/1055-9965.EPI-17-0169
29. He J., Wilkens L.R., Stram D.O., et al. Generalizability and epidemiologic characterization of eleven colorectal cancer GWAS hits in multiple populations. *Cancer Epidemiology, Biomarkers & Prevention*. 2011; 20 (1): 70-81. <https://doi.org/10.1158/1055-9965.EPI-10-0892>
30. Marigorta U.M., Navarro A. High trans-ethnic replicability of GWAS results implies common causal variants. *PLoS Genet.* 2013; 9 (6): e1003566. <https://doi.org/10.1371/journal.pgen.1003566>
31. Hong J., Lunetta K.L., Cupples L.A., Dupuis J., Liu C. Evaluation of a two-stage approach in trans-ethnic meta-analysis in genome-wide association studies. *Genetic Epidemiology*. 2016; 40 (4): 284-292. <https://doi.org/10.1002/gepi.21963>
32. Lin M., Park D.S., Zaitlen N.A., Henn B.M., Gignoux C.R. Admixed populations improve power for variant discovery and portability in genome-wide association studies. *Front Genet.* 2021; 12: 673167. <https://doi.org/10.3389/fgene.2021.673167>
33. Mägi R., Horikoshi M., Sofer T., et al. Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Human Molecular Genetics*. 2017; 26 (18): 3639-3650. <https://doi.org/10.1093/hmg/ddx280>
34. Shi H., Gazal S., Kanai M., et al. Population-specific causal disease effect sizes in functionally important regions impacted by selection. *Nat. Commun.* 2021; 12 (1): 1098. <https://doi.org/10.1038/s41467-021-21286-1>
35. Seidually M., Blazyte A., Jeon S., et al. Decoding a highly mixed Kazakh genome. *Hum. Genet.* 2020; 139 (5): 557-568. <https://doi.org/10.1007/s00439-020-02132-8>

36. Kairov U., Molkenov A., Sharip A., et al. Whole-genome sequencing and genomic variant analysis of Kazakh individuals. *Front Genet.* 2022; 13: 902804. <https://doi.org/10.3389/fgene.2022.902804>
37. Xie T., Shen C., Liu C., et al. Ancestry inference and admixture component estimations of Chinese Kazak group based on 165 AIM-SNPs via NGS platform. *J. Hum. Genet.* 2020; 65 (5): 461-468. <https://doi.org/10.1038/s10038-020-0725-y>
38. Liu Q., Luo L., Liang Q., et al. Efficiency of ancestral inference and characterization of genetic structure of Chinese Kazakh group with a novel InDel panel. *J. Mol. Evol.* 2025; 93 (5): 682-697. <https://doi.org/10.1007/s00239-025-10275-3>
39. Lei C., Liu J., Zhang R., et al. Ancestral origins and admixture history of Kazakhs. *Molecular Biology and Evolution.* 2024; 41 (7): msae144. <https://doi.org/10.1093/molbev/msae144>
40. Raghavan M., Skoglund P., Graf K.E., et al. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature.* 2014; 505 (7481): 87-91. <https://doi.org/10.1038/nature12736>
41. Phipps A.I., Passarelli M.N., Chan A.T., et al. Common genetic variation and survival after colorectal cancer diagnosis: a genome-wide analysis. *CARCIN.* 2016; 37 (1): 87-95. <https://doi.org/10.1093/carcin/bgv161>
42. Buniello A., MacArthur J.A.L., Cerezo M., et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research.* 2019; 47 (D1): D1005-D1012. <https://doi.org/10.1093/nar/gky1120>
43. Pagel K.A., Kim R., Moad K., et al. Integrated informatics analysis of cancer-related variants. *JCO Clinical Cancer Informatics.* 2020; 4: 310-317. <https://doi.org/10.1200/CCI.19.00132>
44. McLaren W., Gil L., Hunt S.E., et al. The Ensembl Variant Effect Predictor. *Genome Biol.* 2016; 17 (1): 122. <https://doi.org/10.1186/s13059-016-0974-4>
45. Purcell S., Neale B., Todd-Brown K., et al. Plink: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics.* 2007; 81 (3): 559-575. <https://doi.org/10.1086/519795>
46. 1000 Genomes Project Consortium; Auton A., Brooks L.D., Durbin R.M., Garrison E.P., Kang H.M., Korbel J.O., Marchini J.L., McCarthy S., McVean G.A., Abecasis G.R. A global reference for human genetic variation. *Nature.* 2015; 526 (7571): 68-74. <https://doi.org/10.1038/nature15393>
2. Yamagishi H., Kuroda H., Imai Y., Hirai-shi H. Molecular pathogenesis of sporadic colorectal cancers. *Chin. J. Cancer.* 2016; 35 (1): 4. <https://doi.org/10.1186/s40880-015-0066-y>
3. Li J., Ma X., Chakravarti D., Shalapour S., DePinho R.A. Genetic and biological hallmarks of colorectal cancer. *Genes Dev.* 2021; 35 (11-12): 787-820. <https://doi.org/10.1101/gad.348226.120>
4. Matsuda T., Fujimoto A., Igarashi Y. Colorectal cancer: epidemiology, risk factors, and public health strategies. *Digestion.* 2025; 106 (2): 91-99. <https://doi.org/10.1159/000543921>
5. Sawicki T., Ruszkowska M., Danielewicz A., Niedźwiedzka E., Arłukowicz T., Przybyłowicz K.E. A review of colorectal cancer in terms of epidemiology, risk factors, development, symptoms and diagnosis. *Cancers.* 2021; 13 (9): 2025. <https://doi.org/10.3390/cancers13092025>
6. Приказ Министра здравоохранения Республики Казахстан от 16 марта 2011 года №145. «О внесении изменений в приказ и. о. Министра здравоохранения Республики Казахстан от 10 ноября 2009 года №685 «Об утверждении Правил проведения профилактических медицинских осмотров целевых групп населения». <https://adilet.zan.kz/rus/docs/V1100006902>
7. Cui R., Okada Y., Jang S.G. Common variant in 6q26-q27 is associated with distal colon cancer in an Asian population. *Gut.* 2011; 60 (6): 799-805. <https://doi.org/10.1136/gut.2010.215947>
8. Peters U., Jiao S., Schumacher F.R. Identification of genetic susceptibility loci for colorectal tumors in a genome-wide meta-analysis. *Gastroenterology.* 2013; 144 (4): 799-807. <https://doi.org/10.1053/j.gastro.2012.12.020>
9. Jia W.H., Zhang B., Matsuo K., Shin A., Xiang Y.B., Jee S.H., Kim D.H., Ren Z., Cai Q., Long J., et al. Genome-wide association analyses in East Asians identify new susceptibility loci for colorectal cancer. *Nat. Genet.* 2013; 45 (2): 191-196. <https://doi.org/10.1038/ng.2505>
10. Zhang B., Jia W.H., Matsuda K., Kweon S.S., Matsuo K., Xiang Y.B., Shin A., Jee S.H., Kim D.H., Cai Q., et al. Large-scale genetic study in East Asians identifies six new loci associated with colorectal cancer risk. *Nat. Genet.* 2014; 46 (6): 533-542. <https://doi.org/10.1038/ng.2985>
11. Al-Tassan N.A., Whiffin N., Hosking F.J. A new GWAS and meta-analysis with 1000Genomes imputation identifies novel risk variants for colorectal cancer. *Sci. Rep.* 2015; 5 (1): 10442. <https://doi.org/10.1038/srep10442>
12. Schumacher F.R., Schmit S.L., Jiao S. Genome-wide association study of colorectal cancer identifies six new susceptibility loci. *Nat. Commun.* 2015; 6 (1): 7138. <https://doi.org/10.1038/ncomms8138>
13. Zeng C., Matsuda K., Jia W.H. Identification of susceptibility loci and genes for colorectal

TRANSLITERATION

1. Sung H., Filho A.M., Laversanne M. Global cancer statistics 2024: GLOBOCAN estimates of incidence and mortality worldwide for 34 cancers in 186 countries. *CA A Cancer J. Clinicians.* 2026; 76 (4): e70090. <https://doi.org/10.3322/caac.70090>

cancer risk. *Gastroenterology*. 2016; 150 (7): 1633-1645. <https://doi.org/10.1053/j.gastro.2016.02.076>

14. Schmit S.L., Edlund C.K., Schumacher F.R. Novel common genetic susceptibility loci for colorectal cancer. *Journal of the National Cancer Institute*. 2019; 111 (2): 146-157. <https://doi.org/10.1093/jnci/djy099>

15. Law P.J., Timofeeva M., Fernandez-Rozadilla C., et al. Association analyses identify 31 new risk loci for colorectal cancer susceptibility. *Nat. Commun.* 2019; 10 (1): 2154. <https://doi.org/10.1038/s41467-019-09775-w>

16. Hofer P., Hagmann M., Brezina S., et al. Bayesian and frequentist analysis of an Austrian genome-wide association study of colorectal cancer and advanced adenomas. *Oncotarget*. 2017; 8 (58): 98623-98634. <https://doi.org/10.18632/oncotarget.21697>

17. Tanikawa C., Kamatani Y., Takahashi A., et al. GWAS identifies two novel colorectal cancer loci at 16q24.1 and 20q13.12. *Carcinogenesis*. 2018;39(5):652-660. <https://doi.org/10.1093/carcin/bgy026>

18. Fernandez-Rozadilla C., Timofeeva M., Chen Z., et al. Deciphering colorectal cancer genetics through multi-omic analysis of 100,204 cases and 154,587 controls of European and East Asian ancestries. *Nat. Genet.* 2023; 55 (1): 89-99. <https://doi.org/10.1038/s41588-022-01222-9>

19. Lu X., Cao M., Han S., Yang Y., Zhou J. Colorectal cancer risk genes are functionally enriched in regulatory pathways. *Sci. Rep.* 2016; 6 (1): 25347. <https://doi.org/10.1038/srep25347>

20. Montazeri Z., Li X., Nyiraneza C., et al. Systematic meta-analyses, field synopsis and global assessment of the evidence of genetic association studies in colorectal cancer. *Gut*. 2020; 69 (8): 1460-1471. <https://doi.org/10.1136/gutjnl-2019-319313>

21. Lachance J. Disease-associated alleles in genome-wide association studies are enriched for derived low frequency alleles relative to HapMap and neutral expectations. *BMC Med. Genomics*. 2010; 3 (1): 57. <https://doi.org/10.1186/1755-8794-3-57>

22. Wang M., Gu D., Du M., et al. Common genetic variation in ETV6 is associated with colorectal cancer susceptibility. *Nat. Commun.* 2016; 7 (1): 11478. <https://doi.org/10.1038/ncomms11478>

23. Schmit S.L., Schumacher F.R., Edlund C.K., et al. Genome-wide association study of colorectal cancer in Hispanics. *CARCIN*. 2016; 37 (6): 547-556. <https://doi.org/10.1093/carcin/bgw046>

24. Tanskanen T., Van Den Berg L., Välimäki N., et al. Genome-wide association study and meta-analysis in Northern European populations replicate multiple colorectal cancer risk loci. *Intl. Journal of Cancer*. 2018; 142 (3): 540-546. <https://doi.org/10.1002/ijc.31076>

25. Lu Y., Kweon S.S., Tanikawa C., et al. Large-scale genome-wide association study of East Asians identifies loci associated with risk for colorectal cancer. *Gastroenterology*. 2019; 156 (5): 1455-1466. <https://doi.org/10.1053/j.gastro.2018.11.066>

26. Lu Y., Kweon S.S., Cai Q., et al. Identification of novel loci and new risk variant in known loci for colorectal cancer risk in East Asians. *Cancer Epidemiology, Biomarkers & Prevention*. 2020; 29 (2): 477-486. <https://doi.org/10.1158/1055-9965.EPI-19-0755>

27. Wong M.C., Ding H., Wang J., Chan P.S., Huang J. Prevalence and risk factors of colorectal cancer in Asia. *Intest. Res.* 2019; 17 (3): 317-329. <https://doi.org/10.5217/ir.2019.00021>

28. Park S.L., Cheng I., Haiman C.A. Genome-wide association studies of cancer in diverse populations. *Cancer Epidemiology, Biomarkers & Prevention*. 2018;27(4):405-417. doi:10.1158/1055-9965.EPI-17-0169

29. He J., Wilkens L.R., Stram D.O., et al. Generalizability and epidemiologic characterization of eleven colorectal cancer GWAS hits in multiple populations. *Cancer Epidemiology, Biomarkers & Prevention*. 2011; 20 (1): 70-81. <https://doi.org/10.1158/1055-9965.EPI-10-0892>

30. Marigorta U.M., Navarro A. High trans-ethnic replicability of GWAS results implies common causal variants. *PLoS Genet.* 2013; 9 (6): e1003566. <https://doi.org/10.1371/journal.pgen.1003566>

31. Hong J., Lunetta K.L., Cupples L.A., Dupuis J., Liu C. Evaluation of a two-stage approach in trans-ethnic meta-analysis in genome-wide association studies. *Genetic Epidemiology*. 2016; 40 (4): 284-292. <https://doi.org/10.1002/gepi.21963>

32. Lin M., Park D.S., Zaitlen N.A., Henn B.M., Gignoux C.R. Admixed populations improve power for variant discovery and portability in genome-wide association studies. *Front Genet.* 2021; 12: 673167. <https://doi.org/10.3389/fgene.2021.673167>

33. Mägi R., Horikoshi M., Sofer T., et al. Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Human Molecular Genetics*. 2017; 26 (18): 3639-3650. <https://doi.org/10.1093/hmg/ddx280>

34. Shi H., Gazal S., Kanai M., et al. Population-specific causal disease effect sizes in functionally important regions impacted by selection. *Nat. Commun.* 2021; 12 (1): 1098. <https://doi.org/10.1038/s41467-021-21286-1>

35. Seidually M., Blazyte A., Jeon S., et al. Decoding a highly mixed Kazakh genome. *Hum. Genet.* 2020; 139 (5): 557-568. <https://doi.org/10.1007/s00439-020-02132-8>

36. Kairov U., Molkenov A., Sharip A., et al. Whole-genome sequencing and genomic variant analysis of Kazakh individuals. *Front Genet.* 2022; 13: 902804. <https://doi.org/10.3389/fgene.2022.902804>

37. Xie T., Shen C., Liu C., et al. Ancestry inference and admixture component estimations of Chinese Kazak group based on 165 AIM-SNPs via NGS platform. *J. Hum. Genet.* 2020; 65 (5): 461-468. <https://doi.org/10.1038/s10038-020-0725-y>

38. Liu Q., Luo L., Liang Q., et al. Efficiency of ancestral inference and characterization of genetic structure of Chinese Kazakh group with a novel InDel panel. *J. Mol. Evol.* 2025; 93 (5): 682-697. <https://doi.org/10.1007/s00239-025-10275-3>

39. Lei C., Liu J., Zhang R., et al. Ancestral origins and admixture history of Kazakhs. *Molecular Biology and Evolution*. 2024; 41 (7): msae144. <https://doi.org/10.1093/molbev/msae144>
 40. Raghavan M., Skoglund P., Graf K.E., et al. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature*. 2014; 505 (7481): 87-91. <https://doi.org/10.1038/nature12736>
 41. Phipps A.I., Passarelli M.N., Chan A.T., et al. Common genetic variation and survival after colorectal cancer diagnosis: a genome-wide analysis. *CARCIN*. 2016; 37 (1): 87-95. <https://doi.org/10.1093/carcin/bgv161>
 42. Bunieello A., MacArthur J.A.L., Cerezo M., et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research*. 2019; 47 (D1): D1005-D1012. <https://doi.org/10.1093/nar/gky1120>
 43. Pagel K.A., Kim R., Moad K., et al. Integrated informatics analysis of cancer-related variants. *JCO Clinical Cancer Informatics*. 2020; 4: 310-317. <https://doi.org/10.1200/CCI.19.00132>
 44. McLaren W., Gil L., Hunt S.E., et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016; 17 (1): 122. <https://doi.org/10.1186/s13059-016-0974-4>
 45. Purcell S., Neale B., Todd-Brown K., et al. Plink: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*. 2007; 81 (3): 559-575. <https://doi.org/10.1086/519795>
 46. 1000 Genomes Project Consortium; Auton A., Brooks L.D., Durbin R.M., Garrison E.P., Kang H.M., Korbel J.O., Marchini J.L., McCarthy S., McVean G.A., Abecasis G.R. A global reference for human genetic variation. *Nature*. 2015; 526 (7571): 68-74. <https://doi.org/10.1038/nature15393>
- Received 10.01.2026
Sent for revision 28.02.2026, 12.04.2026
Accepted 25.07.2026
Published online 30.09.2026

И. В. Коршуков^{1*}, И. А. Кадырова¹, Е. А. Колесникова², С. И. Колесниченко³, Н. А. Кабилдина¹, Д. Ж. Тайжанова¹, А. А. Турмухамбетова⁴

АЛЛЕЛЬНЫЕ ЧАСТОТЫ ОДНОНУКЛЕОТИДНЫХ ПОЛИМОРФИЗМОВ, АССОЦИИРОВАННЫХ С КОЛОРЕКТАЛЬНЫМ РАКОМ, В КАЗАХСКОЙ ПОПУЛЯЦИИ ЦЕНТРАЛЬНОГО КАЗАХСТАНА

¹НАО «Карагандинский медицинский университет» (100008, Республика Казахстан, г. Караганда, ул. Гоголя, 40; e-mail: info@qmu.kz)

²Федеральное государственное автономное образовательное учреждение высшего образования «Российский национальный исследовательский медицинский университет имени Н. И. Пирогова» (117513, Российская Федерация, г. Москва, ул. Островитянова, 1; e-mail: kolesnikova_ea6@rsmu.ru)

³Университетский госпитальный центр Сент-Жюстин, исследовательский центр, Университет Монреаля (QC H3S 2G4, Канада, г. Монреаль, ул. Кот-Сент-Катрин, 3175; e-mail: svetlana.kolesnichenko@umontreal.ca)

⁴НАО «Медицинский университет Астана» (010000, Республика Казахстан, г. Астана, ул. Бейбітшілік, 49а; e-mail: mail@amu.kz)

*Илья Васильевич Коршуков – НАО «Карагандинский медицинский университет»; 100008, Республика Казахстан, г. Караганда, ул. Гоголя, 40; e-mail: Korshukov@qmu.kz

Цель. Оценить аллельные частоты однонуклеотидных полиморфизмов (ОНП), ассоциированных с риском развития колоректального рака (КРР), в казахской популяции Центрального Казахстана и сопоставить полученные результаты с международными референсными данными.

Материалы и методы. В исследование были включены 97 условно здоровых лиц без установленного КРР и семейного анамнеза злокачественных новообразований, определявшие себя, своих биологических родителей, а также бабушек и дедушек по материнской и отцовской линии как казахов. Генотипирование выполняли методом TaqMan OpenArray. После контроля пропусков данных и проверки соответствия распределения генотипов закону Харди – Вайнберга проводили гармонизацию с данными 1000 Genomes Project Phase 3. Популяционное положение локальной выборки оценивали методом анализа главных компонент и иерархической кластеризации. Для каждого ОНП рассчитывали абсолютные разности частот аллелей локальной группы относительно референсных суперпопуляций Евразии и определяли ближайшую референсную суперпопуляцию. Статистические выбросы исключали по правилу Тьюки. В итоговый анализ были включены 55 ОНП.

Результаты и обсуждение. Анализ главных компонент показал промежуточное положение локальной группы в евразийском генетическом пространстве, преимущественно между европейской и восточноазиатской суперпопуляциями. При иерархической кластеризации локальная группа была представлена во всех четырех кластерах, преимущественно распределяясь между двумя со смешанным популяционным составом. Частотный профиль исследованных ОНП характеризовался

выраженной локус-зависимой вариабельностью. Для 29 из 55 ОНП (52,7%) абсолютное различие частот с ближайшей референсной суперпопуляцией не превышало 0,025, тогда как для 12 вариантов (21,8%) оно составляло более 0,05. Наиболее близкой суперпопуляцией для 19 ОНП являлась EAS, для 18 – EUR и для 18 – SAS.

Выводы. Результаты согласуются с представлениями о смешанном характере казахской популяции и демонстрируют локус-зависимую мозаичность частотного профиля исследованной панели ОНП. Практически равное распределение сходства с евразийскими суперпопуляциями указывает на отсутствие единственной референсной суперпопуляции, наиболее близкой к локальной выборке по всей панели вариантов, что необходимо учитывать при переносе результатов исследований по полногеномному поиску ассоциаций, использовании полигенных шкал риска и планировании медико-генетических исследований.

Ключевые слова: колоректальный рак; однонуклеотидные полиморфизмы; аллельные частоты; казахская популяция

И. В. Коршуков^{1*}, И. А. Кадырова¹, Е. А. Колесникова², С. И. Колесниченко³, Н. А. Кабилдина¹, Д. Ж. Тайжанова¹, А. А. Тұрмұхамбетова⁴

ОРТАЛЫҚ ҚАЗАҚСТАННЫҢ ҚАЗАҚ ПОПУЛЯЦИЯСЫНДА КОЛОРЕКТАЛДЫ ОБЫРМЕН АССОЦИАЦИЯЛАНҒАН БІРНУКЛЕОТИДТІ ПОЛИМОРФИЗМДЕРДІҢ АЛЛЕЛЬДІК ЖИІЛІКТЕРІ

¹«Қарағанды медицина университеті» КеАҚ (100008, Қазақстан Республикасы, Қарағанды қ., Гоголь к., 40; e-mail: info@qmu.kz)

²«Н. И. Пирогов атындағы Ресей ұлттық медициналық зерттеу университеті» жоғары білім берудің федералды мемлекеттік автономды білім беру мекемесі (117513, Ресей Федерациясы, Мәскеу қ., Островитянов к-сі, 1; e-mail: kolesnikova_ea6@rsmu.ru)

³Университеттік аурухана орталығы Сент-Жюстин, зерттеу орталығы, Монреаль университеті (QC H3S 2G4, Канада, Монреаль қ., Кот-Сент-Кэтрин к., 3175; e-mail: svetlana.kolesnichenko@umontreal.ca)

⁴«Астана медицина университеті» КеАҚ (010000, Қазақстан Республикасы, Астана қ., Бейбітшілік к., 49а; e-mail: mail@amu.kz)

***Илья Васильевич Коршуков** – «Қарағанды медицина университеті» КеАҚ; 100008, Қазақстан Республикасы, Қарағанды қ., Гоголь к. 40; e-mail: Korshukov@qmu.kz

Зерттеудің мақсаты. Орталық Қазақстанның қазақ популяциясында колоректалды обырдың (КРО) даму қаупімен ассоциацияланған бірнуклеотидті полиморфизмдердің (БНП) аллельдік жиіліктерін бағалау және алынған нәтижелерді халықаралық референстік деректермен салыстыру.

Материалдар және әдістер. Зерттеуге КРО диагнозы қойылмаған және отбасылық анамнезінде қатерлі жаңа түзілімдер болмаған, өзін, биологиялық ата-аналарын, сондай-ақ анасы мен әкесі жағынан ата-әжелерін қазақ деп көрсеткен шартты түрде дені сау 97 адам енгізілді. Генотиптеу TaqMan OpenArray әдісімен жүргізілді. Деректердегі өткізіп алынған мәндерге бақылау жүргізіліп, генотиптер таралуының Харди-Вайнберг тепе-теңдігіне сәйкестігі тексерілгеннен кейін деректер 1000 Genomes Project Phase 3 деректерімен үйлестірілді. Жергілікті іріктеменің популяциялық орны негізгі компоненттерді талдау әдісі және иерархиялық кластерлеу арқылы бағаланды. Әрбір БНП үшін жергілікті топтың аллель жиіліктері мен Еуразияның референстік суперпопуляцияларындағы аллель жиіліктері арасындағы абсолютті айырмашылықтар есептеліп, ең жақын референстік суперпопуляция анықталды. Статистикалық ауытқымалы мәндер Тьюки ережесі бойынша алынып тасталды. Қорытынды талдауға 55 БНП енгізілді.

Нәтижелер және талқылау. Негізгі компоненттерді талдау жергілікті топтың еуразиялық генетикалық кеңістікте, негізінен еуропалық және шығысаязиялық суперпопуляциялар арасында аралық орын алатынын көрсетті. Иерархиялық кластерлеу нәтижесінде жергілікті топ барлық төрт кластерде ұсынылып, негізінен популяциялық құрамы аралас екі кластер арасында бөлінді. Зерттелген БНП-лардың жиілік профилі локуста тәуелді айқын вариабельділікпен сипатталды. 55 БНП-ның 29-ында (52,7%) ең жақын референстік суперпопуляциямен аллель жиіліктерінің абсолютті айырмашылығы 0,025-тен аспады, ал 12 нұсқада (21,8%) бұл көрсеткіш 0,05-тен жоғары болды. 19 БНП үшін ең жақын суперпопуляция EAS, 18 БНП үшін EUR және 18 БНП үшін SAS болды.

Қорытындылар. Алынған нәтижелер қазақ популяциясының аралас сипаты туралы қалыптасқан түсініктерге сәйкес келеді және зерттелген БНП панелінің жиілік профилінің локуста тәуелді мозаикалылығын көрсетеді. Еуразиялық суперпопуляциялармен ұқсастықтың іс жүзінде біркелкі

бөлінуі нұсқалардың бүкіл панелі бойынша жергілікті іріктемеге ең жақын бірыңғай референстік суперпопуляцияның жоқ екенін көрсетеді. Бұл жағдайды геном бойынша ассоциацияларды іздеу зерттеулерінің нәтижелерін популяциялар арасында көшіру кезінде, полигендік қауіп шкалаларын пайдалануда және медициналық-генетикалық зерттеулерді жоспарлауда ескеру қажет.

Кілт сөздер: колоректалды обыр; бірнуклеотидті полиморфизмдер; аллельдік жиіліктер; қазақ популяциясы

© КОЛЛЕКТИВ АВТОРОВ, 2026

УДК 615.322:616-092.9:612.82:

DOI 10.59598/ME-2305-6053-2026-120-3-116-124

N. S. Abugaliyeva^{1*}, M. K. Iztleuov², Y. M. Iztleuov³, S. A. Ryzhkin⁴, G. A. Smagulova⁵, S. S. Saparbaev⁶

PROTECTIVE EFFECTS OF URTICA DIOICA L. OIL ON OXIDATIVE STRESS AND INFLAMMATION IN THE RAT BRAIN FOLLOWING GAMMA IRRADIATION AND HEXAVALENT CHROMIUM EXPOSURE

¹Aktobe Regional Multidisciplinary Hospital (030017, Republic of Kazakhstan, Aktobe c., Zhanakonys str., 8E; e-mail: aktobeob@mail.ru)

²Department of Natural sciences, NC JSC «Marat Ospanov West Kazakhstan Medical University» (030000, Republic of Kazakhstan, Aktobe c., Mareseva str., 68; e-mail: info@zkmk.kz)

³Department of Radiation diagnostics, NC JSC «Marat Ospanov West Kazakhstan Medical University» (030000, Republic of Kazakhstan, Aktobe c., Mareseva str., 68; e-mail: info@zkmk.kz)

⁴Department of Radiology, Radiotherapy, Radiation Hygiene and Radiation Safety, Russian Medical Academy of Continuous Professional Education (125993, Russian Federation, Moscow c., Barrikadnaya str., 2/1; e-mail: rsa777@inbox.ru)

⁵Department of Pharmacology, NC JSC «Marat Ospanov West Kazakhstan Medical University» (030000, Republic of Kazakhstan, Aktobe c., Mareseva str., 68; e-mail: info@zkmk.kz)

⁶Medical Center Al-Jhami (020000, Republic of Kazakhstan, Astana c., Beibit Mailin str., 31; e-mail: al-zhami@mail.ru)

***Nazerke Serikkyzy Abugaliyeva** – Aktobe Regional Multidisciplinary Hospital; 030017, Republic of Kazakhstan, Aktobe c., Zhanakonys str., 8E; e-mail: nazerkeabugaliyevaaa@gmail.com

Aim. To evaluate the neuroprotective effects of *Urtica dioica* L. oil on oxidative stress and inflammatory responses in the brains of rats exposed to ionizing radiation alone or in combination with hexavalent chromium.

Materials and methods. Sixty-five white adult male rats were randomly assigned to five groups (n=13/group): control, irradiation, irradiation + *Urtica dioica* L. oil, irradiation + Cr(VI), and irradiation + Cr(VI) + *Urtica dioica* L. oil. Seven days after irradiation, brain tissue was collected for biochemical analysis. Levels of malondialdehyde (MDA), reduced glutathione (GSH), superoxide dismutase (SOD), catalase (CAT), tumor necrosis factor- α (TNF- α), and interleukin-6 (IL-6) were determined.

Results and discussion. Exposure to ionizing radiation, particularly in combination with Cr(VI), significantly impaired antioxidant defenses and increased lipid peroxidation and inflammatory cytokine levels in brain tissue. Pretreatment with *Urtica dioica* L. oil significantly increased SOD, CAT, and GSH levels while reducing MDA, TNF- α , and IL-6 compared with the corresponding untreated exposed groups (p<0.05).

Conclusions. *Urtica dioica* L. oil attenuated oxidative stress and neuroinflammation induced by ionizing radiation alone or combined with hexavalent chromium. These findings suggest that nettle oil has promising neuroprotective potential through enhancement of endogenous antioxidant defenses and suppression of inflammatory responses.

Key words: ionizing radiation; brain; oxidative stress; inflammation; nettle oil; *Urtica dioica* L.

INTRODUCTION

Medicinal plants have historically served as a fundamental component of healthcare systems worldwide. Beyond their traditional use as remedies, many plant species are also consumed as food ingredients and spices, providing a rich source of bioactive compounds for pharmaceutical research. In recent decades, plant-based nutraceuticals and functional foods have attracted considerable scientific attention due to their potential protective effects against diseases associated with modern lifestyles and environmental factors, including inflammation, hypercholesterolemia, atherosclerosis, endothelial dysfunction, and cancer progression.

Among medicinal plants, *Urtica dioica* L., commonly known as stinging nettle, has attracted considerable scientific interest due to its wide range of biological activities. *Urtica dioica* L. is a perennial herbaceous plant belonging to the genus *Urtica*, characterized by serrated leaves covered with stinging hairs. Its leaves contain a rich composition of nutrients and bioactive compounds, including dietary fiber, essential fatty acids (linoleic, linolenic, palmitic, stearic, and oleic acids), essential amino acids, and pigments such as β -carotene, lutein, and other carotenoids. In addition, nettle is abundant in antioxidant compounds, including polyphenols (kaempferol, quercetin, caffeic acid, coumarins, and other flavo-

noids). It also contains numerous minerals such as calcium, zinc, magnesium, sodium, iron, phosphorus, cobalt, potassium, nickel, and molybdenum, as well as vitamins A, C, K, and B-complex vitamins. Due to this balanced nutritional composition and high content of bioactive compounds, stinging nettle has gained increasing recognition as a valuable medicinal and functional plant [1].

Numerous studies have demonstrated that *Urtica dioica* L. exhibits a broad spectrum of pharmacological properties, including antioxidant, anti-inflammatory, antiproliferative, analgesic, immunomodulatory, antihypertensive, anti-infective, and anti-ulcer activities. Moreover, nettle has been suggested to contribute to the prevention of cardiovascular and neurodegenerative diseases [2].

Oxidative stress plays a central role in the pathogenesis of many neurological disorders. During incomplete reduction of molecular oxygen (O_2), reactive oxygen species (ROS) such as superoxide anion ($O_2^{\bullet-}$), hydroperoxyl radical ($HO_2^{\bullet}$), and hydroxyl radical ($HO^{\bullet}$) are generated [3]. These highly reactive molecules can damage lipids, proteins, and DNA, thereby disrupting cellular homeostasis. ROS-mediated oxidative stress can activate multiple intracellular signaling pathways, including the nuclear factor kappa-B (NF- κ B) pathway. Activation of NF- κ B promotes neuroinflammation, microglial activation, neuronal apoptosis and necrosis, and secondary neuronal damage. Furthermore, it stimulates the transcription of genes encoding pro-inflammatory cytokines such as interleukin-1 β (IL-1 β), interleukin-6 (IL-6), and tumor necrosis factor- α (TNF- α) [4].

Reactive oxygen species particularly affect polyunsaturated fatty acids in cellular membranes, initiating lipid peroxidation chain reactions that ultimately lead to the formation of malondialdehyde (MDA), a widely recognized marker of oxidative damage [5]. To counteract oxidative stress, the body relies on endogenous antioxidant defense systems, including enzymatic antioxidants such as superoxide dismutase (SOD), glutathione peroxidase (GPx), catalase (CAT), and glutathione reductase (GR), as well as non-enzymatic antioxidants such as reduced glutathione (GSH), cysteine, coenzyme Q, and vitamins A, C, and E, metals (Se, Zn) [6].

Plant-derived compounds have attracted significant attention for their ability to modulate oxidative stress and inflammatory responses. Herbal extracts may exert anti-inflammatory effects by targeting various cellular signaling pathways or endogenous enzymes involved in inflammatory processes. In particular, extracts of *Urtica dioica* L. have been reported to modulate the inflammatory cascade by inhibiting cyclooxygenase enzymes (COX-1 and COX-2), thereby reducing the synthesis of inflammatory prostaglandins, and by suppressing mast cell tryptase activity to prevent degranulation [7].

In addition to these mechanisms, bioactive compounds present in nettle, such as rutin, kaempferol, and vitamin A, have been shown to possess strong antioxidant and anti-inflammatory properties. The phenolic and flavonoid constituents of *Urtica dioica* L. contribute to the scavenging of reactive oxygen species and the inhibition of lipid peroxidation, thereby protecting cells from oxidative damage [8].

Despite the growing body of evidence supporting the biological activities of *Urtica dioica* L., further research is needed to identify its active phytochemicals, clarify their mechanisms of action, and evaluate their therapeutic potential in conditions associated with oxidative stress and inflammation.

Therefore, the study aimed to investigate the effects of *Urtica dioica* L. oil on oxidative stress and inflammatory responses in the brain following combined exposure to ionizing radiation and hexavalent chromium, two environmental factors known to induce severe oxidative and inflammatory damage in neural tissues.

MATERIALS AND METHODS

White adult male rats (170-190 g) were obtained from the vivarium of the Scientific and Practical Center of NC JSC «Marat Ospanov West Kazakhstan Medical University» (Aktobe c., Republic of Kazakhstan). Prior to the experiment, the animals underwent a 14-day acclimatization period.

All experimental procedures were approved by the Institutional Animal Ethics Committee of NC JSC «Marat Ospanov West Kazakhstan Medical University» and were conducted in accordance with the European Convention for the Protection of Vertebrate Animals Used for Experimental and Other Scientific Purposes (Strasbourg, 1986).

A total of sixty-five rats were randomly assigned to five experimental groups ($n=13/\text{group}$): group I – control group; group II – rats exposed to a single total-body γ -irradiation dose of 4 Gy; group III – rats treated daily with *Urtica dioica* L. oil at a dose of 2.5 mL/kg body weight for 14 days, followed by total-body γ -irradiation (4 Gy); group IV – rats administered potassium dichromate ($K_2Cr_2O_7$) in drinking water at a concentration equivalent to 180 mg/L Cr(VI) for 14 days, followed by total-body γ -irradiation (4 Gy); group V – rats administered potassium dichromate ($K_2Cr_2O_7$) in drinking water at a concentration equivalent to 180 mg/L Cr(VI) and treated with *Urtica dioica* L. (2.5 mL/kg body weight) for 14 days, followed by total-body γ -irradiation (4 Gy).

Seven days after irradiation, animals were euthanized under light anesthesia by cervical decapitation.

The brain was rapidly excised and placed in ice-cold buffer containing 10 mM Tris-HCl (pH 7.4). Brain tissue samples were centrifuged, and the supernatants were stored at -20°C until biochemical analysis.